Caractérisation des problèmes conjoints de détection et d'estimation

Eric Chaumette

ONERA/DEMR **Ecole Doctorale de l'ENS-CACHAN**

Caractérisation des problèmes conjoints de détection et d'estimation

Habilitation à Diriger des Recherches

(spécialité Sciences et Technologie de l'Information et de la Communication)

présentée et soutenue publiquement le 13/01/2014

par

Eric Chaumette

Composition du Jury

Rapporteurs :	Sylvie MARCOS	Directeur de Recherche CNRS, Gif sur Yvette
	Cédric RICHARD	Professeur des Universités, Nice
	Yannick BERTHOUMIEU	Professeur des Universités, Bordeaux
Examinateurs :	Jean-Yves TOURNERET	Professeur des Universités, Toulouse
	Pascal CHEVALIER	Professeur des Universités, Paris
	Pascal LARZABAL	Professeur des Universités, Cachan (garant)

CONTENTS

LIST OF FIGURES

Remerciements

C’est un très grand honneur pour moi que d’avoir pu défendre cette HDR devant un tel jury : je vous remercie d’avoir pris la peine de lire du ”Chaumette en version originale et sans sous-titrage”.
Je me suis en effet fait plaisir dans la rédaction de ce manuscrit synthétisant ma vision du thème de recherche : mes acquis, mes questionnements, mes axes futurs d’investigation.
Vision qui n’aurait été possible sans la dynamique de la recherche collaborative et partagée, que ce soit avec des thésard(e)s ou des chercheurs confirmés.

Une dédicace spéciale à Pascal Larzabal, mon mentor, qui m’a permis de comprendre la différence entre un scientifique compétent techniquement et un directeur de thèse à la hauteur de ses responsabilités...Merci à toi.
Ton enseignement restera pour moi un phare dans la brume de mon ambivalence.

Sans compter tous les ”...; tu verras, ce sera bon pour ton HDR.” depuis l’obtention de ma thèse. Sans tes encouragements continuels, cette page n’aurait jamais existée, pas plus que l’opportunité qui s’ouvre à moi (poste d’enseignant chercheur à l’ISAE) concrétisant un projet personnel de 25 ans ...

A mes affinités électives, passées, présentes et futures.

A la névrose... ma maîtresse, ma muse, mon aliénation.

A tous ceux qui m'ont inspiré et permis d'en desserrer l'étreinte : j'ai grandi à travers vous.

I. CURRICULUM VITÆ

A. *Etat civil*

Eric Chaumette

Date et lieu de naissance	13 Octobre 1965, Chartres (28)
Nationalité	Française
Situation de Famille	Célibataire, divorcé
Adresse professionnelle	Onera, Centre de Palaiseau Chemin de la Hunière et des Joncherettes BP 80100 FR-91123 Palaiseau Cedex ○ Numéro de téléphone 01 80 38 62 96 ○ Adresses électroniques eric.chaumette@onera.fr eric.chaumette@satie.ens-cachan.fr
Fonctions Actuelles	○ Onera Ingénieur Spécialiste (Position III.B) ○ SATIE - Ecole Normale Supérieure de Cachan Chercheur associé

B. *Parcours académique*

2004 Doctorat en traitement du signal de L'ENS-CACHAN :
Contribution à la caractérisation des performances des problèmes conjoints de détection et d'estimation.
Jury : Philippe Loubaton (Président), Pascal Larzabal (Directeur de Thèse),
Jean Jacques Fuchs (Rapporteur), Jean Yves Tourneret (Rapporteur), Philippe Forster (Examinateur),
Jean Claude Guillerot (Examinateur)

1989 DEA de Traitement du Signal et de l'Image, ENSHEEIT-INPT, Toulouse (Mention Bien)

1989 Ingénieur de l'Ecole Nationale de l'Aviation Civile (ENAC), Filière Electronique, Toulouse

1983 Baccalauréat série C au Lycée Jean Rostand à Chantilly (Mention Bien)

C. *Expérience Professionnelle*

Onera (11/2007-)

Ingénieur Spécialiste en traitement du signal
○ Contribution au simulateur générique radar développé dans le cadre du Plan d'Etudes Amonts (PEA) Evolutions ASTRAD : spécification et développement de la simulation du Front-End émission-propagation-réception, expertise du module traitement du signal.
○ Contribution au simulateur radar développé dans le cadre du contrat SIPRE (simulation de la perturbation des radar par les éoliennes) pour l'ADEME : spécification et développement de la simulation du Front-End émission-réception et de l'interface Front-End radar avec les calculs de champs électromagnétiques réalisés par OKTAL-SE.
○ Spécification des bibliothèques de Font-End en réception et du traitement du signal pour un simulateur radar SSA

(space situation awareness) pour l'ESA (European Space Agency), programmation de la bibliothèque de Front-End.
○ Réalisation d'un simulateur du radar expérimental HYCAM développé par l'ONERA pour la DGA, pour aide à la spécification du sous-ensemble de la chaîne de réception et évaluations des performances globales attendues.
○ Etude de l'apport des techniques haute résolution pour la réduction les temps de pose alloués à l'analyse des cibles aérobies (reconnaissance de cibles non coopératives), PEA NECTAR.

Thales Air Système (TR6) (07/2006 - 10/2007)

Responsable technique équipe Simulation Front-End Radar (4 personnes), client Service Programme TR6
○ Simulation fonctionnelle du Front-End des radars du Portfolio TR6
○ Supervision de l'encapsulation de la RSL (Radar Simulation Library) dans ASTRAD

Thales Naval France (TNF) (07/2000 - 06/2006)

Responsable technique équipe Simulation Front-End radar (4 personnes), client Service Programme TNF
Simulation fonctionnelle du Front-End des radars du Portfolio TNF avec pour objectifs:
○ l'évaluation des performances de la partie traitement du signal de la chaine radar (probabilité de détection, de fausse alarme, précision mesures angulaire, vitesse, distance, ...), pour les radars en production, leurs évolutions ou les concepts futurs,
○ la validation du hardware traitement du signal (jeux d'essais fonctionnels),
○ la validation du traitement de données (données IQ pour l'extraction, le pistage, ...).

Encadrement d'équipe : coordination des taches, garantie de la tenue des jalons
Responsabilité technique : garantie de la représentativité modélisation fonctionnelle / système réel
Responsabilité logicielle : décisions des évolutions du code de la RSL, supervision technique des évolutions (analyse mathématique, règles d'implémentation, définition de scénarios d'essais et analyse des résultats).

Expert Modélisation scène radar, clients Service Projet TNF et Danish Defence Research Establishment
○ modélisation du radar de mesure de SER du DDRE, radar CW à code de phase, mise en place d'un scénario ASMD complexe (radar CW et pulses, brouilleurs DRFM, capacité SAR, ISAR, tir chaff, cinématiques, ...), 30 semaines à Copenhague sur 2 ans.

Responsable technique Simulation Front-End Radar, client Service Programme TNF
Conception et programmation de la RSL :
○ bibliothèque (Pv-Wave, 30 000 lignes, code vectorisé) de simulation fonctionnelle du Front-End de la chaîne radar : modèle d'émetteur, modèle d'antennes (gain-polarisation), propagation en champs lointain, réflexion sur une surface sphérique, points brillants, matrice de rétrodiffusion, cinématiques libres (translation-rotation), modèle de récepteur (modulateur, filtre analogique, filtre numérique, échantillonneur, gvt, limiteur, bruit thermique), modélisation temporelle ou stochastique des signaux.

Mise en œuvre de la RSL dans les études et programmes TNF :
○ Etude Eurofinder WAGNER, client MoD français et néerlandais : responsable de la génération des données IQ pour les 2 scénarios complexes de référence de l'étude (terre et zone littorale)
○ Programme radar Héraklès, client Singapour : évaluation des performances de la mesure angulaire à site bas en présence du trajet réfléchi sur la surface terrestre.
○ Programme radar MRR SAN PC, client Afrique du Sud : spécification des défauts différentiels admissibles de la voie auxiliaire du dispositif d'antibrouillage "Opposition par Lobes Secondaires" (OLS), support aux départements de production industriel, support à la recette client.

Expert Technique en Traitement du Signal Radar, client Service Programme TNF
○ Etude Eurofinder WAGNER : audit technique des traitements avancés proposés (analyse haute résolution Doppler, Track Before Detect, levée d'ambiguïté, antibrouillage)
○ Programme radar Héraklès : analyse théorique (implémentation, prédiction de performances) des algorithmes de mesure angulaire à site bas, compatibles avec le radar Héraklès.
○ Programme radar MRR SAN PC : analyse théorique (implémentation, prédiction de performances) de l'antibrouillage OLS et des algorithmes de mesure angulaire à site bas, compatibles avec le MRR.

Thomson-Csf - Systèmes détection et contrôle (SDC) (09/1990 – 06/2000)

Ingénieur Etudes Amont Radar, service "Radar Algorithmes et Nouveaux Concepts"
○ Etude et définition de traitements adaptatifs de réjection de fouillis polarisé : modélisation, performance en détection, en régulation de fausse alarme (Ku polar, DGA/DRET)
○ Etude des méthodes de détection de cibles furtives: intégration longue (Track Before Detect) non antibrouillée (radar V26D, Arabel, étude DGA/DCN) et antibrouillée (complément d'étude RIAS).
○ Etude des méthodes de calibration sans phare (autocalibration, DGA/DRET)
○ Expertise technique de l'antibrouillage AMSAR (radar aéroporté, mission à Thomson-Detexis)
○ Expertise technique de l'antibrouillage de la poursuite du radar SAFRAN (démonstrateur radar de champ de bataille à FFC, DGA)
○ Etude et réalisation d'un dispositif d'antibrouillage large bande pour le RIAS (démonstrateur radar FFC multi-émetteurs, multi-récepteurs, contrat ONERA-THOMSON, client DGA/STTE).
○ Application des algorithmes de séparation de sources aux Ordres Supérieurs au radar : degarbling en radar secondaire (Etude Autofinancée)

Expert technique auprès du Collège Scientifique et Technique (CST) de THOMSON-CSF en Traitement du Signal (1996 - 1999).

Abréviations et Acronymes

ADEME	Agence de l'Environnement et de la Maîtrise de l'Energie
ASMD	Anti-ship missile defence
ASTRAD	Atelier de Simulation et de Traitement appliqué au Radar et aux Autodirecteurs (PEA DGA - ONERA - THALES)
CCTP	Cahier des clauses techniques particulières
DCN	Direction des Constructions Navales
DGA	Direction Générale pour l'Armement
DRET	Direction des Recherches et Etudes Techniques
DRFM	Digital RF Memory
EDA	European Defense Agency
ESA	European Space Agency
ISAR	Inverse Synthethic Aperture Radar
MIMO	Multiple Input Multiple Output
PEA	Plan d'Etudes Amonts
ONERA	Office National d'Etudes et Recherches Aérospatiales
RADAR	Radio Detection And Ranging
RSL	Radar Simulation Library (bibliothèque logicielle de simulation du Front-End radar)
SAR	Synthethic Aperture Radar
SBU	System Business Unit
TBU	Technical Business Unit
TAD	Thales Air Defence
TNF	Thales Naval France
TR6	Thales Air Système

D. Production scientifique et technique

Revues :

- 14 articles de revue internationale avec comité de lecture publiés (dont 13 depuis 2005)
- 1 article de revue nationale avec comité de lecture publié
- 2 article de revue internationale avec comité de lecture soumis

Conférences :

- 29 articles de conférence internationale avec comité de lecture et actes publiés (dont 19 depuis 2004)
- 5 articles de conférence nationale (GRETSI) avec comité de lecture et actes publiés (dont 1 depuis 2004)

1) Revues internationales avec comité de lecture:

[J14] T. Menni, J. Galy, E. Chaumette, P. Larzabal,"On the versatility of constrained Cramér-Rao bound for estimation performance analysis and design of a system of measurement", IEEE Trans. on AES, accepté le 08/09/2013

[J13] F. Vincent, O. Besson, E. Chaumette, "Computationally efficient approximate maximum likelihood estimation of two closely spaced sources", Signal Processing (Elsevier), 97:83-90, 2014

[J12] U.R.O. Nickel, E. Chaumette, P. Larzabal, "Estimation of extended targets using the Generalized Monopulse Estimator: extension to a mixed target model", IEEE Trans. on AES, 49(3): 2084 2096, 2013

[J11] E. Chaumette, U.R.O. Nickel, P. Larzabal, "Detection and Parameter Estimation of Extended Targets Using the Generalized Monopulse Estimator", IEEE Trans. on AES, 48(4): 3389-3417, 2012

[J10] T. Menni, E. Chaumette, P. Larzabal, J.P. Barbot, "New Results on Deterministic Cramér–Rao Bounds for Real and Complex Parameters", IEEE Trans. on SP, 60(3): 1032-1049, 2012

[J9] U.R.O. Nickel, E. Chaumette, P. Larzabal, "Statistical Performance Prediction of Generalized Monopulse Estimation", IEEE Trans. on AES, 47(1): 381-404, 2011

[J8] J. Galy, E. Chaumette, P. Larzabal, "Joint Detection Estimation Problem of Monopulse Angle Measurement", IEEE Trans. on AES, 46(1): 397-413, 2010

[J7] E. Chaumette, P. Larzabal, "Monopulse-radar tracking of swerling III-IV targets using multiple observations", IEEE Trans. on AES, 44(2): 520-537, 2008

[J6] E. Chaumette, J. Galy, A. Quinlan, P. Larzabal, "A New Barankin Bound Approximation for the Prediction of the Threshold Region Performance of Maximum Likelihood Estimators", IEEE Trans. on SP, 56(11): 5319-5333, 2008

[J5] E. Chaumette, P. Larzabal, "Cramér-Rao Bound Conditioned by the Energy Detector", IEEE SP Letters, 14(7): 477-480, 2007

[J4] E. Chaumette, P. Larzabal, "Statistical prediction of monopulse angle measurement", IET Proc. RSN, 1(5): 377-387, 2007

[J3] A. Renaux, P. Forster, E. Chaumette, P. Larzabal, "On the High-SNR Conditional Maximum-Likelihood Estimator Full Statistical Characterization", IEEE Trans. on SP, 54(12): 4840-4843, 2006

[J2] E. Chaumette, P. Larzabal, P. Forster, "On the influence of a detection step on lower bounds for deterministic parameter estimation", IEEE Trans. on SP, 53(11): 4080-4090, 2005

[J1] E. Chaumette, P.; Comon, D. Muller, "ICA-based technique for radiating sources estimation: application to airport surveillance", IEE Proc. F, 140(6): 395-401, 1993

2) Revues nationales avec comité de lecture:

[JN1] N. Bertaux, P. Larzabal, C. Adnet, E. Chaumette, ”Maximum de vraisemblance paramétrée (PML) : application à la localisation spatio-temporelle radar”, Revue Traitement du Signal, 16(3): 187-201, 1999

3) Revues internationales avec comité de lecture soumises:

[J15] C. Ren, M. N. El Korso, J. Galy, E. Chaumette, P. Larzabal, A. Renaux, ”On the Accuracy and Resolvability of Vector Parameter Estimates”, soumis à IEEE Trans. on SP, re-soumis le 16/10/2013 ((R) Reject and resubmit as a regular paper).

[J16] E. Chaumette, F. Vincent, O. Besson, ”Second-order statistical prediction of L-statistics from multivariate Gaussian distribution”, soumis à IEEE Trans. on IT le 19/11/2013

4) Conférences internationales avec comité de lecture et actes:

[C29] F. Vincent, O. Besson, E. Chaumette, ’Approximate maximum likelihood direction of arrival estimation for two closely spaced sources’, in Proc. IEEE CAMSAP, 2013

[C28] C. Ren, J. Galy, E. Chaumette, P. Larzabal, A. Renaux,”High résolution techniques for radar: myth or reality ?”, in Proc. Eurasip EUSIPCO, 2013

[C27] C. Ren, J. Galy, E. Chaumette, P. Larzabal, A. Renaux, ”Hybrid lower bound on the mse based on the barankin and weiss-weinstein bounds”, in Proc. IEEE ICASSP, 2013

[C26] T. Menni, E. Chaumette, P. Larzabal, ”Reparameterization and constraints for CRB: duality and a major inequality for system analysis and design in the asymptotic region”, in Proc. IEEE ICASSP, 2012

[C25] U.R.O. Nickel, E. Chaumette, P. Larzabal, ”Monopulse Estimation of swerling I–II extended targets”, in Proc. IEEE Int. Conf. on Radar, 2011

[C24] T. Menni, E. Chaumette; P. Larzabal, J.P. Barbot, ”Crb for Active Radar”, in Proc. Eurasip EUSIPCO, 2011

[C23] L. Constancias, P. Brouard, E. Chaumette, A. Brun, S. Attia, ”HYCAM - a software-defined testbed for experimentations of new S band surface radar concepts”, in Proc. EuRAD, 2010

[C22] E. Chaumette, A. Renaux, P. Larzabal, ”New trends in deterministic lower bounds and SNR threshold estimation: From derivable bounds to conjectural bounds”, in Proc. IEEE SAM, 2010

[C21] U.R.O. Nickel, E. Chaumette, P. Larzabal, ”Characterization of the performance of generalized monopulse estimation”, in Proc. IEEE Int. Conf. on Radar, 2009

[C20] E. Chaumette, A. Renaux, P. Larzabal, ”Lower bounds on the mean square error derived from mixture of linear and non-linear transformations of the unbiasness definition”, in Proc. IEEE ICASSP, 2009

[C19] M. Tria; M. Benidir, E. Chaumette, ”Blind séparation of secondary radar signals using time-frequency analysis”, in Proc. Eurasip EUSIPCO, 2008

[C18] E. Chaumette, J. Galy, F. Vincent, A. Renaux, P. Larzabal, ”Mse lower bounds conditioned by the energy detector”, in Proc. Eurasip EUSIPCO, 2007

[C17] F. Vincent, B. Mouton, E. Chaumette, C. Nouals, O. Besson, ”Synthetic Aperture Radar Demonstration Kit for Signal Processing Education”, in Proc. IEEE ICASSP, 2007

[C16] E. Chaumette, J. Galy, F. Vincent, P. Larzabal, ”Computable Lower Bounds for Deterministic Parameter Estimation”, in Proc. IEEE CAMSAP, 2007

[C15] E. Chaumette, F. Vincent, J. Galy, P. Larzabal, ”On the influence of detection tests on deterministic parameters estimation”, in Proc. Eurasip EUSIPCO, 2006

[C14] A. Quinlan, E. Chaumette, P. Larzabal, ”A Direct Method to Generate Approximations of the Barankin Bound”, in Proc. IEEE ICASSP, 2006

[C13] E. Chaumette, P. Saulais, N. Colin, ”Modern Monopulse Tracking”, in Proc. Int. Conf. on Radar, 2004

[C12] E. Chaumette, P. Larzabal, ”Optimal monopulse tracking of signal source of unknown amplitude”, in Proc. Eurasip EUSIPCO, 2004

[C11] E. Chaumette, P. Larzabal, ”Optimal detection theory applied to monopulse antennas”, in Proc. IEEE ICASSP, 2004

[C10] J. Galy, C. Adnet, E. Chaumette, G. Gelle, ”Blind séparation of non-circular sources”, in Proc. IEEE SSAP, 2000

[C9] F. Vincent, E. Chaumette, C. Nouals, D. Muller, ”CRB for Space Situational Awareness Radar (Etude d'un système radar de détection des débris spatiaux)”, in Proc. Int. Conf. on Radar, 1999

[C8] E. Chaumette, P. Calvary, J.M. Ferrier, D. Muller, G. Desodt, ”Low RCS target detection (Détection des cibles de faibles ser)”, in Proc. Int. Conf. on Radar, 1999

[C7] N. Bertaux, P. Larzabal, C. Adnet, E. Chaumette, ”A parameterized maximum likelihood method for multipaths channels estimation”, in Proc. IEEE SPAWC, 1999

[C6] N. Bertaux, P. Larzabal, C. Adnet, E. Chaumette, ”Robust 2D estimation in presence of severe model errors”, in Proc. Int. Symp. PSIP, 1999

[C5] J. Galy, C. Adnet, E. Chaumette, ”Blind methods for interference cancellation in array processing”, in Proc. IEEE DSP, 1997

[C4] J. Galy, C. Adnet, E. Chaumette, ”Narrow band source séparation in wide band context. Applications to array signal processing”, in Proc. IEEE HOS, 1997

[C3] J. Galy, G.; Gelle, C. Adnet, E. Chaumette, ”Narrow Band Source Separation in Wide Band Context”, in Proc. IEEE SPA, 1997

[C2] E. Chaumette, P. Comon, D. Muller, ”ICA-based technique for radiating sources estimation : application to airport surveillance”, Iin Proc. Int. Conf. on Radar, 1994

[C1] E. Chaumette, P. Comon, D. Muller, ”Application of ICA to airport surveillance”, in Proc. IEEE HOS, 1993

5) Conférences nationales (GRETSI) avec comité de lecture et actes:

[CN5] C. Ren, J. Galy, E. Chaumette, P. Larzabal, A. Renaux, ”Une borne inférieure de l'erreur quadratique moyenne pour l'estimation simultanée de paramètres aléatoires et non-aléatoires”, in Proc. GRETSI, 2013

[CN4] N. Bertaux, P. Larzabal, C. Adnet, E. Chaumette, ”Généralisation du Filtre adapté au cas multi-cibles en présence de brouilleurs pour une localisation radar”, in Proc. GRETSI, 1999

[CN3] J. Galy, C. Adnet, E. Chaumette, G. Gelle, ”Séparation de sources non-circulaires”, in Proc. GRETSI, 1999

[CN2] J. Galy, C. Adnet, E. Chaumette, ”Séparation de sources bande étroite dans un contexte large bande en traitement d'antenne”, in Proc. GRETSI, 1997

[CN1] N. Bertaux, P. Larzabal, E. Chaumette, D.Muller, ”Localisation 2D à l'aide d'un réseau phasé très perturbé”, in Proc. GRETSI, 1997

6) Brevet international:

[B1] E. Chaumette, .G Desodt, D. Muller, ”Radar secondaire apte à séparer toutes réponses indépendantes”, EP0521750 A1, Février 1993

7) Rapports d'étude et de contrat:

[R12] Onera : ”Etude d'améliorations des radars contre les cibles a faible signature. Thème 2.1 : Amélioration des Traitements NCTR - La technique de super-résolution”, (61 pages + 26 pages d'annexe technique) PEA Nectar, DGA/UM-AERO, 2010.

[R11] Thales : ”Analyse théorique de l'antibrouillage OLS et des algorithmes de mesure angulaire à site bas,

compatibles avec le radar XXX”, (35 pages) étude interne, 2003

[R10] Thales : ”Radar Simulation Library : Modélisation électromagnétique de la scène Radar/Télécom”, (119 pages) rapport de synthèse interne, 2002

[R9] Thales : ”Analyse théorique des algorithmes de mesure angulaire à site bas compatibles avec le radar XXX”, (30 pages) étude interne, 2001

[R8] Thales : ”Etude et définition de traitements adaptatifs de réjection de fouillis polarisé”, (25 pages) PEA Ku-Polar, DGA/DRET.

[R7] Thales : ”Etude des méthodes de détection de cibles furtives”, (70 pages) étude DGA/DCN, 2000

[R6] Thales : ”Moyennes et variances analytiques du rapport d'écartometrie”, (25 pages) étude interne, 2000

[R5] Thales : ”Imagerie spatio-fréquentielle : écartometrie large bande antibrouillée du RIAS” (11 pages + 17 pages d'annexe technique), contribution au groupe de travail ”Traitements des sources large bande” du Collège Scientifique et Technique (CST) de Thomson, 1998

[R4] Thales : ”Filtrage adapte spatio-fréquentiel : antibrouillage du RIAS” (11 pages + 17 pages d'annexe technique), contribution au groupe de travail ”Traitements des sources large bande” du Collège Scientifique et Technique (CST) de Thomson, 1998

[R3] Thales : ”Etude des méthodes de calibration sans phare. Thème : Analyse des solutions.” (27 pages), PEA Autocalibration, DGA/DRET, 1996

[R2] Thales : ”Etude et réalisation d'un dispositif d'antibrouillage adapté au démonstrateur RIAS”, (292 pages + 64 pages d'annexes) marché DGA-STTE, 1994

[R1] Thales : ”Analyse en composantes indépendantes”, (56 pages + 78 pages d'annexes techniques) étude interne, 1992

8) Polycopié pour l'enseignement:

[P1] Introduction à la Transformée de Fourier et à ses applications, IFIPS Formation Continue (106 pages)

E. Gestion de contrats de recherche

Une première phase professionnelle entièrement consacrée à une activité d'ingénieur de recherche et développement radar dans une TBU (Technical Business Unit) me permit d'acquérir une double expertise technique en traitement du signal radar (valorisée par une nomination au Collège Scientifique et Technique de THOMSON-CSF en 1996) et en simulation de la scène électromagnétique. Ayant pu évaluer cette double compétence lors de la sous-traitance d'une affaire [R7], TNF me proposa une mutation interne en 2000 afin de renforcer son expertise technique. En effet TNF était une SBU (Strategical Business Unit), c'est à dire une division orientée vers la gestion des contrats industriels, notamment ceux en rapport avec la fourniture de radar navals, et qui sous-traitait la quasi-totalité des études (théoriques et simulations) du thème naval aux TBU compétentes. Ma mission était donc de développer une activité de support technique à destination du Service Programme à la fois pour la spécification du front-end des radar et le traitement du signal. Pour développer cette activité, il m'incombait de participer à la recherche de financements internes et externes sous la forme de proposition d'études ou de réponses à appels d'offre.
J'ai ainsi participé à la sollicitation des Ministères de la Défense anglais, hollandais, français (présentation en anglais) pour la proposition de contrats internationaux pluriannuels (programme Eurofinder) et rédigé les clauses techniques (en anglais) de la part TNF de ces programmes (WAGNER, CAESAR, NEMESIS). J'étais également en charge de la proposition (sollicitation et rédactions des clauses techniques) d'études internes à destination des programmes TNF en cours (programme MRR SAN PC pour l'Afrique du Sud, programme Herakles pour Singapour, ...) ou futur (Coast Watcher 100).
J'ai poursuivi cette activité de recherche de contrats à l'ONERA que ce soit à destination de la DGA (PEA ASTRAD Phase 2, PEA DEMPERE (en cours de négociation)), de l'ADEME (SIPRE) ou de l'ESA (simulateur radar SSA). Ma tâche dans cette démarche est toujours relative à l'expertise technique nécessaire à la formulation du besoin

(aide à la rédaction de CCTP pour la DGA) où à la réponse à l'appel d'offre (incluant la cotation (temps/euros) des tâches à réaliser).
J'ai également occupé le rôle de chef d'affaires (1 à 3 ingénieurs à coordonner) pour certains contrats techniques relevant entièrement de mon expertise (MRR SAN PC (TNF), Coast Watcher 100 (TNF), simulateur radar SSA pour l'ESA (ONERA)).

F. Encadrement scientifique

1) Encadrement de thèses:

[T6] ChengFang Ren, "Bornes inférieures de l'erreur quadratique moyenne : théorie et applications en traitement du signal", débutée en Octobre 2012 à l'université Paris-Sud (J. Galy 35%, E. Chaumette 35%, A. Renaux (directeur) 30%)
Cette thèse a donné lieu aux publications [C27] [C28] [CN5] et à la soumission [J15].

[T5] Tarek Menni, "Borne de Cramér-Rao déterministe pour l'analyse des performances asymptotiques en estimation d'un radar actif", soutenue à l'ENS Cachan en Septembre 2012 (E. Chaumette 70%, P. Larzabal (directeur) 30%).
Cette thèse a donné lieu aux publications [J10] [J13] [C26] [C24].
Tarek Menni (militaire de carrière) a été affecté à un poste d'enseignant-chercheur au Centre de Recherche et de Développement (établissement militaire) d'Alger.

[T4] Angela Quinlan, "Model Order Determination and Characterisation of Direction of Arrival (DOA) Estimators in the Acoustic Context", defended at Trinity College Dublin, September 2006 (F. Boland (directeur) 20%, E. Chaumette 30%, J.P. Barbot 30%, P. Larzabal 20%).
La partie que j'ai encadrée concernait les bornes inférieures pour l'estimation déterministe et a donné lieu aux publications [J6] [C14].
Angela Quinlan est maintenant "Associate" chez Carpmael&Ransford, Intellectual Property Management, London.

[T3] Nicolas Bertaux, "Contribution à l'utilisation des méthodes du maximum de vraisemblance en traitement radar actif", soutenue à l'ENS Cachan en Janvier 2000 (E. Chaumette 35%, C. Adnet 35%, P. Larzabal (directeur) 30%).
Cette thèse a donné lieu aux publications [JN1] [CN1] [CN4] [C6] [C7].
Nicolas Bertaux est actuellement enseignant-chercheur (maître de conférence, HDR) a l'Institut Fresnel de Marseille.

[T2] François Vincent, "Etude d'un système radar de détection des débris spatiaux et d'estimation de leur orbite", soutenue à l'Université Paul Sabatier de Toulouse en Novembre 1999 (J. Paillé (directeur) 30%, O. Besson 35%, E. Chaumette 35%).
La partie que j'ai encadrée concernait le calcul de la Borne de Cramér-Rao pour les paramètres orbitaux et a donné lieu à la publication [C9].
François Vincent est actuellement enseignant-chercheur (HDR) à l'ISAE de Toulouse.

[T1] Jérôme Galy, "Antenne adaptative : du second ordre aux ordres supérieurs. Application aux signaux de télécommunications", soutenue à l'Université Paul Sabatier de Toulouse en Octobre 1998 (J. Paillé (directeur) 30%, E. Chaumette 35%, C. Adnet 35%).
Cette thèse a donné lieu aux publications [C3] [C4] [C5] [CN2] [CN3].
Jérôme Galy est actuellement enseignant-chercheur (maître de conférence, HDR en préparation) à l'IUT de Béziers.

2) Encadrement de stages DEA/MASTER de 6 mois:

[S4] Alphonse Bytha, "Etude de la Borne de Todros-Tabrikian", Master 2 mathématiques - option statistiques / Université Paris VI en 2009 (E. Chaumette 100%)

[S3] Ludovic Fontaine, "Applications des algorithmes d'aci au contexte radar", DEA Signal Image Parole / ENSIEG en 1993 (E. Chaumette 100%)

[S2] Frédéric Solera, ”Radar secondaire et algorithmes de séparation des sources bases sur les moments d'ordre supérieur: une nouvelle approche de la résolution du garbling”, DEA Propagation Télédétection Télécommunications / Université De Nice en 1992 (E. Chaumette 100%)

[S1] Christophe Saglio, ”Les moments d'ordre supérieur appliqués à la séparation de sources”, DEA Signal Image Parole / ENSIEG en 1992 (E. Chaumette 100%)

G. Jury de thèse

Marc OUDIN, Thèse de l'Université Paris 6. Thèse soutenue le 1er février 2008 :

ETUDE D'ALGORITHMES DE TRAITEMENT D'ANTENNE SUR SIGNAUX LARGE BANDE ET SIGNAUX RADAR BANDE ETROITE A ANTENNE TOURNANTE

Composition du jury :

Rapporteurs	Sylvie MARCOS	LSS
	Nikolaos LIMNIOS	UTC
Directeurs de thèse	Paul DEHEUVELS	Université Paris 6
	Jean Pierre DELMAS	Telecom Sud Paris
Examinateurs	François LE CHEVALIER	THALES Airborne Systems
	Eric CHAUMETTE	ONERA
Invité, co-encadrant	Frédéric BARBARESCO	THALES Air Systems

H. Activités d'enseignement

Enseignements (total : 476h)

2009 - 2010	Chargé du cours ”Mathématiques du Signal : Introduction à la Transformée de Fourier et à ses applications” à l'IFIPS, formation continue (28h/an).
2000 - 2004	Chargé du cours de Mathématiques-Probabilités en maîtrise GEII de l'IUP de Cachan (30h/an)
1994 - 2002	Chargé de TD de Probabilités-Processus aléatoires, cycle ingénieur de l'ISEP (30h/an)

Séminaires invités

Nov-2011	Kuang Chi Institute (Shenzen, China)	MSE Lower Bounds for Deterministic Parameter Estimation
Nov-2011	Kuang Chi Institute (Shenzen, China)	Monopulse Angle Measurement
Juin-2010	Séminaires Farman de l'Ens Cachan	Bornes Inférieures de l'EQM pour l'estimation de paramètres déterministes
Fév-2010	Séminaires SONDRA (Supélec)	MSE Lower Bounds for Deterministic Parameter Estimation
Jan-2010	LSTA Paris VI	Bornes Inférieures de l'EQM pour l'estimation de paramètres déterministes

I. Administration et responsabilité collective

- Co-organisateur de la session spéciale ”High résolution techniques for radar : myth or reality?” à la conférence EUSIPCO 2013.
- Reviewer régulier pour les revues : IEEE Transactions on Signal Processing (21), IEEE Signal Processing Letters (1), IEEE Transactions on Aerospace and Electronic Systems (14), Elsevier Signal Processing (8), Springer Signal Image and Video Processing (4).

- Participations aux réseaux d'excellence européen NEWCOM, NEWCOM++, NEWCOM#. Ces réseaux ont pour objectifs de faciliter les relations entre les chercheurs européens sur la thématique générale des systèmes de communications radio-cellulaire. Ils font intervenir un grand nombre de sous thèmes (Work Package WP) tel que l'analyse paramétrique d'un canal de propagation ou encore les études de performances des estimateurs des paramètres. Le réseau NEWCOM# fait suite aux réseaux NEWCOM et NEWCOM++ dont SATIE était l'un des participants et auquel j'ai contribué en tant que chercheur associé [J13] [J12] [J11] [J10] [J8] [J7] [J6] [C27] [C28] [C26] [C24] [C22] [C20].
- Expert technique auprès du Collège Scientifique et Technique (CST) de THOMSON-CSF en Traitement du Signal (1996 - 1999) [R4] [R5]

II. Synthèse des activités de recherche

En 23 années d'activité professionnelle en tant qu'ingénieur (en recherche et développement radar) passionné par l'ensemble de la problématique modélisation, traitement et analyse de performance, j'ai capitalisé une double expertise scientifique en traitement du signal radar[1] et en simulation de la scène électromagnétique, concrétisée à la fois par la participation à de nombreuses études techniques internes ou DGA et la conception sur contrats de bibliothèques logicielles de référence dédiées à la simulation du "Front-End" de la scène radar. Cette double expertise inclus les thèmes suivants (cf. §I-C-p6) :

• probabilité de détection, régulation de fausse alarme [R7] [R8].

• développement d'algorithmes et étude de performance pour l'estimation paramétrique déterministe : bornes d'estimation [R12], estimateurs du maximum de vraisemblance [R2] [R3] [R5] [R12] (cf. IV-B-p66), haute précision [R5] [R6] [R9] [R11], haute résolution [R12], track before detect [R7], antibrouillage [R2] [R8] [R11], calibration/autocalibration [R3], séparation de sources aux ordres supérieurs [R1].

• modélisation émission-réception (modulateur, filtre analogique/numérique, échantillonneur, limiteur, amplificateur,. . .), antenne (gain/polarisation), propagation en champs lointain, matrice de rétrodiffusion, matrice de réflexion [R10] (contrats Eurofinder (WAGNER, CAESAR), contrat DGA (ASTRAD), contrat ADEME (SIPRE), contrat ESA (SSA),...).

J'ai par conséquent acquis une bonne compréhension des éléments (interférences, imperfections, largeur de bande, forme d'onde, propagation, diagrammes d'antenne ou de réseau, cinématiques, ...) impactant les performances des problèmes conjoints de détection et d'estimation en traitement du signal radar.

Mon intérêt scientifique pour ce thème de recherche portant sur la "caractérisation des problèmes conjoints de détection et d'estimation" s'est donc initialement (jusqu'en 2000) construit au fil des études théoriques conduites sur contrats ou lors de co-encadrements de thèse (**[T1] [T2] [T3]**) dédiées à l'étude des performances en traitement du signal pour le radar actif où l'approche à paramètres déterministes est privilégiée.

J'ai abordé ce thème par le problème de l'estimation des paramètres déterministes lors d'études concomitantes relatives à la séparation aveugle de sources (analyse en composantes indépendantes) [R1] [S1] [S2] [S3] et à l'estimateur du maximum de vraisemblance (EMV) des paramètres d'une source pour un modèle d'observation déterministe en présence d'interférences (brouilleurs) large bande [R2]. Ces travaux furent complétés par deux thèses que j'ai co-encadrées : **[T1]** et **[T3]**.

La thèse de Jérôme Galy **[T1]** me confirma (présomption soulevée par les stages préalables [S1] [S2] [S3]) que la séparation aveugle de sources avait peu d'avenir en radar car les signaux radar émis n'ont pas les propriétés statistiques requises (indépendance).

La thèse de Nicolas Bertaux **[T3]**, portant sur la formulation et les performances de l'EMV pour des scénarios à sources multiples et des modèles d'observation multiples indépendants, permit de me sensibiliser à la problématique générale de l'estimation des paramètres déterministes pour un modèle d'observation non linéaire : les 3 zones de fonctionnement des estimateurs, l'intérêt de la BCR dans la zone asymptotique comme estimateur de l'EQM, la signification intrinsèque d'une borne inférieure de l'EQM en terme de performances asymptotiques pour l'aide à la conception et au dimensionnement des systèmes (notamment radar).

Ce qui me surprit à l'époque (1996-1999) et qui fut un thème de discussion récurrente avec P. Larzabal, fut le manque de résultat relativement à l'évaluation de l'EQM en dehors de la zone asymptotique (hormis dans le cas d'une cisoïde [RB74]). Une de mes premières conjectures fut d'ailleurs de penser que l'introduction d'un test de détection prolongerait la zone asymptotique et la prédictibilité de l'EQM par la BCR pour le modèle déterministe : une conjecture confirmée pour le modèle stochastique et une application spécifique (écartométrie monopulse, cf. §II-B1-p55) mais infirmée pour le modèle déterministe général et la BCR sans biais (cf. §II-B2-p59)[2].

Puis l'étude des performances de la mesure angulaire par écartométrie [R9] me révéla l'influence d'un test de détection préalable (test d'hypothèses binaires en radar (1)) sur les performances en estimation.

[1]En radar, la partie traitement du signal s'arrête généralement à la construction des détections élémentaires (présences échos ("hits")), c'est à dire en entrée de l'extracteur (fusion des détections élémentaires pour la construction des plots) et en amont du pistage ("tracking", poursuite et mise à jour des plots).

[2]Comme le chantent si bien les Rolling Stones "You can't always get what you want".

La prise de conscience de l'universalité de ce thème de recherche (la quasi totalité des dispositifs de mesure ou de transmission conçus relevant de ce thème) et de sa complexité, tant théorique que calculatoire, en a fait pour moi un sujet inépuisable d'investigations et de questionnements auquel je consacre ma recherche personnelle depuis 2000. En effet, désireux de vouloir approfondir mes connaissances sur ce thème dans une perspective à long terme découplée de la versatilité des études contractuelles, j'effectue, depuis le début de ma thèse, cette recherche en parallèle à mon activité professionnelle.

D'un point de vue formel, le thème général de ma recherche est donc la caractérisation des problèmes conjoints détection-estimation les plus fréquemment rencontrés en écoute active ou passive (radar, télécoms, sonar, ...) : l'estimation des paramètres déterministes (non aléatoires) d'un signal d'intérêt intermittent en présence d'un environnement permanent. Ce problème peut être modélisé par le test d'hypothèses binaires suivant :

$$\begin{array}{rcl} H_0 \text{ (environnement seul)} & : & \mathbf{x} = \mathbf{n}\left(\boldsymbol{\theta}_{\mathbf{n}}\right) \\ H_1 \text{ (environnement et signal)} & : & \mathbf{x} = \mathbf{n}\left(\boldsymbol{\theta}_{\mathbf{n}}\right) + \mathbf{s}\left(\boldsymbol{\theta}_{\mathbf{s}}\right) \end{array} \quad (1)$$

où $\boldsymbol{\theta}_{\mathbf{n}}$ et $\boldsymbol{\theta}_{\mathbf{s}}$ sont respectivement les paramètres déterministes inconnus de l'environnement et du signal d'intérêt, $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_{\mathbf{n}}^T, \boldsymbol{\theta}_{\mathbf{s}}^T\right)^T$ représentant l'ensemble des paramètres inconnus des problèmes conjoints. Je me restreins au cas de l'estimation paramétrique, c'est à dire que le vecteur d'observation $\mathbf{x}$ peut être modélisé par une fonction mathématique de forme connue dépendant de paramètres inconnus au moment de l'observation. Le modèle d'observation est basé sur les connaissances de l'utilisateur concernant le processus physique considéré [VT68]. Il dépend généralement d'une fonction non-linéaire des paramètres et incorpore un modèle statistique (mécanisme de transition probabiliste [VT68]). On remarquera que le modèle (1) ne préjuge pas de la forme du signal d'intérêt, lequel peut donc être mono-source ou multi-sources.

L'intérêt premier pour l'estimation des paramètres déterministes provient de mon domaine d'application professionnel : le radar actif où l'approche à paramètres déterministe est privilégiée. En effet, l'idée sous-jacente est que cette approche est la plus réaliste[3] pour l'analyse d'une scène radar où les paramètres des cibles dépendent de façon déterministe du radar (forme d'onde émise) et du scénario considéré (cinématique des cibles, amplitudes rétrodiffusées par les cibles, ...).

Dans le cadre de paramètres déterministes, cette recherche peut s'aborder graduellement (en terme de difficulté théorique et calculatoire) sous deux axes :

• l'étude des performances en estimation non conditionnelle (sans test de détection préalable) par le biais des bornes de performance en estimation. Dans ce cas il n'y a qu'un seul modèle d'observation H_1 :

$$H_1 \text{ (environnement et signal) : } \mathbf{x} = \mathbf{n}\left(\boldsymbol{\theta}_{\mathbf{n}}\right) + \mathbf{s}\left(\boldsymbol{\theta}_{\mathbf{s}}\right)$$

• l'étude des performances en estimation conditionnelle (avec test de détection préalable), c'est à dire la caractérisation des problèmes conjoints détection-estimation. Cette étude est réalisée par le biais des bornes de performance conditionnelle appliquées à deux modèles d'observation particuliers (l'antenne monopulse et le modèle d'observation déterministe) pour lesquels certains calculs analytiques sont accessibles. Dans ce cas il y a deux modèles d'observation H_0 et H_1 :

$$\begin{array}{rcl} H_0 \text{ (environnement seul)} & : & \mathbf{x} = \mathbf{n}\left(\boldsymbol{\theta}_{\mathbf{n}}\right) \\ H_1 \text{ (environnement et signal)} & : & \mathbf{x} = \mathbf{n}\left(\boldsymbol{\theta}_{\mathbf{n}}\right) + \mathbf{s}\left(\boldsymbol{\theta}_{\mathbf{s}}\right) \end{array}$$

Mon intégration au laboratoire SATIE (thésard puis chercheur associé), unité mixte de recherche CNRS (UMR 8029), m'a permis de me sensibiliser aux règles d'évaluation des chercheurs, que ce soit dans l'enseignement supérieur ou au CNRS, notamment en ce qui concerne les publications.
Selon mon expérience professionnelle, les ingénieurs de recherche et développement ont tendance à privilégier les publications en conférence, notamment dans les conférences liées à leur domaine professionnel. La raison est simple : le temps d'écriture et de révision d'un papier de revue n'est en général pas provisionné dans un contrat.

[3]Ou la moins arbitraire, relativement à la prise en compte de lois a priori requises pour l'approche bayésienne.

La prise en compte des règles d'évaluation (CNU, CNRS, AERES) est devenu pour moi un paramètre fondamental dans la calibration d'un sujet de recherche pour une thèse. Pour ma part, il relève de la responsabilité du directeur de recherche de s'assurer que le sujet proposé permettra (avec une probabilité assez grande) au moins la soumission d'un article de conférence internationale (référencée dans le domaine) accompagnée d'un article de revue internationale (référencée dans le domaine) afin de permettre au futur diplômé de pouvoir postuler à un post-doc ou être qualifié par le CNU.
La capacité à évaluer un sujet de recherche en ces termes requiert une certaine vision du domaine de recherche : la littérature déjà publiée, les problèmes ouverts, les perspectives de recherche.
Si l'on part du principe qu'une thèse est une "danse" à trois temps : l'apprentissage, la compréhension puis l'innovation, sur un tempo de 2.5 années (6 derniers mois pour la rédaction du manuscrit), il reste peu de place pour une recherche purement spéculative sans que certaines pistes ne soient déjà balisées.
Par conséquent, parmi tous les problèmes ouverts et les perspectives que j'évoque (cf. §II-A5-p40,41,44,48,52 et §II-B3-62,63), certains pourront faire l'objet d'un travail de thèse, d'autres d'un sujet de post-doc ou d'une collaboration entre chercheurs confirmés (et en poste).

L'objectif des 2 sections suivantes est donc d'exposer de façon synthétique (sans aucune prétention d'exhaustivité) ma vision de ce thème de recherche : mes acquis, mes questionnements (problèmes ouverts), mes axes futurs d'investigation et de collaboration (conjectures et perspectives). Vision qui n'aurait été possible sans la dynamique de la recherche collaborative et partagée, que ce soit avec des thésard(e)s **(**repérée par **[T*])** ou des chercheurs confirmés.

A. *Bornes de performance en estimation paramétrique déterministe*

1) *Analyse du problème:*

Dans un souci de simplicité, nous considérons essentiellement le cadre de l'estimation paramétrique de vecteurs de paramètres réels. L'extension aux vecteurs de paramètres mixtes, c.-à-d. à composantes réelles ou complexes n'est abordée en détail que pour l'étude de la borne de Cramér-Rao §II-A6-p44 **[T5]** (pour la formulation des autres bornes se référer à [Chau04, § III-M, III-N]).
Dans ce cadre, les observations disponibles sont des vecteurs aléatoires $\mathbf{x} = (x_1, \ldots, x_N)^T \in \mathbb{C}^N$ constituant un sous-ensemble Ω de $\mathbb{C}^N$ appelé espace d'observation ou univers de l'expérience aléatoire. Cet univers Ω est caractérisé par une densité de probabilité conditionnelle $p(\mathbf{x} \mid \boldsymbol{\theta})$ - notée également $p(\mathbf{x}; \boldsymbol{\theta})$ - paramétrée par un vecteur $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_P)^T$ de P valeurs réelles inconnues au moment de l'expérience et appartenant à un sous-ensemble $\Theta \subset \mathbb{R}^P$ appelé espace des paramètres. Dans la suite, $\boldsymbol{\theta}^i, i \in \mathbb{N}$, désigne une valeur particulière de $\boldsymbol{\theta}$.
Nous qualifions de "**clairvoyante**" toute grandeur dont la définition dépend du vecteur de paramètres $\boldsymbol{\theta}$ et nous qualifions de "**réalisable**" toute grandeur dont la définition ne dépend que des observations $\mathbf{x}$ et de constantes indépendantes du vecteur de paramètres $\boldsymbol{\theta}$ [Kay93].
Notre problème est le suivant : nous désirons, à partir des seules observations $\mathbf{x}$, déterminer la valeur d'un vecteur de Q fonctions des paramètres inconnus $\boldsymbol{\theta}$: $\mathbf{g}(\boldsymbol{\theta}) = (g_1(\boldsymbol{\theta}), \ldots, g_Q(\boldsymbol{\theta}))^T \in \mathbb{R}^Q$ en la valeur particulière $\boldsymbol{\theta}^0$ associée aux observations en cours. Il s'agit donc de déterminer une fonction réalisable $\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}(\mathbf{x}) = \left(\widehat{g_1(\boldsymbol{\theta}^0)}(\mathbf{x}), \ldots, \widehat{g_Q(\boldsymbol{\theta}^0)}(\mathbf{x})\right)^T$ de Ω dans $\mathbb{R}^Q$ - notée également $\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}$ - appelée "estimateur" de $\mathbf{g}(\boldsymbol{\theta}^0)$ dont le résultat, pour chaque valeur de $\mathbf{x}$, constituera une valeur estimée de $\mathbf{g}(\boldsymbol{\theta}^0)$.
Puisque $\mathbf{x}$ est un vecteur aléatoire, toute fonction de Ω dans $\mathbb{R}^Q$, notamment $\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}$, non réduite à une constante constitue un vecteur aléatoire. Or la qualité d'estimation de $\mathbf{g}(\boldsymbol{\theta}^0)$ par $\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}$ ne peut être en général arbitraire et dépend étroitement de l'application considérée et de la grandeur recherchée, ce qui pose le problème de sa mesure.

2) *Qualité locale d'un estimateur. Mesure et bornes:*

Une mesure de précision d'estimation locale (c.-à-d. pour une valeur particulière $\boldsymbol{\theta}^0$ de $\boldsymbol{\theta}$) *exhaustive* (précision *exhaustive)* est obtenue naturellement par la probabilité [REKGCLR13] **[T6]** :

$$\mathcal{O}_{\boldsymbol{\theta}^0}\left(\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}, \boldsymbol{\xi}^-, \boldsymbol{\xi}^+\right) = \mathcal{P}\left(\bigcap_{q=1}^{Q}\left(\widehat{g_q(\boldsymbol{\theta}^0)}(\mathbf{x}) \in \left]g_q(\boldsymbol{\theta}^0) - \xi_q^-, g_q(\boldsymbol{\theta}^0) + \xi_q^+\right[\right); \boldsymbol{\theta}^0\right) \tag{2}$$

où $\mathcal{P}(\mathcal{D}; \boldsymbol{\theta}^0)$ désigne la probabilité de l'évènement $\mathcal{D} \subset \Omega$ lorsque $\boldsymbol{\theta}$ prend la valeur particulière $\boldsymbol{\theta}^0$, $\boldsymbol{\xi}^- = \left(\xi_1^-, \ldots, \xi_Q^-\right)^T$ et $\boldsymbol{\xi}^+ = \left(\xi_1^+, \ldots, \xi_Q^+\right)^T$ désignent les erreurs d'estimation admissibles à gauche (par valeur inférieure) et à droite (par valeur supérieure)[4] pour chacune des composantes de $\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}$.
Cependant d'un point de vue de sa mise en oeuvre, il est préférable de reformuler (2) sous la forme "symétrisée":

$$\mathcal{O}_{\boldsymbol{\theta}^0}\left(\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}, \boldsymbol{\xi}^-, \boldsymbol{\xi}^+\right) = \mathcal{P}\left(\bigcap_{q=1}^{Q}\left(\left|\widehat{g_q(\boldsymbol{\theta}^0)}(\mathbf{x}) - g_q(\boldsymbol{\theta}^0) - \frac{\xi_q^+ - \xi_q^-}{2}\right| < \frac{\xi_q^+ + \xi_q^-}{2}\right); \boldsymbol{\theta}^0\right) \tag{3}$$

permettant de faire facilement apparaître les bornes (inférieure et supérieure) suivantes, $\forall s > 1$ [REKGCLR13] **[T6]** :

$$\begin{aligned}\mathcal{P}\left(\left\|\mathbf{D}^{-1}_{\frac{\boldsymbol{\xi}^+ + \boldsymbol{\xi}^-}{2}}\left(\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}(\mathbf{x}) - \mathbf{g}(\boldsymbol{\theta}^0) - \frac{\boldsymbol{\xi}^+ - \boldsymbol{\xi}^-}{2}\right)\right\|_s^s < 1; \boldsymbol{\theta}^0\right) &\leq \mathcal{O}_{\boldsymbol{\theta}^0}\left(\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}, \boldsymbol{\xi}^-, \boldsymbol{\xi}^+\right) \\ &\leq \mathcal{P}\left(\left\|\mathbf{D}^{-1}_{\frac{\boldsymbol{\xi}^+ + \boldsymbol{\xi}^-}{2}}\left(\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}(\mathbf{x}) - \mathbf{g}(\boldsymbol{\theta}^0) - \frac{\boldsymbol{\xi}^+ - \boldsymbol{\xi}^-}{2}\right)\right\|_s^s < Q; \boldsymbol{\theta}^0\right)\end{aligned} \tag{4}$$

[4] $\boldsymbol{\xi}^- \neq \boldsymbol{\xi}^+$ intègre le fait que les d.d.p. de chaque $\widehat{g_q(\boldsymbol{\theta}^0)}(\mathbf{x})$ n'ont aucune raison d'être symétriques autour de $g_q(\boldsymbol{\theta}^0)$ dans le cas général. Par conséquent, la définition ou la recherche de la meilleure précision peut aboutir à des intervalles d'erreur non symétriques.

où $\mathbf{D}_{\boldsymbol{\alpha}} \triangleq \mathbf{diag}(\boldsymbol{\alpha})$ et :

$$\|\mathbf{f}(\mathbf{x})\|_s^u = \left(\|\mathbf{f}(\mathbf{x})\|_s\right)^u, \quad \|\mathbf{f}(\mathbf{x})\|_s = \left(\sum_{q=1}^{Q} |f_q(\mathbf{x})|^s\right)^{\frac{1}{s}}. \tag{5}$$

Dans le cas particulier où $s = 2$ et d'un estimateur gaussien : $\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}(\mathbf{x}) - \mathbf{g}(\boldsymbol{\theta}^0) \sim \mathcal{N}\left(\mathbf{b}(\boldsymbol{\theta}^0), \mathbf{C}(\boldsymbol{\theta}^0)\right)$, alors (4) s'écrit également :

$$\mathcal{P}\left(e\chi_Q^2\left(\delta(\boldsymbol{\theta}^0), \boldsymbol{\sigma}^2(\boldsymbol{\theta}^0)\right) < 1; \boldsymbol{\theta}^0\right) \leq \mathcal{O}_{\boldsymbol{\theta}^0}\left(\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}, \boldsymbol{\xi}^-, \boldsymbol{\xi}^+\right) \leq \mathcal{P}\left(e\chi_Q^2\left(\delta(\boldsymbol{\theta}^0), \boldsymbol{\sigma}^2(\boldsymbol{\theta}^0)\right) < Q; \boldsymbol{\theta}^0\right) \tag{6}$$

$$\left.\begin{array}{r} \delta(\boldsymbol{\theta}^0) = \left\|\mathbf{U}^T(\boldsymbol{\theta}^0)\mathbf{D}^{-1}_{\frac{\boldsymbol{\xi}^+ + \boldsymbol{\xi}^-}{2}}\left(\mathbf{b}(\boldsymbol{\theta}^0) - \frac{\boldsymbol{\xi}^+ - \boldsymbol{\xi}^-}{2}\right)\right\|_2^2 \\ \mathbf{U}(\boldsymbol{\theta}^0)\mathbf{D}_{\boldsymbol{\sigma}^2(\boldsymbol{\theta}^0)}\mathbf{U}^T(\boldsymbol{\theta}^0) = \mathbf{D}^{-1}_{\frac{\boldsymbol{\xi}^+ + \boldsymbol{\xi}^-}{2}}\mathbf{C}(\boldsymbol{\theta}^0)\mathbf{D}^{-1}_{\frac{\boldsymbol{\xi}^+ + \boldsymbol{\xi}^-}{2}} \end{array}\right|, \ \mathbf{U}(\boldsymbol{\theta}^0)\mathbf{U}^T(\boldsymbol{\theta}^0) = \mathbf{U}^T(\boldsymbol{\theta}^0)\mathbf{U}(\boldsymbol{\theta}^0) = \mathbf{I}$$

où $e\chi_Q^2\left(\delta, \boldsymbol{\sigma}^2\right)$ désigne une forme quadratique décentrée [RP49] [Jam64], c.-à-d. une généralisation d'une loi du chi-2 décentrée à Q degrés de liberté et de paramètre δ :

$$e\chi_Q^2\left(\delta = \sum_{q=1}^{Q}\delta_q, \boldsymbol{\sigma}^2\right) \sim \sum_{q=1}^{Q}\sigma_q^2\left|z_q \pm \sqrt{\delta_q}\right|^2, \quad \boldsymbol{\sigma}^2 = \left(\sigma_1^2, \dots, \sigma_Q^2\right)^T, \quad \mathbf{z} = (z_1, \dots, z_Q)^T \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{7}$$

dont le calcul de la fonction de répartition peut être obtenue par calcul numérique [RP49].
Dans le cas général, $\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}$ n'est pas gaussien et $\mathcal{O}_{\boldsymbol{\theta}^0}\left(\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}, \boldsymbol{\xi}^-, \boldsymbol{\xi}^+\right)$ ou ses bornes (4) sont rarement accessibles analytiquement. Dans la perspective de calculs analytiques plus accessibles, la borne inférieure alternative suivante est préférée pour $s > 1$ (cf. §IV-A-p66) [REKGCLR13] **[T6]** :

$$1 - E_{\boldsymbol{\theta}^0}\left[\left\|\mathbf{D}^{-1}_{\frac{\boldsymbol{\xi}^+ + \boldsymbol{\xi}^-}{2}}\left(\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}(\mathbf{x}) - \mathbf{g}(\boldsymbol{\theta}^0) - \frac{\boldsymbol{\xi}^+ - \boldsymbol{\xi}^-}{2}\right)\right\|_s^s\right] \leq$$
$$\mathcal{P}\left(\left\|\mathbf{D}^{-1}_{\frac{\boldsymbol{\xi}^+ + \boldsymbol{\xi}^-}{2}}\left(\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}(\mathbf{x}) - \mathbf{g}(\boldsymbol{\theta}^0) - \frac{\boldsymbol{\xi}^+ - \boldsymbol{\xi}^-}{2}\right)\right\|_s^s < 1; \boldsymbol{\theta}^0\right) \leq \mathcal{O}_{\boldsymbol{\theta}^0}\left(\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}, \boldsymbol{\xi}^-, \boldsymbol{\xi}^+\right) \leq 1 \tag{8}$$

borne inférieure informative uniquement si $E_{\boldsymbol{\theta}^0}\left[\left\|\mathbf{D}^{-1}_{\frac{\boldsymbol{\xi}^+ + \boldsymbol{\xi}^-}{2}}\left(\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}(\mathbf{x}) - \mathbf{g}(\boldsymbol{\theta}^0) - \frac{\boldsymbol{\xi}^+ - \boldsymbol{\xi}^-}{2}\right)\right\|_s^s\right] < 1$ où $E_{\boldsymbol{\theta}}[f(\mathbf{x})] = \int_{\Omega} f(\mathbf{x})\, p(\mathbf{x}; \boldsymbol{\theta})\, d\mathbf{x}$.
En effet (8) s'écrit également :

$$1 - \left\|\mathbf{D}^{-1}_{\frac{\boldsymbol{\xi}^+ + \boldsymbol{\xi}^-}{2}}\left(\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}(\mathbf{x}) - \mathbf{g}(\boldsymbol{\theta}^0) - \frac{\boldsymbol{\xi}^+ - \boldsymbol{\xi}^-}{2}\right)\right\|_{s;\boldsymbol{\theta}^0}^s \leq \mathcal{O}_{\boldsymbol{\theta}^0}\left(\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}, \boldsymbol{\xi}^-, \boldsymbol{\xi}^+\right) \leq 1, \quad s > 1 \tag{9}$$

$$\|\mathbf{f}(\mathbf{x})\|_{s;\boldsymbol{\theta}}^u = \left(\|\mathbf{f}(\mathbf{x})\|_{s;\boldsymbol{\theta}}\right)^u, \quad \|\mathbf{f}(\mathbf{x})\|_{s;\boldsymbol{\theta}} = E_{\boldsymbol{\theta}}\left[\|\mathbf{f}(\mathbf{x})\|_s^s\right]^{\frac{1}{s}} = E_{\boldsymbol{\theta}}\left[\sum_{q=1}^{Q}|f_q(\mathbf{x})|^s\right]^{\frac{1}{s}} \tag{10}$$

ce qui ramène la mesure de précision (du moins une borne inférieure) à un calcul de norme et permet de bénéficier de nombreux résultats connus en algèbre, notamment l'inégalité de Hölder généralisée :

$$\|\mathbf{f}(\mathbf{x}) \odot \mathbf{g}(\mathbf{x})\|_{1;\boldsymbol{\theta}} = E_{\boldsymbol{\theta}}\left[\sum_{q=1}^{Q}|f_q(\mathbf{x})\, g_q(\mathbf{x})|\right] \leq \|\mathbf{f}(\mathbf{x})\|_{s;\boldsymbol{\theta}}\, \|\mathbf{g}(\mathbf{x})\|_{r;\boldsymbol{\theta}}, \ \frac{1}{s} + \frac{1}{r} = 1, \ s > 1 \tag{11}$$

avec égalité ssi $\exists\alpha \in \mathbb{R}^* \ / \ \forall q \in [1, Q], \ f_q(\mathbf{x}) = \alpha sgn(g_q(\mathbf{x}))\, |g_q(\mathbf{x})|^{\frac{r}{s}}$.

La minoration (9) permet de définir une mesure de précision d'estimation locale *a priori* (précision *a priori*), en ce sens qu'elle fournit pour tout estimateur $\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}$ une borne inférieure de précision (pas très précise en général) qui n'est réellement informative que pour $\left\|\mathbf{D}_{\frac{\boldsymbol{\xi}^{+}+\boldsymbol{\xi}^{-}}{2}}^{-1}\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x})-\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)-\frac{\boldsymbol{\xi}^{+}-\boldsymbol{\xi}^{-}}{2}\right)\right\|_{s;\boldsymbol{\theta}^{0}}^{s} \ll 1$. En effet, la précision *a priori* ne permet pas en général de comparer directement la précision *exhaustive* entre deux estimateurs $\widehat{\mathbf{g}_1\left(\boldsymbol{\theta}^{0}\right)}$ et $\widehat{\mathbf{g}_2\left(\boldsymbol{\theta}^{0}\right)}$:

$$\left\|\mathbf{D}_{\frac{\boldsymbol{\xi}^{+}+\boldsymbol{\xi}^{-}}{2}}^{-1}\left(\widehat{\mathbf{g}_1\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x})-\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)-\frac{\boldsymbol{\xi}^{+}-\boldsymbol{\xi}^{-}}{2}\right)\right\|_{s;\boldsymbol{\theta}^{0}}^{s} \leq \left\|\mathbf{D}_{\frac{\boldsymbol{\xi}^{+}+\boldsymbol{\xi}^{-}}{2}}^{-1}\left(\widehat{\mathbf{g}_2\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x})-\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)-\frac{\boldsymbol{\xi}^{+}-\boldsymbol{\xi}^{-}}{2}\right)\right\|_{s;\boldsymbol{\theta}^{0}}^{s}$$
$$\nRightarrow \mathcal{O}_{\boldsymbol{\theta}^{0}}\left(\widehat{\mathbf{g}_2\left(\boldsymbol{\theta}^{0}\right)},\boldsymbol{\xi}^{-},\boldsymbol{\xi}^{+}\right) \leq \mathcal{O}_{\boldsymbol{\theta}^{0}}\left(\widehat{\mathbf{g}_1\left(\boldsymbol{\theta}^{0}\right)},\boldsymbol{\xi}^{-},\boldsymbol{\xi}^{+}\right)$$

Par conséquent, la recherche d'un estimateur le plus précis *a priori*, c'est à dire qui minimise $\left\|\mathbf{D}_{\frac{\boldsymbol{\xi}^{+}+\boldsymbol{\xi}^{-}}{2}}^{-1}\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x})-\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)-\frac{\boldsymbol{\xi}^{+}-\boldsymbol{\xi}^{-}}{2}\right)\right\|_{s;\boldsymbol{\theta}^{0}}^{s}$ ne garantit pas que l'estimateur obtenu soit le plus précis *exhaustivement* (cf. exemples §III-C, §III-D et §VII-A dans [Chau04]).
Néanmoins, la recherche de la valeur minimale de $\left\|\mathbf{D}_{\frac{\boldsymbol{\xi}^{+}+\boldsymbol{\xi}^{-}}{2}}^{-1}\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x})-\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)-\frac{\boldsymbol{\xi}^{+}-\boldsymbol{\xi}^{-}}{2}\right)\right\|_{s;\boldsymbol{\theta}^{0}}^{s}$ permet d'établir une borne inférieure de précision de référence (comparable à un cahier des charges).

3) *Recherche d'un estimateur (a priori) localement le meilleur:*

Cette recherche peut être conduite en faisant appel à l'inégalité de Hölder (11) laquelle constitue la pierre angulaire des travaux de Barankin [Bar49] relatifs à la formulation d'une borne inférieure sur tout moment absolu d'ordre $s > 1$, dans le cas particulier où $\boldsymbol{\xi}^{-} = \boldsymbol{\xi}^{+} = \boldsymbol{\xi}$ pour lequel [REKGCLR13] **[T6]** :

$$\left\|\mathbf{D}_{\frac{\boldsymbol{\xi}^{+}+\boldsymbol{\xi}^{-}}{2}}^{-1}\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x})-\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)-\frac{\boldsymbol{\xi}^{+}-\boldsymbol{\xi}^{-}}{2}\right)\right\|_{s;\boldsymbol{\theta}^{0}}^{s} = \left\|\mathbf{D}_{\boldsymbol{\xi}}^{-1}\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x})-\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)\right)\right\|_{s;\boldsymbol{\theta}^{0}}^{s}. \tag{12}$$

En effet, pour tout estimateur $\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}$:

$$E_{\boldsymbol{\theta}}\left[\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x})\right] = \mathbf{m}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}^{0}}\left[\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x})\,\upsilon_{\boldsymbol{\theta}^{0}}(\mathbf{x};\boldsymbol{\theta})\right], \quad \upsilon_{\boldsymbol{\theta}^{0}}(\mathbf{x};\boldsymbol{\theta}) = \frac{p(\mathbf{x};\boldsymbol{\theta})}{p\left(\mathbf{x};\boldsymbol{\theta}^{0}\right)}. \tag{13}$$

Par conséquent, pour tout estimateur $\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}$, pour tout ensemble de valeurs particulières de $\boldsymbol{\theta}$ $\left\{\boldsymbol{\theta}^{i}\right\}_{1}^{I} \triangleq \left\{\boldsymbol{\theta}^{1},\boldsymbol{\theta}^{2},\ldots,\boldsymbol{\theta}^{I}\right\} \in \Theta$ et pour tout $\left\{\mathbf{w}^{i}\right\}_{1}^{I} \in \mathbb{R}^{Q}$:

$$E_{\boldsymbol{\theta}^{0}}\left[\left(\sum_{i=1}^{I}\mathbf{w}^{i}\upsilon_{\boldsymbol{\theta}^{0}}\left(\mathbf{x};\boldsymbol{\theta}^{i}\right)\right)^{T}\left(\mathbf{D}_{\boldsymbol{\xi}}^{-1}\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x})-\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)\right)\right)\right] = \sum_{i=1}^{I}\left(\mathbf{w}^{i}\right)^{T}\left(\mathbf{D}_{\boldsymbol{\xi}}^{-1}\left(\mathbf{m}\left(\boldsymbol{\theta}^{i}\right)-\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)\right)\right) \tag{14}$$

ce qui conduit, par combinaison de l'inégalité

$$\left|E_{\boldsymbol{\theta}}\left[\mathbf{f}(\mathbf{x})^{T}\mathbf{g}(\mathbf{x})\right]\right| \leq E_{\boldsymbol{\theta}}\left[\left|\mathbf{f}(\mathbf{x})^{T}\mathbf{g}(\mathbf{x})\right|\right] \leq \left\|\mathbf{f}(\mathbf{x})\odot\mathbf{g}(\mathbf{x})\right\|_{1;\boldsymbol{\theta}}$$

et de l'inégalité de Hölder (11), à l'inégalité suivante $(s > 1)$:

$$\frac{\left|\sum_{i=1}^{I}\left(\mathbf{w}^{i}\right)^{T}\left(\mathbf{D}_{\boldsymbol{\xi}}^{-1}\left(\mathbf{m}\left(\boldsymbol{\theta}^{i}\right)-\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)\right)\right)\right|^{s}}{\left\|\sum_{i=1}^{I}\mathbf{w}^{i}\upsilon_{\boldsymbol{\theta}^{0}}\left(\mathbf{x};\boldsymbol{\theta}^{i}\right)\right\|_{\boldsymbol{\theta}^{0};1+\frac{1}{s-1}}^{s}} \leq \left\|\mathbf{D}_{\boldsymbol{\xi}}^{-1}\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x})-\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)\right)\right\|_{s;\boldsymbol{\theta}^{0}}^{s} \tag{15}$$

En pratique on choisit souvent (implicitement) un critère de précision "isotrope" : $\boldsymbol{\xi} = \xi\mathbf{1}_{Q}$, conduisant plus simplement à $(s > 1)$:

$$1-\frac{1}{\xi}\left\|\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x})-\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)\right\|_{s;\boldsymbol{\theta}^{0}}^{s} \leq \mathcal{O}_{\boldsymbol{\theta}^{0}}\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)},\xi\mathbf{1}_{Q},\xi\mathbf{1}_{Q}\right) \leq 1 \tag{16}$$

avec [Bar49] [REKGCLR13] **[T6]** :

$$\sup_{\{\boldsymbol{\theta}^i\}_1^I, \{\mathbf{w}^i\}_1^I, I \in \mathbb{N}} \left\{ \frac{\left| \sum_{i=1}^{I} \left(\mathbf{w}^i\right)^T \left(\mathbf{m}\left(\boldsymbol{\theta}^i\right) - \mathbf{g}\left(\boldsymbol{\theta}^0\right)\right) \right|^s}{\left\| \sum_{i=1}^{I} \mathbf{w}^i \upsilon_{\boldsymbol{\theta}^0}\left(\mathbf{x}; \boldsymbol{\theta}^i\right) \right\|^s_{\boldsymbol{\theta}^0; 1+\frac{1}{s-1}}} \right\} \leq \left\| \widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x}) - \mathbf{g}\left(\boldsymbol{\theta}^0\right) \right\|^s_{s;\boldsymbol{\theta}^0}, \mathbf{m}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\left[\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})\right] \tag{17}$$

ce qui définit la plus grande des bornes inférieures ("supremum" or "least upper bound") du moment absolu d'ordre s ("sth absolute moment"), et que nous désignerons par la suite comme borne de Barankin (BB).

La BB (17) constitue le point de départ de nombreux travaux sur les bornes inférieures [Kie52] [BTM69] [MS69] [MH71] [Gla72] [Abe93] [Kno97] [FL02] [CGQL08] **[T4]** [CRL09] [TT10-I], et plus particulièrement sur les bornes inférieures de l'erreur quadratique moyenne (EQM) correspondant au cas particulier où $s = 2$ (cf. §II-A4-p23).

Cet ensemble de résultats est donc basé uniquement sur l'exploitation d'une hypothèse *a priori* sur le comportement de la valeur moyenne d'un estimateur $\left(E_{\boldsymbol{\theta}}\left[\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})\right] = \mathbf{m}(\boldsymbol{\theta})\right)$ pour un ensemble dénombrable de valeurs particulières du vecteur de paramètres inconnus $\left\{\boldsymbol{\theta}^i\right\}_1^I, I \in \mathbb{N}$, ce qui peut également être vu comme une contrainte linéaire continue exprimée (discrétisée) en un ensemble de points "test" (de la contrainte, "test points") :

$$E_{\boldsymbol{\theta}}\left[\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})\right] = \mathbf{m}(\boldsymbol{\theta}),\ \forall \boldsymbol{\theta} \in \Theta \quad \Rightarrow \quad \forall I \in \mathbb{N},\ \forall \left\{\boldsymbol{\theta}^i\right\}_1^I \in \Theta \quad E_{\boldsymbol{\theta}^i}\left[\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})\right] = \mathbf{m}\left(\boldsymbol{\theta}^i\right). \tag{18}$$

Que ce soit sous sa forme continue ou sous sa forme discrète, la contrainte (18) définit un sous-ensemble convexe de l'espace vectoriel des fonctions de $\Omega \to \mathbb{R}^Q$ que nous appelons par la suite "classe" d'estimateurs. En reformulant (18) sous la forme suivante :

$$E_{\boldsymbol{\theta}}\left[\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})\right] = \mathbf{m}(\boldsymbol{\theta}) = \mathbf{g}(\boldsymbol{\theta}) + \mathbf{b}(\boldsymbol{\theta}), \quad \mathbf{b}(\boldsymbol{\theta}) = \mathbf{m}(\boldsymbol{\theta}) - \mathbf{g}(\boldsymbol{\theta}) \tag{19}$$

on fait apparaitre naturellement la notion de biais ("bias") d'estimation représenté par $\mathbf{b}(\boldsymbol{\theta})$ ainsi que la classe des estimateurs sans biais ("unbiased") de $\mathbf{g}(\boldsymbol{\theta})$ vérifiant :

$$E_{\boldsymbol{\theta}}\left[\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})\right] = \mathbf{g}(\boldsymbol{\theta}), \tag{20}$$

classe pour laquelle la BB (17) s'écrit (plus simplement) :

$$\sup_{\{\boldsymbol{\theta}^i\}_1^I, \{\mathbf{w}^i\}_1^I, I \in \mathbb{N}} \left\{ \frac{\left| \sum_{i=1}^{I} \left(\mathbf{w}^i\right)^T \left(\mathbf{g}\left(\boldsymbol{\theta}^i\right) - \mathbf{g}\left(\boldsymbol{\theta}^0\right)\right) \right|^s}{\left\| \sum_{i=1}^{I} \mathbf{w}^i \upsilon_{\boldsymbol{\theta}^0}\left(\mathbf{x}; \boldsymbol{\theta}^i\right) \right\|^s_{\boldsymbol{\theta}^0; 1+\frac{1}{s-1}}} \right\} \leq \left\| \widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x}) - \mathbf{g}\left(\boldsymbol{\theta}^0\right) \right\|^s_{s;\boldsymbol{\theta}^0}, \mathbf{g}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\left[\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})\right] \tag{21}$$

La BB (21) pour estimateur sans biais est la forme usitée (pratique) de (17). Elle définit la plus grande borne inférieure d'erreur d'estimation à l'ordre s de $\mathbf{g}\left(\boldsymbol{\theta}^0\right)$ pour la classe des estimateurs sans biais sur Θ. L'estimateur associé (dont Barankin a démontré l'existence [Bar49]) constitue l'estimateur sans biais de $\mathbf{g}(\boldsymbol{\theta})$ (*a priori*) localement le meilleur en $\boldsymbol{\theta}^0$ (autrement dit lorsque $\boldsymbol{\theta}$ prend la valeur particulière $\boldsymbol{\theta}^0$, "locally best unbiased estimate"). Malheureusement cet estimateur n'a pas d'expression analytique [Bar49]. Le cas particulier où $s = 2$ permet cependant de formuler :

• une procédure de recherche itérative ne faisant intervenir que les points tests $\left\{\boldsymbol{\theta}^i\right\}_1^I$, réduisant ainsi de façon non négligeable la complexité de la recherche du supremum dans (21),

• une équation intégrale ayant pour solution l'estimateur sans biais (*a priori*) localement le meilleur.

4) Recherche d'un estimateur (a priori) localement le meilleur à l'ordre 2 :

a) La théorie:

D'un point de vue de la conduite des calculs, le choix $s = 2$ s'impose naturellement puisque c'est la plus petite valeur de s pour laquelle le moment absolu est développable. De plus, la norme obtenue peut être associée à un produit scalaire :

$$\langle \mathbf{g}(\mathbf{x}) \mid \mathbf{h}(\mathbf{x}) \rangle_{\boldsymbol{\theta}} = E_{\boldsymbol{\theta}}\left[\mathbf{h}^T(\mathbf{x})\,\mathbf{g}(\mathbf{x})\right], \quad \|\mathbf{g}(\mathbf{x})\|^2_{2;\boldsymbol{\theta}} = \langle \mathbf{g}(\mathbf{x}) \mid \mathbf{g}(\mathbf{x}) \rangle_{\boldsymbol{\theta}} \tag{22}$$

dont la propriété de bilinéarité permet d'optimiser la recherche d'un estimateur sans biais localement le meilleur en $\boldsymbol{\theta}^0$ sur l'espace vectoriel $S_\Omega\left(\boldsymbol{\theta}^0\right) = \left\{\mathbf{g}(\mathbf{x}) \text{ de } \Omega \to \mathbb{R}^Q \text{ tq } \|\mathbf{g}(\mathbf{x})\|^2_{\boldsymbol{\theta}^0;2} < \infty\right\}$.

En effet, considérons à titre didactique le cas où $\boldsymbol{\theta}$ se réduit à un seul paramètre $(P = 1)$: $\boldsymbol{\theta} \triangleq \theta$, et $\mathbf{g}(\boldsymbol{\theta})$ se réduit à une seule fonction $(Q = 1)$: $\mathbf{g}(\boldsymbol{\theta}) \triangleq g(\theta)$. Dans ce cas (21) devient [MS69] :

$$\sup_{\{\theta^i\}_1^I, \{w^i\}_1^I, I \in \mathbb{N}} \left\{\frac{\left(\mathbf{w}^T \boldsymbol{\Delta} g\right)^2}{\mathbf{w}^T \mathbf{R}_{\upsilon_{\theta^0}} \mathbf{w}}\right\} \leq \left\|\widehat{g\left(\theta^0\right)}(\mathbf{x}) - g\left(\theta^0\right)\right\|^2_{2;\theta^0}, \quad \left| \begin{array}{l} \left(\boldsymbol{\Delta}_g\right)_i = g\left(\theta^i\right) - g\left(\theta^0\right) \\ (\mathbf{w})_i = w^i \\ \left(\mathbf{R}_{\upsilon_{\theta^0}}\right)_{i,j} = E_{\theta^0}\left[\upsilon_{\theta^0}\left(\mathbf{x};\theta^i\right)\upsilon_{\theta^0}\left(\mathbf{x};\theta^j\right)\right] \end{array}\right. \tag{23}$$

où $\left\|\widehat{g\left(\theta^0\right)}(\mathbf{x}) - g\left(\theta^0\right)\right\|^2_{2;\theta^0}$ est l'erreur quadratique moyenne (EQM) commise lors de l'estimation de $g\left(\theta^0\right)$ par $\widehat{g\left(\theta^0\right)}$. Or $\frac{\left(\mathbf{w}^T \boldsymbol{\Delta}_g\right)^2}{\mathbf{w}^T \mathbf{R}_{\upsilon_{\theta^0}} \mathbf{w}}$ est maximal pour $\mathbf{w} = \lambda \mathbf{R}^{-1}_{\upsilon_{\theta^0}} \boldsymbol{\Delta}_g$ et vaut alors $\boldsymbol{\Delta}_g^T \mathbf{R}^{-1}_{\upsilon_{\theta^0}} \boldsymbol{\Delta}_g$, ce qui conduit à la forme "utile" ("useful") de la borne de Barankin [MS69] :

$$\sup_{\{\theta^i\}_1^I, I \in \mathbb{N}} \left\{\boldsymbol{\Delta}_g^T \mathbf{R}^{-1}_{\upsilon_{\theta^0}} \boldsymbol{\Delta}_g\right\} \leq \left\|\widehat{g\left(\theta^0\right)}(\mathbf{x}) - g\left(\theta^0\right)\right\|^2_{2;\theta^0}, \quad \left| \begin{array}{l} \left(\boldsymbol{\Delta}_g\right)_i = g\left(\theta^i\right) - g\left(\theta^0\right) \\ \left(\mathbf{R}_{\upsilon_{\theta^0}}\right)_{i,j} = E_{\theta^0}\left[\upsilon_{\theta^0}\left(\mathbf{x};\theta^i\right)\upsilon_{\theta^0}\left(\mathbf{x};\theta^j\right)\right] \end{array}\right. \tag{24}$$

Par ailleurs, le problème de minimisation sous contraintes linéaires :

$$\min\left\{\left\|\widehat{g\left(\theta^0\right)}(\mathbf{x}) - g\left(\theta^0\right)\right\|^2_{2;\theta^0}\right\} \text{ sous } E_{\theta^i}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\right] = g\left(\theta^i\right), \quad \left\{\theta^i\right\}_1^I \in \Theta \tag{25}$$

est équivalent à :

$$\min\left\{\left\|\widehat{g\left(\theta^0\right)}(\mathbf{x}) - g\left(\theta^0\right)\right\|^2_{2;\theta^0}\right\} \text{ sous } E_{\theta^0}\left[\left(\widehat{g\left(\theta^0\right)}(\mathbf{x}) - g\left(\theta^0\right)\right)\upsilon_{\theta^0}\left(\mathbf{x};\theta^i\right)\right] = g\left(\theta^i\right) - g\left(\theta^0\right), \quad \left\{\theta^i\right\}_1^I \in \Theta,$$

et a pour solution [MS69] [Gla72] :

$$\boldsymbol{\Delta}_g^T \mathbf{R}^{-1}_{\upsilon_{\theta^0}} \boldsymbol{\Delta}_g = \left\|\widehat{g\left(\theta^0\right)}_{opt}(\mathbf{x}) - g\left(\theta^0\right)\right\|^2_{2;\theta^0}, \quad \widehat{g\left(\theta^0\right)}_{opt}(\mathbf{x}) = g\left(\theta^0\right) + \sum_{i=1}^{I} w^i \upsilon_{\theta^0}\left(\mathbf{x};\theta^i\right), \quad \mathbf{R}_{\upsilon_{\theta^0}} \mathbf{w} = \boldsymbol{\Delta}_g, \tag{26}$$

ce qui traduit le fait que la plus grande borne inférieure de l'EQM pour un nombre fini de points test $\left\{\theta^i\right\}_1^I$ est obtenue simplement en exprimant la contrainte "sans biais" en les points tests [MS69].
Ce résultat fondamental permet non seulement de réduire de façon non négligeable la complexité de la recherche de bornes inférieures (plus de maximisation relativement aux pondérations w^i) mais également de restreindre le sous ensemble des fonctions contenant l'estimateur sans biais localement le meilleur. En effet, à partir de ce résultat, Glave [Gla72] a démontré qu'un maillage uniforme de Θ $\left(\theta^{i+1} - \theta^i = d\theta\right)$ permet d'assurer une convergence - au sens de l'EQM - de $\boldsymbol{\Delta}_g^T \mathbf{R}^{-1}_{\upsilon_{\theta^0}} \boldsymbol{\Delta}_g$ vers la BB (24). Il suffit pour cela d'exprimer (26) sous la forme :

$$\boldsymbol{\Delta}_g^T \mathbf{R}^{-1}_{\upsilon_{\theta^0}} \boldsymbol{\Delta}_g = \boldsymbol{\Delta}_g^T \mathbf{w}\, d\theta, \; \widehat{g\left(\theta^0\right)}_{opt}(\mathbf{x}) = g\left(\theta^0\right) + \sum_{i=1}^{I} \upsilon_{\theta^0}\left(\mathbf{x};\theta^i\right) w\left(\theta^i\right) d\theta, \; \mathbf{R}_{\upsilon_{\theta^0}} \mathbf{w}\, d\theta = \boldsymbol{\Delta}_g, \; (\mathbf{w})_i = w\left(\theta^i\right) \tag{27}$$

et d'en considérer la limite lorsque $I \to \infty$:

$$\left\| \widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right) - g\left(\theta^0\right) \right\|^2_{2;\theta^0} \geq \int\limits_{\Theta} \left(g\left(\theta\right) - g\left(\theta^0\right)\right) w\left(\theta\right) d\theta \tag{28a}$$

$$\widehat{g\left(\theta^0\right)}_{opt}\left(\mathbf{x}\right) = g\left(\theta^0\right) + \frac{1}{p\left(\mathbf{x};\theta^0\right)} \int\limits_{\Theta} p\left(\mathbf{x};\theta\right) w\left(\theta\right) d\theta \tag{28b}$$

$$\int\limits_{\Theta} R_{\upsilon_{\theta^0}}\left(\theta,\theta'\right) w\left(\theta'\right) d\theta' = g\left(\theta\right) - g\left(\theta^0\right), \quad R_{\upsilon_{\theta^0}}\left(\theta,\theta'\right) = E_{\theta^0}\left[\upsilon_{\theta^0}\left(\mathbf{x};\theta\right)\upsilon_{\theta^0}\left(\mathbf{x};\theta'\right)\right] \tag{28c}$$

ce qui définit une équation intégrale (28c) [Mar97] [FL02] [CGQL08] **[T4]** dont la solution analytique n'a pas (encore) été formulée dans le cas général, ni même pour les modèles d'observations gaussiens classiques en traitement du signal ("conditional or unconditional model" [SN90]). Seuls certains cas particuliers impliquant la loi de Poisson ou la loi exponentielle ont permis d'obtenir une solution analytique [Kie52] [Mor83] [Mar97] [Pom03].

Les limitations pratiques de la forme utile de la borne de Barankin (24), que ce soit pour la formulation d'une solution analytique (28a) ou pour la minimisation du coût de calcul d'une recherche itérative (27), ont conduit certains auteurs [Kie52] [TT10-I] à reformuler le problème de minimisation (25) dans d'autre bases de fonctions en suivant l'approche générale suivante initiée par Kiefer [Kie52].

Considérons $\{\mathbf{w}_k\}_1^K \in \mathbb{R}^I$ une famille de K vecteurs indépendants et $\mathbf{W} = [\mathbf{w}_1 \ \ldots \ \mathbf{w}_K] \in \mathcal{M}_{\mathbb{R}}\left(I,K\right)$. Puisque $E_{\theta^i}\left[h\left(\mathbf{x}\right)\right] = E_{\theta^0}\left[h\left(\mathbf{x}\right)\upsilon_{\theta^0}\left(\mathbf{x};\theta^i\right)\right]$, alors, $\forall K \leq I$:

$$E_{\theta^i}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right] = g\left(\theta^i\right), \left\{\theta^i\right\}_1^I \in \Theta \Rightarrow E_{\theta^0}\left[\left(\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right) - g\left(\theta^0\right)\right)\mathbf{w}_k^T \boldsymbol{\upsilon}_{\theta^0}\left(\mathbf{x},\left\{\theta^i\right\}_1^I\right)\right] = \mathbf{w}_k^T \boldsymbol{\Delta}_g, \left\{\mathbf{w}_k\right\}_1^K \tag{29}$$

où $\boldsymbol{\upsilon}_{\theta^0}\left(\mathbf{x},\left\{\theta^i\right\}_1^I\right) = \left(\upsilon_{\theta^0}\left(\mathbf{x};\theta^1\right),\ldots,\upsilon_{\theta^0}\left(\mathbf{x};\theta^I\right)\right)^T$ et (cf. §II-A4b-p26) :

$$\boldsymbol{\Delta}_g^T \mathbf{R}_{\upsilon_{\theta^0}}^{-1} \boldsymbol{\Delta}_g \geq \min\left\{\left\| \widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right) - g\left(\theta^0\right) \right\|^2_{2;\theta^0}\right\} = \boldsymbol{\Delta}_g^T \left(\mathbf{W}\left(\mathbf{W}^T \mathbf{R}_{\upsilon_{\theta^0}} \mathbf{W}\right)^{-1} \mathbf{W}^T\right) \boldsymbol{\Delta}_g \tag{30}$$

L'inégalité (30) est l'application d'un résultat général d'algèbre : la projection orthogonale d'un vecteur sur un s.e.v. ne dépend pas de la base du s.e.v. choisie. Appliqué à notre problème de minimisation d'une norme sous-contrainte linéaire (25) qui n'est autre qu'un problème de projection orthogonale (cf. Lemme 1-p27) sur le s.e.v. engendré par la famille de vecteurs $\left\{\boldsymbol{\upsilon}_{\theta^0}\left(\mathbf{x};\theta^1\right),\ldots,\boldsymbol{\upsilon}_{\theta^0}\left(\mathbf{x};\theta^I\right)\right\} \in S_{\Omega}\left(\theta^0\right)$, ce résultat général d'algèbre nous rappelle que :

1) la borne obtenue (norme de la projection orthogonale) est invariante par transformation bijective des I contraintes. En effet si $K = I$ et $\mathbf{W}$ est inversible, alors : $\mathbf{W}\left(\mathbf{W}^T \mathbf{R}_{\upsilon_{\theta^0}} \mathbf{W}\right)^{-1} \mathbf{W}^T = \mathbf{R}_{\upsilon_{\theta^0}}^{-1}$.

2) toute transformation injective des I contraintes réduit la borne obtenue.

Le point 1) suggère que la famille $\left\{\boldsymbol{\upsilon}_{\theta^0}\left(\mathbf{x};\theta^1\right),\ldots,\boldsymbol{\upsilon}_{\theta^0}\left(\mathbf{x};\theta^I\right)\right\}, I \in \mathbb{N}$, n'est pas nécessairement la plus pertinente pour implémenter la recherche (27) de l'estimateur sans biais localement le meilleur ou trouver sa forme analytique (28c) (illustré par Kiefer [Kie52] dans 2 cas particuliers).

Le point 2) suggère la prise en compte d'un compromis possible entre complexité de mise en oeuvre et "pertinence" d'une borne de complexité moindre : certaines transformations linéaires peuvent "concentrer" les I contraintes initiales en $K \ll I$ nouvelles contraintes dont la borne associée $\left(\boldsymbol{\Delta}_g^T \left(\mathbf{W}\left(\mathbf{W}^T \mathbf{R}_{\upsilon_{\theta^0}} \mathbf{W}\right)^{-1} \mathbf{W}^T\right) \boldsymbol{\Delta}_g\right)$ sera proche de la borne initiale $\left(\boldsymbol{\Delta}_g^T \mathbf{R}_{\upsilon_{\theta^0}}^{-1} \boldsymbol{\Delta}_g\right)$, avec une complexité de mise en oeuvre (notamment l'inversion de matrice) nettement moindre (conjecture illustrée par Todros et Tabrikian dans un cas particulier [TT08] [TT10-I]).

En posant $\mathbf{w}_k \triangleq \mathbf{w}\left(\tau_k\right) = \left(w\left(\theta^1,\tau_k\right),\ldots,w\left(\theta^I,\tau_k\right)\right)^T$, c'est à dire en considérant que $\mathbf{w}_k$ est un vecteur d'échantillons d'une fonction paramétrique $w\left(\theta,\tau\right) : \Theta \times \Lambda \to \mathbb{R}$ où $\Lambda \subset \mathbb{R}$, alors lorsque $\left\{\theta^i\right\}_1^I, I \to \infty$, décrit un maillage uniforme de Θ, (29) devient :

$$E_{\theta^0}\left[\left(\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right) - g\left(\theta^0\right)\right)\eta\left(\mathbf{x},\tau_k,\theta^0\right)\right] = \Gamma_w\left(\tau_k,\theta^0\right), \quad k \in [1,K] \tag{31}$$

$$\eta_w\left(\mathbf{x};\tau,\theta^0\right) = \frac{1}{p\left(\mathbf{x};\theta^0\right)} \int\limits_{\Theta} w\left(\tau,\theta\right) p\left(\mathbf{x};\theta\right) d\theta, \quad \Gamma_w\left(\tau,\theta^0\right) = \int\limits_{\Theta} w\left(\tau,\theta\right)\left(g\left(\theta\right) - g\left(\theta^0\right)\right) d\theta \tag{32}$$

Par analogie avec les résultats précédents depuis (25) jusqu'à (28b), en substituant $I \to K$, $\upsilon_{\theta^0}\left(\mathbf{x};\theta^i\right) \to \eta_w\left(\mathbf{x},\tau_k,\theta^0\right)$, $\mathbf{\Delta}_g = \left(g\left(\theta^1\right) - g\left(\theta^0\right), \ldots, g\left(\theta^I\right) - g\left(\theta^0\right)\right)^T \to \mathbf{\Gamma}_w\left(\theta^0\right) = \left(\Gamma_w\left(\tau_1,\theta^0\right), \ldots, \Gamma_w\left(\tau_K,\theta^0\right)\right)^T$, lorsque $\{\tau_k\}_1^K, K \to \infty$, décrit un maillage uniforme de Λ, alors :

$$\left\|\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right) - g\left(\theta^0\right)\right\|_{2;\theta^0}^2 \geq \int_\Lambda \Gamma_w\left(\tau,\theta^0\right)\beta\left(\tau\right)d\tau \tag{33a}$$

$$\widehat{g\left(\theta^0\right)}_{opt}\left(\mathbf{x}\right) = g\left(\theta^0\right) + \int_\Lambda \eta_w\left(\mathbf{x};\tau,\theta^0\right)\beta\left(\tau\right)d\tau \tag{33b}$$

$$\int_\Lambda R_{\eta_w}\left(\tau,\tau',\theta^0\right)\beta\left(\tau'\right)d\tau' = \Gamma_w\left(\tau,\theta^0\right) \tag{33c}$$

$$R_{\eta_w}\left(\tau,\tau',\theta^0\right) = E_{\theta^0}\left[\eta_w\left(\mathbf{x},\tau,\theta^0\right)\eta_w\left(\mathbf{x},\tau',\theta^0\right)\right] = \int_\Theta\int_\Theta w\left(\tau,\theta\right)R_{\upsilon_{\theta^0}}\left(\theta,\theta'\right)w\left(\tau',\theta'\right)d\theta d\theta' \tag{33d}$$

On remarque que la forme discrète (29) ou la forme continue (32) introduisent une transformation linéaire de la contrainte sans biais et de la d.d.p... Comme mentionné précédemment , la généralisation de l'équation intégrale (28c) par (33c) [TT08] [TT10-I] n'a pas permis à ce jour de formuler une solution analytique dans le cas général, ni même pour les modèles d'observations gaussiens classiques en traitement du signal ("conditional or unconditional model" [SN90]). Seuls certains cas particuliers impliquant la loi de Poisson ou la loi exponentielle ont permis d'obtenir une solution analytique [Kie52] [Mor83] [Mar97] [Pom03].
De plus, quand bien même cette solution analytique serait accessible, il apparait évident que dans le cas général, l'estimateur optimal (33b) dépend de la valeur du paramètre à estimer θ^0 : l'estimateur (*a priori*) localement le meilleur à l'ordre 2 est clairvoyant et ne peut donc pas être obtenu en pratique, c'est à dire à partir des seules observations $\mathbf{x}$.
Néanmoins, conceptuellement la BB (33a) fournit une valeur de référence ("benchmark") qui permet de juger de la qualité d'un estimateur sans biais au sens de l'EQM (i.e. la précision *a priori* à l'ordre 2) :
• une proximité à cette borne peut être considérée comme un critère de qualité de l'estimateur (efficacité d'un estimateur),
• un éloignement notable à cette borne peut suggérer que l'estimateur considéré pourrait être remplacé par un estimateur (à découvrir) d'EQM inférieure.
D'autre part, la BB peut également être considérée comme un outil d'aide à la conception ("design") d'un système devant respecter un cahier des charges défini en terme de précision d'estimation (au sens de l'EQM) de certains paramètres d'intérêt [MS69] [MH71]:
• soit la BB est inférieure à l'EQM requise : alors il peut éventuellement exister un estimateur sans biais réalisable à partir du système considéré qui permette de satisfaire les spécifications,
• soit la BB est supérieure à l'EQM requise : alors il ne peut exister d'estimateur sans biais réalisable à partir du système considéré qui permette de satisfaire les spécifications; il faut donc faire évoluer le système de façon ad hoc afin d'abaisser la BB.

C'est pourquoi depuis les premiers travaux applicatifs en traitement du signal sur la BB initiés par McAulay et Seidman [MS69], de nombreux auteurs ont proposé diverses approximations calculables de la BB, approximations essayant de conjuguer la possibilité d'une mise en oeuvre par calcul informatique et une proximité à la BB ("tightness").

Transformations non linéaires de la contrainte sans biais

Une extension de l'approche initiée par [Kie52] et généralisée par [TT10-I] est possible pour les d.d.p.. $p\left(\mathbf{x};\theta\right)$ vérifiant [BTM69] [CRL09] :

$$t\left(p\left(\mathbf{x};\theta\right)\right) = k\left(\theta,t\right)p\left(\mathbf{x};\gamma\left(\theta,t\right)\right), \quad k\left(\theta,t\right) = \int_\Omega t\left(p\left(\mathbf{x};\theta\right)\right)d\mathbf{x} \tag{34}$$

où $t\left(\ \right)$ est une fonction de $\mathbb{R}\rightarrow\mathbb{R}$, $\gamma\left(\theta,t\right)$ est une fonction de $\Theta\rightarrow\Theta$ et $k\left(\theta,t\right)$ est une fonction de $\Theta\rightarrow\mathbb{R}$ dépendantes de $t\left(\ \right)$. En effet dans ce cas, un estimateur sans biais (53) vérifie également :

$$E_{\theta^0}\left[\left(\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)-g\left(\theta^0\right)\right)\frac{t\left(p\left(\mathbf{x};\theta\right)\right)}{p\left(\mathbf{x};\theta^0\right)}\right]=k\left(\theta,t\right)\left[g\left(\gamma\left(\theta,t\right)\right)-g\left(\theta^0\right)\right],\quad\forall\theta\in\Theta. \tag{35}$$

Dans le cas le plus général, s'il existe un ensemble de fonctions $t_\theta\left(\ \right)$ vérifiant (34), alors tous les résultats développés lors d'une transformation linéaire de $p\left(\mathbf{x};\theta\right)$ et $g\left(\theta\right)-g\left(\theta^0\right)$ s'appliquent en substituant simplement :

$$\nu\left(\mathbf{x};\theta\right)\triangleq\frac{p\left(\mathbf{x};\theta\right)}{p\left(\mathbf{x};\theta^0\right)}\rightarrow\nu\left(\mathbf{x};\theta\right)\triangleq\frac{t_\theta\left(p\left(\mathbf{x};\theta\right)\right)}{p\left(\mathbf{x};\theta^0\right)},\qquad g\left(\theta\right)-g\left(\theta^0\right)\rightarrow k\left(\theta,t_\theta\right)\left[g\left(\gamma\left(\theta,t_\theta\right)\right)-g\left(\theta^0\right)\right].$$

Ainsi (32) devient :

$$\eta_w\left(\mathbf{x};\tau,\theta^0\right)=\int_\Theta\frac{w\left(\tau,\theta\right)t_\theta\left(p\left(\mathbf{x};\theta\right)\right)}{p\left(\mathbf{x};\theta^0\right)}d\theta,\quad\Gamma_w\left(\tau,\theta^0\right)=\int_\Theta w\left(\tau,\theta\right)k\left(\theta,t_\theta\right)\left[g\left(\gamma\left(\theta,t_\theta\right)\right)-g\left(\theta^0\right)\right]d\theta, \tag{36}$$

définissant le cas général d'un mélange de transformations linéaire et non-linéaire.
A titre d'exemple, considérons le modèle gaussien complexe circulaire $\mathbf{x}\sim\mathcal{CN}\left(\mathbf{m}\left(\boldsymbol{\theta}\right),\mathbf{C}\left(\boldsymbol{\theta}\right)\right)$ de d.d.p.. :

$$p\left(\mathbf{x};\boldsymbol{\theta}\right)=p\left(\mathbf{x};\mathbf{m}\left(\boldsymbol{\theta}\right),\mathbf{C}\left(\boldsymbol{\theta}\right)\right)=\frac{e^{-\left(\mathbf{x}-\mathbf{m}\left(\boldsymbol{\theta}\right)\right)^H\mathbf{C}\left(\boldsymbol{\theta}\right)^{-1}\left(\mathbf{x}-\mathbf{m}\left(\boldsymbol{\theta}\right)\right)}}{\pi^M\left|\mathbf{C}\left(\boldsymbol{\theta}\right)\right|}$$

Alors la transformation $t_q\left(y\right)=y^q$ peut être appliquée au modèle d'observation pour lequel (mélange de modèle déterministe et stochastique) :

$$\mathbf{m}\left(\boldsymbol{\theta}\right)=\mathbf{m}\left(\boldsymbol{\varepsilon}\right),\quad\mathbf{C}\left(\boldsymbol{\theta}\right)=\boldsymbol{\Psi}\left(\boldsymbol{\zeta}\right)\mathbf{C_s}\boldsymbol{\Psi}\left(\boldsymbol{\zeta}\right)^H+\mathbf{C_n},\quad\boldsymbol{\theta}=\left[\boldsymbol{\varepsilon}^T,\boldsymbol{\zeta}^T,\mathbf{vec}\left(\mathbf{C_s}\right)^T,\mathbf{vec}\left(\mathbf{C_n}\right)^T\right]^T, \tag{37}$$

et conduit à :

$$t_q\left(p\left(\mathbf{x};\boldsymbol{\theta}\right)\right)=k\left(\boldsymbol{\theta},t_q\right)p\left(\mathbf{x};\boldsymbol{\gamma}\left(\boldsymbol{\theta},t_q\right)\right),\quad\left|\begin{array}{l}k\left(\boldsymbol{\theta},t_q\right)=\frac{\pi^{M(1-q)}}{q^{qM}}\left|\frac{\mathbf{C}(\boldsymbol{\theta})}{q}\right|^{1-q}\\ \boldsymbol{\gamma}\left(\boldsymbol{\theta},t_q\right)=\left[\boldsymbol{\varepsilon}^T,\boldsymbol{\zeta}^T,\frac{\mathbf{vec}(\mathbf{C_s})}{q}^T,\frac{\mathbf{vec}(\mathbf{C_n})^T}{q}\right]^T\end{array}\right. \tag{38}$$

b) La pratique :

Chronologiquement, les principaux contributeurs à la formulation de bornes inférieures (déterministe) de l'EQM, mentionnés ou publiés dans la littérature "classique" en traitement du signal (IEEE, IET, Elsevier) sont : [Fre43] [Dar45] [Rao45] [Cra46] [Bat46] [Bar49] [Ham50] [CR51] [Kie52] [FG52] [MS69] [BTM69] [MH71] [Gla72] [Alb73] [Abe93] [CGQL08] **[T4]** [CRL09] [TT10-I].
Indépendamment de la chronologie, l'ensemble des approximations pratiques de la BB s'obtient à partir de la généralisation des travaux de Mcaulay-Seidman (25), généralisation qui s'obtient aisément à l'aide des deux formes duales d'un lemme d'algèbre linéaire :
• la minimisation d'une norme sous contraintes linéaires (Lemme 1-p27) [Gla72] [FL02] [CGQL08] **[T4]**,
• la généralisation de l'inégalité de Cauchy-Schwartz aux matrices de Gram (généralement connue sous le nom d'inégalité de covariance [Bly74] [Abe93]).

Nous utilisons le Lemme 1 en prenant $\left\{\mathbf{u}_q\right\}_1^Q$ / $\mathbf{u}_q\triangleq\widehat{g_q\left(\boldsymbol{\theta}^0\right)}\left(\mathbf{x}\right)-g_q\left(\boldsymbol{\theta}^0\right)$ (cas général des fonction multiples de paramètres multiples) car il permet une meilleure compréhension des hypothèses (contraintes) associées aux différentes bornes de l'EQM [Gla72] [FL02] [CGQL08] **[T4]**. Pour des raisons de lisibilité, nous écrivons $\min\left\{\mathbf{G}\left(\left\{\mathbf{u}_q\right\}_1^Q\right)\right\}$ pour désigner $\min\limits_{\left\{\mathbf{u}_q\right\}_1^Q}\left\{\mathbf{G}\left(\left\{\mathbf{u}_q\right\}_1^Q\right)\right\}$[5]. Enfin, la compréhension des principaux résultats de ce chapitre est facilitée par la connaissance de deux lemmes complémentaires également joints.

[5] Au sens de la relation d'ordre partiel définie sur les matrices symétriques [HJ99, §7.7] par : $\mathbf{A}\geq\mathbf{B}$ si $\mathbf{A}-\mathbf{B}$ est positive.

Lemme 1

Soit $\mathbb{U}$ *un espace vectoriel euclidien*[6] *de dimension quelconque (finie ou infinie) sur le corps des réels* $\mathbb{R}$. *Soit* $\{\mathbf{c}_i\}_1^I$ *une famille de* I *vecteurs de* $\mathbb{U}$ *telle que* $\mathbf{G}\left(\{\mathbf{c}_i\}_1^I\right) = \mathbf{MDM}^T$, $\mathbf{M} \in \mathcal{M}_{\mathbb{R}}\left(I, \widetilde{I}\right)$, $\mathbf{M}^T\mathbf{M} = \mathbf{I}_{\widetilde{I}}$, $\mathbf{D} = \mathbf{Diag}\left(d_1, \ldots, d_{\widetilde{I}}\right)$, $d_1 \geq \ldots \geq d_{\widetilde{I}} > 0$, $\widetilde{I} \leq I$. *Soit* $\{\mathbf{u}_q\}_1^Q$ *une famille de* Q *vecteurs inconnus de* $\mathbb{U}$. *Alors la solution de:*

$$\min\left\{\mathbf{G}\left(\{\mathbf{u}_q\}_1^Q\right)\right\} \textit{ sous } \mathbf{G}\left(\{\mathbf{u}_q\}_1^Q, \{\mathbf{c}_i\}_1^I\right) = \mathbf{V} \in \mathcal{M}_{\mathbb{R}}(I, Q)$$

est:

$$\min\left\{\mathbf{G}\left(\{\mathbf{u}_q\}_1^Q\right)\right\} = \left(\mathbf{M}^T\mathbf{V}\right)^T \mathbf{D}^{-1}\left(\mathbf{M}^T\mathbf{V}\right), \quad \mathbf{u}_q = \sum_{i=1}^{I} (\mathbf{A})_{i,q}\, \mathbf{c}_i, \ \mathbf{A} = \mathbf{MD}^{-1}\mathbf{M}^T\mathbf{V} \tag{39}$$

si et seulement si $\mathbf{V}$ *vérifie* $\left(\mathbf{MM}^T\right)\mathbf{V} = \mathbf{V}$, *i.e. si et seulement si* $\mathbf{V} \in im\left\{\mathbf{G}\left(\{\mathbf{c}_i\}_1^I\right)\right\}$.

Si $\mathbf{G}\left(\{\mathbf{c}_i\}_1^I\right)$ *est une matrice de rang plein, alors* (39) *se réduit à:*

$$\min\left\{\mathbf{G}\left(\{\mathbf{u}_q\}_1^Q\right)\right\} = \mathbf{V}^T\mathbf{G}\left(\{\mathbf{c}_i\}_1^I\right)^{-1}\mathbf{V}, \quad \mathbf{u}_q = \sum_{i=1}^{I} (\mathbf{A})_{i,q}\, \mathbf{c}_i, \ \mathbf{A} = \mathbf{G}\left(\{\mathbf{c}_i\}_1^I\right)^{-1}\mathbf{V}$$

Lemme 2

Soit $\{\mathbf{c}_i\}_1^{I+1}$ *une famille de* $I+1$ *vecteurs indépendants de* $\mathbb{U}$ *et soit* $\mathbf{V}$ *une matrice de* $\mathcal{M}_{\mathbb{R}}(I, Q)$. *Alors, le problème de minimisation de* $\mathbf{G}\left(\{\mathbf{u}_q\}_1^Q\right)$ *vérifie l'équivalence suivante :*

$$\begin{gathered}\min\left\{\mathbf{G}\left(\{\mathbf{u}_q\}_1^Q\right)\right\} \textit{ sous } \begin{cases}\mathbf{G}\left(\{\mathbf{u}_q\}_1^Q, \{\mathbf{c}_i\}_1^{I+1}\right) = \begin{bmatrix}\mathbf{V}\\ \mathbf{0}^T\end{bmatrix}\\ \mathbf{G}\left(\{\mathbf{c}_i\}_1^I, \mathbf{c}_{I+1}\right) = \mathbf{0}^T\end{cases}\\ \Updownarrow\\ \min\left\{\mathbf{G}\left(\{\mathbf{u}_q\}_1^Q\right)\right\} \textit{ sous } \mathbf{G}\left(\{\mathbf{u}_q\}_1^Q, \{\mathbf{c}_i\}_1^I\right) = \mathbf{V}\end{gathered} \tag{40}$$

Lemme 3

Soit $\{\mathbf{c}_i\}_1^I$ *une famille de* I *vecteurs indépendants de* $\mathbb{U}$ *et soit* $\mathbf{V}$ *une matrice de* $\mathcal{M}_{\mathbb{R}}(I, Q)$. *Alors, le problème de minimisation de* $\mathbf{G}\left(\{\mathbf{u}_q\}_1^Q\right)$ *vérifie l'équivalence suivante :*

$$\begin{gathered}\min\left\{\mathbf{G}\left(\{\mathbf{u}_q\}_1^Q\right)\right\} \textit{ sous } \mathbf{G}\left(\{\mathbf{u}_q\}_1^Q, \{\mathbf{c}_i\}_1^I\right) = \mathbf{V} \in \mathcal{M}_{\mathbb{R}}(I, Q)\\ \Updownarrow\\ \min\left\{\mathbf{G}\left(\{\mathbf{u}_q\}_1^Q\right)\right\} \textit{ sous } \begin{cases}\mathbf{W}^T\mathbf{G}\left(\{\mathbf{u}_q\}_1^Q, \{\mathbf{c}_i\}_1^I\right) = \mathbf{W}^T\mathbf{V} \in \mathcal{M}_{\mathbb{R}}(I, Q)\\ \mathbf{W} \in \mathcal{M}_{\mathbb{R}}(I, I), \quad |\mathbf{W}| \neq 0\end{cases}\\ \Updownarrow\\ \min\left\{\mathbf{G}\left(\{\mathbf{u}_q\}_1^Q\right)\right\} \textit{ sous } \begin{cases}\mathbf{G}\left(\{\mathbf{u}_q\}_1^Q, \{\mathbf{c}'_i\}_1^I\right) = \mathbf{W}^T\mathbf{V} \in \mathcal{M}_{\mathbb{R}}(I, Q)\\ \mathbf{c}'_i = \sum_{l=1}^{I} W_{i,l}\mathbf{c}_l, \quad \mathbf{W} \in \mathcal{M}_{\mathbb{R}}(I, I), \quad |\mathbf{W}| \neq 0\end{cases}\end{gathered}, \tag{41}$$

i.e. le problème de minimisation est invariant par une transformation bijective des I *contraintes linéaires.*

La généralisation des travaux de Mcaulay-Seidman (25) correspond à la recherche de :

$$\min\left\{\left\|\widehat{g\left(\theta^0\right)}(\mathbf{x}) - g\left(\theta^0\right)\right\|_{2;\theta^0}^2\right\} \text{ sous } E_{\theta^0}\left[\left(\widehat{g\left(\theta^0\right)}(\mathbf{x}) - g\left(\theta^0\right)\right)\mathbf{w}_k^T\boldsymbol{\upsilon}_{\theta^0}\left(\mathbf{x}, \{\theta^i\}_1^I\right)\right] = \mathbf{w}_k^T\boldsymbol{\Delta}_g, \ \{\mathbf{w}_k\}_1^K, \ K \leq I$$

[6]Pour la généralisation aux espaces hermitiens permettant de traiter le cas des paramètres mixtes (réels et complexes), se référer à [Men12] **[T5]** [MCLB12].

laquelle s'écrit également, en posant $\mathbf{u}_1 = \widehat{g\left(\theta^0\right)}(\mathbf{x}) - g\left(\theta^0\right)$, $\mathbf{W} = [\mathbf{w}_1 \ \ldots \ \mathbf{w}_K]$ et en adoptant une notation adaptée aux Lemmes $1, 2, 3$:

$$\min\left\{\mathbf{G}_{\theta^0}(\mathbf{u}_1)\right\} \text{ sous } \mathbf{W}^T\mathbf{G}_{\theta^0}\left(\mathbf{u}_1, \{\mathbf{c}_i\}_1^I\right) = \mathbf{W}^T\mathbf{v}, \quad \mathbf{c}_i \triangleq \upsilon_{\theta^0}\left(\mathbf{x}, \theta^i\right), \quad \mathbf{v} = \boldsymbol{\Delta}_g$$

$$\Updownarrow$$

$$\min\left\{\mathbf{G}_{\theta^0}(\mathbf{u}_1)\right\} \text{ sous } \mathbf{G}_{\theta^0}\left(\mathbf{u}_1, \{\mathbf{c}'_k\}_1^K\right) = \mathbf{v}', \qquad \mathbf{c}'_k = \sum_{i=1}^{I} W_{i,k}\mathbf{c}_i, \quad \mathbf{v}' = \mathbf{W}^T\boldsymbol{\Delta}_g$$

dont la solution est (Lemmes $1+3$) :

$$\boldsymbol{\Delta}_g^T \mathbf{R}_{\upsilon_{\theta^0}}^{-1} \boldsymbol{\Delta}_g \geq \min\left\{\mathbf{G}_{\theta^0}(\mathbf{u}_1)\right\} = \left(\mathbf{W}^T\boldsymbol{\Delta}_g\right)^T \left(\mathbf{W}^T\mathbf{R}_{\upsilon_{\theta^0}}\mathbf{W}\right)^{-1}\left(\mathbf{W}^T\boldsymbol{\Delta}_g\right)$$

où $\langle \ | \ \rangle_{\theta^0}$ défini par (22) est le produit scalaire considéré.

Chaque borne est entièrement déterminée (Lemme 1) par le choix particulier des contraintes résultant de la combinaison du choix des points test $\left\{\theta^i\right\}_1^I$ associé éventuellement à une transformation linéaire injective $\mathbf{W}$:

- la borne de Mcaulay-Seidman (BMS) [MS69] :

$$\min\left\{\mathbf{G}(\mathbf{u}_1)\right\} \text{ sous } \left| \begin{array}{l} \left\{\begin{array}{l} E_{\theta^0}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\right] = g\left(\theta^0\right) \\ E_{\theta^i}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\right] = g\left(\theta^i\right), i \in [1, I] \end{array}\right. \\ \qquad\qquad \Updownarrow \\ \left\{\begin{array}{l} \langle \mathbf{u}_1 \mid \mathbf{c}_0\rangle_{\theta^0} = 0 \\ \langle \mathbf{u}_1 \mid \mathbf{c}_i\rangle_{\theta^0} = g\left(\theta^i\right) - g\left(\theta^0\right), i \in [1, I] \end{array}\right. \text{ où } \left| \begin{array}{l} \mathbf{c}_0 = 1(\mathbf{x}) \\ \mathbf{c}_i = \upsilon_{\theta^0}\left(\mathbf{x}, \theta^i\right), i \in [1, I] \end{array}\right. \end{array}\right. \tag{42}$$

avec recherche du supremum :

$$BMS = \sup_{\{\theta^i\}_1^I, I \in \mathbb{N}} \left\{\min\left\{\mathbf{G}(\mathbf{u}_1)\right\}\right\} = \sup_{\{\theta^i\}_1^I, I \in \mathbb{N}} \left\{\boldsymbol{\Delta}_g^T \mathbf{R}_{\upsilon_{\theta^0}}^{-1} \boldsymbol{\Delta}_g\right\}$$

- la borne de Hammersley-Chapman-Robbins (BHCR) [Ham50] [CR51] est une version simplifiée à 1 point test de la BMS :

$$\min\left\{\mathbf{G}(\mathbf{u}_1)\right\} \text{ sous } \left| \begin{array}{l} \left\{\begin{array}{l} E_{\theta^0}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\right] = g\left(\theta^0\right) \\ E_{\theta^0+d\theta}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\right] = g\left(\theta^0 + d\theta\right) \end{array}\right. \\ \qquad\qquad \Updownarrow \\ \left\{\begin{array}{l} \langle \mathbf{u}_1 \mid \mathbf{c}_0\rangle_{\theta^0} = 0 \\ \langle \mathbf{u}_1 \mid \mathbf{c}_1\rangle_{\theta^0} = g\left(\theta^0 + d\theta\right) - g\left(\theta^0\right) \end{array}\right. \text{ où } \left| \begin{array}{l} \mathbf{c}_0 = 1(\mathbf{x}) \\ \mathbf{c}_1 = \upsilon_{\theta^0}\left(\mathbf{x}, \theta^0 + d\theta\right) \end{array}\right. \end{array}\right. \tag{43}$$

$$BHCR = \sup_{d\theta}\left\{\boldsymbol{\Delta}_g^T \mathbf{R}_{\upsilon_{\theta^0}}^{-1} \boldsymbol{\Delta}_g\right\} = \sup_{d\theta}\left\{\frac{\left(g\left(\theta^0 + d\theta\right) - g\left(\theta^0\right)\right)^2}{\displaystyle\int_\Omega \frac{p(\mathbf{x};\theta^0+d\theta)^2}{p(\mathbf{x};\theta^0)} d\mathbf{x} - 1}\right\}$$

- la borne de (Frechet-Darmois) Cramer-Rao (BCR) [Fre43] [Dar45] [Rao45] [Cra46] est la limite locale de la BHCR :

$$\lim_{d\theta \to 0} \min\left\{\mathbf{G}(\mathbf{u}_1)\right\} \text{ sous } \left| \begin{array}{l} \left\{\begin{array}{l} E_{\theta^0}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\right] = g\left(\theta^0\right) \\ E_{\theta^0+d\theta}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\right] = g\left(\theta^0 + d\theta\right) \end{array}\right. \\ \qquad\qquad \Updownarrow \\ \left\{\begin{array}{l} \langle \mathbf{u}_1 \mid \mathbf{c}_0\rangle_{\theta^0} = 0 \\ \langle \mathbf{u}_1 \mid \mathbf{c}_1\rangle_{\theta^0} = g\left(\theta^0 + d\theta\right) - g\left(\theta^0\right) \end{array}\right. \text{ où } \left| \begin{array}{l} \mathbf{c}_0 = 1(\mathbf{x}) \\ \mathbf{c}_1 = \upsilon_{\theta^0}\left(\mathbf{x}, \theta^0 + d\theta\right) \end{array}\right. \end{array}\right. \tag{44}$$

$$BCR=\lim_{d\theta\rightarrow 0}\left\{\frac{\left(g\left(\theta^{0}+d\theta\right)-g\left(\theta^{0}\right)\right)^{2}}{\int\limits_{\Omega}\frac{p(\mathbf{x};\theta^{0}+d\theta)^{2}}{p(\mathbf{x};\theta^{0})}d\mathbf{x}-1}\right\}=\frac{\left(\frac{\partial g(\theta^{0})}{\partial\theta}\right)^{2}}{E_{\theta^{0}}\left[\left(\frac{\partial\ln p(\mathbf{x};\theta^{0})}{\partial\theta}\right)^{2}\right]}$$

ce qui est équivalent à :

$$\min\left\{\mathbf{G}\left(\mathbf{u}_{1}\right)\right\}\ \text{sous}\ \left|\begin{array}{c}E_{\theta^{0}+d\theta}\left[\widehat{g\left(\theta^{0}\right)}\left(\mathbf{x}\right)\right]=g\left(\theta^{0}+d\theta\right)+o\left(d\theta\right)\\ \Updownarrow\\ \left\{\begin{array}{l}\langle\mathbf{u}_{1}\mid\mathbf{c}_{0}'\rangle_{\theta^{0}}=0\\ \langle\mathbf{u}_{1}\mid\mathbf{c}_{1}'\rangle_{\theta^{0}}=\frac{\partial g(\theta^{0})}{\partial\theta}\end{array}\right.\ \text{où}\left|\begin{array}{l}\mathbf{c}_{0}'=1\left(\mathbf{x}\right)\\ \mathbf{c}_{1}'=\frac{\partial\ln p(\mathbf{x};\theta^{0})}{\partial\theta}\end{array}\right.\end{array}\right. \tag{45}$$

• la borne de Battacharayya [Bat46] (BBa) à l'ordre $L\geq 2$ est une extension de la BCR en ce sens qu'elle est la solution de :

$$\min\left\{\mathbf{G}\left(\mathbf{u}_{1}\right)\right\}\ \text{sous}\ \left|\begin{array}{c}E_{\theta^{0}+d\theta}\left[\widehat{g\left(\theta^{0}\right)}\left(\mathbf{x}\right)\right]=g\left(\theta^{0}+d\theta\right)+o\left(d\theta^{L}\right)\\ \Updownarrow\\ \left\{\begin{array}{l}\langle\mathbf{u}_{1}\mid\mathbf{c}_{0}'\rangle_{\theta^{0}}=0\\ \langle\mathbf{u}_{1}\mid\mathbf{c}_{1}'\rangle_{\theta^{0}}=\frac{\partial g(\theta^{0})}{\partial\theta}\\ \langle\mathbf{u}_{1}\mid\mathbf{c}_{l}'\rangle_{\theta^{0}}=\frac{\partial^{l}g(\theta^{0})}{\partial^{l}\theta},l\in[2,L]\end{array}\right.\ \text{où}\left|\begin{array}{l}\mathbf{c}_{0}'=1\left(\mathbf{x}\right)\\ \mathbf{c}_{1}'=\frac{\partial\ln p(\mathbf{x};\theta^{0})}{\partial\theta}\\ \mathbf{c}_{l}'=\frac{1}{p(\mathbf{x};\theta^{0})}\frac{\partial^{l}p(\mathbf{x};\theta^{0})}{\partial^{l}\theta},l\in[2,L]\end{array}\right.\end{array}\right. \tag{46}$$

La borne de Fraser-Gutman [FG52] (BFG) est une généralisation la borne de Battacharayya (46) visant à relâcher les conditions d'existence de cette borne. Elle peut aussi être vue comme une réécriture (transformation linéaire bijective) de la BMS dont l'objectif est de faire apparaitre les différences finies convergeant vers les dérivées successive de $p\left(\mathbf{x};\theta\right)$ et $g\left(\theta\right)$ au voisinage de θ^{0} : la BFG est à la BBa ce que la BHCR est à la BCR.

• la borne de McAulay-Hofstetter [MH71] (BMH) consiste à combiner la BCR et la BMS sous la forme du système de contraintes suivantes :

$$\min\left\{\mathbf{G}\left(\mathbf{u}_{1}\right)\right\}\ \text{sous}\ \left|\begin{array}{c}\left\{\begin{array}{l}E_{\theta^{0}+d\theta}\left[\widehat{g\left(\theta^{0}\right)}\left(\mathbf{x}\right)\right]=g\left(\theta^{0}+d\theta\right)+o\left(d\theta\right)\\ E_{\theta^{i}}\left[\widehat{g\left(\theta^{0}\right)}\left(\mathbf{x}\right)\right]=g\left(\theta^{i}\right),i\in[1,I]\end{array}\right.\\ \Updownarrow\\ \left\{\begin{array}{ll}\langle\mathbf{u}_{1}\mid\mathbf{c}_{0}'\rangle_{\theta^{0}}&=0\\ \langle\mathbf{u}_{1}\mid\mathbf{c}_{1}'\rangle_{\theta^{0}}&=\frac{\partial g(\theta^{0})}{\partial\theta}\\ \left\langle\mathbf{u}_{1}\mid\mathbf{c}_{1+i}'\right\rangle_{\theta^{0}}&=g\left(\theta^{i}\right)-g\left(\theta^{0}\right),i\in[1,I]\end{array}\right.\ \text{où}\left|\begin{array}{l}\mathbf{c}_{0}'=1\left(\mathbf{x}\right)\\ \mathbf{c}_{1}'=\frac{\partial\ln p(\mathbf{x};\theta^{0})}{\partial\theta}\\ \mathbf{c}_{1+i}'=\upsilon_{\theta^{0}}\left(\mathbf{x},\theta^{i}\right),i\in[1,I]\end{array}\right.\end{array}\right. \tag{47}$$

• la borne de Glave (BG) [Gla72] [CGQL08][7] **[T4]** est une extension de la (BMH) :

$$\min\left\{\mathbf{G}\left(\mathbf{u}_{1}\right)\right\}\ \text{sous}\ \left|\begin{array}{c}\left\{\begin{array}{l}E_{\theta^{0}+d\theta}\left[\widehat{g\left(\theta^{0}\right)}\left(\mathbf{x}\right)\right]=g\left(\theta^{0}+d\theta\right)+o\left(d\theta\right)\\ E_{\theta^{i}+d\theta}\left[\widehat{g\left(\theta^{0}\right)}\left(\mathbf{x}\right)\right]=g\left(\theta^{i}+d\theta\right)+o\left(d\theta\right),i\in[1,I]\end{array}\right.\\ \Updownarrow\\ \left\{\begin{array}{ll}\langle\mathbf{u}_{1}\mid\mathbf{c}_{0}'\rangle_{\theta^{0}}&=0\\ \langle\mathbf{u}_{1}\mid\mathbf{c}_{1}'\rangle_{\theta^{0}}&=\frac{\partial g(\theta^{0})}{\partial\theta}\\ \left\langle\mathbf{u}_{1}\mid\mathbf{c}_{1+i}'\right\rangle_{\theta^{0}}&=g\left(\theta^{i}\right)-g\left(\theta^{0}\right),i\in[1,I]\\ \left\langle\mathbf{u}_{1}\mid\mathbf{c}_{(1+I)+i}'\right\rangle_{\theta^{0}}&=\frac{\partial g(\theta^{i})}{\partial\theta},i\in[1,I]\end{array}\right.\ \text{où}\left|\begin{array}{ll}\mathbf{c}_{0}'&=1\left(\mathbf{x}\right)\\ \mathbf{c}_{1}'&=\frac{\partial\ln p(\mathbf{x};\theta^{0})}{\partial\theta}\\ \mathbf{c}_{1+i}'&=\upsilon_{\theta^{0}}\left(\mathbf{x},\theta^{i}\right),i\in[1,I]\\ \mathbf{c}_{(1+I)+i}'&=\frac{\frac{\partial p\left(\mathbf{x};\theta^{i}\right)}{\partial\theta}}{p(\mathbf{x};\theta^{0})},i\in[1,I]\end{array}\right.\end{array}\right. \tag{48}$$

[7] Lors de la soumission de [CGQL08] nous n'avions pas lu la section (applicative) V de [Gla72] qui mentionne [Gla72, (40)] la borne proposée dans [CGQL08].

• la borne d'Abel [Abe93] (BA) consiste à combiner la BBa et la BMS sous la forme du système de contraintes suivantes:

$$\min\left\{\mathbf{G}\left(\mathbf{u}_1\right)\right\} \text{ sous } \left| \begin{array}{l} \left\{ \begin{array}{r} E_{\theta^0+d\theta}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\right] = g\left(\theta^0+d\theta\right)+o\left(d\theta^L\right) \\ E_{\theta^i}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\right] = g\left(\theta^i\right), i\in[1,I] \end{array}\right. \\ \Updownarrow \\ \left\{ \begin{array}{ll} \langle\mathbf{u}_1 \mid \mathbf{c}'_0\rangle_{\theta^0} & = 0 \\ \langle\mathbf{u}_1 \mid \mathbf{c}'_1\rangle_{\theta^0} & = \frac{\partial g(\theta^0)}{\partial\theta} \\ \langle\mathbf{u}_1 \mid \mathbf{c}'_l\rangle_{\theta^0} & = \frac{\partial^l g(\theta^0)}{\partial^l\theta}, l\in[2,L] \\ \left\langle\mathbf{u}_1 \mid \mathbf{c}'_{L+i}\right\rangle_{\theta^0} & = g\left(\theta^i\right)-g\left(\theta^0\right), i\in[1,I] \end{array}\right. \text{ où } \left| \begin{array}{ll} \mathbf{c}'_0 & = 1\left(\mathbf{x}\right) \\ \mathbf{c}'_1 & = \frac{\partial\ln p(\mathbf{x};\theta^0)}{\partial\theta} \\ \mathbf{c}'_l & = \frac{\frac{\partial^l p(\mathbf{x};\theta^0)}{\partial^l\theta}}{p(\mathbf{x};\theta^0)}, l\in[2,L] \\ \mathbf{c}'_{L+i} & = \upsilon_{\theta^0}\left(\mathbf{x},\theta^i\right), i\in[1,I] \end{array}\right. \end{array}\right. \tag{49}$$

• la borne de Todros et Tabrikian [TT08] [TT10-I] (BTT) est une variante de la BMH (cf. [MH71, (7)] et [TT08, (9)]) introduisant une transformation linéaire injective des contraintes de la BMS :

$$\min\left\{\mathbf{G}\left(\mathbf{u}_1\right)\right\} \text{ sous } \left| \begin{array}{l} \left\{ \begin{array}{l} E_{\theta^0+d\theta}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\right] = g\left(\theta^0+d\theta\right)+o\left(d\theta\right) \\ E_{\theta^0}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\left(\mathbf{W}^T\boldsymbol{\upsilon}_{\theta^0}\left(\mathbf{x},\left\{\theta^i\right\}_1^I\right)\right)\right] = \mathbf{W}^T\boldsymbol{\Delta}_g \end{array}\right. , \mathbf{W} = \left[\mathbf{w}_1 \ \ldots \ \mathbf{w}_K\right], K\leq I \\ \Updownarrow \\ \left\{ \begin{array}{ll} \langle\mathbf{u}_1 \mid \mathbf{c}'_0\rangle_{\theta^0} & = 0 \\ \langle\mathbf{u}_1 \mid \mathbf{c}'_1\rangle_{\theta^0} & = \frac{\partial g(\theta^0)}{\partial\theta} \\ \left\langle\mathbf{u}_1 \mid \mathbf{c}'_{1+k}\right\rangle_{\theta^0} & = \mathbf{w}_k^T\boldsymbol{\Delta}_g, k\in[1,K] \end{array}\right. \text{ où } \left| \begin{array}{l} \mathbf{c}'_0 = 1\left(\mathbf{x}\right) \\ \mathbf{c}'_1 = \frac{\partial\ln p(\mathbf{x};\theta^0)}{\partial\theta} \\ \mathbf{c}'_{1+i} = \mathbf{w}_k^T\boldsymbol{\upsilon}_{\theta^0}\left(\mathbf{x},\left\{\theta^i\right\}_1^I\right), k\in[1,K] \end{array}\right. \end{array}\right. \tag{50}$$

où $\mathbf{w}_k$ est un vecteur colonne de la matrice de TFD (Transformée de Fourrier Discrète). Dans le cas où $K = I$, alors $\mathbf{W}$ est la matrice de TFD (bijective) et BTT = BMH. Par conséquent l'intérêt de la BTT est de proposer une optimisation du compromis puissance de calcul - proximité à la BB (ici la BMH) en choisissant de manière ad hoc les K transformations linéaires $\{\mathbf{w}_k\}_1^K$ de façon à ce que $K \ll I$ et BTT→BMH. Si le principe mérite d'être souligné, il n'existe pas à ce jour de stratégie générale connue pour choisir les K transformations linéaires $\{\mathbf{w}_k\}_1^K$. De plus, une lecture attentive des propriétés attendues de la matrice $\mathbf{W}^T\mathbf{R}_{\upsilon_{\theta^0}}\mathbf{W}$ [TT08, Fig 1] [TT10-I] [EKRBM13] montre qu'elles ne sont pas aisément transposables au cas d'un mélange de 2 sources ... ce qui fait de cette borne une "piste" théorique de recherche et non pas une borne applicative générale (il est alors plus sur de prendre directement la BMH).

L'avantage de l'approche basée sur la minimisation d'une norme sous contraintes linéaires (Lemme 1-p27) est que la borne obtenue ne dépend que des contraintes : il suffit donc de comparer les ensembles de contraintes pour chaque borne pour pouvoir les classer. Notons symboliquement $\mathcal{C}_B$ l'ensemble des contraintes associées à la borne B. Puisque, pour un nombre maximal de I points test $\left\{\theta^i\right\}_1^I$:

$$\left| \begin{array}{l} \mathcal{C}_{BCR} \subset \mathcal{C}_{BHCR} \subset \mathcal{C}_{BMS} = \mathcal{C}_{BFG} \subset \mathcal{C}_{BMH} \subset \mathcal{C}_{BG} \\ \mathcal{C}_{BCR} \subset \mathcal{C}_{BTT} \subset \mathcal{C}_{BMH} \subset \mathcal{C}_{BG} \\ \mathcal{C}_{BCR} \subset \mathcal{C}_{BBa} \subset \mathcal{C}_{BMS} = \mathcal{C}_{BFG} \subset \mathcal{C}_{BA} \end{array}\right.$$

par conséquent, pour un nombre maximal de I points test $\left\{\theta^i\right\}_1^I$ (partiellement obtenu dans [TT10-ISIT]) :

$$\left| \begin{array}{r} BG \geq BMH \geq BMS = BFG \geq BHCR \geq BCR \\ BG \geq BMH \geq BTT \geq BCR \\ BA \geq BMS = BFG \geq BBa \geq BCR \end{array}\right. \tag{51}$$

ce qui fait apparaître simplement que la BCR est la plus "petite" des bornes inférieures de l'EQM.
On peut distinguer les bornes inférieures (BCR, BBa) qui caractérisent les estimateurs sans biais localement au voisinage de θ^0 ("locally best locally unbiased") et qui correspondent à la satisfaction de la contrainte :

$$E_{\theta^0+d\theta}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\right] = g\left(\theta^0+d\theta\right)+o\left(d\theta^L\right) = g(\theta^0)+\sum_{l=1}^{L}\frac{\partial^l g(\theta^0)}{\partial\theta^l}\frac{d\theta^l}{l!}+o_{\theta^0}\left(d\theta^L\right), \tag{52}$$

des autres bornes (BHCR, BMS, BFG, BMH, BG, BA, BTT) qui cherchent à caractériser les estimateurs uniformément sans biais sur tout l'espace des paramètres Θ ("uniformly unbiased" or "unbiased").
En fait toutes les bornes existantes peuvent être vues [Gla72] [Abe93] [CGQL08] **[T4]** comme une tentative d'approximer la contrainte (uniformément) sans biais :

$$E_\theta\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right]=g\left(\theta\right),\quad\forall\theta\in\Theta \tag{53}$$

par morceau sur Θ : $\Theta=\bigcup_{i=1}^{I}S^i$, $S^i=\left[\theta^i-\left(\frac{\theta^i-\theta^{i-1}}{2}\right),\theta^i+\frac{\theta^{i+1}-\theta^i}{2}\right]$ et :

$$E_{\theta^i+d\theta}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right]=g\left(\theta^i+d\theta\right)+o\left(d\theta^{L^i}\right)\Leftrightarrow\left\{\begin{array}{l}\left\langle\mathbf{u}_1\mid\mathbf{c}_0^i\right\rangle_{\theta^0}=g\left(\theta^i\right)-g\left(\theta^0\right)\\\left\langle\mathbf{u}_1\mid\mathbf{c}_l^i\right\rangle_{\theta^0}=\frac{\partial^l g(\theta^i)}{\partial^l\theta},l\in\left[1,L^i\right]\end{array}\right.\text{ où }\left|\begin{array}{l}\mathbf{c}_0^i=\upsilon_{\theta^0}\left(\mathbf{x},\theta^i\right)\\\mathbf{c}_l^i=\frac{1}{p(\mathbf{x};\theta^0)}\frac{\partial^l p(\mathbf{x};\theta^i)}{\partial^l\theta},l\in\left[1,L^i\right]\end{array}\right.$$

Cette approximation de la contrainte sans biais (53) fournie $L=\sum_{i=1}^{I}\left(L^i+1\right)$ contraintes linéaires conduisant à une suite d'approximations de la BB (Lemme 1-p27) [Gla72] [Abe93] [CGQL08] **[T4]** :

$$\mathrm{BB}_{L^1,\ldots,L^I}^{S^1,\ldots,S^I}=\mathbf{v}^T\mathbf{G}_{\theta^0}^{-1}\mathbf{v},\qquad\mathbf{v}^T=\left[\left(\mathbf{v}^1\right)^T,\ldots,\left(\mathbf{v}^I\right)^T\right],\quad\mathbf{G}_{\theta^0}=E_{\theta^0}\left[\mathbf{c}\mathbf{c}^T\right],\quad\mathbf{c}^T=\left[\left(\mathbf{c}^1\right)^T,\ldots,\left(\mathbf{c}^I\right)^T\right] \tag{54}$$

$$\mathbf{v}^i=\left[g(\theta^i)-g(\theta^0),\frac{\partial g(\theta^i)}{\partial\theta},\ldots,\frac{\partial^{L^i}g(\theta^i)}{\partial\theta^{L^i}}\right]^T,\quad\mathbf{c}^i=\frac{1}{p\left(\mathbf{x};\theta^0\right)}\left[p\left(\mathbf{x};\theta^i\right),\frac{\partial p\left(\mathbf{x};\theta^i\right)}{\partial\theta},\ldots,\frac{\partial^{L^i}p\left(\mathbf{x};\theta^i\right)}{\partial\theta^{L^i}}\right]^T$$

laquelle converge vers la BB lorsque $\underset{1\leq i\leq I}{min}\left(L^i\right)\to\infty$. L'idée sous jacente est de trouver la meilleure approximation de (53) la plus compacte (L minimal), sachant que l'estimateur optimal solution de (54) vérifie :

$$\widehat{g\left(\theta^0\right)}_{opt}\left(\mathbf{x}\right)\;=\;g(\theta^0)+\sum_{i=1}^{I}\sum_{l=0}^{L^i}\alpha_{i,l}\mathbf{c}_l^i\left(\mathbf{x}\right),\quad\boldsymbol{\alpha}=\mathbf{G}_{\theta^0}^{-1}\mathbf{v} \tag{55}$$

$$E_\theta\left[\widehat{g\left(\theta^0\right)}_{opt}\left(\mathbf{x}\right)\right]-g\left(\theta\right)\;=\;m_{opt}\left(\theta\right)-g\left(\theta\right),\quad m_{opt}\left(\theta\right)=\sum_{i=1}^{I}\sum_{l=0}^{L^i}\alpha_{i,l}E_\theta\left[\mathbf{c}_l^i\left(\mathbf{x}\right)\right]. \tag{56}$$

Par conséquent si on désire obtenir une bonne approximation (56) de $g\left(\theta\right)$ par $m_{opt}\left(\theta\right)$ il semble judicieux de doter $m_{opt}\left(\theta\right)$ des mêmes conditions de régularité que $g\left(\theta\right)$: continuité, dérivabilité (éventuellement aux ordres supérieures), ce que traduit la BG (cf. figure 1-p32). Selon les résultats disponibles dans la littérature courante [Abe93] [FL02], la prise en compte des dérivées d'ordres supérieurs pour des valeurs modérées de L^i $\left(2\leq L^i\leq 4\right)$ occasionne une augmentation marginale de l'approximation $\mathrm{BB}_{L^1,\ldots,L^I}^{S^1,\ldots,S^I}$ par rapport à une augmentation du nombre I d'intervalles S^i. Par conséquent à nombre de contraintes L fixé, l'expérience acquise tend à montrer que la meilleure approximation disponible de la BB est la BG. Les composantes de $\mathbf{v}^T\left(\mathbf{G}_{\boldsymbol{\theta}^0}^G\right)^{-1}\mathbf{v}$ (54), à savoir $\mathbf{v}$ et $\mathbf{G}_{\boldsymbol{\theta}^0}^G$, permettant l'évaluation de la BG ont été calculées [CGQL08, Appendix] dans le cas général d'un vecteur $\mathbf{g}\left(\boldsymbol{\theta}\right)$ de Q fonctions dépendant d'un vecteur $\boldsymbol{\theta}$ de P paramètres inconnus (62), lorsque le vecteur d'observation est gaussien réel :

$$\mathbf{x}'\sim\mathcal{N}_{2N}\left(\mathbf{m}_{\mathbf{x}'}\left(\boldsymbol{\theta}\right),\mathbf{C}_{\mathbf{x}'}\left(\boldsymbol{\theta}\right)\right),\quad p\left(\mathbf{x}';\mathbf{m}_{\mathbf{x}'}\left(\boldsymbol{\theta}\right),\mathbf{C}_{\mathbf{x}'}\left(\boldsymbol{\theta}\right)\right)=\frac{e^{-\frac{1}{2}\left(\mathbf{x}'-\mathbf{m}_{\mathbf{x}'}(\boldsymbol{\theta})\right)^T\mathbf{C}_{\mathbf{x}'}^{-1}(\boldsymbol{\theta})\left(\mathbf{x}'-\mathbf{m}_{\mathbf{x}'}(\boldsymbol{\theta})\right)}}{\sqrt{2\pi}^{2N}\sqrt{\left|\mathbf{C}_{\mathbf{x}'}\left(\boldsymbol{\theta}\right)\right|}}, \tag{57}$$

ce qui inclut le modèle d'observation gaussien complexe circulaire $\left(\mathbf{x}'=\left(\mathrm{Re}\left\{x_1\right\},\ldots,\mathrm{Re}\left\{x_N\right\},\mathrm{Im}\left\{x_1\right\},\ldots,\mathrm{Im}\left\{x_N\right\}\right)^T\right)$:

$$\mathbf{x}\sim\mathcal{CN}_N\left(\mathbf{m}_{\mathbf{x}}\left(\boldsymbol{\theta}\right),\mathbf{C}_{\mathbf{x}}\left(\boldsymbol{\theta}\right)\right),\quad p\left(\mathbf{x};\mathbf{m}_{\mathbf{x}}\left(\boldsymbol{\theta}\right),\mathbf{C}_{\mathbf{x}}\left(\boldsymbol{\theta}\right)\right)=\frac{e^{-\left(\mathbf{x}-\mathbf{m}_{\mathbf{x}}(\boldsymbol{\theta})\right)^H\mathbf{C}_{\mathbf{x}}^{-1}(\boldsymbol{\theta})\left(\mathbf{x}-\mathbf{m}_{\mathbf{x}}(\boldsymbol{\theta})\right)}}{\pi^M\left|\mathbf{C}_{\mathbf{x}}\left(\boldsymbol{\theta}\right)\right|}. \tag{58}$$

Les différents termes contenus dans $\mathbf{G}_{\boldsymbol{\theta}^0}^G$ conduisant à la BG [CGQL08, Appendix] **[T4]** permettent également de formuler pour les différents modèles gaussiens (57)(58) la BCR, la BHCR, la BMS (64)[8], la BFG, la BMH (64) et la BTT (cf. Annexe IV-F-p94).

[8][TK99] contient le calcul des composantes de la BMS uniquement pour les modèles Gaussiens ciculaires déterministe $\left(\mathbf{x}\sim\mathcal{CN}_N\left(\mathbf{m}_{\mathbf{x}}\left(\boldsymbol{\theta}\right),\mathbf{C}_{\mathbf{x}}\right)\right)$ ou stochastique $\left(\mathbf{x}\sim\mathcal{CN}_N\left(\mathbf{m}_{\mathbf{x}},\mathbf{C}_{\mathbf{x}}\left(\boldsymbol{\theta}\right)\right)\right)$

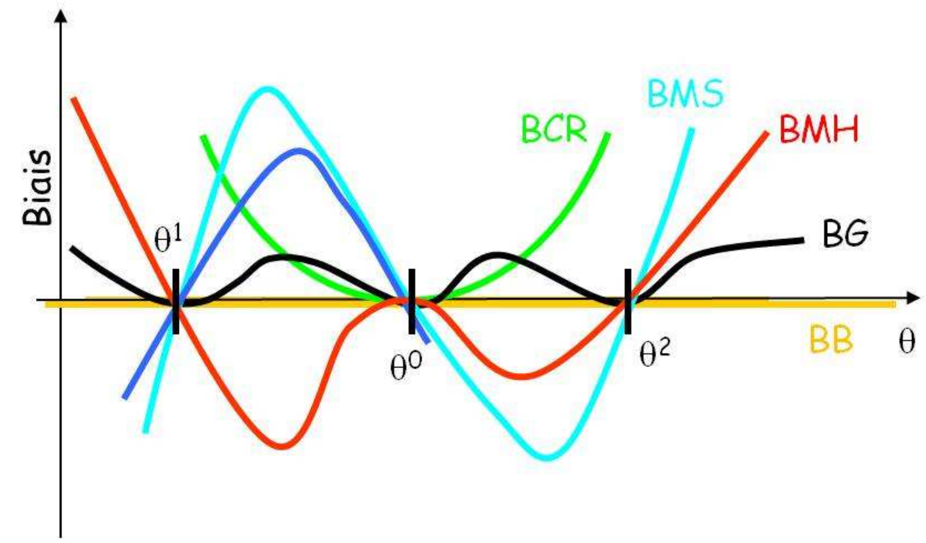


Fig. 1. Illustration des contraintes sur le biais prises en compte par différentes bornes inférieures en θ^0 pour 2 points test $\{\theta^1, \theta^2\}$

c) Extensions :

Cas multiparamètres

Une particularité fondamentale de la recherche d'un estimateur (*a priori*) localement le meilleur à l'ordre 2 est la bilinéarité du produit scalaire associé à la norme (22), laquelle à travers l'égalité :

$$\forall \boldsymbol{\lambda} \in \mathbb{R}^Q,\ \mathbf{G}_{\boldsymbol{\theta}^0}\left(\{\mathbf{u}_1'\}\right) = \left\| \sum_{q=1}^{Q} \lambda_q \mathbf{u}_q \right\|_{2;\boldsymbol{\theta}^0}^2 = \boldsymbol{\lambda}^T \mathbf{G}_{\boldsymbol{\theta}^0}\left(\{\mathbf{u}_q\}_1^Q\right) \boldsymbol{\lambda} \text{ où } \mathbf{u}_1' = \sum_{q=1}^{Q} \lambda_q \mathbf{u}_q,\ \left\{\mathbf{u}_q \triangleq \widehat{g_q\left(\boldsymbol{\theta}^0\right)}(\mathbf{x}) - g_q\left(\boldsymbol{\theta}^0\right)\right\}_1^Q,$$

exprime la possibilité d'une minimisation de la norme de toute combinaison linéaire, permettant ainsi d'accéder à la minimisation conjointe de la norme de chaque composante de $\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x}) - \mathbf{g}\left(\boldsymbol{\theta}^0\right)$, ce que traduisent les lemmes $1, 2, 3$-p27 relatifs à la minimisation des matrices de Gram $\mathbf{G}_{\boldsymbol{\theta}^0}\left(\{\mathbf{u}_q\}_1^Q\right)$ (résultat dont l'extension au cas général $s > 1$ n'est pas disponible à ma connaissance).
La généralisation des approximations de la BB (BCR, BHCR, BMS, ...) au cas général se fait par adaptation des contraintes les définissant. Par exemple, la contrainte définissant la BCR (45) devient naturellement :

$$E_{\boldsymbol{\theta}^0+d\boldsymbol{\theta}}\left[\widehat{g_q\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})\right] = g_q\left(\boldsymbol{\theta}^0 + d\boldsymbol{\theta}\right) + o\left(\|d\boldsymbol{\theta}\|\right), \quad \forall q \in [1, Q]$$

soit encore :

$$E_{\boldsymbol{\theta}^0+d\boldsymbol{\theta}}\left[\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})\right] = \mathbf{g}\left(\boldsymbol{\theta}^0 + d\boldsymbol{\theta}\right) + \mathbf{o}\left(\|d\boldsymbol{\theta}\|\right) = \mathbf{g}\left(\boldsymbol{\theta}^0\right) + \frac{\partial \mathbf{g}(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}^T} d\boldsymbol{\theta} + \mathbf{o}\left(\|d\boldsymbol{\theta}\|\right)$$

$$\Updownarrow$$

$$\begin{cases} E_{\boldsymbol{\theta}^0}\left[\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x}) - \mathbf{g}\left(\boldsymbol{\theta}^0\right)\right)^T\right] = \mathbf{0}^T \\ E_{\boldsymbol{\theta}^0}\left[\left(\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}}\right)\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x}) - \mathbf{g}\left(\boldsymbol{\theta}^0\right)\right)^T\right] = \left(\frac{\partial \mathbf{g}(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}^T}\right)^T \end{cases} \tag{59}$$

conduisant à (Lemmes $1, 2$-p27) :

$$\mathbf{BCR}_{\mathbf{g}|\boldsymbol{\theta}}\left(\boldsymbol{\theta}^0\right) = \frac{\partial \mathbf{g}\left(\boldsymbol{\theta}^0\right)}{\partial \boldsymbol{\theta}^T}\left(\mathbf{G}_{\boldsymbol{\theta}^0}^{CR}\right)^{-1}\left(\frac{\partial \mathbf{g}\left(\boldsymbol{\theta}^0\right)}{\partial \boldsymbol{\theta}^T}\right)^T, \quad \mathbf{G}_{\boldsymbol{\theta}^0}^{CR} = E_{\boldsymbol{\theta}^0}\left[\frac{\partial \ln p\left(\mathbf{x};\boldsymbol{\theta}^0\right)}{\partial \boldsymbol{\theta}}\frac{\partial \ln p\left(\mathbf{x};\boldsymbol{\theta}^0\right)}{\partial \boldsymbol{\theta}^T}\right] \triangleq \mathbf{F}_{\boldsymbol{\theta}^0} \tag{60}$$

où $\mathbf{F}_{\boldsymbol{\theta}^0}$ est la matrice d'information de Fisher (MIF) [Fis21]. De même les contraintes définissant la BG (48) deviennent (59) et (61) :

$$E_{\boldsymbol{\theta}^i+d\boldsymbol{\theta}}\left[\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})\right] = \mathbf{g}\left(\boldsymbol{\theta}^i + d\boldsymbol{\theta}\right) + \mathbf{o}_i\left(\|d\boldsymbol{\theta}\|\right)$$
$$\Updownarrow$$
$$\begin{cases} E_{\boldsymbol{\theta}^0}\left[\frac{p(\mathbf{x};\boldsymbol{\theta}^i)}{p(\mathbf{x};\boldsymbol{\theta}^0)}\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x}) - \mathbf{g}\left(\boldsymbol{\theta}^0\right)\right)^T\right] = \left(\mathbf{g}\left(\boldsymbol{\theta}^i\right) - \mathbf{g}\left(\boldsymbol{\theta}^0\right)\right)^T \\ E_{\boldsymbol{\theta}^0}\left[\left(\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta}^i)}{\partial \boldsymbol{\theta}}\frac{p(\mathbf{x};\boldsymbol{\theta}^i)}{p(\mathbf{x};\boldsymbol{\theta}^0)}\right)\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x}) - \mathbf{g}\left(\boldsymbol{\theta}^0\right)\right)^T\right] = \left(\frac{\partial \mathbf{g}(\boldsymbol{\theta}^i)}{\partial \boldsymbol{\theta}^T}\right)^T \end{cases} \quad (61)$$

conduisant à (Lemmes 1-p27) [CGQL08] **[T4]** :

$$\mathbf{BG}_{\mathbf{g}|\boldsymbol{\theta}}\left(\boldsymbol{\theta}^0\right) = \mathbf{V}^T\left(\mathbf{G}^G_{\boldsymbol{\theta}^0}\right)^{-1}\mathbf{V}, \quad \mathbf{G}^G_{\boldsymbol{\theta}^0} = \begin{bmatrix} \mathbf{G}^{MS}_{\boldsymbol{\theta}^0} & \mathbf{H}^T_{\boldsymbol{\theta}^0} \\ \mathbf{H}_{\boldsymbol{\theta}^0} & \mathbf{FE}_{\boldsymbol{\theta}^0} \end{bmatrix}, \left| \begin{array}{l} \mathbf{V} = \left[\boldsymbol{\Delta}\mathbf{g}, \frac{\partial \mathbf{g}(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}^T}, \ldots, \frac{\partial \mathbf{g}(\boldsymbol{\theta}^I)}{\partial \boldsymbol{\theta}^T}\right]^T \\ \boldsymbol{\Delta}\mathbf{g} = \left[\mathbf{g}\left(\boldsymbol{\theta}^0\right) - \mathbf{g}\left(\boldsymbol{\theta}^0\right), \ldots, \mathbf{g}\left(\boldsymbol{\theta}^N\right) - \mathbf{g}\left(\boldsymbol{\theta}^I\right)\right] \end{array}\right. \quad (62)$$

et :

$$\mathbf{G}^{MS}_{\boldsymbol{\theta}^0} = E_{\boldsymbol{\theta}^0}\left[\begin{pmatrix} \frac{p(\mathbf{x};\boldsymbol{\theta}^0)}{p(\mathbf{x};\boldsymbol{\theta}^0)} \\ \vdots \\ \frac{p(\mathbf{x};\boldsymbol{\theta}^I)}{p(\mathbf{x};\boldsymbol{\theta}^0)} \end{pmatrix}\begin{pmatrix} \frac{p(\mathbf{x};\boldsymbol{\theta}^0)}{p(\mathbf{x};\boldsymbol{\theta}^0)} \\ \vdots \\ \frac{p(\mathbf{x};\boldsymbol{\theta}^I)}{p(\mathbf{x};\boldsymbol{\theta}^0)} \end{pmatrix}^T\right], \quad \mathbf{H}_{\boldsymbol{\theta}^0} = E_{\boldsymbol{\theta}^0}\left[\begin{pmatrix} \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}}\frac{p(\mathbf{x};\boldsymbol{\theta}^0)}{p(\mathbf{x};\boldsymbol{\theta}^0)} \\ \vdots \\ \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta}^I)}{\partial \boldsymbol{\theta}}\frac{p(\mathbf{x};\boldsymbol{\theta}^I)}{p(\mathbf{x};\boldsymbol{\theta}^0)} \end{pmatrix}\begin{pmatrix} \frac{p(\mathbf{x};\boldsymbol{\theta}^0)}{p(\mathbf{x};\boldsymbol{\theta}^0)} \\ \vdots \\ \frac{p(\mathbf{x};\boldsymbol{\theta}^I)}{p(\mathbf{x};\boldsymbol{\theta}^0)} \end{pmatrix}^T\right]$$

$$\mathbf{FE}_{\boldsymbol{\theta}^0} = E_{\boldsymbol{\theta}^0}\left[\begin{pmatrix} \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}}\frac{p(\mathbf{x};\boldsymbol{\theta}^0)}{p(\mathbf{x};\boldsymbol{\theta}^0)} \\ \vdots \\ \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta}^I)}{\partial \boldsymbol{\theta}}\frac{p(\mathbf{x};\boldsymbol{\theta}^I)}{p(\mathbf{x};\boldsymbol{\theta}^0)} \end{pmatrix}\begin{pmatrix} \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}}\frac{p(\mathbf{x};\boldsymbol{\theta}^0)}{p(\mathbf{x};\boldsymbol{\theta}^0)} \\ \vdots \\ \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta}^I)}{\partial \boldsymbol{\theta}}\frac{p(\mathbf{x};\boldsymbol{\theta}^I)}{p(\mathbf{x};\boldsymbol{\theta}^0)} \end{pmatrix}^T\right]$$

La BMS et la BMH se déduisent directement de (62) :

$$\mathbf{BMS}_{\mathbf{g}|\boldsymbol{\theta}}\left(\boldsymbol{\theta}^0\right) = \boldsymbol{\Delta}\mathbf{g}\left(\mathbf{G}^{MS}_{\boldsymbol{\theta}^0}\right)^{-1}\boldsymbol{\Delta}\mathbf{g}^T, \quad \begin{array}{c} \mathbf{BMH}_{\mathbf{g}|\boldsymbol{\theta}}\left(\boldsymbol{\theta}^0\right) = \left[\boldsymbol{\Delta}\mathbf{g}, \frac{\partial \mathbf{g}(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}^T}\right]\left(\mathbf{G}^{MH}_{\boldsymbol{\theta}^0}\right)^{-1}\left[\boldsymbol{\Delta}\mathbf{g}, \frac{\partial \mathbf{g}(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}^T}\right]^T \\ \mathbf{G}^{MH}_{\boldsymbol{\theta}^0} = \begin{bmatrix} \mathbf{G}^{MS}_{\boldsymbol{\theta}^0} & \mathbf{H}^T_{\boldsymbol{\theta}^0} \\ \mathbf{H}_{\boldsymbol{\theta}^0} & \mathbf{G}^{CR}_{\boldsymbol{\theta}^0} \end{bmatrix} \end{array} \quad (64)$$

Estimateurs avec biais

La prise en compte d'un biais découle naturellement de la définition générale de la précision *a priori* d'estimation (9) vérifiant pour $s = 2$:

$$1 - \left\|\mathbf{D}^{-1}_{(\boldsymbol{\xi}^+ + \boldsymbol{\xi}^-)/2}\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x}) - \mathbf{g}\left(\boldsymbol{\theta}^0\right) - \frac{\boldsymbol{\xi}^+ - \boldsymbol{\xi}^-}{2}\right)\right\|^2_{2;\boldsymbol{\theta}^0} \leq \mathcal{O}_{\boldsymbol{\theta}^0}\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}, \boldsymbol{\xi}^-, \boldsymbol{\xi}^+\right) \leq 1 \quad (65)$$

soit encore, en intégrant $\mathbf{D}_{(\boldsymbol{\xi}^+ + \boldsymbol{\xi}^-)/2}$ à la définition de la norme :

$$1 - \left\|\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x}) - \mathbf{g}\left(\boldsymbol{\theta}^0\right) - \frac{\boldsymbol{\xi}^+ - \boldsymbol{\xi}^-}{2}\right\|^2_{2;(\boldsymbol{\xi}^+ + \boldsymbol{\xi}^-)/2,\boldsymbol{\theta}^0} \leq \mathcal{O}_{\boldsymbol{\theta}^0}\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}, \boldsymbol{\xi}^-, \boldsymbol{\xi}^+\right) \leq 1 \quad (66)$$

$$\|\mathbf{h}(\mathbf{x})\|_{2;(\boldsymbol{\xi}^+ + \boldsymbol{\xi}^-)/2,\boldsymbol{\theta}^0} = E_{\boldsymbol{\theta}^0}\left[\|\mathbf{h}(\mathbf{x})\|^2_{2;(\boldsymbol{\xi}^+ + \boldsymbol{\xi}^-)/2}\right]^{\frac{1}{2}}, \quad \|\mathbf{h}(\mathbf{x})\|_{2;(\boldsymbol{\xi}^+ + \boldsymbol{\xi}^-)/2} = \left\|\mathbf{D}^{-1}_{(\boldsymbol{\xi}^+ + \boldsymbol{\xi}^-)/2}\mathbf{h}(\mathbf{x})\right\|_2$$

Comme déjà mentionné, les d.d.p.. de chaque $\widehat{g_q\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})$ n'ont aucune raison d'être symétriques autour de $g_q\left(\boldsymbol{\theta}^0\right)$ dans le cas général (cf. les estimateurs efficaces Annexe IV-B-p66 ou les estimateurs atteignant la BBa [Mor83] [Pom03]). Par conséquent, la définition ou la recherche de la meilleure précision peut nécessiter la prise en compte d'intervalles d'erreur non symétriques $\left(\boldsymbol{\xi}^{-} \neq \boldsymbol{\xi}^{+}\right)$ et aboutir à la recherche d'un estimateur localement le meilleur biaisé :

$$\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}_{opt}(\mathbf{x}) = \arg\min\left\{\left\|\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x}) - \mathbf{g}\left(\boldsymbol{\theta}^{0}\right) - \frac{\boldsymbol{\xi}_{\boldsymbol{\theta}^{0}}^{+} - \boldsymbol{\xi}_{\boldsymbol{\theta}^{0}}^{-}}{2}\right\|_{2;\left(\boldsymbol{\xi}_{\boldsymbol{\theta}^{0}}^{+}+\boldsymbol{\xi}_{\boldsymbol{\theta}^{0}}^{-}\right)/2,\boldsymbol{\theta}^{0}}^{2}\right\} \tag{67}$$

puisque rien n'interdit de définir une exigence de précision exhaustive dépendant de la valeur $\boldsymbol{\theta}^{0}$ du vecteur de paramètres $\boldsymbol{\theta}$: $\left(\boldsymbol{\xi}_{\boldsymbol{\theta}^{0}}^{+}, \boldsymbol{\xi}_{\boldsymbol{\theta}^{0}}^{-}\right)$. D'un point de vue théorique, il est important de noter que tous les résultats présentés préalablement pour $s = 2$ ne dépendent que des propriétés générales des normes associées à un produit scalaire (22); ils s'appliquent donc également à (65)(67) sous la forme :

$$\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}_{opt}(\mathbf{x}) = \arg\min\left\{\left\|\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x}) - \mathbf{h}\left(\boldsymbol{\theta}^{0}\right)\right\|_{2;\left(\boldsymbol{\xi}_{\boldsymbol{\theta}^{0}}^{+}+\boldsymbol{\xi}_{\boldsymbol{\theta}^{0}}^{-}\right)/2,\boldsymbol{\theta}^{0}}^{2}\right\}, \quad \mathbf{h}(\boldsymbol{\theta}) = \mathbf{g}(\boldsymbol{\theta}) + \mathbf{b}(\boldsymbol{\theta}), \ \mathbf{b}(\boldsymbol{\theta}) = \frac{\boldsymbol{\xi}_{\boldsymbol{\theta}}^{+} - \boldsymbol{\xi}_{\boldsymbol{\theta}}^{-}}{2}$$

avec :

$$\left\|\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x}) - \mathbf{h}\left(\boldsymbol{\theta}^{0}\right)\right\|_{2;\left(\boldsymbol{\xi}_{\boldsymbol{\theta}^{0}}^{+}+\boldsymbol{\xi}_{\boldsymbol{\theta}^{0}}^{-}\right)/2,\boldsymbol{\theta}^{0}}^{2} = \begin{array}{c} \left\|E_{\boldsymbol{\theta}^{0}}\left[\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x})\right] - \mathbf{h}\left(\boldsymbol{\theta}^{0}\right)\right\|_{2;\left(\boldsymbol{\xi}_{\boldsymbol{\theta}^{0}}^{+}+\boldsymbol{\xi}_{\boldsymbol{\theta}^{0}}^{-}\right)/2}^{2} + \\ \left\|\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x}) - E_{\boldsymbol{\theta}^{0}}\left[\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x})\right]\right\|_{2;\left(\boldsymbol{\xi}_{\boldsymbol{\theta}^{0}}^{+}+\boldsymbol{\xi}_{\boldsymbol{\theta}^{0}}^{-}\right)/2,\boldsymbol{\theta}^{0}}^{2} \end{array},$$

ce qui fait apparaître le compromis "biais-variance" dans la recherche de (67), en ce sens qu'il n'existe pas de résultat général permettant d'affirmer que $\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}_{opt}(\mathbf{x})$ est obtenu pour $E_{\boldsymbol{\theta}^{0}}\left[\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x})\right] - \mathbf{h}\left(\boldsymbol{\theta}^{0}\right) = \mathbf{0}$ ou pour $E_{\boldsymbol{\theta}^{0}}\left[\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x})\right] - \mathbf{g}\left(\boldsymbol{\theta}^{0}\right) = \mathbf{0}$.
Au contraire, depuis l'exemple célèbre de Stein [Ste56] généralisé dans [JS61] (et exemples pratiques dans [EM77]), il est acquis que pour certaines familles de d.d.p.. $p(\mathbf{x};\boldsymbol{\theta})$ ce n'est pas l'estimateur sans biais $\left(E_{\boldsymbol{\theta}}\left[\widehat{\boldsymbol{\theta}^{0}}(\mathbf{x})\right] = \boldsymbol{\theta}\right)$ qui fourni l'EQM $\left(\left\|\widehat{\boldsymbol{\theta}^{0}}(\mathbf{x}) - \boldsymbol{\theta}^{0}\right\|_{2;\boldsymbol{\theta}^{0}}^{2}\right)$ minimale lorsque $\boldsymbol{\xi}_{\boldsymbol{\theta}}^{+} = \boldsymbol{\xi}_{\boldsymbol{\theta}}^{-} = \boldsymbol{\xi} = \xi\mathbf{1}_{Q}$.

Par conséquent différents auteurs ont essayé de caractériser l'estimateur biaisé localement le meilleur au sens de :

$$\begin{aligned} \widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}_{opt}(\mathbf{x}) &= \arg\min\left\{\left\|\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x}) - \mathbf{g}\left(\boldsymbol{\theta}^{0}\right)\right\|_{2;\boldsymbol{\theta}^{0}}^{2}\right\} \\ &= \arg\min\left\{\left\|E_{\boldsymbol{\theta}^{0}}\left[\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x})\right] - \mathbf{g}\left(\boldsymbol{\theta}^{0}\right)\right\|_{2}^{2} + \left\|\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x}) - E_{\boldsymbol{\theta}^{0}}\left[\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x})\right]\right\|_{2;\boldsymbol{\theta}^{0}}^{2}\right\} \end{aligned}$$

que ce soit en prenant en compte une contrainte de biais uniforme $\left(E_{\boldsymbol{\theta}}\left[\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x})\right] = \mathbf{g}(\boldsymbol{\theta}) + \mathbf{b}(\boldsymbol{\theta})\right)$ [Alb73] [TT10-II] où une contrainte de biais local $\left(E_{\boldsymbol{\theta}^{0\cdot}+d\boldsymbol{\theta}}\left[\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x})\right] = \mathbf{g}\left(\boldsymbol{\theta}^{0\cdot} + d\boldsymbol{\theta}\right) + \mathbf{b}\left(\boldsymbol{\theta}^{0\cdot} + d\boldsymbol{\theta}\right) + \mathbf{o}\left(\|d\boldsymbol{\theta}\|\right)\right)$ en exploitant la BCR avec biais (60) [HFU96] [Eld06] [KE08] :

$$\left\|\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}(\mathbf{x}) - \mathbf{g}\left(\boldsymbol{\theta}^{0}\right)\right\|_{2;\boldsymbol{\theta}^{0}}^{2} \geq \left\|\mathbf{b}\left(\boldsymbol{\theta}^{0}\right)\right\|_{2}^{2} + tr\left(\left(\mathbf{I}_{Q} + \frac{\partial\mathbf{b}\left(\boldsymbol{\theta}^{0}\right)}{\partial\boldsymbol{\theta}^{T}}\right)\mathbf{F}_{\boldsymbol{\theta}^{0}}^{-1}\left(\mathbf{I}_{Q} + \frac{\partial\mathbf{b}\left(\boldsymbol{\theta}^{0}\right)}{\partial\boldsymbol{\theta}^{T}}\right)^{T}\right). \tag{68}$$

Les différents résultats obtenus sont généralement de l'ordre de la démonstration de l'existence d'estimateur $\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}_{opt}(\mathbf{x})$ biaisé, mais ne donnent pas l'expression du biais optimal (éventuellement nul) pour un problème d'estimation donné (sauf cas particuliers triviaux [HFU96] [Eld06] [KE08] [TT10-II]).
Quand bien même l'expression analytique du biais optimal $\mathbf{b}(\boldsymbol{\theta})$ serait connue, il n'existe malheureusement pas (à ma connaissance) de stratégie d'estimation permettant de construire un estimateur réalisable avec un biais donné. Réciproquement, en pratique le biais dépend toujours de l'estimateur spécifique considéré et n'est en général pas connu (ni même calculable).
Par contre si un estimateur est baisé, alors il faut impérativement intégrer ce biais dans le calcul des bornes inférieures de l'EQM sous peine d'exhiber "une violation de la borne" (EQM inférieure à la borne) comme l'illustrent les exemples étudiés dans la section suivante.

d) Liens avec l'estimateur au sens du maximum de vraisemblance (EMV) :

L'étude des propriétés de l'EMV n'étant pas un de mes axes de recherche premiers, une synthèse des résultats classiques sur l'EMV est fournie en Annexe IV-B-p66, notamment diverses relations asymptotiques entre l'EQM de l'EMV et la BCR. Hormis ces résultats asymptotiques, le calcul analytique de l'EQM de l'EMV, valable pour toutes les valeurs du RSB et pour un nombre fini d'échantillons, reste un problème ouvert, même pour des problèmes d'estimation relativement simples (estimation d'un angle ou d'une fréquence normalisée).
Le phénomène de "décrochement" ("threshold effect") de l'EQM de l'EMV, c'est à dire son accroissement rapide relativement à la BCR lorsque le RSB diminue en deçà d'une certaine valeur, semble être bien connu dès le début des années 60 ("it is well known ..." [MS69]). Si les travaux des premiers contributeurs à la recherche d'approximation de la BB (pour estimateur uniformément sans biais) [MS69] [MH71] [Gla72] [CS81] (et par la suite [ZS93] [Kno97] [NBL04]) avaient pour objectif d'exhiber un outil théorique justifiant le phénomène de "décrochement" et permettant d'analyser les facteurs contributifs (dans un objectif d'aide à la conception des systèmes), à partir des années 90 plusieurs auteurs [Abe93] [TK99] [CGQL08] **[T4]** [TT10-I] ont tenté d'utiliser les différentes bornes inférieures de l'EQM comme estimateur de l'EQM vraie de l'EMV.
Il s'est ensuivi une évolution du sens associé à la notion de "proximité de la borne" ("lower bound tightness") : en effet si cette notion faisait initialement référence à une "proximité à la BB" (puisqu'il n'y avait pas de comparaison avec l'EQM de l'EMV), elle est devenue une "proximité à l'EQM de l'EMV" (terminologie employée dans la "Method of Interval Errors" (MIE) [VT68] [RB74] évoquée ci-après).
Lorsque la recherche du paramètre s'effectue sur un support fini (par exemple lors de l'estimation d'un angle ou d'une fréquence normalisée), ces travaux ont permis :
• de montrer qu'il existe deux classes de bornes : les "Small Errors bounds" (pour estimateurs localement sans biais : BCR, BBa) et les "Larges Errors bounds" (pour estimateurs uniformément sans biais : BHCR, BMS, BA, ...) qui peuvent être reliées au comportement de l'EQM des EMV en fonction du RSB ou du nombre de réalisation indépendantes,
• de mettre en évidence 3 zones comportementales de l'EQM des EMV, que le modèle d'observation soit déterministe ou stochastique.
Lorsque le RSB ou le nombre d'observations est élevé, la zone est dite asymptotique : l'erreur d'estimation est généralement faible et peut être prédite par la BCR ("Small Errors bound") ou par la variance asymptotique lorsqu'elle est calculable [Ric05] (cf. Annexe IV-B-p66 pour plus de détails). Lorsque le RSB ou le nombre d'observations décroît, il apparaît un accroissement rapide de l'EQM (donc de l'erreur) dû à l'apparition d'"outliers" (terme anglo-saxon désignant une observation aberrante) dans le critère du MV (123)(128)(130). On appelle cette région la zone de décrochement ou de transition. Elle apparait dans tous les modèles d'observations non-linéaires donnant lieu à des fonctions de vraisemblance $p(\mathbf{x};\theta)$ possédant des maxima locaux (lobes secondaires de la fonction d'ambiguïté du filtre adapté, par exemple). Enfin, lorsque le RSB ou le nombre d'observations est très faible, le signal observé se réduit principalement à la composante de bruit, ce qui conduit à une distribution de l'estimateur ne dépendant plus des paramètres à estimer (distribution à priori) et à une valeur limite vers laquelle l'EQM convergence asymptotiquement. Elle exhibe alors un comportement asymptotiquement "plat" : c'est la zone de non-information. Ce comportement est illustré par la figure 2-p36 associée au problème de l'estimation de la fréquence ε d'une cisoïde dont le modèle d'observation s'écrit :

$$\mathbf{x} = \mathbf{b}(\varepsilon)\,\sigma + \mathbf{n}, \quad \mathbf{b}(\varepsilon) = \left(1, e^{j2\pi\varepsilon}, ..., e^{j(N-1)2\pi\varepsilon}\right)^T, \quad \varepsilon \in \left]-0.5, 0.5\right[\tag{69}$$

où $\mathbf{n}$ est un bruit blanc centré de matrice de covariance $\mathbf{C_n} = \sigma_{\mathbf{n}}^2 \mathbf{I}_N$, $\sigma \in \mathbb{C}$ est l'amplitude complexe de la source dans le cas d'une observation déterministe, $\frac{|\sigma|^2}{\sigma_{\mathbf{n}}^2}$ étant le RSB.
Lorsque le biais n'est pas pris en compte, seules les "Large Errors bounds" rendent compte du phénomène de décrochement, encore que les bornes inférieures connues ne permettent que d'approximer (sous-évaluer) la valeur du RSB ou du nombre d'observation indépendantes pour laquelle il se produit (cf. figure 4-p39 et §II-A5-p38).

La zone de transition peut également être "douce" (smooth) sans présenter d'effet de décrochement. Dans ce cas, la fonction de vraisemblance du paramètre recherché est une fonction unimodale n'admettant qu'un seul maximum. Ce cas est illustré par la figure 3-p37 associée au problème de l'estimation du rapport d'écartométrie $r(\varepsilon)$ d'une

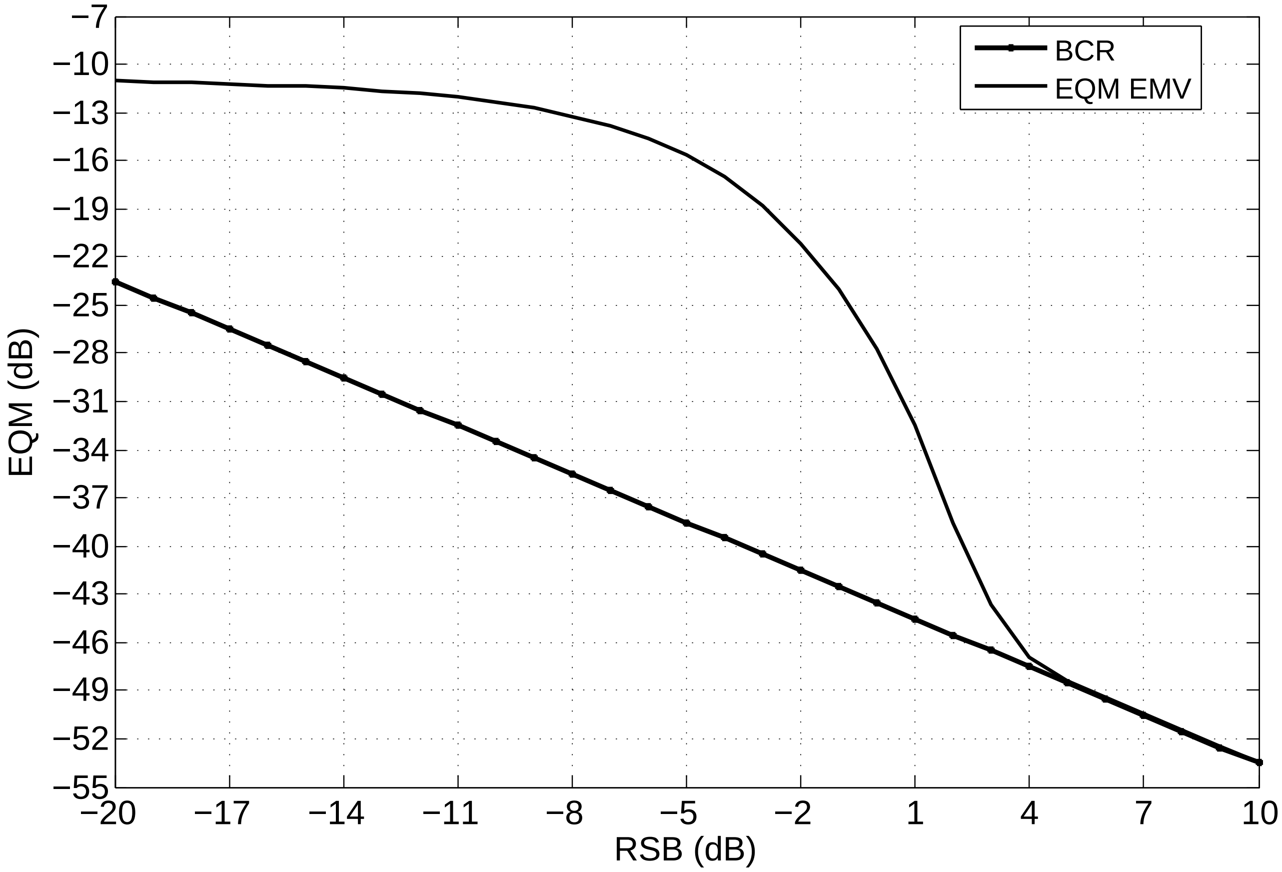


Fig. 2. Comportement de l'EQM de l'EMVD de la fréquence inconnue ε d'une cisoïde ((69), $\sigma_{\mathbf{n}}^2$ et σ connus), $\varepsilon^0 = 0, N = 10$

antenne à 2 voies de réception (voies "somme" Σ et "différence" Δ), dite antenne Monopulse dont le modèle d'observation s'écrit :

$$\mathbf{x}^l = \left(\Sigma^l, \Delta^l\right)^T = \mathbf{b}(\varepsilon)\,\sigma^l + \mathbf{n}^l, \quad \sigma^l = \alpha^l g_\Sigma(\varepsilon), \quad \mathbf{b}(\varepsilon) = (1, r(\varepsilon))^T, \ r(\varepsilon) = \frac{g_\Delta(\varepsilon)}{g_\Sigma(\varepsilon)} \quad l = 1, ..., L \tag{70}$$

où $\mathbf{n}$ est un bruit blanc centré de matrice de covariance connue $\mathbf{C_n} = \mathbf{I}_N$, $\left\{\alpha^l\right\}_{l=1}^L$ sont les amplitudes complexes de la source de signal, $g_\Sigma(\varepsilon)$ et $g_\Delta(\varepsilon)$ sont les gains complexes en amplitude (diagramme de rayonnement) de chaque voie pour la direction d'arrivée ε. L'étude du rapport d'écartométrie au voisinage de la direction principale de rayonnement de l'antenne $(\varepsilon = 0)$ montre que dans le lobe principal à $3dB$: $r(\varepsilon) \approx k\varepsilon$ [She84] [Lev88], ce qui conduit à l'estimateur $\widehat{\varepsilon}_{MV} = \widehat{r(\varepsilon)}_{MV}/k$ dont la caractérisation statistique se déduit naturellement de celle de $\widehat{r(\varepsilon)}_{MV} \approx \mathrm{Re}\left\{\left(\sum_{l=1}^L \left(\Sigma^l\right)^* \Delta^l\right) / \left(\sum_{l=1}^L \left|\Sigma^l\right|^2\right)\right\}$ [Mos69] [GCL10], que le modèle d'observation soit déterministe ou stochastique.

Dans les deux types de fonction de vraisemblance (69)(70), la BCR perd sa propriété de borne inférieure dans la zone de non-information, c'est à dire à nombre d'observations ou à RSB faible. La principale raison est qu'un estimateur localement sans biais d'un paramètre du signal source n'existe généralement pas dans la zone de non-information. Dans ce cas :

$$E_\theta\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\right] = g(\theta) + b(\theta)$$

où $b(\theta)$ est la fonction de biais dépendant du RSB et/ou du nombre d'observations. Il faut alors considérer les bornes inférieures de l'EQM des estimateurs avec biais (cf. §II-A4c-p32). Ce raffinement théorique permet de restituer aux bornes inférieures de l'EQM avec biais leur propriété de borne inférieure, comme le montre la figure

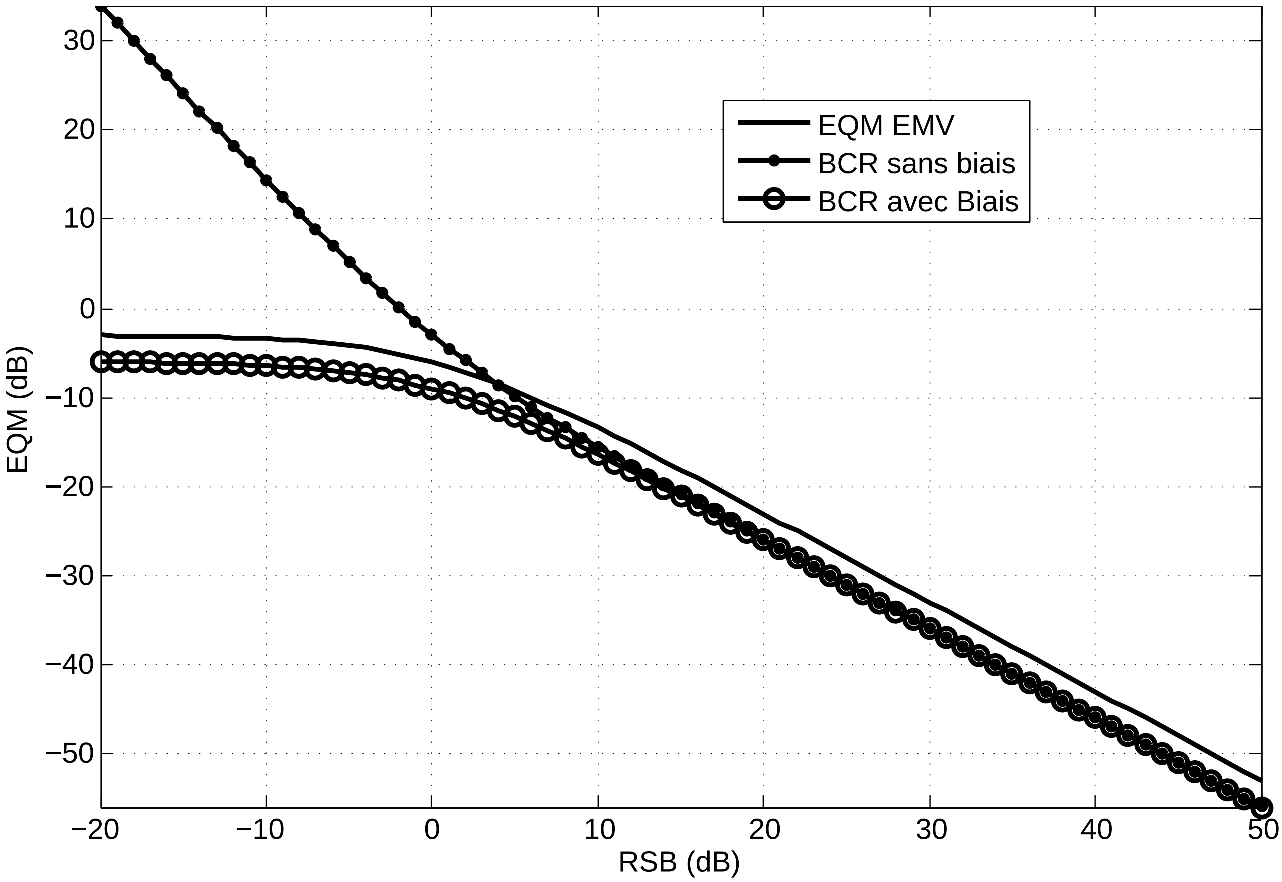


Fig. 3. Comportement de l'EQM de l'EMV du rapport d'écartométrie (70), modèle stochastique, $L = 2$

3-p37 et la figure 5-p42, sans toutefois prolonger leur valeur "prédictive" de l'EQM aux zones de transition ou de non-information. Mais la prise en compte du biais n'est pas toujours possible car il dépend de l'estimateur spécifique considéré et n'est en général ni connu ni même calculable.

Une méthode alternative à la prédiction de l'EQM vraie de l'EMV par les bornes inférieures de l'EQM est la modélisation de l'EQM à partir de son comportement dans les zones asymptotique et de non-information. L'exemple célèbre est celui fourni par Rife et Boorstyn [RB74] dans le cadre de l'estimation spectrale appliquée à une unique fréquence (69) où ils ont montré que l'EQM de l'EMV peut être modélisée par :

$$\left\| \widehat{\varepsilon^0} (\mathbf{x}) - \varepsilon^0 \right\|^2_{2;\boldsymbol{\theta}^0} \simeq (1-p)\sigma^2_{asymptotique} + p\sigma^2_{a\,priori}, \tag{71}$$

où p représente la probabilité d'apparition d'un outlier (l'EMV est en dehors du lobe principal de la fonction d'ambiguïté du filtre adapté), $\sigma^2_{asymptotique}$ est la variance asymptotique de l'estimateur (parfois approximée par la BCR) et $\sigma^2_{a\,priori}$ la variance dans la zone de non-information (généralement approximée par la variance d'une variable aléatoire uniforme sur le support du paramètre, si le support est fini). Selon les hypothèses mises en jeu dans le calcul de la probabilité p, l'équation (71) fournit une approximation plus ou moins précise de l'EQM d'un estimateur sur les trois zones susmentionnées et permet donc une caractérisation de la zone de décrochement. Cette méthode et ses variantes telles que la Method of Interval Errors (MIE) [VT68] ont été appliquées avec succès à divers problèmes d'estimation [RB74] [SB85] [QK94] [TST94] [JAW95] [FLB04] [BFL04-D] [BFL04-S] [NLF05] [Ath05] [Ric05] [RH12] [Can13]. Il est important de remarquer que ces méthodes fournissent une EQM approchée pour un estimateur particulier et, en conséquence, ne rendent pas compte des performances ultimes. De plus, cette méthode souffre d'un défaut majeur : le calcul de la probabilité d'apparition d'un outlier p est un problème non-trivial notamment dans un contexte multi-paramètres [Ath05] [Ric05].

5) Recherche d'un estimateur (a priori) localement le meilleur à l'ordre 2. Problèmes ouverts, conjectures et perspectives :

• **Problème ouvert 1. La mise en oeuvre.**

La principale difficulté dans la mise en oeuvre des approximations de la BB prenant en compte le caractère sans biais de façon uniforme (donc hors BCR et BBa) est à la fois de l'ordre de la précision numérique des calculs et du nombre rapidement exponentiel de calculs à effectuer. Pour illustrer cette problématique considérons le cas particulier du modèle d'observation paramétrique "bande étroite" (cf. Annexe IV-B-p66) gaussien déterministe à une seule source :

$$\mathbf{x} = \mathbf{b}(\boldsymbol{\varepsilon})\,\sigma + \mathbf{n}, \quad p(\mathbf{x};\boldsymbol{\theta}) = \frac{1}{\left(\pi\sigma_{\mathbf{n}}^2\right)^N} e^{-\frac{\|\mathbf{x}-\mathbf{b}(\boldsymbol{\varepsilon})\sigma\|^2}{\sigma_{\mathbf{n}}^2}}, \quad \boldsymbol{\theta} = \left(\boldsymbol{\varepsilon}^T, \mathrm{Re}\{\sigma\}, \mathrm{Im}\{\sigma\}, \sigma_{\mathbf{n}}^2\right)^T \tag{72}$$

dans lequel $\mathbf{n}$ est un bruit blanc centré de matrice de covariance $\mathbf{C_n} = \sigma_{\mathbf{n}}^2 \mathbf{I}_N$, $\sigma \in \mathbb{C}$ est l'amplitude complexe de la source et $\mathbf{b}(\boldsymbol{\varepsilon})$ est un vecteur (de fonctions de transfert) dépendant d'un vecteur $\boldsymbol{\varepsilon}$ de P' paramètres inconnus d'intérêt.
Le cas "minimaliste", mais pas très réaliste dans la pratique, consiste à considérer que $\sigma_{\mathbf{n}}^2$ et σ sont connus.
Dans ce cas, un problème de référence est l'estimation de la fréquence d'une cisoïde, problème pour lequel $\boldsymbol{\theta} \triangleq \boldsymbol{\varepsilon}$ se réduit à 1 paramètre et (72) devient (69).
On désire alors comparer (au sens de (51)) les bornes suivantes : BMS, BMH, BA, BG, BTT. Si on se base sur les exemples de performance de la BTT fournis dans [TT08] [TT10-I] établis pour $N = 10$, un maillage uniforme de $]-0.5, 0.5[$ avec un nombre de points test ε^i compris entre 512 et 1024 et pour $K \in \{1, 4, 32\}$, la comparaison est tout simplement impossible avec un calcul en double précision, car l'inversion des matrices $\mathbf{G}_{\theta^0}$ (54) n'est pas stable numériquement, ce que contournent les exemples de la BTT en réduisant la dimension des matrices à inverser à $K \leq 32$.
Il faut donc passer en quadruple précision complexe (non disponible sous Matlab, Scilab, Pv-Wave, ...) pour pouvoir conduire cette comparaison.
Une des conséquences de ce problème numérique est qu'à l'heure actuelle, avec les langages usuels de programmation ne supportant que la double précision complexe, il est nécessaire de remplacer l'approche proposée par Glave, à savoir l'utilisation d'un maillage uniforme dont le pas diminue, par un calcul de supremum sur une collection de sous-ensembles de points test $\left\{\varepsilon^i\right\}_1^I$. Si ce calcul doit être fait pour de nombreuses valeurs de $RSB = \frac{|\sigma|^2}{\sigma_{\mathbf{n}}^2}$, un compromis entre le nombre de points test I et le nombre d'élements de la collection doit être effectuée pour borner le temps de calcul, limitant ainsi la valeur informative du supremum obtenu.
A titre d'exemple, si nous reprenons le modèle (72) dans le cas de la cisoïde où le seul paramètre inconnu est la fréquence ε, alors la comparaison des implémentations suivantes pour $\varepsilon^0 = 0$:
• BHCR : 2 points test $\{0, \delta\}$ + supremum sur δ où δ balaie $]-0.5, 0.5[$ avec un pas de $\frac{1}{1024}$,
• BMS : 3 points test $\{0, \delta, -\delta\}$ + supremum sur δ où δ balaie $]0, 0.5[$ avec un pas de $\frac{1}{1024}$,
• BG : 3 points test $\{0, \delta, -\delta\}$ + supremum sur δ où δ balaie $]0, 0.5[$ avec un pas de $\frac{1}{1024}$,
• BG2 : 3 points test $\{0, \delta, -\delta\}$ + supremum sur δ où δ balaie $]0, 0.5[$ avec un pas de $\frac{1}{1024\times128}$,
• BTT : 1024 points test uniformément distribués sur Θ, $\mathbf{W}$ est une matrice composée de $K = 32$ colonnes
. ad hoc de la matrice de DFT,
est fournie par la figure 4-p39. Il est intéressant de remarquer que la BG est soit inférieure soit égale à la BTT (BG2) en fonction du maillage sur lequel le supremum à 3 points test est calculé, ce qui est bien loin de l'inégalité théorique (51) obtenue pour un jeu de I points test $\left\{\varepsilon^i\right\}_1^I$ fixé. Autrement dit, il faut être conscient que les différentes comparaisons disponibles dans la littérature courante sur les approximations de la BB sont en fait des comparaisons entre des implémentation données, lesquelles peuvent être très éloignées des implémentations optimales (au sens de la convergence vers la BB) difficiles à mettre en oeuvre encore à ce jour.

Mais l'approche de Glave n'a de sens que si Θ se limite à un intervalle; elle n'est donc clairement pas envisageable pour le paramètre complexe σ dont les deux composantes $\left(\sigma_r \triangleq \mathrm{Re}\{\sigma\}, \sigma_i \triangleq \mathrm{Im}\{\sigma\}\right)$ appartiennent a priori à $\mathbb{R}$.

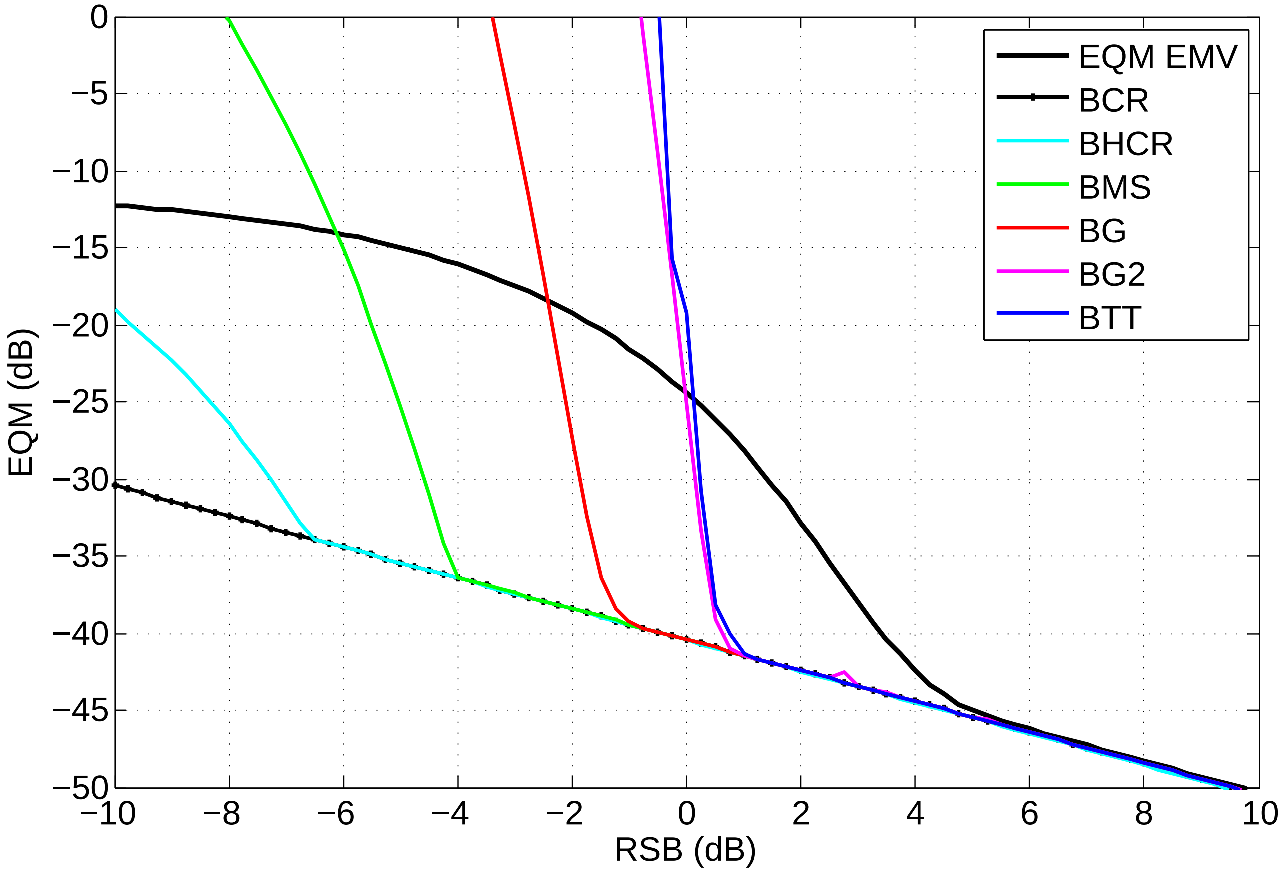


Fig. 4. Comparaison d'implémentations des bornes BHCR, BMS, BG, BTT dans le cas d'une cisoïde de fréquence inconnue ε, $\varepsilon^0 = 0$, $N = 10$

En fait la recherche d'un supremum d'une approximation de la BB incorporant des paramètres dont le domaine de définition n'est pas borné $\left(\boldsymbol{\theta} = (\varepsilon, \sigma_r, \sigma_i)^T\right)$ est de même nature que la formulation générale initiale de la BB (21) où la recherche du supremum s'écrit $\sup_{\{\boldsymbol{\theta}^i\}_1^I, \{\mathbf{w}^i\}_1^I, I \in \mathbb{N}} \left\{ C'\left(\{\boldsymbol{\theta}^i\}_1^I, \{\mathbf{w}^i\}_1^I\right) \right\} \equiv \sup_{\{\boldsymbol{\theta}^i\}_1^I, I \in \mathbb{N}} \left\{ \sup_{\{\mathbf{w}^i\}_1^I \in \mathbb{R}^Q} \left\{ C'\left(\{\boldsymbol{\theta}^i\}_1^I, \{\mathbf{w}^i\}_1^I\right) \right\} \right\}$ où $C'\left(\{\boldsymbol{\theta}^i\}_1^I, \{\mathbf{w}^i\}_1^I\right)$ est un critère à maximiser.

Comme nous l'avons évoqué précédemment , l'un des apports fondamentaux du choix du moment d'ordre 2 est précisément de fournir une solution analytique à $\sup_{\{\mathbf{w}^i\}_1^I \in \mathbb{R}^Q} \left\{ C'\left(\{\boldsymbol{\theta}^i\}_1^I, \{\mathbf{w}^i\}_1^I\right) \right\} = C\left(\{\boldsymbol{\theta}^i\}_1^I\right)$ afin de réduire la recherche à $\sup_{\{\boldsymbol{\theta}^i\}_1^I, I \in \mathbb{N}} \left\{ C\left(\{\boldsymbol{\theta}^i\}_1^I\right) \right\}$. La transposition "idéale" à notre problème serait de trouver une solution analytique à la recherche de certains supremums, notamment pour les paramètres dont le domaine de définition n'est pas borné : $\sup_{\{\boldsymbol{\theta}^i\}_1^I, I \in \mathbb{N}} \left\{ C\left(\{\boldsymbol{\theta}^i\}_1^I\right) \right\} \equiv \sup_{\{\varepsilon^i\}_1^I, I \in \mathbb{N}} \left\{ \sup_{\{\sigma^i\}_1^I \in \mathbb{C}} \left\{ C\left(\{\varepsilon^i\}_1^I, \{\sigma^i\}_1^I\right) \right\} \right\}$.

C'est l'approche proposée dans [TK99] (reprise dans [LTN11]) pour la BMS (approximation la plus simple de la BB) appliquée aux modèles gaussiens complexes circulaires déterministe ou stochastique. Comme la résolution exacte de $\sup_{\{\sigma^i\}_1^I \in \mathbb{C}} \left\{ C\left(\{\varepsilon^i\}_1^I, \{\sigma^i\}_1^I\right) \right\}$ semble difficile (voire impossible) pour le modèle gaussien d'observation bande étroite dans le cas général, [TK99] propose une solution sous optimale (mais analytique) basée sur la minimisation de $tr\left(\mathbf{G}_{\boldsymbol{\theta}^0}^{MS}\right)$ relativement à $\{\sigma^i\}_1^I$. Dans notre cas d'étude, la matrice $\mathbf{G}_{\boldsymbol{\theta}^0}^{MS}$ pour le calcul de la BMS associée à

$\{\boldsymbol{\theta}^0\} \cup \{\boldsymbol{\theta}^i\}_1^I \left(\boldsymbol{\theta} = (\varepsilon, \sigma_r, \sigma_i)^T\right)$ s'écrit (64) :

$$\mathbf{G}_{\boldsymbol{\theta}^0}^{MS} = \mathbf{B}_{\boldsymbol{\theta}^0} - \mathbf{1}_I \mathbf{1}_I^T \ / \ (\mathbf{B}_{\boldsymbol{\theta}^0})_{i,i'} = e^{\frac{2|\sigma^0|^2}{\sigma_{\mathbf{n}}^2} \mathrm{Re}\left\{\left(\mathbf{b}(\varepsilon^i)\frac{\sigma^i}{\sigma^0} - \mathbf{b}(\varepsilon^0)\right)^H \left(\mathbf{b}(\varepsilon^{i'})\frac{\sigma^{i'}}{\sigma^0} - \mathbf{b}(\varepsilon^0)\right)\right\}} \tag{73}$$

et [TK99] :

$$\left\{\sigma_{opt}^i\right\}_1^I = \arg \inf_{\{\sigma^i\}_1^I \in \mathbb{C}} \left\{tr\left(\mathbf{G}_{\boldsymbol{\theta}^0}^{MS}\right)\right\} = \arg \inf_{\{\sigma^i\}_1^I \in \mathbb{C}} \left\{tr\left(\mathbf{B}_{\boldsymbol{\theta}^0}\right)\right\} = \left\{\sigma_{opt}^i = \sigma^0 \frac{\mathbf{b}^H(\varepsilon^i)\,\mathbf{b}(\varepsilon^0)}{\|\mathbf{b}(\varepsilon^i)\|_2 \|\mathbf{b}(\varepsilon^0)\|_2}\right\}_1^I \tag{74}$$

En fait, du fait de la forme particulière de $\mathbf{G}_{\boldsymbol{\theta}^0}^{MS}$ (73), les $\left\{\sigma_{opt}^i\right\}_1^I$ fournis par $\inf_{\{\sigma^i\}_1^I \in \mathbb{C}} \left\{tr\left(\mathbf{G}_{\boldsymbol{\theta}^0}^{MS}\right)\right\}$ sont aussi les solutions de :

$$\left\{\sigma_{opt}^i\right\}_1^I = \arg \inf_{\{\sigma^i\}_1^I \in \mathbb{C}} \left\{\left\|\mathbf{vec}\left(\mathbf{G}_{\boldsymbol{\theta}^0}^{MS}\right)\right\|_2^2\right\} = \arg \inf_{\{\sigma^i\}_1^I \in \mathbb{C}} \left\{\left\|\mathbf{vec}\left(\mathbf{B}_{\boldsymbol{\theta}^0} - \mathbf{1}_I \mathbf{1}_I^T\right)\right\|_2^2\right\} \tag{75}$$

En effet, puisque :

$$\left\{\begin{array}{c} \left|\mathrm{Re}\left\{\left(\mathbf{b}(\varepsilon^i)\frac{\sigma^i}{\sigma^0} - \mathbf{b}(\varepsilon^0)\right)^H \left(\mathbf{b}(\varepsilon^{i'})\frac{\sigma^{i'}}{\sigma^0} - \mathbf{b}(\varepsilon^0)\right)\right\}\right| \leq \left\|\mathbf{b}(\varepsilon^i)\frac{\sigma^i}{\sigma^0} - \mathbf{b}(\varepsilon^0)\right\|_2 \left\|\mathbf{b}(\varepsilon^{i'})\frac{\sigma^{i'}}{\sigma^0} - \mathbf{b}(\varepsilon^0)\right\|_2 \\ \left\|\mathbf{b}(\varepsilon)\frac{\sigma}{\sigma^0} - \mathbf{b}(\varepsilon^0)\right\|_2^2 \leq \left\|\mathbf{b}(\varepsilon^0)\right\|_2 \left(1 - \frac{|\mathbf{b}^H(\varepsilon)\mathbf{b}(\varepsilon^0)|^2}{\|\mathbf{b}(\varepsilon)\|_2^2 \|\mathbf{b}(\varepsilon^0)\|_2^2}\right) \text{ avec égalité pour } \frac{\sigma}{\sigma^0} = \frac{\mathbf{b}^H(\varepsilon)\mathbf{b}(\varepsilon^0)}{\|\mathbf{b}(\varepsilon)\|_2 \|\mathbf{b}(\varepsilon^0)\|_2} \end{array}\right.$$

on en déduit que non seulement les $\left\{\sigma_{opt}^i\right\}_1^I$ sont données par (74) mais qu'également le meilleur choix pour les points test ε^i sont les valeurs de ε qui maximisent $\frac{|\mathbf{b}^H(\varepsilon)\mathbf{b}(\varepsilon^0)|^2}{\|\mathbf{b}(\varepsilon)\|_2^2 \|\mathbf{b}(\varepsilon^0)\|_2^2}$, autrement dit les valeurs ε correspondant aux lobes secondaires de la fonction d'ambiguïté (initialement suggéré dans [MS69]).
Cette approche nous ramène à la recherche précédente, c'est à dire le cas où σ est connu avec en plus un choix préférentiel des points tests $\left\{\varepsilon^i\right\}_1^I$. Mais $\sigma_{\mathbf{n}}^2$ est encore supposé connu.
Par conséquent, à ma connaissance, il n'existe pas à l'heure actuelle dans la littérature courante d'article proposant une implémentation du calcul de la BMS pour le modèle canonique à 1 sources (72) lorsque $\boldsymbol{\theta} = \left(\varepsilon, \sigma_r, \sigma_i, \sigma_{\mathbf{n}}^2\right)^T$. De plus, le fait que la maximisation proprosée par soit sous-optimale ne plaide pas pour son incorporation dans les comparaisons entre approximations de la BB (BMS, BMH, BA, BG, BTT) puisque l'effet de cette sous-optimalité sur le supremum pour les paramètres d'intérêts (ε dans notre exemple) n'est pas quantifiable ... et pourrait être préjudiciable en matière de proximité "montrée" à la BB. Par exemple, la forme générale de la matrice $\mathbf{G}_{\boldsymbol{\theta}^0}$ pour la BG étant (62) :

$$\mathbf{G}_{\boldsymbol{\theta}^0}^{G} = \begin{bmatrix} \mathbf{G}_{\boldsymbol{\theta}^0}^{MS} & \mathbf{H}_{\boldsymbol{\theta}^0}^T \\ \mathbf{H}_{\boldsymbol{\theta}^0} & \mathbf{FE}_{\boldsymbol{\theta}^0} \end{bmatrix}$$

les conséquences de la minimisation de $\left\|\mathbf{vec}\left(\mathbf{G}_{\boldsymbol{\theta}^0}^{MS}\right)\right\|_2^2$ sur $\left\|\mathbf{vec}\left(\mathbf{G}_{\boldsymbol{\theta}^0}^{G}\right)\right\|_2^2$ sont difficiles à évaluer a priori. Le même problème se pose pour la BA, la BMH (64) et la BTT.
Enfin, quand bien même la solution théorique de $\sup_{\{\sigma^i\}_1^I \in \mathbb{C}} \left\{C\left(\left\{\varepsilon^i\right\}_1^I, \left\{\sigma^i\right\}_1^I\right)\right\}$ serait trouvée pour une approximation de la BB dans le cas particulier du modèle d'observation bande étroite, il reste à implémenter $\sup_{\{\boldsymbol{\varepsilon}^i\}_1^I, I \in \mathbb{N}} \{\ \}$ qui reste un problème ouvert pour les critères ne dépendant pas linéairement des $\left\{\boldsymbol{\varepsilon}^i\right\}_1^I$ comme le montre la littérature sur le sujet appliquée à la mise en oeuvre du MV (cf. Annexe IV-B-p66). Même le choix "préférentiel" des valeurs de ε correspondant aux lobes secondaires de la fonction d'ambiguïté devient un problème non trivial (voir intractable) lorsque le nombre de sources augmentent conduisant à une fonction d'ambiguïté généralisée.

En conséquence, à ma connaissance, nous ne savons toujours pas ce que vaut la BB même dans le cas trivial d'un seul paramètre inconnu appartenant à un domaine de définition borné.

Perspectives :

1) Il manque la formulation d'un algorithme itératif (en le nombre de points test I) de type CLEAN [TS88] ou AP [ZW88] pour pouvoir mettre en oeuvre le calcul des approximations de la BB.

2) Quel que soit cet algorithme, il sera d'autant plus rapide que le nombre de points test I requis pour atteindre la

BB sera faible. Et ce nombre de points I dépend directement de la qualité de l'approximation de la contrainte sans biais uniforme par l'estimateur optimal associé à la borne (56) $\left(E_{\theta}\left[\widehat{g\left(\theta^{0}\right)_{opt}}\left(\mathbf{x}\right)\right] \simeq g\left(\theta\right)\right)$, lequel est entièrement déterminé à partir de l'ensemble des I vecteurs (de contrainte linéaire) $\mathbf{c}^{i} \triangleq \mathbf{c}\left(\mathbf{x},\theta^{i}\right)$ considéré. Il pourrait être opportun de pousser plus avant les travaux présentés dans [TT10-I] afin de prendre en compte des approximations non polynomiales (principe d'approximation de la BBa, BA, BMH, BG, BTT).

• Problème ouvert 2. La sensibilité au biais.

Depuis l'article fondateur de [MS69] mentionnant :
"*Our intuitive examination of the threshold effect has already led us to the conclusion that it is the sidelobe peaks of the autocorrelation function that are the critical points. That is, it is these peaks that lead to estimates exhibiting large excursions from the true parameter value. Since it is these so-called anomalous errors that lead to system threshold it is reasonable to choose the test points to be the critical points, i.e., the points at which the sidelobe peaks occur.*"
il est implicitement admis par la communauté que c'est la possibilité de prendre en compte la contrainte sans biais uniforme (sous sa forme discrétisée en I points test) qui permet aux "Large Errors bounds" (BHCR, BMS, BMH, BG, BA, BTT) de rendre compte du phénomène de décrochement. A contrario, les "Small Errors bounds" (BCR,BBa) ne prenant en compte qu'une contrainte sans biais locale, ne peuvent rendre compte ni de la zone de décrochement ni de la zone de non-information.
Il est vrai que cette intuition a toujours été vérifiée (puisqu'on ne sait pas calculer la BB) dans les diverses études des approximations de la BB pour estimateurs (uniformément) sans biais.

Par contre cette intuition ne semble pas s'appliquer à l'EQM intégrant un biais comme le montrent deux exemples, la cisoïde pondérée

$$\mathbf{x} = \left(\mathbf{b}\left(\varepsilon\right) \odot \mathbf{w}\right)\sigma + \mathbf{n}, \quad \mathbf{w} \in \mathbb{C}^{N}, \quad \mathbf{b}\left(\varepsilon\right) = \left(1, e^{j2\pi\varepsilon}, ..., e^{j(N-1)2\pi\varepsilon}\right)^{T}, \quad \varepsilon \in \left]-0.5, 0.5\right[\tag{76}$$

et l'écartométrie monopulse (70), illustrés par les figure 5-p42 et 3-p37. Par ailleurs, ces deux exemples couvrent les deux cas théoriques possibles pour les fonctions de vraisemblance $p\left(\mathbf{x};\theta\right)$: celles possédant des maxima locaux et celles ne possédant qu'un seul maximum.
La figure 5-p42 montre clairement que la BCR intégrant le biais (courbe CRB_{b}) :

$$\left\|\widehat{\varepsilon^{0}}\left(\mathbf{x}\right) - \varepsilon^{0}\right\|_{2;\theta^{0}}^{2} \geq b^{2}\left(\varepsilon^{0}\right) + \frac{1}{F_{\theta^{0}}} tr\left(1 + \frac{\partial b\left(\varepsilon^{0}\right)}{\partial\varepsilon}\right)^{2}$$

reproduit le phénomène de décrochement quasiment à l'identique de la BHCR intégrant le biais (courbe ChRB_{b}), ce qui suggère que la prise en compte (précise) du biais localement est le facteur déterminant dans l'évaluation d'une borne inférieure, la prise en compte de ce biais uniformément (en 2 points test pour la BHCR) n'étant qu'un facteur marginal.
Ce résultat est confirmé par la courbe 3-p37 où la BCR avec biais reproduit fidèlement l'évolution de l'EQM de l'EMV.
Et dans les deux cas, la BCR avec biais retrouve sa propriété de borne inférieure même en zone de non-information où la notion d'estimateur localement sans biais est un peu difficile à interpréter ...

Perspectives :

Comme nous ne savons pas exprimer analytiquement la BB pour les modèles d'observation gaussiens non-linéaires (en les paramètres inconnus), la distance entre la BCR et la BB pour une classe d'estimateur de biais $b\left(\theta\right)$ peut s'obtenir par l'encadrement suivant :

$$BCR_{\theta|\theta}\left(\theta^{0}\right) \leq BB_{\theta|\theta}\left(\theta^{0}\right) \leq \left\|\widehat{\theta^{0}}\left(\mathbf{x}\right) - \theta^{0}\right\|_{2;\theta^{0}}^{2}, \quad \forall\widehat{\theta^{0}}\left(\mathbf{x}\right) \ / \ E_{\theta}\left[\widehat{\theta^{0}}\left(\mathbf{x}\right)\right] = \theta + b\left(\theta\right)$$

$$\Downarrow$$

$$BB_{\theta|\theta}\left(\theta^{0}\right) - BCR_{\theta|\theta}\left(\theta^{0}\right) \leq \min_{\widehat{\theta^{0}}(\mathbf{x})/E_{\theta}\left[\widehat{\theta^{0}}(\mathbf{x})\right]=\theta+b(\theta)} \left\{\left\|\widehat{\theta^{0}}\left(\mathbf{x}\right) - \theta^{0}\right\|_{2;\theta^{0}}^{2}\right\} - BCR_{\theta|\theta}\left(\theta^{0}\right) \tag{77}$$

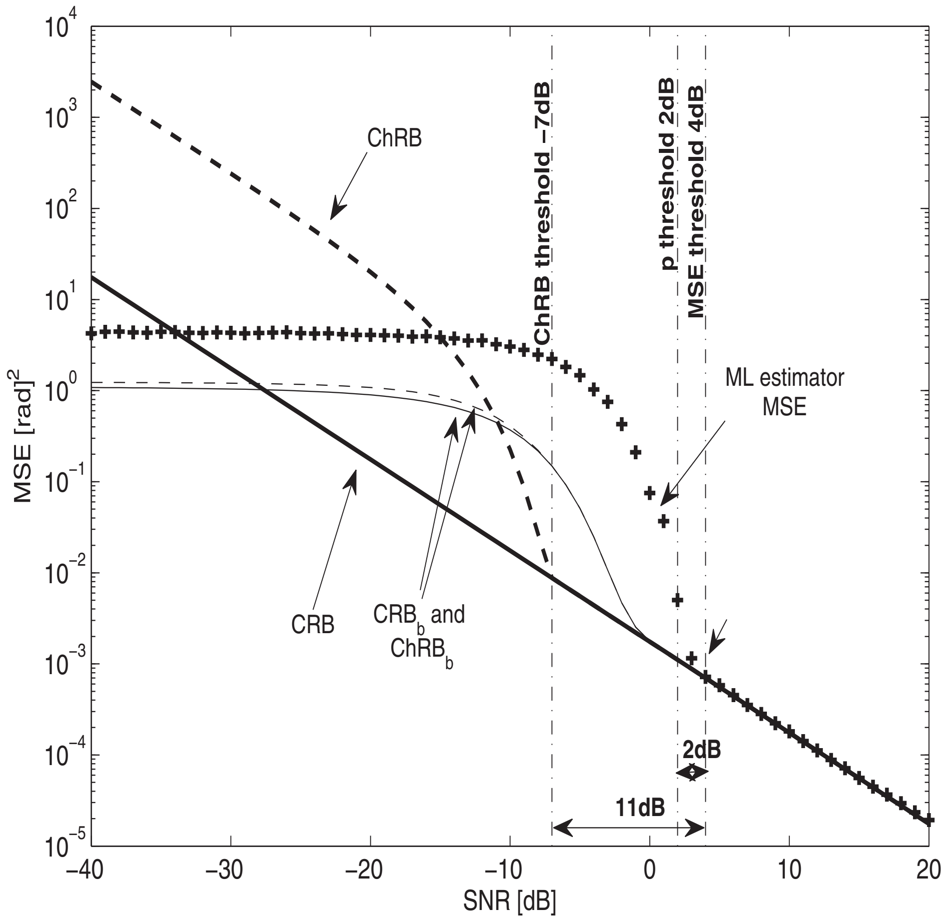


Fig. 5. Comportement de l'EQM de l'EMVD de la fréquence inconnue ε d'une cisoïde pondérée ((76), $\sigma_{\mathbf{n}}^2$ et σ connus), $\varepsilon^0 = \frac{\pi}{3}, N = 10$. Comparaison avec la BCR et la BHCR avec et sans biais [NFBL09, Fig3].

à condition de pouvoir exhiber au moins un estimateur (réalisable ou clairvoyant) de θ présentant le biais $b(\theta)$.
C'est en appliquant (77) et les résultats asymptotiques de l'EMV (cf. Annexe IV-B-p66) que nous pouvons affirmer que sous certaines conditions la BCR converge vers la BB en zone asymptotique pour les estimateurs sans biais. Cette propriété se prolonge aisément aux estimateurs avec biais $\left(E_\theta\left[\widehat{\theta^0}(\mathbf{x})\right] = \theta + b(\theta)\right)$. En effet puisque :

$$E_\theta\left[\widehat{\theta^0}(\mathbf{x})\right] = \theta + b(\theta) \;\Rightarrow\; \forall\theta^0 \in \Theta,\; \left\|\widehat{\theta^0}(\mathbf{x}) - \theta^0\right\|_{2;\theta^0}^2 = b^2\left(\theta^0\right) + \left\|\widehat{\theta^0}(\mathbf{x}) - \left(\theta^0 + b\left(\theta^0\right)\right)\right\|_{2;\theta^0}^2$$

par conséquent :

$$\widehat{\theta_{opt}^0}(\mathbf{x}) = \arg\min\left\{\left\|\widehat{\theta^0}(\mathbf{x}) - \theta^0\right\|_{2;\theta^0}^2 \text{ sous } E_{\theta^0+d\theta}\left[\widehat{\theta^0}(\mathbf{x})\right] = \theta^0 + d\theta + o(d\theta)\right\}$$

$$\Downarrow$$

$$\widehat{\theta_{opt}^0}(\mathbf{x}) + b(\theta) = \arg\min\left\{\left\|\widehat{\theta^0}(\mathbf{x}) - \theta^0\right\|_{2;\theta^0}^2 \text{ sous } E_{\theta^0+d\theta}\left[\widehat{\theta^0}(\mathbf{x})\right] = \theta^0 + d\theta + b\left(\theta^0 + d\theta\right) + o(d\theta)\right\}$$

Les résultats présentés par les figures 4-p39 et 5-p42 montrent que cette distance varie notablement en zone de décrochement et de non-information en fonction de la présence d'un biais ou non.
D'un point de vue théorique il serait donc intéressant d'étendre les résultats obtenus dans [NFBL09] à l'aide des travaux de [Ath05] pour obtenir un calcul plus général du biais de l'EMV $\widehat{\boldsymbol{\theta}}$ de $\boldsymbol{\theta}$ sous la forme [NFBL09, (38)] :

$$\mathbf{b}(\boldsymbol{\theta}) = p\left(E_{\boldsymbol{\theta}}\left[\widehat{\boldsymbol{\theta}} \mid \text{outlier}\right] - \boldsymbol{\theta}\right) + (1-p)\left(E_{\boldsymbol{\theta}}\left[\widehat{\boldsymbol{\theta}} \mid \text{no outlier}\right] - \boldsymbol{\theta}\right), \quad p = P\left(outlier\right).$$

Si la généralisation au cas multiparamètres pour les modèles d'observation déterministe et stochastique confirme la proximité de la BCR avec biais à la BB avec biais par l'intermédiaire de (77), alors on pourrait également vérifier si une approximation au 1er (ou 2nd) ordre du biais [Ber99, §8.3] **[T3]** n'est pas suffisante pour obtenir une bonne estimation de la zone de décrochement en lieu et place des diverses approximations de la BB que nous ne savons pas mettre en oeuvre en pratique.

• Problème ouvert 3. Estimations déterministe et bayésienne : une liaison cachée ?

Puisque :

$$\left\|\widehat{\theta^0}(\mathbf{x})-\theta^0\right\|_{2;\theta^0}^2=\left\|E_{\theta^0}\left[\widehat{\theta^0}(\mathbf{x})\right]-\theta^0\right\|_2^2+\left\|\widehat{\theta^0}(\mathbf{x})-E_{\theta^0}\left[\widehat{\theta^0}(\mathbf{x})\right]\right\|_{2;\theta^0}^2,$$

Barankin choisit naturellement d'explorer la solution pour laquelle :

$$\left\|E_{\theta^0}\left[\widehat{\theta^0}(\mathbf{x})\right]-\theta^0\right\|_2^2=0 \;\Leftrightarrow\; E_{\theta^0}\left[\widehat{\theta^0}(\mathbf{x})\right]=\theta^0,$$

et qui définit un sous-ensemble convexe $\mathcal{C}_{\theta^0}$ sous la forme d'un sous espace affine (s.e.a.) :

$$E_{\theta^0}\left[\widehat{\theta^0}(\mathbf{x})\right]=\theta^0 \quad\Leftrightarrow\quad \widehat{\theta^0}(\mathbf{x})=\theta^0+\varphi(\mathbf{x}),\; \varphi(\mathbf{x})\in\mathcal{S}_{\theta^0}=\left\{\varphi(\mathbf{x})\;/\;E_{\theta^0}\left[\varphi(\mathbf{x})\right]=0\right\}.$$

On peut facilement ramener (par translation) la recherche de la solution sur le s.e.a. $\mathcal{C}_{\theta^0}$ à la recherche d'une solution sur la direction $\mathcal{S}_{\theta^0}$:

$$\min_{\mathcal{C}_{\theta^0}}\left\{\left\|\widehat{\theta^0}(\mathbf{x})-\theta^0\right\|_{2;\theta^0}^2\right\} \;\Leftrightarrow\; \min_{\mathcal{S}_{\theta^0}}\left\{\left\|\varphi(\mathbf{x})\right\|_{2;\theta^0}^2\right\},$$

ce qui permet d'appliquer les résultats disponibles sur les e.v., notamment sur les normes. L'avantage de cette approche est qu'elle garantit l'existence d'une solution non triviale puisque, par exemple :

$$E_{\theta^0}\left[\frac{\partial\ln p\left(\mathbf{x};\theta^0\right)}{\partial\theta}\right]=0 \;\Rightarrow\; \frac{\partial\ln p\left(\mathbf{x};\theta^0\right)}{\partial\theta}\in\mathcal{S}_{\theta^0} \;\Rightarrow\; \left\{\widehat{\theta^0}(\mathbf{x})=\theta^0+\alpha\frac{\partial\ln p\left(\mathbf{x};\theta^0\right)}{\partial\theta},\alpha\in\mathbb{R}\right\}\in\mathcal{C}_{\theta^0}$$

L'élimination de la solution triviale $\widehat{\theta^0}(\mathbf{x})=\theta^0 \;\Leftrightarrow\; \varphi(\mathbf{x})=0$ est obtenue en ajoutant une ou plusieurs contraintes que la solution triviale ne peut vérifier. La famille de contraintes proposée par Barankin est de type linéaire et correspond à la recherche d'estimateurs "uniformément sans biais" :

$$E_{\theta^0}\left[\varphi(\mathbf{x})\,\upsilon_{\theta^0}\left(\mathbf{x},\theta^i\right)\right]=\theta^i-\theta^0 \;\Leftrightarrow\; E_{\theta^i}\left[\widehat{\theta^0}(\mathbf{x})\right]=E_{\theta^i}\left[\theta^0+\varphi(\mathbf{x})\right]=\theta^i,$$

c'est à dire se décomposant sur le s.e.v. généré par $\left\{\upsilon_{\theta^0}\left(\mathbf{x},\theta^i\right)\right\}_{i=1}^{I}$. Par nature, la solution proposée par Barankin n'impose pas de contrainte sur $im\left\{\varphi(\mathbf{x})\right\}$ puisqu'elle dépend de $\left\{im\left\{\frac{p(\mathbf{x},\theta^i)}{p(\mathbf{x},\theta^0)}\right\}\right\}_{i=1}^{I}$ et peut donc parcourir $\mathbb{R}$.

La prise en compte d'une contrainte sur $im\left\{\widehat{\theta^0}(\mathbf{x})\right\}$ relève d'une tout autre approche. Pour amorcer la discussion, considérons le cas où Θ est un sous ensemble borné de $\mathbb{R}$, par exemple $\Theta=\left]a,b\right[$. Alors le sous ensemble $\mathcal{C}^{]a,b[}=\left\{\psi(\mathbf{x})\;/\;im\left(\psi(\mathbf{x})\right)\subset\left]a,b\right[\right\}$ est convexe et nous pourrions nous intéresser à la recherche :

$$\psi_{opt}(\mathbf{x})=\arg\min_{\mathcal{C}^{]a,b[}}\left\{\left\|\psi(\mathbf{x})-\theta^0\right\|_{2;\theta^0}^2\right\} \tag{78}$$

qui reste un problème de minimisation d'un critère convexe sur un sous-ensemble convexe, mais qui n'est pas un s.e.v..

De façon surprenante, l'estimation bayésienne fourni un sous-ensemble convexe des estimateurs appartenant à $\mathcal{C}^{]a,b[}$, à savoir l'ensemble des moyennes *a posteriori* :

$$\mathcal{C}_w^{]a,b[}=\left\{\widehat{\theta_w}(\mathbf{x})\triangleq E_{\theta|\mathbf{x}}\left[\theta;w\right]=\frac{\int_\Theta \theta p(\mathbf{x},\theta)\,d\theta}{\int_\Theta p(\mathbf{x},\theta)\,d\theta},\; p(\mathbf{x},\theta)=p(\mathbf{x};\theta)\,w(\theta),\; w(\theta)\text{ d.d.p..}\right\}$$

lesquelles sont également caractérisées par un biais vérifiant :

$$b_w(\theta) = \int_\Omega \widehat{\theta_w}(\mathbf{x})\, p(\mathbf{x};\theta)\, d\mathbf{x} - \theta, \quad E_\theta[b_w(\theta); w] = \int_\Theta b_w(\theta)\, w(\theta)\, d\theta = 0.$$

Perspectives :

Dans un premier temps il serait intéressant de savoir si $\mathcal{C}_w^{]a,b[} = \mathcal{C}^{]a,b[}$. Si tel était le cas, il semblerait judicieux d'explorer les résultats disponibles en estimation bayésienne pour définir une contrainte permettant d'éliminer la solution triviale tout en fournissant une solution calculable à (78).

Cette possibilité d'une "relation cachée" entre l'estimation déterministe et l'estimation bayésienne est soutenue par la borne "conjecturelle" de l'EQM (dite borne de Weiss-Weinstein déterministe) proposée dans [Ren06] et illustrée dans [CRL10]. Cette borne fut conjecturée au titre du principe de transposition des différentes bornes de l'EQM déterministe (BCR, BHCR, BMS, ...) en une forme analogue dans le cadre de l'estimation bayésienne [RFLRN08]. L'idée fut d'appliquer un principe réciproque à la borne de Weiss-Weinstein (BWW) établie en estimation bayésienne, principe qui conduit à la formulation de la contrainte :

$$E_{\theta^0}\left[\left(\widehat{\theta^0}(\mathbf{x}) - \theta^0\right)\left(\upsilon_{\theta^0}^q\left(\mathbf{x}, \theta^0 + d\theta\right) - \upsilon_{\theta^0}^{1-q}\left(\mathbf{x}, \theta^0 - d\theta\right)\right)\right] = -\, d\theta E_{\theta^0}\left[\upsilon_{\theta^0}^{1-q}\left(\mathbf{x}, \theta^0 - d\theta\right)\right], \quad q \in [0,1] \tag{79}$$

et à la borne :

$$BWW_{\theta|\theta}\left(\theta^0\right) \geq \sup_{q,d\theta}\left\{\frac{d\theta^2 E_{\theta^0}^2\left[\upsilon_{\theta^0}^{1-q}\left(\mathbf{x}, \theta^0 - d\theta\right)\right]}{E_{\theta^0}\left[\left(\upsilon_{\theta^0}^q\left(\mathbf{x}, \theta^0 + d\theta\right) - \upsilon_{\theta^0}^{1-q}\left(\mathbf{x}, \theta^0 - d\theta\right)\right)^2\right]}\right\} \tag{80}$$

dont un cas particulier est la BHCR pour $q = 0$ et $q = 1$. Appliqué au modèle d'observation déterministe d'une cisoïde (69) la borne "conjecturelle" BWW (80) se trouve être une borne inférieure estimant de façon très précise le début de la zone de transition en comparaison avec les bornes connues [CRL10] (cf. Annexe IV-K-p208). Nous l'avons qualifié de conjecturelle car à ce jour nous n'avons pas été en mesure de montrer que la contrainte (79) relevait d'un mélange de transformations linéaire et non linéaire (35)(36) de la contrainte uniformément sans biais. Donc nous ne savons pas a priori pour quel classe d'estimateur elle est valide.

6) La BCR. Un outil d'aide à la conception des systèmes de mesure complexes :

Comme nous l'avons évoqué précédemment, la BCR est la plus petite et la plus simple des bornes inférieures locales de l'EQM. Rappelons que la BCR est une borne pour les estimateurs $\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}$ de $\mathbf{g}(\boldsymbol{\theta}^0)$ localement sans biais au voisinage de $\boldsymbol{\theta}^0$ vérifiant (59) :

$$E_{\boldsymbol{\theta}^0 + d\boldsymbol{\theta}}\left[\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}(\mathbf{x})\right] = \mathbf{g}\left(\boldsymbol{\theta}^0 + d\boldsymbol{\theta}\right) + \mathbf{o}(\|d\boldsymbol{\theta}\|) = \mathbf{g}\left(\boldsymbol{\theta}^0\right) + \frac{\partial \mathbf{g}(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}^T} d\boldsymbol{\theta} + \mathbf{o}(\|d\boldsymbol{\theta}\|). \tag{81}$$

Son expression est donc (60) :

$$\mathbf{BCR}_{\mathbf{g}|\boldsymbol{\theta}}\left(\boldsymbol{\theta}^0\right) = \frac{\partial \mathbf{g}(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}^T} \mathbf{F}_{\boldsymbol{\theta}^0}^{-1} \left(\frac{\partial \mathbf{g}(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}^T}\right)^T, \quad \mathbf{F}_{\boldsymbol{\theta}^0} = E_{\boldsymbol{\theta}^0}\left[\frac{\partial \ln p\left(\mathbf{x}; \boldsymbol{\theta}^0\right)}{\partial \boldsymbol{\theta}} \frac{\partial \ln p\left(\mathbf{x}; \boldsymbol{\theta}^0\right)}{\partial \boldsymbol{\theta}^T}\right], \tag{82}$$

expression particulièrement simple à évaluer pour les modèles complexes gaussiens circulaires :

$$\mathbf{x} \sim \mathcal{CN}_N\left(\mathbf{m}_\mathbf{x}(\boldsymbol{\theta}), \mathbf{C}_\mathbf{x}(\boldsymbol{\theta})\right), \quad p\left(\mathbf{x}; \mathbf{m}_\mathbf{x}(\boldsymbol{\theta}), \mathbf{C}_\mathbf{x}(\boldsymbol{\theta})\right) = \frac{e^{-(\mathbf{x} - \mathbf{m}_\mathbf{x}(\boldsymbol{\theta}))^H \mathbf{C}_\mathbf{x}^{-1}(\boldsymbol{\theta})(\mathbf{x} - \mathbf{m}_\mathbf{x}(\boldsymbol{\theta}))}}{\pi^M \left|\mathbf{C}_\mathbf{x}(\boldsymbol{\theta})\right|}, \tag{83}$$

pour lesquels (formule de Slepian Bangs) [VT02, 8.34] :

$$\left(\mathbf{F}_{\boldsymbol{\theta}^0}\right)_{p,p'} = tr\left(\mathbf{C}_\mathbf{x}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{C}_\mathbf{x}(\boldsymbol{\theta})}{\partial \theta_p} \mathbf{C}_\mathbf{x}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{C}_\mathbf{x}(\boldsymbol{\theta})}{\partial \theta_{p'}}\right) + 2\,\mathrm{Re}\left\{\left(\frac{\partial \mathbf{m}_\mathbf{x}(\boldsymbol{\theta})}{\partial \theta_p}\right)^H \mathbf{C}_\mathbf{x}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{m}_\mathbf{x}(\boldsymbol{\theta})}{\partial \theta_{p'}}\right\} \tag{84}$$

Sa simplicité de formulation et ses connections asymptotiques avec l'EQM de l'EMV (cf. Annexe IV-B-p66) en font un outil privilégié d'analyse des performances asymptotiques en estimation de tout système de mesure satisfaisant au modèle d'observation (83) (ce qui reste vrai si le modèle d'observation est gaussien réel ou complexe non-circulaire (57)).
L'idéal serait de pouvoir savoir si pour la valeur $\boldsymbol{\theta}^0$ considérée, le modèle d'observation (83) a atteint la zone asymptotique pour les paramètres d'intérêt, ou s'il a atteint la zone de transition, zone à partir de laquelle la BCR (pour estimateur sans biais) devient une borne un peu trop optimiste par rapport au comportement de l'EMV. Un début de réponse est formulé pour le modèle d'observation déterministe dans §II-A7.

La plupart des problèmes faisant appel à la BCR a été initialement formulé pour des paramètres réels [Fre43] [Dar45] [Rao45] [Cra46]. Toutefois, certaines applications telles que le radar, le sonar ou les télécoms sont confrontées au problème d'estimation de paramètres mixtes complexes [YB92] [VDB94] et réels éventuellement sous contraintes [GO90] [Mar93] [SNg98] [SKM01] [JR04] [YG05] [BHE10]. La forme de la contrainte définissant la BCR (81) basée uniquement sur la différentiabilité de $p(\mathbf{x};\boldsymbol{\theta})$ et $\mathbf{g}(\boldsymbol{\theta})$ en $\boldsymbol{\theta}^0$ peut être facilement étendue au cas où le modèle d'observation est fonction de P_c paramètres complexes et P_r paramètres réels $(2P_c + P_r = P)$:

$$p(\mathbf{x};\boldsymbol{\theta}), \quad \boldsymbol{\theta} = (\theta_1, \ldots, \theta_P)^T = \left(\mathrm{Re}\left\{\boldsymbol{\theta}_c^T\right\}, \mathrm{Im}\left\{\boldsymbol{\theta}_c^T\right\}, \boldsymbol{\theta}_r^T\right)^T \in \mathbb{R}^P,\ \boldsymbol{\theta}_c \in \mathbb{C}^{P_c},\ \boldsymbol{\theta}_r \in \mathbb{R}^{P_r}, \tag{85}$$

conduisant à une paramétrisation duale de $p(\mathbf{x};\boldsymbol{\theta})$ et $\mathbf{g}(\boldsymbol{\theta})$:

$$\mathbf{g}(\boldsymbol{\theta}), p(\mathbf{x};\boldsymbol{\theta}) \quad : \quad \boldsymbol{\theta} = \left(\mathrm{Re}\left\{\boldsymbol{\theta}_c^T\right\}, \mathrm{Im}\left\{\boldsymbol{\theta}_c^T\right\}, \boldsymbol{\theta}_r^T\right)^T \in \mathbb{R}^P \tag{86}$$

$$\mathbf{g}(\underline{\boldsymbol{\theta}}), p(\mathbf{x};\underline{\boldsymbol{\theta}}) \quad : \quad \underline{\boldsymbol{\theta}} = \left(\boldsymbol{\theta}_c^T, (\boldsymbol{\theta}_c^*)^T, \boldsymbol{\theta}_r^T\right)^T \in \mathbb{C}^{2P_c} \times \mathbb{R}^{P_r} \tag{87}$$

et une différentiabilité duale [MCLB12] (cf. Annexe IV-H-p142) [Men12] **[T5]** :

$$\left\{\begin{array}{l} \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} d\boldsymbol{\theta} = \frac{\partial \mathbf{g}(\underline{\boldsymbol{\theta}})}{\partial (\boldsymbol{\theta}_c)^T} d\boldsymbol{\theta}_c + \frac{\partial \mathbf{g}(\underline{\boldsymbol{\theta}})}{\partial (\boldsymbol{\theta}_c^*)^T} d\boldsymbol{\theta}_c^* + \frac{\partial \mathbf{g}(\underline{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}_r^T} d\boldsymbol{\theta}_r = \frac{\partial \mathbf{g}(\underline{\boldsymbol{\theta}})}{\partial \underline{\boldsymbol{\theta}}^T} d\underline{\boldsymbol{\theta}} \\ \frac{\partial p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} d\boldsymbol{\theta} = \frac{\partial p(\mathbf{x};\underline{\boldsymbol{\theta}})}{\partial (\boldsymbol{\theta}_c)^T} d\boldsymbol{\theta}_c + \frac{\partial p(\mathbf{x};\underline{\boldsymbol{\theta}})}{\partial (\boldsymbol{\theta}_c^*)^T} d\boldsymbol{\theta}_c^* + \frac{\partial p(\mathbf{x};\underline{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}_r^T} d\boldsymbol{\theta}_r = \frac{\partial p(\mathbf{x};\underline{\boldsymbol{\theta}})}{\partial \underline{\boldsymbol{\theta}}^T} d\underline{\boldsymbol{\theta}} \end{array}\right., \quad d\boldsymbol{\theta}_c^* = (d\boldsymbol{\theta}_c)^*. \tag{88}$$

L'extension naturelle des Lemmes $1, 2, 3$-p27 au cas des produits hermitiens [MCLB12] [Men12] **[T5]** et la forme duale de (81) :

$$E_{\underline{\boldsymbol{\theta}}^0 + d\underline{\boldsymbol{\theta}}}\left[\widehat{\mathbf{g}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x})\right] = \mathbf{g}\left(\underline{\boldsymbol{\theta}}^0 + d\underline{\boldsymbol{\theta}}\right) + \mathbf{o}(d\underline{\boldsymbol{\theta}}) = \mathbf{g}\left(\underline{\boldsymbol{\theta}}^0\right) + \frac{\partial \mathbf{g}\left(\underline{\boldsymbol{\theta}}^0\right)}{\partial \underline{\boldsymbol{\theta}}^T} d\underline{\boldsymbol{\theta}} + \mathbf{o}(d\underline{\boldsymbol{\theta}}), \tag{89}$$

conduisent à la forme générale suivante de la BCR :

$$\mathbf{BCR}_{\underline{\mathbf{g}}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}^0\right) = \frac{\partial \underline{\mathbf{g}}^*\left(\underline{\boldsymbol{\theta}}^0\right)}{\partial \underline{\boldsymbol{\theta}}^H} \mathbf{F}_{\underline{\boldsymbol{\theta}}^0}^{-1} \frac{\partial \underline{\mathbf{g}}^T\left(\underline{\boldsymbol{\theta}}^0\right)}{\partial \underline{\boldsymbol{\theta}}} \tag{90}$$

$$\mathbf{F}_{\underline{\boldsymbol{\theta}}^0} = E_{\underline{\boldsymbol{\theta}}^0}\left[\frac{\partial \ln p(\mathbf{x};\underline{\boldsymbol{\theta}})}{\partial \underline{\boldsymbol{\theta}}} \left(\frac{\partial \ln p(\mathbf{x};\underline{\boldsymbol{\theta}})}{\partial \underline{\boldsymbol{\theta}}}\right)^H\right], \quad \mathbf{F}_{\underline{\boldsymbol{\theta}}^0} = \mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\left(\frac{\partial \ln p(\mathbf{x};\underline{\boldsymbol{\theta}})}{\partial \underline{\boldsymbol{\theta}}}\right)^*\right\}\right) \tag{91}$$

$$\left(\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x}) - \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)\right)_{\text{eff}}^T = \frac{\partial \ln p(\mathbf{x};\underline{\boldsymbol{\theta}})}{\partial \underline{\boldsymbol{\theta}}^H} \mathbf{F}_{\underline{\boldsymbol{\theta}}^0}^{-1} \frac{\partial \underline{\mathbf{g}}^T\left(\underline{\boldsymbol{\theta}}^0\right)}{\partial \underline{\boldsymbol{\theta}}} \tag{92}$$

Une autre particularité de la BCR est qu'elle permet d'intégrer facilement la prise en compte d'un ensemble de K contraintes égalités fonction du vecteur de paramètres inconnus $\underline{\boldsymbol{\theta}}$ (K_c contraintes complexes et K_r contraintes réelles) non redondantes :

$$\underline{\mathbf{f}}(\underline{\boldsymbol{\theta}}) = \mathbf{0} \in \mathbb{C}^{2K_c} \times \mathbb{R}^{K_r},\ 2K_c + K_r = K,\ 1 \leq K < P \tag{93}$$

où non redondantes signifie que la matrice $\frac{\partial \underline{\mathbf{f}}(\underline{\boldsymbol{\theta}})}{\partial \underline{\boldsymbol{\theta}}^T} \in \mathcal{M}_{\mathbb{C}}(K,P)$ est de rang K. Dans ce cas, la BCR contrainte [GO90] [Mar97] [SNg98] [BHE09] [JR04] [ZJLJ11] [MCLB12] **[T5]** s'écrit :

$$\mathbf{BCR}^c_{\underline{\mathbf{g}}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}^0\right) = \frac{\partial \underline{\mathbf{g}}^*\left(\underline{\boldsymbol{\theta}}^0\right)}{\partial \underline{\boldsymbol{\theta}}^H} \mathbf{U}^*_{\underline{\boldsymbol{\theta}}^0}\left(\mathbf{F}^c_{\underline{\boldsymbol{\theta}}^0}\right)^{-1}\mathbf{U}^T_{\underline{\boldsymbol{\theta}}^0}\frac{\partial \underline{\mathbf{g}}^T\left(\underline{\boldsymbol{\theta}}^0\right)}{\partial \underline{\boldsymbol{\theta}}} \quad (94)$$

$$\mathbf{F}^c_{\underline{\boldsymbol{\theta}}^0} = \mathbf{U}^T_{\underline{\boldsymbol{\theta}}^0}\mathbf{F}_{\underline{\boldsymbol{\theta}}^0}\mathbf{U}^*_{\underline{\boldsymbol{\theta}}^0}, \quad \mathbf{F}^c_{\underline{\boldsymbol{\theta}}^0} = \mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\mathbf{U}^H_{\underline{\boldsymbol{\theta}}^0}\left(\frac{\partial \ln p\left(\mathbf{x};\underline{\boldsymbol{\theta}}\right)}{\partial \underline{\boldsymbol{\theta}}}\right)^*\right\}\right) \quad (95)$$

$$\left(\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}\left(\mathbf{x}\right) - \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)\right)^T_{\text{eff}} = \frac{\partial \ln p\left(\mathbf{x};\underline{\boldsymbol{\theta}}\right)}{\partial \underline{\boldsymbol{\theta}}^H}\mathbf{U}^*_{\underline{\boldsymbol{\theta}}^0}\left(\mathbf{F}^c_{\underline{\boldsymbol{\theta}}^0}\right)^{-1}\mathbf{U}^T_{\underline{\boldsymbol{\theta}}^0}\frac{\partial \underline{\mathbf{g}}^T\left(\underline{\boldsymbol{\theta}}^0\right)}{\partial \underline{\boldsymbol{\theta}}}$$

où $\mathbf{U}_{\underline{\boldsymbol{\theta}}^0}$ est une base $\ker\left\{\frac{\partial \underline{\mathbf{f}}(\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\theta}}^T}\right\}$ et $\mathbf{F}^c_{\underline{\boldsymbol{\theta}}^0}$ représente la MIF contrainte (on remarquera que la BCR contrainte ne dépend pas du choix de $\mathbf{U}_{\underline{\boldsymbol{\theta}}^0}$).
De plus il a été montré [YG05] [MKS07] [MCLB12] **[T5]** qu'un ensemble de contraintes égalités non redondantes (93):

$$\underline{\mathbf{f}}\left(\underline{\boldsymbol{\theta}}\right) = \mathbf{0} \in \mathbb{C}^K, \ \underline{\boldsymbol{\theta}} \in \mathbb{C}^P, \ 1 \leq K < P,$$

est équivalente à une reparamétrisation implicite (théorème des fonctions implicites [Haz02]) de la forme $\underline{\boldsymbol{\theta}} = \underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}\right)$, $\dim\left\{\underline{\boldsymbol{\omega}}\right\} = P - K$, vérifiant l'inégalité de reparamétrisation [MCLB12] **[T5]** :

$$\mathbf{BCR}_{\underline{\mathbf{g}}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^0\right)\right) \geq \mathbf{BCR}_{\underline{\mathbf{g}}(\underline{\boldsymbol{\theta}})|\underline{\boldsymbol{\omega}}}\left(\underline{\boldsymbol{\omega}}^0\right) \quad (96)$$

où :

$$\mathbf{BCR}_{\underline{\mathbf{g}}(\underline{\boldsymbol{\theta}})|\underline{\boldsymbol{\omega}}}\left(\underline{\boldsymbol{\omega}}^0\right) = \frac{\partial \underline{\mathbf{g}}^*\left(\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^0\right)\right)}{\partial \underline{\boldsymbol{\theta}}^H}\mathbf{BCR}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\omega}}}\left(\underline{\boldsymbol{\omega}}^0\right)\frac{\partial \underline{\mathbf{g}}^T\left(\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^0\right)\right)}{\partial \underline{\boldsymbol{\theta}}},$$

$$\mathbf{BCR}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\omega}}}\left(\underline{\boldsymbol{\omega}}^0\right) = \frac{\partial \underline{\boldsymbol{\theta}}^*\left(\underline{\boldsymbol{\omega}}^0\right)}{\partial \underline{\boldsymbol{\omega}}^H}\left(\frac{\partial \underline{\boldsymbol{\theta}}^T\left(\underline{\boldsymbol{\omega}}^0\right)}{\partial \underline{\boldsymbol{\omega}}}\mathbf{F}_{\underline{\boldsymbol{\theta}}(\underline{\boldsymbol{\omega}}^0)}\frac{\partial \underline{\boldsymbol{\theta}}^*\left(\underline{\boldsymbol{\omega}}^0\right)}{\partial \underline{\boldsymbol{\omega}}^H}\right)^{-1}\frac{\partial \underline{\boldsymbol{\theta}}^T\left(\underline{\boldsymbol{\omega}}^0\right)}{\partial \underline{\boldsymbol{\omega}}},$$

et $\frac{\partial \underline{\boldsymbol{\theta}}(\underline{\boldsymbol{\omega}}^0)}{\partial \underline{\boldsymbol{\omega}}^T} \triangleq \mathbf{U}_{\underline{\boldsymbol{\theta}}^0}$ où $\mathbf{U}_{\underline{\boldsymbol{\theta}}^0} \in \mathcal{M}_{\mathbb{C}}(P, P-K)$ est une base de $\ker\left\{\frac{\partial \underline{\mathbf{f}}(\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\theta}}^T}\right\}$. $\mathbf{U}_{\underline{\boldsymbol{\theta}}^0}$ peut toujours être évaluée sous la forme (après réarrangement des composantes de $\underline{\boldsymbol{\theta}}$) suivante [MKS07] [MCLB12] **[T5]** :

$$\mathbf{U}_{\underline{\boldsymbol{\theta}}^0} = \begin{bmatrix} \mathbf{I}_{P-K} \\ -\left(\frac{\partial \underline{\mathbf{f}}(\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\varepsilon}}^T}\right)^{-1}\frac{\partial \underline{\mathbf{f}}(\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\omega}}^T}\end{bmatrix}, \quad \underline{\boldsymbol{\theta}} = \begin{pmatrix} \underline{\boldsymbol{\omega}} \\ \underline{\boldsymbol{\varepsilon}} \end{pmatrix} \quad (97)$$

où $\underline{\boldsymbol{\varepsilon}}$ est un vecteur de K composantes de $\underline{\boldsymbol{\theta}}$ pour lesquelles les K colonnes de dérivées partielles - colonnes de la matrice $\frac{\partial \underline{\mathbf{f}}(\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\theta}}^T}$ - sont indépendantes.

D'un point de vue pratique la combinaison de (94) et de (96)(97) fournit un outil algébrique puissant pour l'analyse des performance en estimation d'un système de mesure de paramètres d'intérêts constitués de L modèles d'observations [MGCL12] **[T5]**.
A titre illustratif, considérons un système de mesure produisant L modèles d'observation bande étroite (125) associés à l'observation de M sources :

$$\mathbf{x}^l\left(\underline{\boldsymbol{\theta}}^l\right) = \mathbf{B}^l\left(\underline{\boldsymbol{\Xi}}^l, \underline{\boldsymbol{\delta}}^l\right)\boldsymbol{\sigma}^l + \mathbf{n}^l\left(\underline{\boldsymbol{\theta}}^l_{\mathbf{n}}, \underline{\boldsymbol{\delta}}^l\right), \quad \left(\underline{\boldsymbol{\theta}}^l\right)^T = \left(\left(\underline{\boldsymbol{\sigma}}^l\right)^T, \left(\underline{\boldsymbol{\Xi}}^l\right)^T, \left(\underline{\boldsymbol{\delta}}^l\right)^T, \left(\underline{\boldsymbol{\theta}}^l_{\mathbf{n}}\right)^T\right), \quad l = 1, ..., L, \quad (98)$$

où pour chaque observation l :

· $\mathbf{x}^l \in \mathcal{M}_{\mathbb{C}}\left(N^l, 1\right)$ est le vecteur d'observations instantanées de dimension N^l,

· $\boldsymbol{\sigma}^l$ est le vecteur des amplitudes des M signaux sources,

· $\underline{\boldsymbol{\delta}}^l$ est un vecteur de paramètres indépendants des sources de bruit ou de signal, c'est-à-dire typiquement des paramètres du dispositif physique produisant les observations (position des capteurs, gains divers, ...),

· $\mathbf{n}^l\left(\underline{\boldsymbol{\theta}}^l_{\mathbf{n}}, \underline{\boldsymbol{\delta}}^l\right) \in \mathcal{M}_{\mathbb{C}}\left(N^l, 1\right)$ est le vecteur des signaux de bruit (bruits thermiques, brouilleurs, ...) additifs dépendant des vecteurs de paramètres $\underline{\boldsymbol{\theta}}^l_{\mathbf{n}}$ associées aux sources de bruit et des paramètres $\underline{\boldsymbol{\delta}}^l$,

· $\underline{\boldsymbol{\Xi}}^l = \left(\left(\underline{\boldsymbol{\varepsilon}}_1^l\right)^T, \ldots, \left(\underline{\boldsymbol{\varepsilon}}_M^l\right)^T \right)^T$ est l'ensemble des paramètres inconnus d'intérêt des M sources,

· $\mathbf{B}^l\left(\underline{\boldsymbol{\Xi}}^l, \underline{\boldsymbol{\delta}}^l\right) = \left[\mathbf{b}^l\left(\underline{\boldsymbol{\varepsilon}}_1^l, \underline{\boldsymbol{\delta}}^l\right), \mathbf{b}^l\left(\underline{\boldsymbol{\varepsilon}}_2^l, \underline{\boldsymbol{\delta}}^l\right), \cdots, \mathbf{b}^l\left(\underline{\boldsymbol{\varepsilon}}_M^l, \underline{\boldsymbol{\delta}}^l\right)\right] \in \mathcal{M}_{\mathbb{C}}\left(N^l, M\right)$ est la matrice des fonctions de transfert vectorielles dépendant des paramètres inconnus d'intérêt des M sources et des paramètres $\underline{\boldsymbol{\delta}}^l$.

Notons alors $\mathbf{X}^T = \left(\left(\mathbf{x}^1\right)^T, \ldots, \left(\mathbf{x}^L\right)^T \right)$, $\underline{\boldsymbol{\Theta}}^T = \left(\left(\underline{\boldsymbol{\theta}}^1\right)^T, \ldots, \left(\underline{\boldsymbol{\theta}}^L\right)^T \right)$ et supposons qu'une expression analytique de $p\left(\mathbf{X}; \underline{\boldsymbol{\Theta}}\right)$ soit disponible de telle sorte que la FIM (non contrainte) :

$$\mathbf{F}_{\underline{\boldsymbol{\Theta}}^0} = E_{\underline{\boldsymbol{\Theta}}^0}\left[\frac{\partial \ln p\left(\mathbf{X}; \underline{\boldsymbol{\Theta}}^0\right)}{\partial \underline{\boldsymbol{\Theta}}} \left(\frac{\partial \ln p\left(\mathbf{X}; \underline{\boldsymbol{\Theta}}^0\right)}{\partial \underline{\boldsymbol{\Theta}}} \right)^H \right] \tag{99}$$

puisse être calculée analytiquement ou numériquement (ce qui est le cas si $\mathbf{X}\left(\underline{\boldsymbol{\Theta}}\right)$ est gaussien (84)).
Alors l'impact de nombre de facteurs sur la précision en estimation des paramètres d'intérêt des M sources $\left\{\underline{\boldsymbol{\Xi}}^l\right\}_{l=1}^L$ peut être analysé grâce à (94) (96) (97) par simple calcul algébrique d'une BCR contrainte à partir de l'expression (99).

Par exemple, il est bien connu que les performances des stratégies d'estimation paramétriques tel que l'EMV dépendent par essence de la bonne connaissance du modèle paramétrique des observations. Par conséquent un facteur déterminant dans l'évaluation réaliste des performances attendues est la prise en compte d'éventuelles erreurs de modèles [FLV06] [Fer11]. Dans ce cadre, l'effet de la connaissance ou non de la valeur d'un paramètre δ_q^l du dispositif physique produisant les observations peut être évalué en changeant le statut du paramètre, à savoir en le faisant passer d'inconnu à connu (et vice versa), ce qui se traduit en terme de contrainte égalité simplement par :

$$\delta_q^l = \left(\delta_q^l\right)^0 \;\Leftrightarrow\; \delta_q^l - \left(\delta_q^l\right)^0 = 0. \tag{100}$$

Il suffit alors d'exprimer la matrice $\mathbf{U}_{\underline{\boldsymbol{\theta}}^0}$ associée à (100) à l'aide de (97) pour calculer algébriquement la BCR contrainte suivant (96) et pouvoir comparer les performances attendues sur $\left\{\underline{\boldsymbol{\Xi}}^l\right\}_{l=1}^L$ suivant que la valeur de δ_q^l est connue ou inconnue; et décider si la variation de performance (augmentation de la BCR selon (96)) réclame une calibration de δ_q^l (équivalente à connaître δ_q^l).

On peut également mesurer l'effet de la variabilité des paramètres d'intérêt sur l'estimation de fonctions de ces paramètres $\underline{\mathbf{g}}\left(\underline{\boldsymbol{\Theta}}\right) \triangleq \underline{\mathbf{g}}\left(\underline{\boldsymbol{\Xi}}^1, \ldots, \underline{\boldsymbol{\Xi}}^L\right)$ (par exemple leur valeur moyenne : $\underline{\mathbf{g}}\left(\underline{\boldsymbol{\Theta}}\right) = \frac{1}{L}\sum_{l=1}^L \underline{\boldsymbol{\Xi}}^l$), ce qui revient à introduire les contraintes suivantes :

$$\underline{\boldsymbol{\Xi}}^l - \underline{\boldsymbol{\Xi}}^1 = 0,\; 2 \leq l \leq L. \tag{101}$$

En effet cette variabilité peut dépendre du scénario de M sources considéré : par exemple, si $\left\{\underline{\boldsymbol{\Xi}}^l\right\}_{l=1}^L$ sont des paramètres cinématiques de cibles observées par un système de mesure radar, leur variabilité dépend de leur nature (écho fixe, cible lente, cible rapide, cible manoeuvrante, ...) et de la durée des observations. L'étude de l'effet de la variabilité devient une étude de la sensibilité des performances en estimation de $\underline{\mathbf{g}}\left(\underline{\boldsymbol{\Xi}}^1, \ldots, \underline{\boldsymbol{\Xi}}^L\right)$ en fonction du scénario considéré.

On peut également s'intéresser au gain potentiel en terme de précision d'estimation lorsque l'on raffine la complexité du modèle d'observation, ce qui consiste en général à exhiber une relation entre les différentes observations d'un même paramètre pouvant prendre la forme d'une reparamétrisation (par exemple les algorithmes de type Track Before Detect en poursuite) :

$$\underline{\boldsymbol{\Xi}}^l = \underline{\boldsymbol{\Xi}}^l\left(\boldsymbol{\omega}\right),\; 1 \leq l \leq L. \tag{102}$$

Là encore l'emploi de (100) permet de quantifier ce que prédit (96) et de décider si le gain en performance offert par cette reparamétrisation est suffisant pour implémenter un estimateur de $\boldsymbol{\omega}$ en lieu et place des estimateurs de $\left\{\underline{\boldsymbol{\Xi}}^l\right\}_{l=1}^L$.
Cette idée peut également être appliquée aux paramètres dit de nuisance [Sch91] [Kay93] [VT02], c'est à dire les paramètres inconnus des sources qui ne sont pas les paramètres d'intérêts. Par exemple, la prise en compte d'un effet Doppler pendant les L observations s'écrit :

$$\left(\boldsymbol{\sigma}^l\right)_m = \left(\boldsymbol{\sigma}^l\right)_m e^{j2\pi\omega_m\left(t^l - t^1\right)}, \quad 1 \leq m \leq M, 2 \leq l \leq L, \tag{103}$$

où $\left(t^1,\ldots,t^L\right)$ sont les instants d'observation des L modèles. L'effet Doppler peut être vu soit comme une reparamétrisation (103) $\left\{\boldsymbol{\sigma}^l=\boldsymbol{\sigma}^l\left(\boldsymbol{\sigma}^1,\boldsymbol{\omega}\right)\right\}_{l=1}^L$, soit comme un ensemble de contraintes égalités :

$$\left(\boldsymbol{\sigma}^l\right)_m-\left(\boldsymbol{\sigma}^l\right)_m e^{j2\pi\omega_m\left(t^l-t^1\right)}=0,\quad 1\leq m\leq M, 2\leq l\leq L, \tag{104}$$

dont l'effet sera, en vertu de (96), d'améliorer les performances en estimation des paramètres d'intérêt $\left\{\underline{\boldsymbol{\Xi}}^l\right\}_{l=1}^L$. D'un point de vue de la conception du système de mesure, l'exploitation d'un effet Doppler requiert l'acquisition d'un signal cohérent pendant les L observations, ce qui est une exigence qui a un coût dans sa mise en oeuvre (notamment relativement à la stabilité de tous les oscillateurs).

Les possibilité d'usage la BCR sous contraintes dépassent largement, à la fois théoriquement et en pratique, ces quelques exemples et sont abordés avec beaucoup plus en détails dans [MGCL12] **[T5]**.

Ma conviction est que la BCR sous contrainte est l'outil d'avenir pour l'analyse et la conception des systèmes de mesure. En effet pour un système un peu réaliste, c'est à dire à bande limitée (donc pas bande étroite) et désirant exploiter les "diversités" disponibles (temporelle, fréquentielle, spatiale, codage, ...), la combinaison de plusieurs modèles d'observations (ne serait-ce que l'exploitation de forme d'ondes (codes) différentes [Men12] **[T5]**) rend l'obtention de formule analytique de la BCR pour paramètres d'intérêts de plus en plus difficile. Et quand bien même elle est possible, la formule obtenue est tellement complexe [MGCL12, §IV.B.2] **[T5]** que nous ne savons l'exploiter qu'à travers son évaluation numérique. Sans compter le fait qu'il faut refaire le calcul analytique à chaque fois qu'une hypothèse (contrainte) est introduite, ou s'appuyer sur une "bibliothèque" de formules analytiques existantes.
Et c'est précisément là que réside la puissance de la BCR sous contrainte : il n'y qu'à disposer d'une "bibliothèque" de matrices de contraintes $\mathbf{U}_{\underline{\boldsymbol{\theta}}^0}$ et de les combiner entre elles pour obtenir [MGCL12, §VI] **[T5]**, quelque soit la forme de $\mathbf{F}_{\underline{\boldsymbol{\Theta}}^0}$ (99), les performances en estimations des paramètres d'intérêt par une simple combinaison d'opérations matricielles (multiplication, inversion, multiplication). Ceci pose naturellement le problème de l'optimisation du calcul de la BCR contrainte, problème pour lequel certains résultats existent déjà [HF94] [Tun12].

Perspectives :

Dans [MGCL12, §IV] **[T5]** nous avons fourni l'expression de $\mathbf{F}_{\underline{\boldsymbol{\Theta}}^0}$ (99) pour le modèle d'observation déterministe. Il serait pertinent de fournir l'expression de $\mathbf{F}_{\underline{\boldsymbol{\Theta}}^0}$ pour le modèle d'observation stochastique généralisé (L modèles d'observation à bande limitée) caractérisant les signaux sources lorsque leurs amplitudes sont gaussiennes (modèles Swerling 1 et 2 en radar), ce qui permettrait d'étudier la sensibilité d'un système de mesure au modèle probabiliste des amplitudes des sources.
A terme l'obtention de $\mathbf{F}_{\underline{\boldsymbol{\Theta}}^0}$ pour un modèle d'observation SIRV [Ova11] permettrait de prendre en compte des bruits (fouillis, brouilleurs, bruits thermiques) plus réalistes.

7) Précision exhaustive d'un estimateur :

Considérons le cas particulier des erreurs d'estimation uniformes $\left(\boldsymbol{\xi}^-=\boldsymbol{\xi}^+=\boldsymbol{\xi}=\xi\mathbf{1}_Q\right)$ pour lequel la précision *exhaustive* (3) se réduit à **[T6]** :

$$\mathcal{O}_{\boldsymbol{\theta}^0}\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)},\xi\right)\triangleq\mathcal{O}_{\boldsymbol{\theta}^0}\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)},\xi\mathbf{1}_Q,\xi\mathbf{1}_Q\right)=\mathcal{P}\left(\bigcap_{q=1}^{Q}\left(\left|\widehat{g_q\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})-g_q\left(\boldsymbol{\theta}^0\right)\right|<\xi\right);\boldsymbol{\theta}^0\right)$$

On peut alors définir une relation d'ordre dite "stochastique" [GL01, §3.1.3] entre 2 estimateurs $\widehat{\mathbf{g}_1\left(\boldsymbol{\theta}^0\right)}$ et $\widehat{\mathbf{g}_2\left(\boldsymbol{\theta}^0\right)}$ par :

$$\widehat{\mathbf{g}_1\left(\boldsymbol{\theta}^0\right)}\overset{st}{\geq}\widehat{\mathbf{g}_2\left(\boldsymbol{\theta}^0\right)}\iff\exists\xi_0\ /\ \forall\xi<\xi_0,\quad\mathcal{O}_{\boldsymbol{\theta}^0}\left(\widehat{\mathbf{g}_1\left(\boldsymbol{\theta}^0\right)},\xi\right)\geq\mathcal{O}_{\boldsymbol{\theta}^0}\left(\widehat{\mathbf{g}_2\left(\boldsymbol{\theta}^0\right)},\xi\right),\quad\xi_0,\xi>0, \tag{105}$$

cette relation permettant d'introduire le concept selon lequel l'estimateur $\widehat{\mathbf{g}_1\left(\boldsymbol{\theta}^0\right)}$ est "plus précis que" (p.p.q.) l'estimateur $\widehat{\mathbf{g}_2\left(\boldsymbol{\theta}^0\right)}$. Malheureusement, il est établi [GL01, §3.1.3] que les relations d'ordre de ce type ne sont que

des relations d'ordre partielles : on ne peut pas comparer tous les estimateurs entre eux au sens de la précision *exhaustive*. Cette limitation disparait avec la précision *a priori* (16) où la norme définit naturellement une relation d'ordre totale, mais qui est potentiellement trompeuse car les moments "masquent" le comportement local ($\xi \rightarrow 0$) des estimateurs.
Indépendamment d'arguments prenant en compte la complexité de mise en oeuvre, conceptuellement nous nous retrouvons donc face à un dilemme : soit nous utilisons une mesure précise (*exhaustive*) de la qualité d'un estimateur mais ne permettant pas de définir le meilleur estimateur au sens de cette mesure, soit nous utilisons une mesure imprécise (*a priori*) de la qualité d'un estimateur mais permettant de définir le meilleur estimateur.
Un compromis possible est la mesure de précision *exhaustive* de l'estimateur (de précision) *a priori* le meilleur $\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}_{opt}$, notamment à l'ordre 2 **[T6]**.

Comme l'évaluation analytique de $\mathcal{O}_{\boldsymbol{\theta}^0}\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}_{opt}, \boldsymbol{\xi}^-, \boldsymbol{\xi}^+\right)$ (3) est généralement impossible, on procède par encadrement (expression de bornes), dont un possible est **[T6]** :

$$\mathcal{P}\left(\left\|\mathbf{D}^{-1}_{\frac{\boldsymbol{\xi}^+ + \boldsymbol{\xi}^-}{2}}\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}_{opt}(\mathbf{x}) - \mathbf{g}\left(\boldsymbol{\theta}^0\right) - \frac{\boldsymbol{\xi}^+ - \boldsymbol{\xi}^-}{2}\right)\right\|_2^2 < 1; \boldsymbol{\theta}^0\right) \leq \mathcal{O}_{\boldsymbol{\theta}^0}\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}_{opt}, \boldsymbol{\xi}^-, \boldsymbol{\xi}^+\right)$$
$$\leq \mathcal{P}\left(\left\|\mathbf{D}^{-1}_{\frac{\boldsymbol{\xi}^+ + \boldsymbol{\xi}^-}{2}}\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}_{opt}(\mathbf{x}) - \mathbf{g}\left(\boldsymbol{\theta}^0\right) - \frac{\boldsymbol{\xi}^+ - \boldsymbol{\xi}^-}{2}\right)\right\|_2^2 < Q; \boldsymbol{\theta}^0\right) \quad (106)$$

La définition de bornes visant à simplifier l'approche analytique, il est implicite que l'utilisation de ces bornes génère une perte d'information souvent difficilement quantifiable. Par conséquent, le principal problème dans la définition (puis l'utilisation) de bornes est de savoir a priori si elles sont précises ("tigth"). La pertinence de ces bornes pour un problème donné est le plus souvent vérifiée a posteriori au vue des résultats qu'elles fournissent.
Les bornes considérées (106) reposent sur l'encadrement bien connu d'un hyper-rectangle par l'hyper-ellipse contenue (borne inférieure) et l'hyper-ellipse contenante (borne supérieure), encadrement dont il impossible de prédire l'effet sur l'évaluation des probabilités à mesure que la dimension Q augmente.

Néanmoinssi $\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}_{opt}$ est un estimateur gaussien : $\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}_{opt}(\mathbf{x}) - \mathbf{g}\left(\boldsymbol{\theta}^0\right) \sim \mathcal{N}\left(\mathbf{b}\left(\boldsymbol{\theta}^0\right), \mathbf{C}\left(\boldsymbol{\theta}^0\right)\right)$, alors (106) s'écrit également :

$$\mathcal{P}\left(e\chi^2_Q\left(\delta\left(\boldsymbol{\theta}^0\right), \boldsymbol{\sigma}^2\left(\boldsymbol{\theta}^0\right)\right) < 1\right) \leq \mathcal{O}_{\boldsymbol{\theta}^0}\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}_{opt}, \boldsymbol{\xi}^-, \boldsymbol{\xi}^+\right) \leq \mathcal{P}\left(e\chi^2_Q\left(\delta\left(\boldsymbol{\theta}^0\right), \boldsymbol{\sigma}^2\left(\boldsymbol{\theta}^0\right)\right) < Q\right) \quad (107)$$
$$\left.\begin{array}{r} \delta\left(\boldsymbol{\theta}^0\right) = \left\|\mathbf{U}^T\left(\boldsymbol{\theta}^0\right)\mathbf{D}^{-1}_{\frac{\boldsymbol{\xi}^+ + \boldsymbol{\xi}^-}{2}}\left(\mathbf{b}\left(\boldsymbol{\theta}^0\right) - \frac{\boldsymbol{\xi}^+ - \boldsymbol{\xi}^-}{2}\right)\right\|_2^2 \\ \mathbf{U}\left(\boldsymbol{\theta}^0\right)\mathbf{D}_{\boldsymbol{\sigma}^2\left(\boldsymbol{\theta}^0\right)}\mathbf{U}^T\left(\boldsymbol{\theta}^0\right) = \mathbf{D}^{-1}_{\frac{\boldsymbol{\xi}^+ + \boldsymbol{\xi}^-}{2}}\mathbf{C}\left(\boldsymbol{\theta}^0\right)\mathbf{D}^{-1}_{\frac{\boldsymbol{\xi}^+ + \boldsymbol{\xi}^-}{2}} \end{array}\right|, \ \mathbf{U}\left(\boldsymbol{\theta}^0\right)\mathbf{U}^T\left(\boldsymbol{\theta}^0\right) = \mathbf{U}^T\left(\boldsymbol{\theta}^0\right)\mathbf{U}\left(\boldsymbol{\theta}^0\right) = \mathbf{I},$$

où $\mathcal{P}\left(e\chi^2_Q\left(\delta, \boldsymbol{\sigma}^2\right) < t\right)$ (fonction de répartition) peut être obtenue par approximation numérique [RP49], ce qui représente un intérêt opérationnel certain.

Par conséquent, un parfait exemple d'application du compromis considéré est le modèle d'observation (réel ou complexe) déterministe pour lequel l'estimateur (*a priori*) le meilleur (l'EMV) est asymptotiquement gaussien, sans biais, efficace (il atteint la BCR) et réalisable (il atteint donc aussi la BB) lorsqu'asymptotique signifie au choix : $N \rightarrow \infty$ et/ou $L \rightarrow \infty$ et/ou RSB $\rightarrow \infty$ (cf. Annexe IV-B-p66).

Néanmoins pour être applicable, (106) requiert que $\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}_{opt}$ soit gaussien, ce qui n'est établi que dans la zone asymptotique : il est donc nécessaire de pouvoir déterminer si pour la valeur particulière $\boldsymbol{\theta}^0$, le modèle d'observation de d.d.p.. $p\left(\mathbf{x}; \boldsymbol{\theta}^0\right)$ opère en zone asymptotique.

Un test de fonctionnement en zone asymptotique relativement simple peut être formulé lors de l'estimation des paramètres d'intérêt $\boldsymbol{\Xi}^0$ de M sources de signal pour un modèle bande étroite pour lequel l'EMV s'écrit (cf. Annexe

IV-B-p66) **[T6]** :

$$\widehat{\Xi^0} = \arg\min_{\Xi} \left\{ C\left(\Xi;\Xi^0\right) \right\}, \quad C\left(\Xi;\Xi^0\right) = \sum_{l=1}^{L} \frac{\left\| \mathbf{\Pi}_{\mathbf{B}(\Xi)} \mathbf{x}^l \left(\Xi^0\right) \right\|^2}{LM}, \quad \Xi^T = \left(\boldsymbol{\varepsilon}_1^T, \boldsymbol{\varepsilon}_2^T, \cdots, \boldsymbol{\varepsilon}_M^T\right) \tag{108}$$

$$C\left(\Xi;\Xi^0\right) \sim \mathcal{CX}^2_{ML}\left(\mathcal{F}\left(\Xi;\Xi^0\right), \frac{\sigma^2_{\mathbf{n}}}{LM} \right), \quad \mathcal{F}\left(\Xi;\Xi^0\right) = \sum_{l=1}^{L} \frac{\left\| \mathbf{\Pi}_{\mathbf{B}(\Xi)} \mathbf{B}\left(\Xi^0\right) \boldsymbol{\sigma}^l \right\|^2}{LM}$$

où $\mathcal{F}\left(\Xi;\Xi^0\right)$ est une fonction de corrélation généralisée (filtre adapté généralisé) et $\mathcal{CX}^2_K\left(\delta,\sigma^2\right)$ représente une loi du chi-2 complexe (circulaire) décentrée à K degrés de liberté et de paramètre de décentrement δ.
Ce test consiste à vérifier que toutes les réalisations possibles du critère $C\left(\Xi;\Xi^0\right)$ appartienne au lobe principale à α_{dB} de $C\left(\Xi^0;\Xi^0\right)$, ce qui, pour α_{dB} suffisamment proche de 0, valide les hypothèses sous lesquelles (développement de Gauss-Newton) la gaussianité asymptotique de $\widehat{\Xi^0}$ a été démontrée dans [RFCL06]. Dans le principe (cf. [REKGCLR13] **[T6]** pour plus de détails), il suffit pour cela de vérifier que les supports (relativement à un seuil infinitésimal) des d.d.p.. $C\left(\Xi_\alpha;\Xi^0\right)$ et $C\left(\Xi^0;\Xi^0\right)$, où Ξ_α vérifie $\frac{\mathcal{F}(\Xi_\alpha;\Xi^0)}{\mathcal{F}(\Xi^0;\Xi^0)} = \alpha$, sont sans intersection, ce qui signifie que pour toute valeur $\Xi \notin \Upsilon_{\Xi^0}(\alpha)$, $\Upsilon_{\Xi^0}(\alpha) = \left\{ \Xi \;/\; \frac{\mathcal{F}(\Xi;\Xi^0)}{\mathcal{F}(\Xi^0;\Xi^0)} > \alpha \right\}$, aucune réalisation du critère $C\left(\Xi;\Xi^0\right)$ ne peut être supérieure à une réalisation du critère $C\left(\Xi^0;\Xi^0\right)$ (CQFV).

Outre la caractérisation de la qualité de l'EMV, (106)(107) offre la possibilité de mesurer la "résolvabilité" ("resolvability") de l'EMV au sens de sa capacité à distinguer (résoudre) les estimées $\{\widehat{\boldsymbol{\varepsilon}_m}\}_{m=1}^M$ des paramètres d'intérêt des M sources ("statistical résolution limit" lorsque $M = 2$). Pour cela, il faut se doter d'un critère de séparation, par exemple basé sur la prise en compte d'hyperrectangle $\mathcal{R}_m$ centré sur les valeurs inconnues des paramètres à estimer **[T6]** :

$$\mathcal{R}_m\left(\boldsymbol{\epsilon}_m\right) = \left\{ \boldsymbol{\varepsilon} : \bigcap_{p=1}^{P} \left| (\boldsymbol{\varepsilon})_p - (\boldsymbol{\varepsilon}_m)_p \right| < (\boldsymbol{\epsilon}_m)_p \right\} \tag{109}$$

et définir de la probabilité de résolvabilité associée :

$$\mathcal{O}_{\boldsymbol{\theta}^0}\left(\widehat{\Xi^0}, \boldsymbol{\xi}, \boldsymbol{\xi} \right), \; \xi^T = \left(\boldsymbol{\epsilon}_1^T, \boldsymbol{\epsilon}_2^T, \cdots, \boldsymbol{\epsilon}_M^T\right), \tag{110}$$

$$\mathcal{R}_m\left(\boldsymbol{\epsilon}_m\right) \cap \mathcal{R}_{m'}\left(\boldsymbol{\epsilon}_{m'}\right) = \varnothing, \; \forall m' \neq m \in [1, M] \tag{111}$$

c'est à dire la probabilité que $\widehat{\boldsymbol{\varepsilon}_m} \in \mathcal{R}_m\left(\boldsymbol{\epsilon}_m\right), 1 \leq m \leq M$, lorsque les hyperrectangles $\mathcal{R}_m$ sont disjoints (111). L'idée sous jacente est que nous ne considérons pas comme un succès toute réalisation qui conduit au "saut" d'un estimé d'un hyperrectangle à un autre.
Alors nous pouvons dire que les vecteurs de paramètres $\{\boldsymbol{\varepsilon}_m\}_{m=1}^M$ sont "résolus" par l'EMV $\{\widehat{\boldsymbol{\varepsilon}_m}\}_{m=1}^M$ si :

$$\exists \boldsymbol{\xi} \text{ vérifiant (111) } / \; \mathcal{P}_{\min} \leq \mathcal{O}_{\boldsymbol{\theta}^0}\left(\widehat{\Xi^0}, \boldsymbol{\xi}, \boldsymbol{\xi} \right) \leq \mathcal{P}_{\max} \tag{112}$$

où $\mathcal{P}_{\min}$ et $\mathcal{P}_{\max}$ caractérisent la confiance que nous désirons avoir **[T6]**.

Une caractéristique essentielle de cette approche **[T6]** est d'introduire dans le critère de "résolvabilité" (112) une exigence de précision d'estimation (111), alors que la plupart des critères disponibles dans la littérature courante [Lee92] [YB92W] [Smi05] [SM05] [LN07] [AW08] [EKBRM11] ne prennent en compte que la résolvabilité, sans se soucier d'une éventuelle permutation dans les estimées. Afin de comparer les différentes approches, nous nous sommes intéressés pour $M = 2$ à l'évaluation du seuil de résolution limite (en terme de RSB) de la fréquence de 2 cisoïdes :

$$\mathbf{x} = \mathbf{b}\left(\varepsilon_1\right)\sigma_1 + \mathbf{b}\left(\varepsilon_2\right)\sigma_2 + \mathbf{n}, \quad \mathbf{b}\left(\varepsilon\right) = \left(1, e^{j2\pi\varepsilon}, ..., e^{j(N-1)2\pi\varepsilon}\right)^T, \quad \varepsilon \in \left]-0.5, 0.5\right[.$$

Les résultats présentés figure 6-p51 correspondent à la configuration suivante : $N = 32$, $(\sigma_1, \sigma_2) = \sqrt{\frac{RSB}{N}}\left(1, e^{j\frac{\pi}{8}}\right)$, $\delta\varepsilon = \varepsilon_2 - \varepsilon_1 = \frac{1}{N}\frac{1}{k}$ où $k \in \{2, 4, 8, 16, 32\}$, $\boldsymbol{\xi} = \frac{\delta}{2}\mathbf{1}_M$ (erreur isotrope), $\mathcal{P}_{\min} = 0.95$ et $\mathcal{P}_{\max} = 0.99$. Les courbes présentées correspondent au seuil de résolution limite (SRL) obtenu selon l'approche de [Lee92] [YB92W] (RSB LYB), de [Smi05] (RSB S), de [SM05] [LN07] (RSB SMLN), de [AW08] (RSB AW) et celle proposée (112) (RSB (Pmax) et RSB (Pmin)).

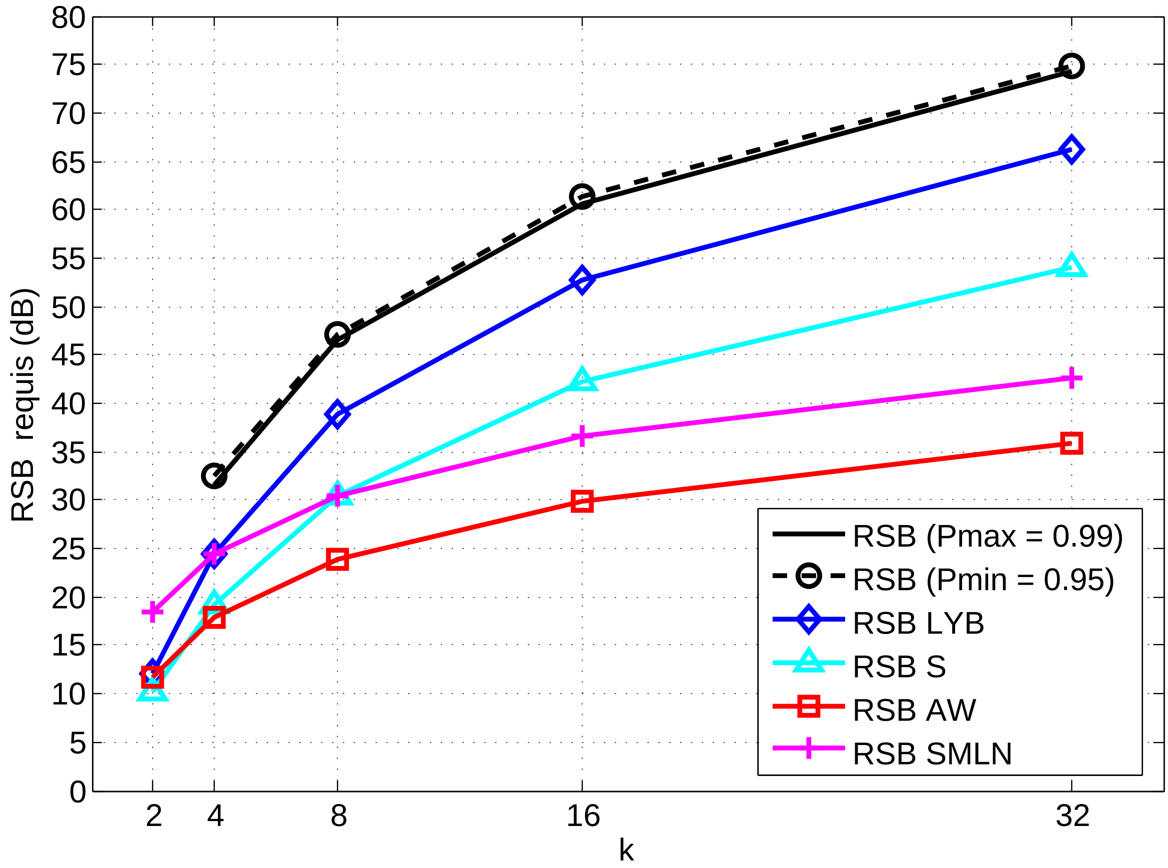


Fig. 6. RSB requis pour atteindre le Seuil de Résolution Limite lors de l'estimation de la fréquence ϵ de 2 cisoides, $\delta\epsilon = \frac{1}{kN}$, $N = 32$, $k \in \{2, 4, 8, 16, 32\}$

Il apparait clairement que l'exigence d'une précision d'estimation accompagnant la résolvabilité aboutit à la définition d'un SRL nettement supérieur au critère de Smith [Smi05] qui constitue depuis quelques années une sorte de "benchmark" pour les publications postérieures.
Plus précisément le critère de Smith correspond à la recherche du RSB (RSB S) pour lequel :

$$\text{Smith : } \mathcal{P}\left(\left|(\widehat{\varepsilon_2} - \widehat{\varepsilon_1}) - \delta\varepsilon\right| \leq \delta\varepsilon\right) = 0.68 \;\Leftrightarrow\; \left\{ \begin{array}{r} \mathcal{P}\left(\widehat{\varepsilon_2} \leq \widehat{\varepsilon_1}\right) = 0.16 \\ \mathcal{P}\left(\widehat{\varepsilon_2} \geq (\widehat{\varepsilon_1} + \delta\varepsilon) + \delta\varepsilon\right) = 0.16 \end{array} \right. \tag{113}$$

et illustre ce que je considère comme une "imprécision" dans la définition communément admise du SRL : la capacité à distinguer (séparer) les paramètres de M sources indépendamment de la possibilité d'une permutation dans l'application (au sens de relation binaire) $\{\widehat{\varepsilon_m}\}_{m=1}^{M} \to \{\varepsilon_m\}_{m=1}^{M}$.
En effet dans le cas du critère Smith (113), je risque de considérer $\widehat{\varepsilon_2}$ en lieu et place de $\widehat{\varepsilon_1}$ dans 16% des cas et $\widehat{\varepsilon_2}$ est peu précise dans également 16% des cas!
La conséquence de la réalisation de l'événement $(\widehat{\varepsilon_2} \leq \widehat{\varepsilon_1} \mid \varepsilon_1 < \varepsilon_2)$ est l'apparition d'un biais et d'une variance mesurée (après tri croissant des M composantes de l'EMV) pouvant violer la BCR sans biais (suivant la valeur de $\delta\varepsilon$), ce qui pose clairement le problème de la mise en oeuvre pratique de l'EMV et des conditions d'identifiabilité de cette mise en oeuvre, un sujet peu abordé dans la littérature courante, à ma connaissance. Et c'est précisément la non comptabilisation de ce type d'occurrence qu'impose la contrainte (111).

Bien que reposant sur des approches différentes, (RSB AW) et (RSB SMLN) comptabilisent également comme succès les occurrences $(\widehat{\varepsilon_2} \leq \widehat{\varepsilon_1})$, ce qui explique en partie (cf. [REKGCLR13, §V] **[T6]**) leur RSL optimiste. Enfin, (RSB LYB) est basé sur l'exploitation des lois marginales (relativement à ε_1 et ε_2) définissant un critère moins exigeant que (112) prenant en compte la loi conjointe; il conduit donc par définition à un SRL plus petit.

La mise en évidence de cette sous-évaluation (ou évaluation optimiste) des conditions nécessaires pour atteindre un SRL donné en présence de $M = 2$ sources estimées avec précision, laissent à penser que lorsque le nombre de sources (et de paramètres) augmentent, les performances en précision (*a priori*) communément admises dans la littérature courante sont trompeuses, au sens de trop optimistes par rapport à l'information de précision exhaustive qu'elles contiennent [RGCLR13] **[T6]** (cf. Annexe IV-L-p213).

Problèmes ouverts et perspectives :
La modélisation précise de la mise en oeuvre de l'EMV reste, à ma connaissance, un problème ouvert souvent ignoré dans les publications de la littérature courante car :
• les exemples d'application présentés n'explorent généralement pas les scénarios où le problème des permutations dans l'application $\{\widehat{\varepsilon_m}\}_{m=1}^{M} \to \{\varepsilon_m\}_{m=1}^{M}$ modifie de façon significative le biais et la variance ($\min\{|\varepsilon_m - \varepsilon_{m'}|\}$ suffisamment grand),
• les exemples d'application présentés évitent d'introduire ce phénomène difficile à relier aux calculs asymptotiques (ce qui est notre cas dans [VBC13]).
Il constitue à ce titre un de mes axes de recherche collaborative futurs.

Il serait pertinent de généraliser la procédure de test de la zone asymptotique proposée dans [REKGCLR13] au modèle déterministe à L observations considéré dans [MGCL12] **[T5]**, ce qui permettrait à terme d'unifier les résultats de [REKGCLR13] **[T6]** (exposés dans la présente section) et [MGCL12] **[T5]** (§II-A6-p44) afin de présenter un outil général d'aide à la conception des systèmes de mesure basé sur une précision exhaustive (avec mise en oeuvre de l'EMV selon [MSK08]).

Dans le cas du modèle stochastique, l'EMVS n'est pas gaussien à nombre fini (petit) d'observations indépendantes L. Les bornes (106) pourraient néanmoins être applicables si la loi de l'EMVS permettait une évaluation numérique de la fonction de répartition d'une forme quadratique du type $\left\|\mathbf{D}^{-1}_{\frac{\boldsymbol{\xi}^{+}+\boldsymbol{\xi}^{-}}{2}}\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^{0}\right)}_{opt}(\mathbf{x}) - \mathbf{g}\left(\boldsymbol{\theta}^{0}\right) - \frac{\boldsymbol{\xi}^{+}-\boldsymbol{\xi}^{-}}{2}\right)\right\|_2^2$. Il serait donc intéressant de dresser un inventaire des lois connues permettant un calcul numérique des bornes (106) et de vérifier si l'une d'elles correspond au cas de l'EMVS.

Enfin, de récents travaux [MVL13] tendent à suggérer que le compromis considéré (précision exhaustive de l'estimateur *a priori* le meilleur) pourrait étendu à une partie de la zone de transition, sous certaines conditions.

8) Contributions:

[J15] [J13] [J10] [J6] [J3] [JN1] - [C28] [C26] [C24] [C22] [C20] [C16] [C14] [C9] [C7] [CN4]

Certains résultats préliminaires établis dans ma thèse [Chau04] m'ont permis d'orienter une partie de la thèse d'Angela Quinlan **[T4]** vers la formulation d'une nouvelle approximation de la borne de Barankin (BB_1^N)[9] suffisamment générale pour être utilisée avec une large classe de d.d.p. $p(\mathbf{x};\theta)$ (celles pour lesquelles la BCR est calculable) et supérieure (tighter) aux approximations existantes [MS69] [MH71] [Gla72] [Abe93] [Kno97] [ZS93] [TK99] à l'époque. Afin que la borne inférieure de l'EQM proposée puisse être utilisée en lieu et place de l'approximation usuelle de complexité calculatoire équivalente (la borne de MacAulay-Seidman), nous avons calculé son expression analytique [J6] [C14] :
• pour le cas général de l'estimation de plusieurs fonctions dépendant d'un vecteur de paramètres,
• lorsque est une $\mathbf{x}$ est un vecteur d'observation gaussien réel ou complexe (circulaire ou non circulaire), en détaillant sa formulation pour les modèles d'observation déterministe et stochastique, réels ou complexes.
Ces travaux ont également permis :
• de présenter une synthèse nouvelle sur les approximations existantes de la borne de Barankin, depuis ses

[9] En fait la borne de Glave proposée dans la section (applicative) V de [Gla72] que nous n'avions pas lue lors de la soumission de [CGQL08].

formulations discrètes jusqu'à sa formulation exacte sous forme d'équation intégrale continue,
• d'introduire une méthode générale d'approximation de la borne de Barankin permettant de revisiter les approximations existantes et de justifier la forme de l'approximation proposée (BB_1^N).

Ces travaux initiaux ont été étendus par la prise en compte de nouveaux types de transformation de la vraisemblance [C20] ou l'introduction d'un biais spécifique dérivé des bornes bayésiennes (paramètres aléatoires) [C22] améliorant la prédiction de la zone de décrochement, sans toutefois diminuer la complexité d'implémentation.

Par ailleurs, l'examen de la littérature courante sur la borne de Cramér-Rao a fait apparaître des lacunes relatives à sa formulation dans le contexte radar actuel, à savoir large bande, multi-émetteurs et multi-récepteurs (MIMO), multi-sources, multi-paramètres à observations multiples. En effet dans la littérature courante, les observations multiples sont définies comme des réalisations indépendantes d'un même modèle d'observation, alors qu'en radar il s'agit en général de la combinaison de modèles d'observation différents (variation de la forme d'onde émise). Ce constat a motivé la partie du travail que j'ai supervisée dans la thèse de François Vincent **[T2]**, à savoir le calcul de la BCR pour paramètres orbitaux de débris spatiaux et récepteurs multiples large bande [C9]. La généralisation de ce travail préliminaire était l'objectif principal de la thèse de Tarek Menni **[T5]** que j'ai encadrée, à savoir l'établissement d'une expression analytique générale de la BCR déterministe pour l'analyse des performances asymptotiques en estimation d'un radar actif [C24]. Sa résolution a soulevé un certain nombre de points théoriques parfois partiellement résolus dans la littérature courante, notamment le calcul de la BCR pour paramètres mixtes (complexes et réels) que nous avons éclairé sous l'angle de la minimisation d'une norme sous contraintes linéaires permettant [J10] :
• une démonstration simplifiée en évitant les transformations matricielles généralement utilisées quand il s'agit de paramètres complexes [JR04] [YB92],
• de rectifier certains résultats antérieurs [VDB94],
• de clarifier quelques conditions de régularité qui sont inutilement restrictives [Abe93] [BHE09] [SM01] [SNg98].
De plus, l'exploitation de L modèles d'observation déterministes indépendants associés à l'émission-réception de L formes d'onde conduit dans la littérature classique à autant de calculs différents de la BCR qu'il existe de combinaison différentes entre le nombre de modèles MIMO distincts (par exemple 1 ou L) et les possibilités de paramétrisation des cibles. Là encore, nous avons proposé **[T5]** une méthode originale de calcul unique basé sur la prise en compte de contraintes égalités sur les paramètres inconnus correspondant au calcul d'une BCR contrainte, dont la dérivation est simplifiée sous l'angle de la minimisation d'une norme sous contraintes linéaires. Nous avons établi également l'équivalence entre une reparamétrisation (injective) et la prise en compte de contraintes égalités, ainsi que l'inégalité de reparamétrisation associée qui se révèle être un outil précieux pour comparer certaines BCR entre elles ou pour effectuer l'analyse des performances asymptotiques d'un système complexe [J13] [C26].
Cette capacité d'analyse est d'autant plus pertinente en pratique lorsqu'on peut déterminer si le modèle d'observation est en zone asymptotique pour les paramètres d'intérêt, ou s'il a atteint la zone de transition, zone à partir de laquelle la BCR (pour estimateur sans biais) devient une borne un peu trop optimiste par rapport au comportement de l'EMV. La recherche d'un tel test a motivé une partie des travaux entrepris par C. Ren dans sa thèse **[T6]** que je co-encadre. De plus, dans le cas du modèle d'observation déterministe, le fonctionnement en zone asymptotique permet de caractériser les performances de l'EMV en probabilité et non plus en moment [C28] [J15].

B. Problèmes conjoints détection-estimation

Dans les monographies disponibles en traitement du signal [VT68] [Sch91] [Kay93] [Kay98] [LC98] [VT02] traitant de l'estimation paramétrique (cf. §II-A), l'espace d'observation $\Omega \subset \mathbb{C}^M$ est généralement caractérisé par une unique densité de probabilité $p(\mathbf{x};\boldsymbol{\theta})$. Or un grand nombre de problèmes opérationnels ne correspondent pas à cette hypothèse implicite car ils sont la combinaison indissociable d'un problème de détection (test d'hypothèses) et d'un problème d'estimation. Le cas le plus simple est le test d'hypothèses binaire modélisant un signal d'intérêt intermittent qui n'est pas toujours présent dans les observations $\mathbf{x}$:

$$\begin{aligned} H_0 \text{ (environnement seul)} &: \quad \mathbf{x} = \mathbf{n} \\ H_1 \text{ (environnement et signal)} &: \quad \mathbf{x} = \mathbf{n} + \mathbf{s} \end{aligned}$$

Il s'agit alors de savoir en premier lieu si le signal observé contient (H_1) ou non (H_0) le signal d'intérêt $\mathbf{s}$ superposé au signal $\mathbf{n}$ provenant de l'environnement permanent; puis de lancer la procédure d'estimation des paramètres inconnus du signal d'intérêt $\mathbf{s}$ seulement s'il est présent.
Ainsi, dans la plupart des applications pratiques, les problèmes de détection (choix du modèle d'observation) et d'estimation sont conjoints : pour estimer de façon appropriée, il faut savoir sélectionner le modèle d'observation contenant le signal d'intérêt, sélection définie par une règle de décision (ou test de détection)[10] dépendant elle-même de paramètres inconnus, lesquels doivent au préalable être estimés également pour aboutir à une règle de décision réalisable.
Un test de détection réalisable (par exemple le TRVG ("GLRT") décrit en Annexe IV-C-p72) définit un sous ensemble $\mathcal{D}$ de Ω déterminant le choix de l'hypothèse H_1 et ne dépendant pas des paramètres. D'un point de vue formel, il s'agit d'un conditionnement des observations par un événement réalisable $\mathcal{D}$. Par conséquent, tous les développements préalablement introduits en estimation (cf. §II-A) s'appliquent : il suffit de changer d'univers d'observation en remplaçant dans toutes les expressions Ω et $p(\mathbf{x};\boldsymbol{\theta})$ respectivement par $\mathcal{D}$ et $p(\mathbf{x} \mid \mathcal{D} \cap H_1;\boldsymbol{\theta}) = p(\mathbf{x} \mid H_1;\boldsymbol{\theta}) / \mathcal{P}(\mathcal{D} \cap H_1;\boldsymbol{\theta})$, $\mathcal{P}(\mathcal{D} \cap H_1;\boldsymbol{\theta}) = \int_{\mathcal{D}} p(\mathbf{x} \mid H_1;\boldsymbol{\theta})\, d\mathbf{x}$. Ainsi la norme (9) définissant la précision *a priori* devient :

$$\|\mathbf{f}(\mathbf{x})\|_{s;\boldsymbol{\theta}|\mathcal{D}\cap H_1} = E_{\boldsymbol{\theta}|\mathcal{D}\cap H_1}\left[\|\mathbf{f}(\mathbf{x})\|_s^s\right]^{\frac{1}{s}} = E_{\boldsymbol{\theta}|\mathcal{D}\cap H_1}\left[\sum_{q=1}^{Q} |f_q(\mathbf{x})|^s\right]^{\frac{1}{s}}, \ E_{\boldsymbol{\theta}|\mathcal{D}\cap H_1}[\mathbf{f}(\mathbf{x})] = \int_{\mathcal{D}} \mathbf{f}(\mathbf{x})\, p(\mathbf{x} \mid \mathcal{D} \cap H_1;\boldsymbol{\theta})\, d\mathbf{x}.$$

Dans le cas particulier où $s = 2$, la forme adaptée du produit scalaire (22) au nouvel univers $\mathcal{D}$ sous H_1 est :

$$\langle \mathbf{g}(\mathbf{x}) \mid \mathbf{h}(\mathbf{x}) \rangle_{\boldsymbol{\theta}|\mathcal{D}\cap H_1} = E_{\boldsymbol{\theta}|\mathcal{D}\cap H_1}\left[\mathbf{h}^T(\mathbf{x})\,\mathbf{g}(\mathbf{x})\right],$$

et la définition d'un estimateur sans biais (20) devient :

$$E_{\boldsymbol{\theta}|\mathcal{D}\cap H_1}\left[\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}(\mathbf{x})\right] = \mathbf{g}(\boldsymbol{\theta}) \ \Leftrightarrow \ \left| \begin{array}{c} \left\langle \widehat{\mathbf{g}(\boldsymbol{\theta}^0)}(\mathbf{x}) - \mathbf{g}(\boldsymbol{\theta}^0) \mid \upsilon_{\boldsymbol{\theta}^0|\mathcal{D}\cap H_1}(\mathbf{x};\boldsymbol{\theta}) \right\rangle_{\boldsymbol{\theta}^0|\mathcal{D}\cap H_1} = \mathbf{g}(\boldsymbol{\theta}) - \mathbf{g}(\boldsymbol{\theta}^0) \\ \upsilon_{\boldsymbol{\theta}^0|\mathcal{D}\cap H_1}(\mathbf{x};\boldsymbol{\theta}) = \frac{p(\mathbf{x}|\mathcal{D}\cap H_1;\boldsymbol{\theta})}{p(\mathbf{x}|\mathcal{D}\cap H_1;\boldsymbol{\theta}^0)} \end{array} \right. .$$

Il est clair que le conditionnement apparaît comme une difficulté majeure car il faut non seulement pouvoir exprimer la d.d.p.. conditionnelle $p(\mathbf{x} \mid \mathcal{D} \cap H_1;\boldsymbol{\theta})$ mais également pouvoir calculer les différentes espérances conditionnelles $E_{\boldsymbol{\theta}|\mathcal{D}\cap H_1}[\,]$ requises pour le calcul des bornes inférieures ou de l'EQM. Par exemple, la MIF (60) devient :

$$\begin{aligned} \mathbf{F}_{\boldsymbol{\theta}|\mathcal{D}\cap H_1} &= E_{\boldsymbol{\theta}|\mathcal{D}\cap H_1}\left[\frac{\partial \ln p(\mathbf{x} \mid H_1;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln p(\mathbf{x} \mid H_1;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right] - \frac{\partial \ln \mathcal{P}(\mathcal{D} \cap H_1;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln \mathcal{P}(\mathcal{D} \cap H_1;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \\ \mathbf{F}_{\boldsymbol{\theta}|\mathcal{D}\cap H_1} &= -E_{\boldsymbol{\theta}|\mathcal{D}\cap H_1}\left[\frac{\partial^2 \ln p(\mathbf{x} \mid H_1;\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right] + \frac{\partial^2 \ln \mathcal{P}(\mathcal{D} \cap H_1;\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \end{aligned} \quad (114)$$

Dans la suite nous notons $\mathcal{P}_D \triangleq \mathcal{P}(\mathcal{D} \cap H_1;\boldsymbol{\theta})$ et $\mathcal{P}_{FA} \triangleq \mathcal{P}(\mathcal{D} \cap H_0;\boldsymbol{\theta})$.

[10]Quelques éléments de la théorie de la détection appliquée au test d'hypothèse binaires sont rappelés en Annexe IV-C-p72

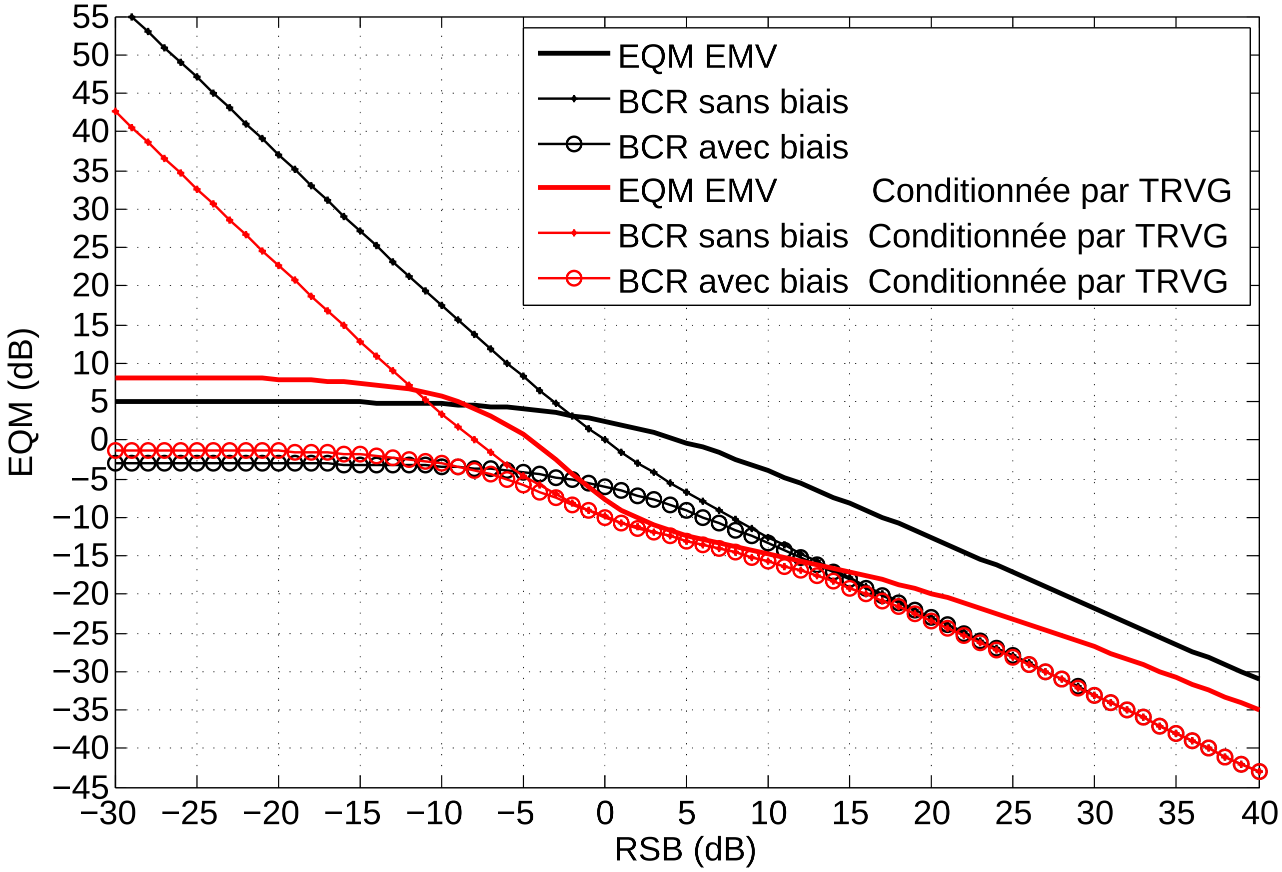


Fig. 7. Influence du TRVG sur la BCR et l'EQM de l'EMV du rapport d'écartométrie (115), modèle stochastique, $L = 1$, $\mathcal{P}_{FA} = 10^{-4}$, $r(\varepsilon) = 0$

Outre la nécessité théorique pour caractériser de façon correcte le problème conjoint détection-estimation, un autre intérêt de la prise en compte du conditionnement est la conjecture selon laquelle le test de détection puisse permettre d'étendre le domaine "prédictif" des bornes pour "Erreur Faible" comme la BCR. En effet, un tel test devrait majoritairement sélectionner les observations à forte énergie constituées de signal et de bruit, et rejeter les observations à faible énergie constituées de bruit seul qui dégradent l'EQM et l'éloigne des bornes pour "Erreur Faible".

1) Exemple de l'écartométrie monopulse :

Une application pratique mettant en oeuvre un test d'hypothèses binaires est par exemple la mesure de la direction d'arrivée par une antenne monopulse déjà évoquée au §II-A4d-p35. Si cette application n'est pas considérée comme une application de référence dans la communauté du traitement du signal (où l'application de référence est l'estimation de la fréquence d'une cisoïde (69)), elle est néanmoins une application fondamentale en localisation angulaire radar (et goniométrie télécoms) puisque tous les radars actuels en service l'utilisent comme technique haute-précision angulaire alimentant les algorithmes de pistage (tracking). Rappelons qu'il s'agit d'un modèle d'observation à 2 voies de réception (voies "somme" Σ et "différence" Δ) s'écrivant :

$$\mathbf{x}^l = \left(\Sigma^l, \Delta^l\right)^T = \mathbf{b}(\varepsilon)\,\sigma^l + \mathbf{n}^l, \quad \sigma^l = \alpha^l g_\Sigma(\varepsilon), \quad \mathbf{b}(\varepsilon) = (1, r(\varepsilon))^T, \; r(\varepsilon) = \frac{g_\Delta(\varepsilon)}{g_\Sigma(\varepsilon)} \quad l = 1, ..., L \qquad (115)$$

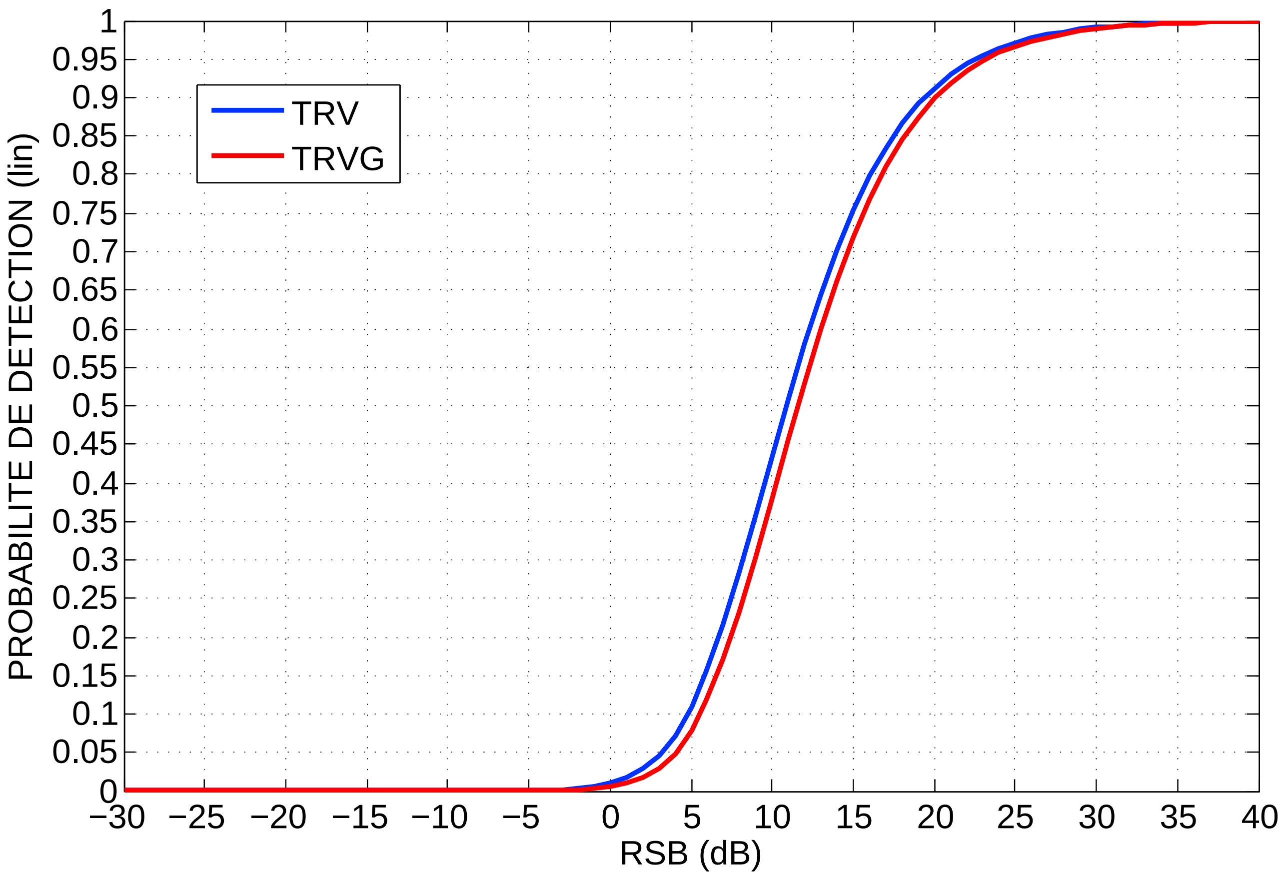


Fig. 8. Probabilité de détection du TRVG et du TRV d'une antenne monopulse (115), modèle stochastique, $L = 1$, $\mathcal{P}_{FA} = 10^{-4}$, $r(\varepsilon) = 0$

où $\mathbf{n}$ est un bruit blanc centré de matrice de covariance connue $\mathbf{C_n} = \mathbf{I}_N$, $\{\alpha^l\}_{l=1}^{L}$ sont les amplitudes complexes de la source de signal, $g_\Sigma(\varepsilon)$ et $g_\Delta(\varepsilon)$ sont les gains complexes en amplitude (diagramme de rayonnement) de chaque voie pour la direction d'arrivée ε. L'étude du rapport d'écartométrie au voisinage de la direction principale de rayonnement de l'antenne $(\varepsilon = 0)$ montre que dans le lobe principal à $3dB$: $r(\varepsilon) \approx k\varepsilon$ [She84] [Lev88], ce qui conduit à l'estimateur $\widehat{\varepsilon}_{MV} = \widehat{r(\varepsilon)}_{MV}/k$ dont la caractérisation statistique se déduit naturellement de celle de $\widehat{r(\varepsilon)}_{MV} \approx \mathrm{Re}\left\{\left(\sum_{l=1}^{L}(\Sigma^l)^*\Delta^l\right)/\left(\sum_{l=1}^{L}|\Sigma^l|^2\right)\right\}$ [Mos69][GCL10], que le modèle d'observation soit déterministe ou stochastique.

Outre son intérêt pratique, l'estimation du rapport d'écartométrie présente la particularité d'avoir tous les calculs analytiques accessibles : EQM, BCR, biais conditionnés par les tests $\left\{\sum_{l=1}^{L}|\Sigma^l|^2 > T\right\}$ ou $\left\{\sum_{l=1}^{L}\left(|\Sigma^l|^2 + |\Delta^l|^2\right) > T\right\}$, pour les modèles d'observations déterministe et stochastique [Chau04] [CLF05] [GCL10].

Dans le cas particulier mono-observation $(L = 1)$ les expressions de l'EMV et du TRVG (143) sont exactes :

$$TRVG : |\Sigma^1|^2 + |\Delta^1|^2 \underset{H_0}{\overset{H_1}{\gtrless}} T \quad \text{et} \quad \widehat{r}_{MV} = \mathrm{Re}\left\{\frac{\Delta^1}{\Sigma^1}\right\}, \tag{116}$$

et lorsque $r(\varepsilon) = 0$ le TRV (140) s'écrit :

$$TRV : |\Sigma^1|^2 \underset{H_0}{\overset{H_1}{\gtrless}} T \tag{117}$$

De plus, le cas $r(\varepsilon) = 0$ correspond à une source dans la direction principale de rayonnement de la voie somme $(\varepsilon = 0)$ et constitue la position angulaire de référence (pour la mesure de performance en détection et en estimation) associée au maximum d'énergie reçue.

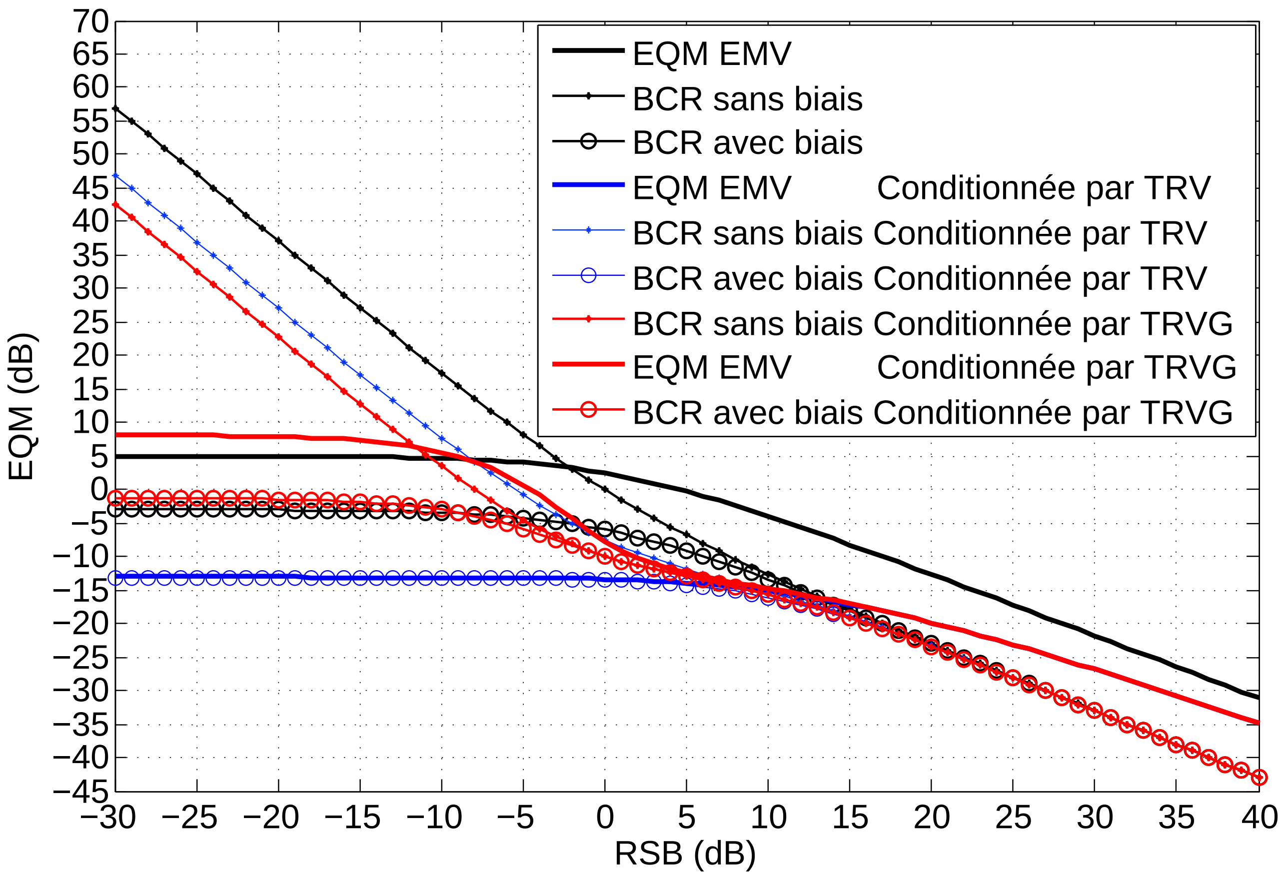


Fig. 9. Influence du TRVG et du TRV sur la BCR et l'EQM de l'EMV du rapport d'écartométrie (115), modèle stochastique, $L = 1$, $\mathcal{P}_{FA} = 10^{-4}$, $r(\varepsilon) = 0$

Quelques résultats significatifs sont fournis par la figure 7-p55 représentant l'EQM et la BCR de $\widehat{r}_{MV}$ (116) pour un estimateur localement avec et sans biais, sans conditionnement ou avec conditionnement par le TRVG (116), en fonction du RSB sur la voie somme Σ, pour une $\mathcal{P}_{FA}$ donnée. En effet, dans le problème du test d'hypothèses binaires, la $\mathcal{P}_{FA}$ est la grandeur d'importance pratique car elle représente (sous H_0) la proportion de bruit seul capable de franchir l'étape de détection et qui sera transmise au processus d'estimation. A chaque P_{FA} correspond un seuil de détection T calculé à l'aide de la relation (pour $L = 1$) :

$$\mathcal{P}_{FA} = e^{-T} (1 + T) \tag{118}$$

Cet exemple, bien qu'un peu particulier puisque la région dite "de transition" ne présente pas d'effet de décrochement en l'absence de conditionnement $(P_{FA} = 1)$, permet d'illustrer certaines considérations d'ordre général.
La première est la nécessité de prendre en compte le conditionnement par le test de détection pour qu'une borne inférieure donnée conserve sa propriété de bonne inférieure pour les problèmes conjoints détection-estimation. Ceci est illustré pour le domaine de RSB $[-10, 15]\,dB$ où seule la forme conditionnée de la BCR est pertinente.
Une mesure du conditionnement, au sens de la restriction des observations, est précisément la $\mathcal{P}_D$ (mesure au sens des probabilités) présentée figure 8-p56 pour les 2 tests : TRV (117) et TRVG (116).
La seconde est la confirmation de notre conjecture : la prise en compte d'un test de détection peut notablement modifier le comportement des bornes pour "Erreur Faible" dans la région de transition. Le test de détection joue le rôle de sélecteur des observations de plus fortes énergies qui contiennent nécessairement le signal utile (tant que $\mathcal{P}_D \gg \mathcal{P}_{FA}$), puisque les observations ne contenant que du bruit ne peuvent franchir ce seuil que dans la proportion fixée par la $\mathcal{P}_{FA}$. Ce mécanisme peut donc avoir pour effet d'améliorer la représentativité de la BCR

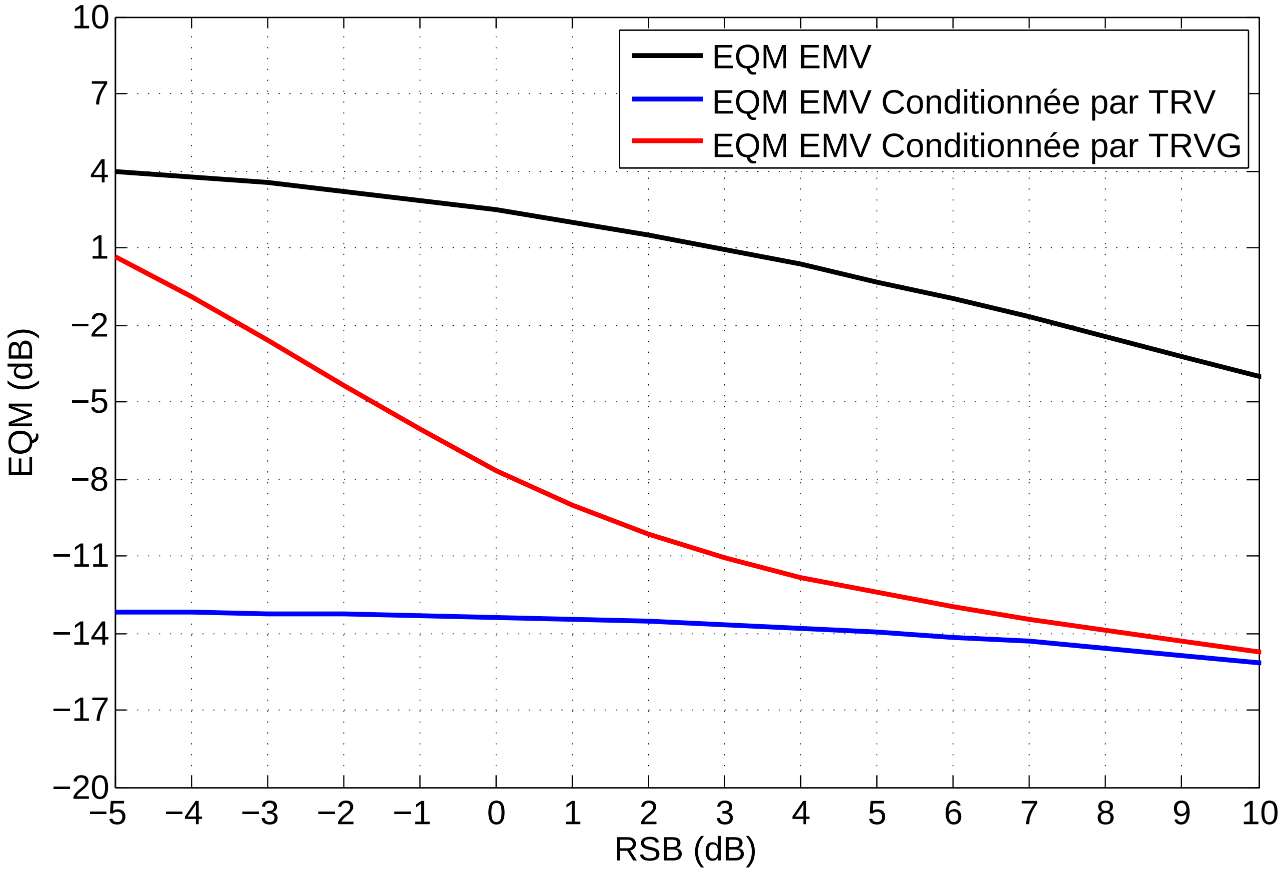


Fig. 10. Influence du TRVG et du TRV sur l'EQM de l'EMV du rapport d'écartométrie (115), modèle stochastique, $L = 1$, $\mathcal{P}_{FA} = 10^{-4}$, $r(\varepsilon) = 0$

conditionnée sur un sous-ensemble d'observations, voire même de modifier drastiquement les conditions d'obtention d'un estimateur efficace. En effet, comme $\widehat{r}_{MV}$ correspond à l'EMV stochastique, il ne peut être efficace à fort RSB; il est par conséquent assez original d'exhiber un estimateur au sens du MV efficace au voisinage de $5dB$! L'analyse théorique de ce phénomène est un sujet pour de futures recherches, car il relève d'un problème plutôt complexe consistant à rechercher les d.d.p. $p(\mathbf{x} \mid \mathcal{D} \cap H_1; \theta)$ vérifiant (cf. §IV-B-p66) :

$$\widehat{\theta}_{opt}(\mathbf{x}) - \theta = \frac{1}{F_{\theta|\mathcal{D}\cap H_1}} \left[\frac{\partial \ln p(\mathbf{x} \mid H_1; \theta)}{\partial \theta} - \frac{\partial \ln \mathcal{P}(\mathcal{D} \cap H_1; \theta)}{\partial \theta} \right]$$

La troisième est qu'il existe une limite à la pertinence de l'information fournie par la BCR même conditionnée pour les estimateurs sans biais à faible RSB. Comme mentionné précédemment, la raison est qu'un estimateur localement sans biais d'un paramètre du signal source n'existe généralement pas lorsque le RSB tend vers 0, que les observations soient conditionnées ou pas.
Ces résultats confirment, qu'indépendamment d'un éventuel conditionnement des observations, la connaissance précise du biais est un facteur déterminant pour la qualité de la prédiction de l'EQM de l'EMV par les bornes inférieures de l'EQM, notamment la BCR.

Ces résultats se retrouvent également lorsque le test de détection devient le TRV (117), comme le montre la figure 9-p57 révélant un autre exemple original où l'EMV est efficace uniquement dans les zones de transition et de non-information.

Il apparait également clairement que l'EQM de l'EMV est très sensible à la nature du conditionnement à mesure

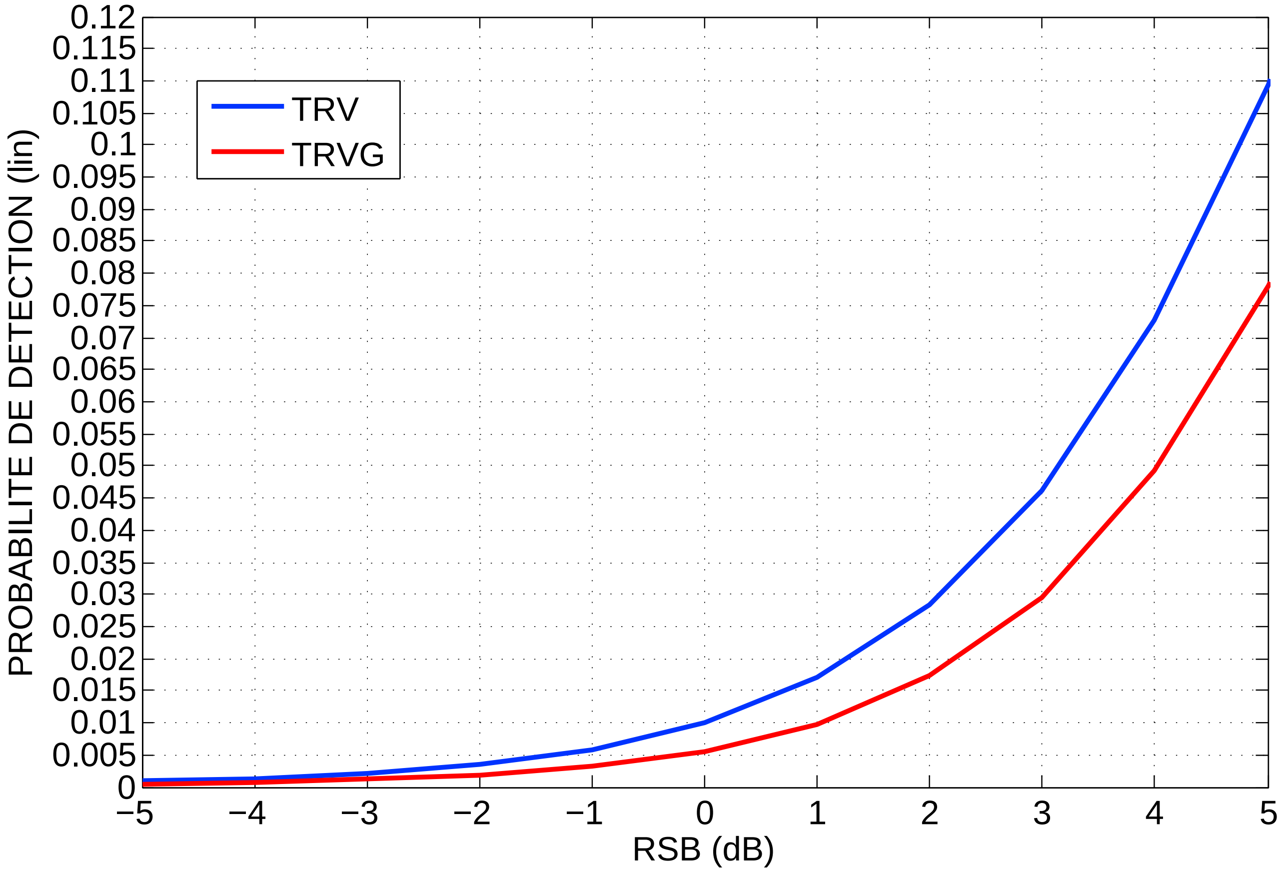


Fig. 11. Probabilité de détection (Zoom) du TRVG et du TRV d'une antenne monopulse (115), modèle stochastique, $L = 1$, $\mathcal{P}_{FA} = 10^{-4}$, $r(\varepsilon) = 0$

que $\mathcal{P}_D \to \mathcal{P}_{FA}$ comme le montre les figures 10-p58 et 11-p59.

Enfin, on peut se demander si l'absence de valeur prédictive de la BCR conditionnée (par le TRVG par exemple) dans la zone de non-information constitue un réel problème pratique pour la caractérisation des problèmes conjoints détection-estimation. En effet, d'un point de vue opérationnel, la caractérisation des performances en estimation n'a d'intérêt que si le signal utile a été détecté avec une probabilité suffisamment grande représentant un objectif opérationnel. Généralement, en deçà de cette probabilité le signal d'intérêt est tout simplement considéré comme non détecté (par exemple pour les $RSB \leq 0dB$ si nous nous restreignons aux $\mathcal{P}_D \geq 0.01$).

Cet exemple montre que les contraintes opérationnelles dans la pratique peuvent exclure la zone de non-information de la zone où la prédiction des performances en estimation présente un intérêt pratique. Par conséquent, il est donc opérationnellement plus intéressant de porter ses efforts de caractérisation sur la prise en compte du conditionnement des observations et de son effet sur la zone de transition que sur la description fine des performances en estimation dans la zone de non-information.

2) Exemple du modèle d'observation déterministe :

La caractérisation de la plupart des bornes inférieures de l'EQM (BCR, BHCR, BMS, BMH, BG, BTT) peut être étendue au conditionnement du modèle d'observation déterministe

$$\mathbf{x} \sim \mathcal{CN}_N\left(\mathbf{m_x}(\boldsymbol{\theta}), \mathbf{C_x}\right), \quad p(\mathbf{x}; \boldsymbol{\theta}) = \frac{e^{-(\mathbf{x}-\mathbf{m_x}(\boldsymbol{\theta}))^H \mathbf{C_x}^{-1}(\mathbf{x}-\mathbf{m_x}(\boldsymbol{\theta}))}}{\pi^M \left|\mathbf{C_x}\right|} \tag{119}$$

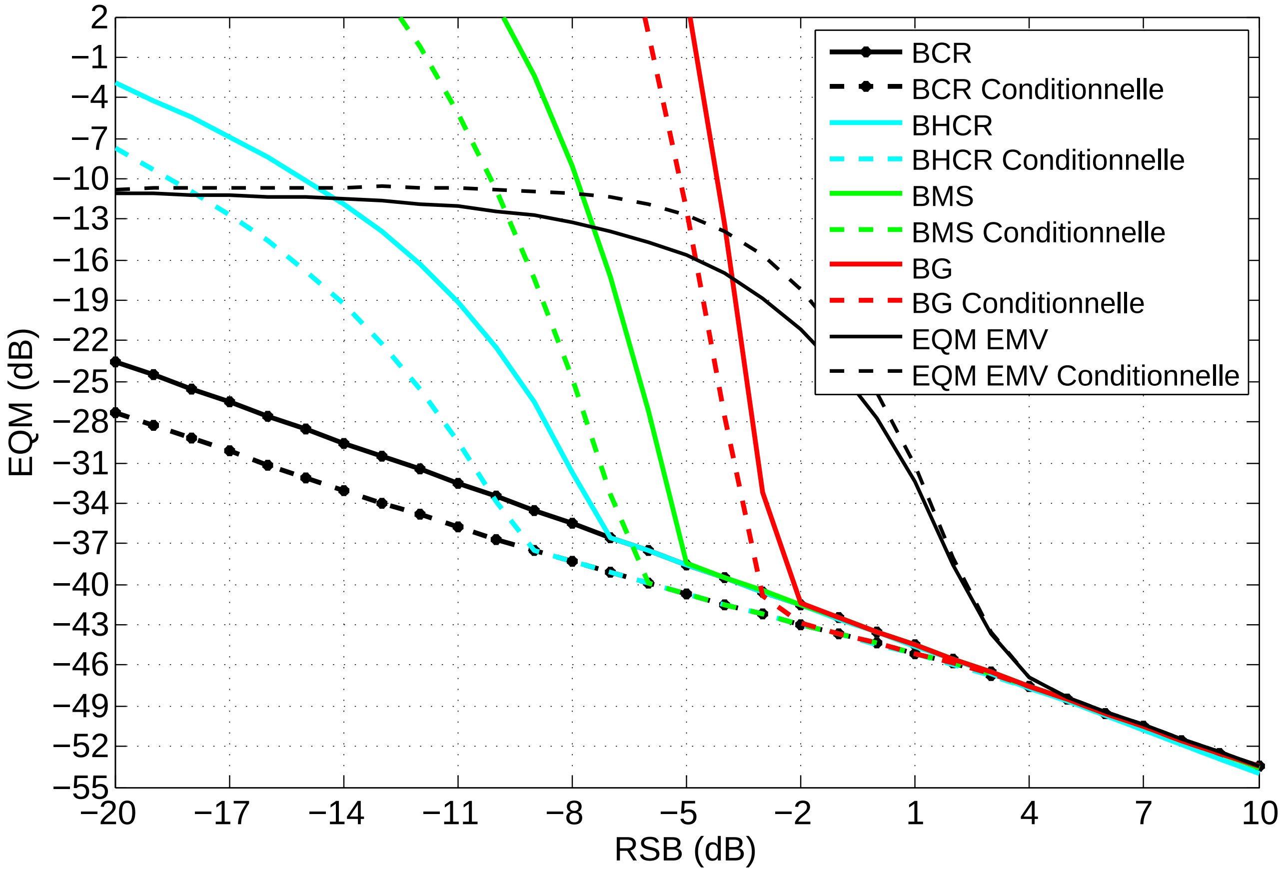


Fig. 12. Influence du détecteur d'énergie (120) sur l'EQM de l'EMV et les bornes BCR, BHCR, BMS, BG, dans le cas d'une cisoïde de fréquence inconnue ε (122), $\varepsilon^0 = 0$, $N = 10, \mathcal{P}_{FA} = 10^{-3}$

lorsque $\mathbf{C_x}$ est connue par le détecteur d'énergie ("energy detector") [Kay98, §7.3] :

$$\mathbf{x}^H \mathbf{C}_\mathbf{x}^{-1} \mathbf{x} \gtrless T. \tag{120}$$

Ce test est très utilisé en pratique (par exemple en écartométrie monopulse) car très facile à mettre en oeuvre. Bien entendu, dans la plupart des cas ses performances en détection seront loin d'approcher les performances optimales du TRV. Il constitue de ce fait un test de référence en ce sens que les performances du TRVG ("GLRT") associé à l'EMV de $\boldsymbol{\theta}$ sont attendues entre le TRV et le détecteur d'énergie [Kay98, §7.3].
La MIF conditionnée (114) devient alors [CL07C] :

$$\begin{aligned}
\mathbf{F}_{\boldsymbol{\theta}|\mathcal{D}\cap H_1} &= \mathbf{F}_{\boldsymbol{\theta}} \left(\frac{1 - P_{N+1}(\boldsymbol{\theta})}{1 - P_N(\boldsymbol{\theta})} \right) + w_N(\boldsymbol{\theta}) \left(\frac{\partial \left(\mathbf{m}_\mathbf{x}^H(\boldsymbol{\theta}) \mathbf{C}_\mathbf{x}^{-1} \mathbf{m}_\mathbf{x}(\boldsymbol{\theta}) \right)}{\partial \boldsymbol{\theta}} \frac{\partial \left(\mathbf{m}_\mathbf{x}^H(\boldsymbol{\theta}) \mathbf{C}_\mathbf{x}^{-1} \mathbf{m}_\mathbf{x}(\boldsymbol{\theta}) \right)}{\partial \boldsymbol{\theta}^T} \right) \qquad (121) \\
w_N(\boldsymbol{\theta}) &= \frac{2P_{N+1}(\boldsymbol{\theta}) - P_N(\boldsymbol{\theta}) - P_{N+2}(\boldsymbol{\theta})}{1 - P_N(\boldsymbol{\theta})} - \left(\frac{P_{N+1}(\boldsymbol{\theta}) - P_N(\boldsymbol{\theta})}{1 - P_N(\boldsymbol{\theta})} \right)^2 \\
P_L(\boldsymbol{\theta}) &= \int_0^T p_{\mathcal{X}_{2L}^2}\left(t; \mathbf{m}_\mathbf{x}^H(\boldsymbol{\theta}) \mathbf{C}_\mathbf{x}^{-1} \mathbf{m}_\mathbf{x}(\boldsymbol{\theta})\right) dt, \quad p_{\mathcal{X}_{2L}^2}(t;\lambda) = e^{-(t+\lambda)} I_{L-1}\left(2\sqrt{\lambda t}\right) \sqrt{\frac{t}{\lambda}}^{(L-1)}
\end{aligned}$$

où $I_L(z)$ est la fonction modifiée de Bessel de première espèce [Kay98].
L'expression (121), pourtant obtenue pour un test simple (120), n'est pas facile à analyser. Ceci illustre à la fois la difficulté non seulement calculatoire mais également d'analyse de l'influence d'un test de détection sur les

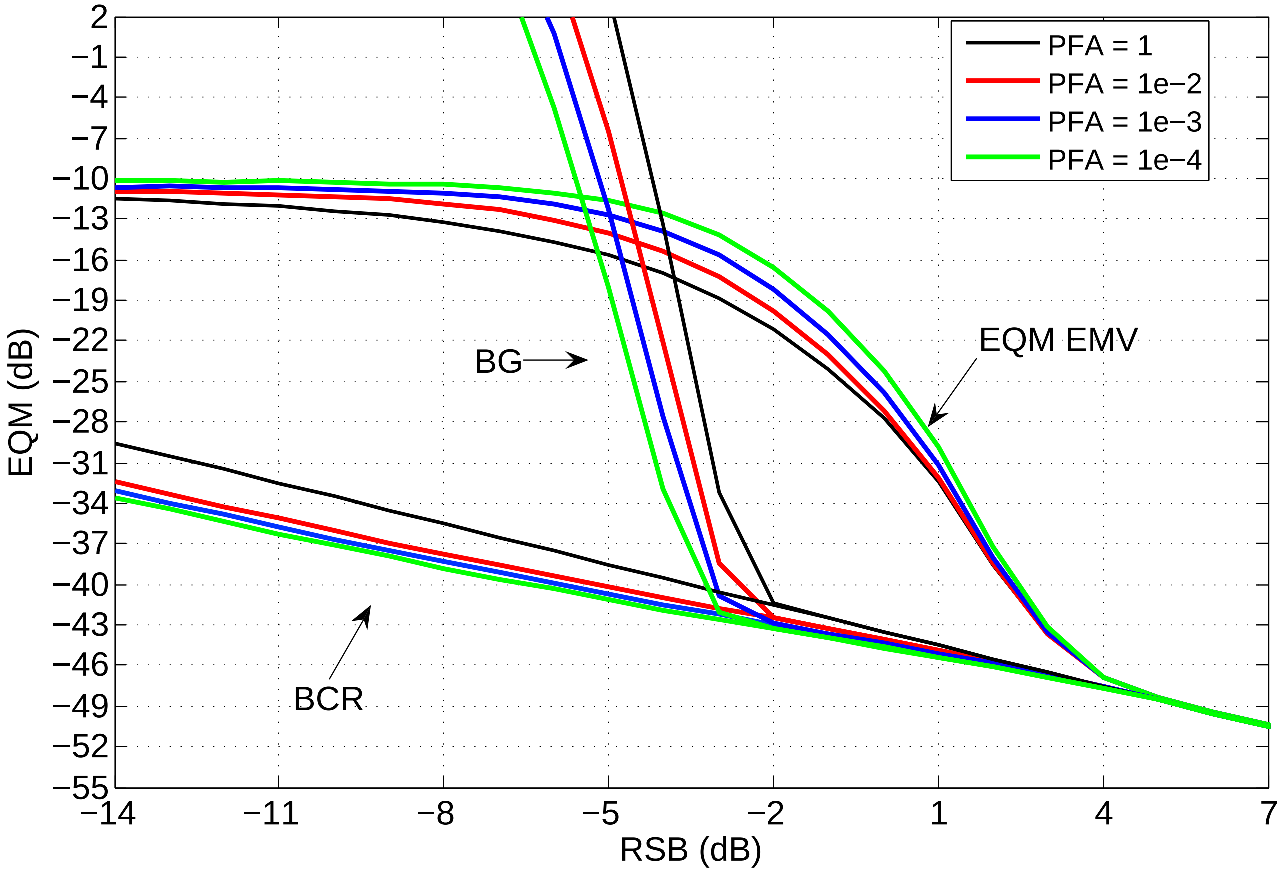


Fig. 13. Influence du détecteur d'énergie (120) sur l'EQM de l'EMV et les bornes BCR, BG, dans le cas d'une cisoïde de fréquence inconnue ε (122), $\varepsilon^0 = 0$, $N = 10, \mathcal{P}_{FA} \in \left\{10^{-2}, 10^{-3}, 10^{-4}\right\}$

performances en estimation. La prise en compte du test (120) pour les autres bornes inférieures de l'EQM (BHCR, BMS, BMH, BG, BTT) est décrite dans [CGVRL07] (cf. Annexe IV-J-p213) et conduit à des expressions encore plus complexes qui ne peuvent être analysées que par simulation numériques.

Par exemple, considérons de nouveau le problème de l'estimation de la fréquence ε d'une cisoïde dont le modèle d'observation s'écrit :

$$\mathbf{x} = \mathbf{b}(\varepsilon)\,\sigma + \mathbf{n}, \quad \mathbf{b}(\varepsilon) = \left(1, e^{j2\pi\varepsilon}, ..., e^{j(N-1)2\pi\varepsilon}\right)^T, \quad \varepsilon \in \left]-0.5, 0.5\right[\tag{122}$$

où $\mathbf{n}$ est un bruit blanc centré de matrice de covariance $\mathbf{C_n} = \sigma_\mathbf{n}^2 \mathbf{I}_N$, $\sigma \in \mathbb{C}$ est l'amplitude complexe de la source dans le cas d'une observation déterministe, $\frac{|\sigma|^2}{\sigma_\mathbf{n}^2}$ étant le RSB. Alors, dans le cas particulier où le seul paramètre inconnu est la fréquence ε, (120) devient $\left\{\|\mathbf{x}\|^2 \gtrless T\right\}$ et la comparaison des implémentations suivantes pour $\varepsilon^0 = 0$:

- BHCR : 2 points test $\{0, \delta\}$ + supremum sur δ où δ balaie $\left]-0.5, 0.5\right[$ avec un pas de $\frac{1}{1024}$,
- BMS : 3 points test $\{0, \delta, -\delta\}$ + supremum sur δ où δ balaie $\left]0, 0.5\right[$ avec un pas de $\frac{1}{1024}$,
- BG : 3 points test $\{0, \delta, -\delta\}$ + supremum sur δ où δ balaie $\left]0, 0.5\right[$ avec un pas de $\frac{1}{1024}$,

est présentée figure 12-p60 laquelle présente un résultat un peu paradoxal. En effet, comme évoqué précédemment dans le cas de l'écartométrie monopulse, l'effet attendu du test d'énergie (120) est de sélectionner les observations de plus fortes énergies qui contiennent nécessairement le signal utile, ce qui devrait avoir pour effet d'améliorer la précision d'estimation sur l'ensemble des observations franchissant le test. Cette intuition est confirmée par le comportement des différentes bornes inférieures de l'EQM pour lesquelles la prise en compte du conditionnement conduit à une décroissance de leur valeur. Mais elle est infirmée par le comportement de l'EQM de l'EMV dont la valeur croit dans la zone de transition, comportement contre-intuitif qui ne semble pas dépendre de la valeur du seuil T (ou de la $\mathcal{P}_{FA}$) comme le montre la figure 13-p61.

3) Problèmes ouverts, conjectures et perspectives:

• **Problème ouvert 1. La caractérisation des performances conjointes.**

La principale difficulté théorique dans l'établissement de formules analytiques décrivant les performances conjointes détection-estimation est d'ordre calculatoire : nombre d'intégrales n'ont pas d'expression analytique (notamment les bornes d'intégration) lorsque le domaine d'intégration $\mathcal{D}$ est quelconque (et défini par une équation implicite) dans un espace de dimension $N \geq 4$. Il parait donc illusoire de penser que ce problème pourra être caractérisé analytiquement pour tout test de détection réalisable. L'approche par simulation de type Monte-Carlo principalement utilisée dans les travaux sur ce thème souffre également d'une limitation bien connue : le temps de calcul pour réaliser le nombre de tirages appartenant à $\mathcal{D}$ nécessaires pour estimer de façon correcte l'EQM conditionnelle. En effet ce nombre de tirages devient rapidement prohibitif si on désire explorer les $\mathcal{P}_{FA}$ faibles $\left(\left[10^{-6}, 10^{-2}\right]\right)$ conduisant à des seuils de détection élevés et des $\mathcal{P}_D$ potentiellement $\ll 1$ en fonction du RSB (cf. figure 8-p56). Ceci conduit en général à une grande difficulté, voire l'impossibilité, d'explorer les 3 zones de fonctionnement de l'EQM pour un nombre de configurations (N, M, RSB, défauts de modèles, ...) satisfaisantes permettant une analyse approfondie du couplage détection-estimation.

Perspective :

A ce titre une étude plus exhaustive des performances conjointes du détecteur d'énergie (120) pourrait apporter une performance de référence. Comme la connaissance du biais semble être un élément déterminant de la pertinence de la BCR, il serait opportun d'étendre la méthode de calcul du biais utilisé dans [NFBL09] lors d'un conditionnement par le détecteur d'énergie pour vérifier si la BCR avec biais obtenue conserve sa capacité à prédire avec précision le comportement de l'EQM (et éventuellement fournir une explication au paradoxe mentionné). Une extension de ces résultats au modèle stochastique permettrait également d'étudier l'influence de la loi de fluctuation des amplitudes des signaux sources sur les performances conjointes.
L'idéal serait de pouvoir établir cette caractérisation pour le TRV, généralement clairvoyant, car cela nous permettrait de déterminer les meilleures performances en estimation sachant le meilleur test de détection. Or nous avons déjà pu démontrer que la BCR (la borne la plus simple) conditionnée par un test de détection clairvoyant est toujours nulle, donc non informative [Chau04, §V.C.2] [CGVL07] ce qui laisse augurer quelques problèmes théoriques en perspective ....

• **Problème ouvert 2. Critère d'optimalité des problèmes conjoints détection-estimation.**

Un problème qui reste pour moi ouvert est la question de la définition d'un critère de performances conjointes optimales.
Si nous restreignons notre univers (au sens de l'ensemble des observations accessibles) à une réalisation d'un problème conjoint détection-estimation, la définition d'un critère d'optimalité tel que la densité d'EQM :

$$\frac{\left\|\widehat{g(\theta_0)}(\mathbf{x}) - g(\theta_0)\right\|^2_{2;\theta_0|\mathcal{D}\cap H_1}}{\mathcal{P}(\mathcal{D}\cap H_1;\theta_0)},$$

proposée dans ma thèse [Chau04, §V-D] peut paraître un candidat possible, car il admet une borne inférieure qui n'est autre que la BCR non-conditionnée. Selon ce critère, il parait opportun de rechercher les meilleures performances en estimation sachant le meilleur test de détection (cf. ci-dessus).
Cependant dans nombre d'applications, le problème conjoint détection-estimation n'est que la première étape d'un modèle à observations multiples beaucoup plus complexe faisant intervenir une rétroaction : les estimées produites sont utilisées pour "recaler" un modèle d'état dynamique dont l'état suivant servira à définir certains paramètres du problème conjoint détection-estimation pour l'observation suivante.

Un exemple classique est la poursuite (radar, sonar, télécoms, ...) d'une ou plusieurs sources de signal : le problème conjoint détection-estimation caractérise l'appareil de mesure produisant les estimées des paramètres cinématiques des sources, puis ces estimées sont utilisées pour mettre à jour un modèle cinématique des sources (filtre de Kalman, ...) lequel réalise une prédiction de la position à venir des sources lors de la prochaine mesure et définit les paramètres de la fenêtre d'observation (fenêtre distance, doppler, angulaire, ...) à acquérir et sur laquelle s'appliquera la stratégie conjointe détection-estimation.
Dans ce contexte, nous pouvons légitimement nous demander quel critère d'optimalité choisir : faut-il détecter plus souvent quitte à être moins précis (seuil de détection plus bas, donc RSB requis plus bas), ou l'inverse, à savoir être très précis (RSB élevé) quitte à détecter peu souvent (seuil élevé pour ne pas transmettre des estimations imprécises). Ou peut être que l'optimal résulte d'un compromis médian ?
Il apparait clairement que le critère d'optimalité doit dépendre du couplage statistique entre l'algorithme d'actualisation du modèle d'état (filtre de Kalman, ...) et la stratégie conjointe détection-estimation (cadence de renouvellement et précision des estimés) : c'est le thème du filtrage non-linéaire au sens bayésien du terme ("the discrete-time nonlinear filtering problem with additive Gaussian process noise and measurement noise" [FRT02] [HRFT04]).

Perspective *:*
Certains travaux ont commencé (cf. [HRFT04] et les références citées) à étudier l'influence du test de détection sur les performances en estimation (BCR a posteriori) du filtre non-linéaire au sens bayésien du terme avec une modélisation des performances en detection sommaire se résumant à la prise en compte d'un ensemble de valeurs de $\mathcal{P}_D$ (une par observation élémentaire). Une première démarche complémentaire pourrait être d'exploiter ces travaux pour rechercher les valeurs de $\mathcal{P}_D$ minimisant la BCR a posteriori, puis de les comparer au performance du TRV pour déterminer s'il existe potentiellement ou pas un test réalisable qui permet de les atteindre.

4) Contributions:

[J2] [J4] [J5] [J7] [J8] [J9] [J11] [J12] - [C11] [C12] [C13] [C15] [C16] [C18] [C21] [C25]

Cet axe de recherche est original, en ce sens que dans les monographies disponibles en traitement du signal [VT68] [Sch91] [Kay93] [Kay98] [LC98] [VT02] les théories de la détection (décision) et de l'estimation sont abordées séparément. La raison de cette dichotomie historique est probablement l'accroissement significatif de la difficulté à caractériser les estimateurs (notamment le calcul de l'EQM) lorsque l'univers d'observation est un sous ensemble de $\Omega = \mathbb{C}^M$ (même dans le cas Gaussien). Hormis certaines contributions antérieures en poursuite radar (radar tracking, cf. [FRT02] et les références citées), l'étude théorique générale des problèmes conjoints détection-estimation ne semble pas avoir été abordée (à ma connaissance) jusqu'à ma thèse [Chau04] et l'article [J2] où nous montrons que les principales propriétés nécessaires au développement des bornes inférieures de l'EQM peuvent être étendues formellement lors du conditionnement des observations par un événement réalisable. Les nombreux travaux publiés dans le domaine (cf. [DNYSF09] et les références citées) abordent la caractérisation des problèmes conjoints détection-simulation par l'association d'un TRVG (143) ("one stage or two-stage GLRT") conditionnant les estimateurs du maximum de vraisemblance associés, les performances conjointes étant généralement évaluées par simulation (toujours pour les performances en estimation, généralement pour les performances en détection).
Pour ma part, j'ai choisi la perspective d'une extension des résultats généraux obtenus en estimation non-conditionnelle, notamment l'obtention d'expressions analytiques générales pour des tests réalisables simples communément employés dans de multiples applications en traitement du signal (radar, sonar, télécoms, ...). Ainsi, une première partie de ces travaux est consacrée à l'établissement des bornes inférieures de l'EQM pour un modèle d'observation déterministe conditionné par le détecteur d'énergie (120) [J5] [C18]. Parallèlement, une seconde partie est consacrée à l'analyse des conséquences du conditionnement (nature du test) sur les performances en estimation (biais, efficacité, zone de décrochement) [C15] ou sur les performances conjointes (détection-estimation) [J8]. Enfin la troisième partie est l'application des deux premières à une technique haute précision angulaire : l'écartométrie monopulse (115), pour laquelle tous les calculs analytiques sont accessibles : EQM, BCR, biais conditionnés par le détecteur d'énergie (120), pour les modèles d'observations déterministe et stochastique. De plus la possibilité d'obtenir une expression analytique de l'EQM d'un estimateur angulaire réalisable (et dérivant de l'EMV) est une motivation

supplémentaire pour la recherche des généralisations possibles du cas cible simple (diffracteur unique) mono-voie d'écartométrie [J4] [J7], à savoir : le cas cible simple (diffracteur unique) multi-voies adaptatives d'écartométrie [J9], puis le cas cible complexe (diffracteurs multiples) multi-voies adaptatives d'écartométrie [J11] [J12] pour arriver peut être à la caractérisation analytique du cas multi-cibles complexes (diffracteurs multiples) multi-voies adaptatives d'écartométrie, proche conceptuellement de l'estimation des fréquences de plusieurs cisoïdes.

Cette partie de ma recherche a été initiée par des travaux [J2] [J4] [J5] [J7] directement issus de ma thèse supervisée par le Professeur Pascal Larzabal, travaux que nous tentons régulièrement d'ouvrir à collaboration, que ce soit avec d'anciens thésards maintenant enseignant-chercheur (Jérôme Gally **[T1]** ou François Vincent **[T2]**) [J8] [C15] [C18] ou avec des chercheurs français (Alexandre Renaux) [C18] ou étrangers (Ulrich Nickel) [J9] [J11] [J12] [C21] [C25].

C. Publications choisies

Je joins en annexe les publications suivantes :

- **[J2]** E. Chaumette, P. Larzabal, P. Forster, "On the influence of a detection step on lower bounds for deterministic parameter estimation", IEEE Trans. on SP, 53(11): 4080-4090, 2005.
  Par ce que c'est le premier. De plus il illustre la problématique du problème conjoint détection-estimation : une théorie simple pour les détecteurs réalisables, mais des calculs complexes en pratique à cause du domaine conditionnel d'intégration.
- **[J6]** E. Chaumette, J. Galy, A. Quinlan, P. Larzabal, "A New Barankin Bound Approximation for the Prediction of the Threshold Region Performance of Maximum Likelihood Estimators", IEEE Trans. on SP, 56(11): 5319-5333, 2008.
  Par ce qu'il étend les développement présentés §II-A4b-p26 et que c'est le seul article de la littérature courante qui permette de calculer en pratique les BHCR, BMS, BMH, BG, BTT pour un modèle d'observation gaussien réel ou complexe.
- **[J7]** E. Chaumette, P. Larzabal, "Monopulse-radar tracking of swerling III-IV targets using multiple observations", IEEE Trans. on AES, 44(2): 520-537, 2008
  Par ce que la formulation analytique du biais et de l'EQM (§IV) est un exemple d'ingéniérie appliquée au calcul (voir §VII.C).
- **[J10]** T. Menni, E. Chaumette, P. Larzabal, J.P. Barbot, "New Results on Deterministic Cramér–Rao Bounds for Real and Complex Parameters", IEEE Trans. on SP, 60(3): 1032-1049, 2012
  Par ce que c'est un exemple de l'adage "plus c'est simple, plus c'est puissant" appliqué à la formulation de la BCR. La minimisation d'une norme sous contrainte linéaire (résultat élémentaire en algèbre) permet de revisiter la formulation de la BCR, corriger certaines inexactitudes, étendre les conditions de régularité et formuler de façon générale les conditions d'identifiabilité.
- **[J11]** E. Chaumette, U.R.O. Nickel, P. Larzabal, "Detection and Parameter Estimation of Extended Targets Using the Generalized Monopulse Estimator", IEEE Trans. on AES, 48(4): 3389-3417, 2012
  Par ce qu'il ne faut pas se laisser décourager par le calcul des 25 termes d'une expression présentant un intérêt théorique et pratique (cf. (50-51) et §IX.D).
- **[C18]** E. Chaumette, J. Galy, F. Vincent, A. Renaux, P. Larzabal, "Mse lower bounds conditioned by the energy detector", in Proc. Eurasip EUSIPCO, 2007
  Pour compléter les résultats fournies §II-B2-p59.
- **[C22]** E. Chaumette, A. Renaux, P. Larzabal, "New trends in deterministic lower bounds and SNR threshold estimation: From derivable bounds to conjectural bounds", in Proc. IEEE SAM, 2010
  Pour illustrer la discussion sur le "Problème ouvert 3" §II-A5-p38.
- **[C28]** C. Ren, J. Galy, E. Chaumette, P. Larzabal, A. Renaux,"High résolution techniques for radar: myth or reality ?", in Proc. Eurasip EUSIPCO, 2013
  Pour illustrer la discussion sur la résolvabilité de l'EMVD et les questions qu'elle soulève II-A7-p48

III. Conclusion et Perspectives

Comme évoqué précédemment, l'universalité du thème de recherche "caractérisation des problèmes conjoints de détection et d'estimation" (la quasi totalité des dispositifs de mesure ou de transmission conçus relevant de ce thème) et sa complexité, tant théorique que calculatoire, en font pour moi un sujet inépuisable d'investigations et de questionnements auquel je compte consacrer ma recherche future.
De plus, j'ai découvert dans la recherche et l'encadrement collaboratif de thésards, une stimulation intellectuelle d'une grande richesse scientifique et humaine (la science avance aussi en conférence ...). Mon ambition est d'ailleurs de trouver à terme un poste d'enseignant-chercheur dans une école d'ingénieur, ce qui constituerait pour moi la solution idéale pour m'assurer une pérennité de l'action intellectuelle et pour capitaliser mon expérience professionnelle d'ingénieur, que ce soit dans l'élaboration de coopérations industrielles ou de projets pédagogiques.
Soucieux de la transmission et de la fertilisation de la connaissance, j'ai toujours saisi les différentes opportunités d'enseignement, d'encadrement ou de séminaires qui se sont présentées à moi.

Les perspectives sont multiples, tant il reste à faire dans ce thème de recherche. Je ferai néanmoins la distinction entre les perspectives immédiates et celles à long termes.
Les perspectives immédiates sont celles que j'ai pu élaborer sur la base de mes connaissance actuelles concernant l'estimation déterministe et que j'ai intégrées à la synthèse de mes travaux de recherches puisqu'elles visent à répondre à certains problèmes (à ma connaissance) toujours ouverts : cf. §II-A5-p40,41,44,48,52 et §II-B3-62,63.

Mon activité d'ingénieur de recherche et développement en traitement du signal radar m'a permis d'explorer les différentes composantes techniques nécessaires à la compréhension fine des systèmes complexes que sont les dispositifs de mesures (radar) ou de transmission de l'information (télécoms) modernes. Il m'a donc naturellement conduit aux domaines de la simulation de la physique ainsi qu'à celui du traitement du signal associé. Mais je n'ai jamais eu l'occasion professionnellement d'être engagé dans une étude appartenant au traitement de données en radar, domaine où sont explorées les techniques de pistage, de fusion, de reconnaissance (logique floue), ... essentiellement basées sur une approche bayésienne appliquée à la théorie du filtrage non-linéaire (bayésien).
L'estimation bayésienne reste donc pour moi l'approche complémentaire à explorer, projet personnel à long terme d'extension de mon expertise en détection-estimation pour paramètres déterministes à un contexte de paramètres bayésiens.
Cette perspective à long terme s'inscrit également en complément des perspectives immédiates, puisqu'il semble exister de fortes connections entre certains problèmes ouverts d'estimation déterministe et l'approche bayésienne.
C'est pourquoi j'ai entrepris le co-encadrement de la thèse de Chengfang Ren **[T6]** dont la thématique est précisément l'exploration des bornes de performances pour les modèles d'observation paramétriques au sens le plus large, c'est à dire pour paramètres hybrides (déterministes et bayésiens) [C27] [C28].
En effet, dans la plupart des applications, un modèle d'observation purement bayésien n'existe pas : les densités a priori ont des paramètres déterministes inconnus au moment de l'expérience. De même, la prise en compte de paramètres bayésiens en estimation déterministe (imperfections, cinématiques, amplitudes des sources, ...) permet d'obtenir une performance moyenne du système pour l'ensemble des paramètres (y compris les déterministes).

Ma recherche à venir sera donc hybride.

IV. Annexes

A. Une inégalité de type Markov-Bienaymé-Tchebychev

Soit $\mathcal{D}_{\boldsymbol{\xi}}(t) = \left\{ \mathbf{x} \in \Omega \;/\; \sum\limits_{q=1}^{Q} \left| \frac{\widehat{g_q(\boldsymbol{\theta}^0)}(\mathbf{x}) - h_q}{\xi_q} \right|^s < t \right\}$.

$$
\begin{aligned}
E_{\boldsymbol{\theta}^0}\left[\sum_{q=1}^{Q}\left|\frac{\widehat{g_q\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})-h_q}{\xi_q}\right|^s\right] &= \int_{\Omega}\sum_{q=1}^{Q}\left|\frac{\widehat{g_q\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})-h_q}{\xi_q}\right|^s p\left(\mathbf{x};\boldsymbol{\theta}^0\right)d\mathbf{x} \\
&= \int_{\mathcal{D}_{\boldsymbol{\xi}}(t)}\sum_{q=1}^{Q}\left|\frac{\widehat{g_q\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})-h_q}{\xi_q}\right|^s p\left(\mathbf{x};\boldsymbol{\theta}^0\right)d\mathbf{x}+\int_{\overline{\mathcal{D}_{\boldsymbol{\xi}}(t)}}\sum_{q=1}^{Q}\left|\frac{\widehat{g_q\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})-h_q}{\xi_q}\right|^s p\left(\mathbf{x};\boldsymbol{\theta}^0\right)d\mathbf{x} \\
&\geq \int_{\overline{\mathcal{D}_{\boldsymbol{\xi}}(t)}}\sum_{q=1}^{Q}\left|\frac{\widehat{g_q\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})-h_q}{\xi_q}\right|^s p\left(\mathbf{x};\boldsymbol{\theta}^0\right)d\mathbf{x} \\
&\geq t\mathcal{P}\left(\overline{\mathcal{D}_{\boldsymbol{\xi}}(t)};\boldsymbol{\theta}^0\right)
\end{aligned}
$$

En particulier :

$$
E_{\boldsymbol{\theta}^0}\left[\sum_{q=1}^{Q}\left|\frac{\widehat{g_q\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})-h_q}{\xi_q}\right|^s\right] \geq 1-\mathcal{P}\left(\mathcal{D}_{\boldsymbol{\xi}}(1);\boldsymbol{\theta}^0\right)
$$

soit encore :

$$
1-E_{\boldsymbol{\theta}^0}\left[\sum_{q=1}^{Q}\left|\frac{\widehat{g_q\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})-h_q}{\xi_q}\right|^s\right] \leq \mathcal{P}\left(\sum_{q=1}^{Q}\left|\frac{\widehat{g_q\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})-h_q}{\xi_q}\right|^s < 1;\boldsymbol{\theta}^0\right)
$$

B. Estimateur au sens du Maximum de Vraisemblance

L'estimation au sens du maximum de vraisemblance est une stratégie d'estimation parmi d'autres stratégies possibles [Cra46] [Sch91] [Kay93], au sens de principe, méthode, algorithme permettant de formuler de façon systématique un estimateur réalisable du paramètre inconnu à estimer. Elle est la réponse de bon sens à la question : sachant que nous observons $\mathbf{x}$ et que nous connaissons le modèle de la d.d.p.. $p(\mathbf{x};\theta)$, quelle est la valeur du paramètre θ qui rend notre observation la plus probable ? Mathématiquement ce concept se traduit donc simplement par la stratégie d'estimation suivante :

$$
\widehat{\theta}_{\text{mv}}(\mathbf{x}) = \arg\max_{\theta}\{p(\mathbf{x};\theta)\} \Rightarrow \widehat{\theta}_{\text{mv}}(\mathbf{x}) = \arg\left\{\frac{\partial p(\mathbf{x};\theta)}{\partial\theta}=0\right\} \tag{123}
$$

où $\widehat{\theta}_{\text{mv}}(\mathbf{x})$ est l'estimateur au sens du maximum de vraisemblance (EMV) de θ.
Bien qu'historiquement introduite par R.A. Fisher [Fis21] bien avant les premières formulations de la BCR [Fre43] [Dar45] [Rao45] [Cra46], cette stratégie découle naturellement de l'étude des bornes inférieures de l'EQM locale. En effet, par construction (Lemme 1-p27)(56), chaque borne inférieure de l'EQM locale est la variance d'un estimateur spécifique. Malheureusement cet estimateur est généralement clairvoyant; par conséquent ce principe ne constitue donc pas à proprement dit une stratégie d'estimation. Néanmoins, il existe certains modèles d'observation pour lesquels l'estimateur atteignant la borne inférieure est réalisable [Fre43] [Dar45] [Rao45] [Cra46] [Bat46] [Mor83] [Pom03]; ce principe mérite donc d'être pris en considération.

A titre d'exemple, considérons le cas le plus simple : l'estimation d'un unique paramètre inconnu θ. Dans ce cas, l'estimateur atteignant la BCR vérifie (cf. Lemme 1-p27 et (45)) :

$$\widehat{\theta}_{\text{eff}}(\mathbf{x}) - \theta = \frac{1}{F(\theta)} \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}, \quad \forall \mathbf{x} \in \Omega, \quad F_\theta = E_\theta\left[\left(\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right)^2\right] \tag{124}$$

Un tel estimateur est dit efficace [Cra46]. La recherche des d.d.p.. $p(\mathbf{x};\theta)$ pour lesquelles l'estimateur efficace $\widehat{\theta}_{\text{eff}}$ est réalisable a été traitée par Fréchet [Fre43] (et généralisé par Darmois [Dar45]); elles sont de la forme :

$$p(\mathbf{x};\theta) = e^{\frac{\partial \mu(\theta)}{\partial \theta}[h(\mathbf{x})-\theta]+\mu(\theta)+l(\mathbf{x})} \text{ avec } h(\mathbf{x}), \, l(\mathbf{x}) \text{ et } \mu(\theta) \text{ telles que } \int_\Omega f_\theta(\mathbf{x})\, d\mathbf{x} = 1,$$

où $\widehat{\theta}_{\text{eff}}(\mathbf{x}) \triangleq h(\mathbf{x})$ constitue l'estimateur efficace de θ, lequel coïncide avec l'EMV puisqu'il est également solution de l'équation (123). Cette propriété a fortement contribué à l'utilisation quasi-systématique des estimateurs au sens du MV pour résoudre les problèmes d'estimation, en ce sens que si un estimateur efficace et réalisable existe, alors c'est l'EMV (la réciproque étant fausse dans le cas général, cf. ci-après).
La popularité de l'estimation au sens du MV provient non seulement de la variété des modèles d'observation auxquels elle s'applique [VT02], mais également de ses bonnes performances statistiques asymptotiques identifiées historiquement très tôt [Fre43] [Dar45] [Rao45] [Cra46]. En effet, l'EMV est :

- consistant, c'est à dire que l'estimateur converge en probabilité vers la vraie valeur,
- asymptotiquement efficace, c'est à dire que la variance de l'estimateur tend vers la BCR,
- asymptotiquement gaussien,

où asymptotique doit être pris au sens d'un nombre d'observations indépendantes tendant vers l'infini.
De plus, pour les modèles d'observation gaussiens [SN90] [OVSN93] [Ber99] **[T3]** [RFCL06] [RFBL07], des propriétés complémentaires d'efficacité asymptotiques peuvent être mis en évidence (cf. ci-après).

1) Le modèle d'observation paramétrique "bande étroite":

Soit le modèle d'observation paramétrique dit bande étroite ("narrow band" [VT02, §8]) associé à N observations complexes simultanées de M sources de signal en présence de signaux de bruit additifs (nuisances) :

$$\mathbf{x}^l \triangleq \mathbf{x}^l(\boldsymbol{\theta}) = \mathbf{B}(\boldsymbol{\Xi})\boldsymbol{\sigma}^l + \mathbf{n}^l \qquad l = 1, ..., L \tag{125}$$

où:

L est le nombre d'observations considérées (temporelles, fréquentielles, après traitements, ...),

$\mathbf{x}^l \in \mathcal{M}_{N\times 1}(\mathbb{C})$ est le vecteur d'observations instantanées,

$\boldsymbol{\sigma}^l \in \mathcal{M}_{M\times 1}(\mathbb{C})$ est le vecteur des amplitudes des signaux sources,

$\mathbf{n}^l \in \mathcal{M}_{N\times 1}(\mathbb{C})$ est le vecteur des signaux de bruit (nuisances: bruits thermiques, brouilleurs, ...) additifs,

$\boldsymbol{\Xi} = [\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_M] \in \mathcal{M}_{P\times M}(\mathbb{R})$ est la matrice réelle concaténation des P paramètres inconnus des M sources,

$\mathbf{B}(\boldsymbol{\Xi}) = [\mathbf{b}(\boldsymbol{\varepsilon}_1), \mathbf{b}(\boldsymbol{\varepsilon}_2), \cdots, \mathbf{b}(\boldsymbol{\varepsilon}_M)] \in \mathcal{M}_{N\times M}(\mathbb{C})$ est la matrice des fonctions de transfert vectorielles $\mathbf{b}(\boldsymbol{\varepsilon}_m) \in \mathcal{M}_{N\times 1}(\mathbb{C})$ dépendant des P paramètres inconnus des sources.

Ce modèle "simple" d'observation est néanmoins relativement général en ce sens qu'il ne présume pas du mode de représentation retenue (temporel, fréquentiel, ...) : il ne fait que supposer une relation linéaire entre les coordonnées des observations et les coordonnées des sources de signal dans un espace vectoriel de dimension N. Aussi, ce modèle est il aussi bien utilisé en Analyse Spectrale (estimation de fréquences) [SM97] qu'en Traitement d'Antenne (estimation des directions d'arrivée et des retards) [VT02].
D'autre part, le processus $\mathbf{x}^l$ étant observé lors de L observations, l'ensemble des observations disponible peut également s'écrire sous la forme d'une matrice d'observations $\mathbf{X}$ [OVSN93] :

$$\mathbf{X} \triangleq \mathbf{X}(\boldsymbol{\theta}) = \mathbf{B}(\boldsymbol{\Xi})\boldsymbol{\Sigma} + \mathbf{N} \quad \left\{ \begin{array}{l} \mathbf{X} = \left[\mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^L\right] \\ \boldsymbol{\Sigma} = \left[\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2, , \cdots, \mathbf{x}^L\right] \\ \mathbf{N} = \left[\mathbf{n}^1, \mathbf{n}^2, \cdots, \mathbf{x}^L\right] \end{array} \right. \tag{126}$$

Les hypothèses sur la nature des signaux observés sont classiquement les suivantes [OVSN93] [VT02] :
• H1S : les signaux $\boldsymbol{\sigma}^l$ sont gaussiens circulaires $\left(E\left[\boldsymbol{\sigma}^l\left(\boldsymbol{\sigma}^l\right)^T\right]=\mathbf{0}\right)$, à moyennes nulles $\left(E\left[\boldsymbol{\sigma}^l\right]=\mathbf{0}\right)$, temporellement blancs, spatialement colorés de matrice de covariance inconnue $E\left[\boldsymbol{\sigma}^l\left(\boldsymbol{\sigma}^l\right)^H\right]=\mathbf{R}_{\boldsymbol{\sigma}}$.
• H1D : les signaux $\boldsymbol{\sigma}^l$ sont déterministes, mais inconnus.
• H2 : le bruit $\mathbf{n}^l$ est indépendant des signaux $\boldsymbol{\sigma}^l$, gaussien circulaire $\left(E\left[\mathbf{n}^l\left(\mathbf{n}^l\right)^T\right]=\mathbf{0}\right)$, à moyenne nulle $\left(E\left[\mathbf{n}^l\right]=\mathbf{0}\right)$, temporellement blanc, spatialement blanc $E\left[\mathbf{n}^l\left(\mathbf{n}^l\right)^H\right]=\mathbf{R}_{\mathbf{n}}=\sigma_{\mathbf{n}}^2\mathbf{I}_N$ de puissance $\left(\sigma_{\mathbf{n}}^2\right)$ inconnue.
• H3 : $\mathbf{B}\left(\boldsymbol{\Xi}\right)$ est supposée de rang plein $(M<N)$.
Par conséquent les observations $\mathbf{x}^l$ sont gaussiennes et indépendantes (H1S/H1D et H2). L'hypothèse H1S définit le modèle d'observation gaussien stochastique (unconditionnal observation model); l'hypothèse H1D définit le modèle d'observation gaussien déterministe (conditionnal observation model) [SN90].
En radar [Swe60] [CL07M, §3.1] :
• le modèle d'observation stochastique correspond à un ensemble de cibles fluctuant mutuellement selon des lois de Swerling 1-2 (selon qu'il y a ou non fluctuation d'observation à observation),
• le modèle d'observation déterministe correspond à un ensemble de cibles mutuellement non fluctuantes, loi de Swerling 0, ou fluctuantes d'observation à observation mais de loi de fluctuation à priori inconnue.
Le cas des fluctuations de type Swerling 3-4 n'a pas été abordé à ce jour dans la littérature courante en tant que modèle d'observation de référence. Il n'existe donc pas de résultats généraux pour ces types de fluctuation.

EMV Déterministe (H1D)

Les signaux sources $\boldsymbol{\sigma}^l$ sont supposés déterministes mais inconnus. Les observations $\mathbf{x}^l$ sont aléatoires gaussiennes non centrées circulaires stationnaires

$$E\left[\mathbf{x}^l\right]=\mathbf{B}\left(\boldsymbol{\Xi}\right)\boldsymbol{\sigma}^l,\ E\left[\left(\mathbf{x}^l-E\left[\mathbf{x}^l\right]\right)\left(\mathbf{x}^{l'}-E\left[\mathbf{x}^{l'}\right]\right)^H\right]=\sigma_{\mathbf{n}}^2\mathbf{I}_N\delta_l^{l'},\ E\left[\left(\mathbf{x}^l-E\left[\mathbf{x}^l\right]\right)\left(\mathbf{x}^{l'}-E\left[\mathbf{x}^{l'}\right]\right)^T\right]=\mathbf{0}.$$

Le problème contient $2TM+PM+1$ paramètres inconnus :

$$\boldsymbol{\theta}_D=\left[\mathbf{vec}^T\left(\boldsymbol{\Sigma}\right)\ \mathbf{vec}^T\left(\boldsymbol{\Xi}\right)\ \sigma_{\mathbf{n}}^2\right]^T,$$

et la vraisemblance (d.d.p..) des observations s'écrit:

$$p\left(\mathbf{X};\boldsymbol{\theta}_D\right)=\frac{1}{\left(\pi\sigma_{\mathbf{n}}^2\right)^{TN}}e^{-\frac{1}{\sigma_{\mathbf{n}}^2}\sum_{l=1}^{L}\left\|\mathbf{x}^l-\mathbf{B}(\boldsymbol{\Xi})\boldsymbol{\sigma}^l\right\|^2}. \tag{127}$$

L'EMV déterministe (EMVD) $\widehat{\boldsymbol{\theta}}_D$ est alors obtenu par maximisation de la vraisemblance (127), problème séparable [OVSN93] :

$$\widehat{\boldsymbol{\theta}}_D=\arg\min_{\boldsymbol{\theta}_D}\left\{TN\ln\left(\sigma_{\mathbf{n}}^2\right)+\sum_{l=1}^{L}\frac{\left\|\mathbf{x}^l-\mathbf{B}\left(\boldsymbol{\Xi}\right)\boldsymbol{\sigma}^l\right\|^2}{\sigma_{\mathbf{n}}^2}\right\}\Rightarrow\left\{\begin{array}{l}\widehat{\sigma_{\mathbf{n}}^2}\left(\boldsymbol{\Xi}\right)=\frac{1}{TN}\sum_{l=1}^{L}\left\|\mathbf{x}^l-\mathbf{B}\left(\boldsymbol{\Xi}\right)\boldsymbol{\sigma}^l\right\|^2\\ \widehat{\boldsymbol{\Sigma}}\left(\boldsymbol{\Xi}\right)=\mathbf{B}^{\dagger}\left(\boldsymbol{\Xi}\right)\mathbf{X}\Leftrightarrow\hat{\mathbf{s}}^l=\mathbf{B}^{\dagger}\left(\boldsymbol{\Xi}\right)\mathbf{x}^l\\ \widehat{\boldsymbol{\Xi}}_D=\arg\min_{\boldsymbol{\Xi}}\left\{tr\left(\boldsymbol{\Pi}_{\mathbf{B}(\boldsymbol{\Xi})}^{\perp}\widehat{\mathbf{R}}_{\mathbf{x}}\right)\right\}\end{array}\right. \tag{128}$$

où $\widehat{\mathbf{R}}_{\mathbf{x}}=\frac{1}{L}\sum_{l=1}^{L}\mathbf{x}^l\left(\mathbf{x}^l\right)^H$, $\boldsymbol{\Pi}_{\mathbf{B}}^{\perp}=\mathbf{I}_N-\boldsymbol{\Pi}_{\mathbf{B}}$, $\boldsymbol{\Pi}_{\mathbf{B}}=\mathbf{B}\mathbf{B}^{\dagger}$, $\mathbf{B}^{\dagger}=\left(\mathbf{B}^H\mathbf{B}\right)^{-1}\mathbf{B}^H$, $\boldsymbol{\Pi}_{\mathbf{B}}$ étant le projecteur orthogonal sur $Im\left\{\mathbf{B}\right\}$.

EMV Stochastique (H1S)

Les signaux sources $\boldsymbol{\sigma}^l$ sont aléatoires gaussiens centrés circulaires stationnaires. Les observations $\mathbf{x}^l$ sont aléatoires gaussiennes centrées circulaires stationnaires :

$$E\left[\mathbf{x}^l\right]=\mathbf{0},\ E\left[\mathbf{x}^l\left(\mathbf{x}^{l'}\right)^H\right]=\mathbf{R}_{\mathbf{x}}\delta_l^{l'}=\left(\mathbf{B}\left(\boldsymbol{\Xi}\right)\mathbf{R}_{\boldsymbol{\sigma}}\mathbf{B}^H\left(\boldsymbol{\Xi}\right)+\sigma_{\mathbf{n}}^2\mathbf{I}_N\right)\delta_l^{l'},\ E\left[\mathbf{x}^l\left(\mathbf{x}^{l'}\right)^T\right]=\mathbf{0},$$

$$E\left[\boldsymbol{\sigma}^l\left(\boldsymbol{\sigma}^{l'}\right)^H\right]=\mathbf{R}_{\boldsymbol{\sigma}}\delta_l^{l'},\ E\left[\boldsymbol{\sigma}^l\left(\boldsymbol{\sigma}^{l'}\right)^T\right]=\mathbf{0}.$$

Le problème contient $\frac{M(M+1)}{2}+PM+1$ paramètres inconnus :

$$\boldsymbol{\theta}_S=\left[\left\{(\mathbf{R}_{\boldsymbol{\sigma}})_{m,m'}\right\}_{m\leq m'},\mathbf{vec}\left(\boldsymbol{\Xi}\right)^T,\sigma_{\mathbf{n}}^2\right]^T,$$

et la vraisemblance (d.d.p..) des observations s'écrit :

$$p\left(\mathbf{X};\boldsymbol{\theta}_S\right)=\frac{1}{\left(\pi^N\left|\mathbf{R}_{\mathbf{x}}\right|\right)^L}e^{-\sum_{l=1}^{L}(\mathbf{x}^l)^H\mathbf{R}_{\mathbf{x}}^{-1}\mathbf{x}^l}=\frac{1}{\left(\pi^N\left|\mathbf{R}_{\mathbf{x}}\right|\right)^L}e^{-Ltr\left(\mathbf{R}_{\mathbf{x}}^{-1}\widehat{\mathbf{R}}_{\mathbf{x}}\right)},\qquad \widehat{\mathbf{R}}_{\mathbf{x}}=\frac{1}{L}\sum_{l=1}^{L}\mathbf{x}^l\left(\mathbf{x}^l\right)^H. \tag{129}$$

L'EMV stochastique (EMVS) $\widehat{\boldsymbol{\theta}}_S$ est obtenu par maximisation de la vraisemblance (129), problème séparable [OVSN93] :

$$\widehat{\boldsymbol{\theta}}_S=\arg\min_{\boldsymbol{\theta}_S}\left\{\ln\left|\mathbf{R}_{\mathbf{x}}\right|+tr\left(\mathbf{R}_{\mathbf{x}}^{-1}\widehat{\mathbf{R}}_{\mathbf{x}}\right)\right\}\Rightarrow\begin{cases}\widehat{\sigma_{\mathbf{n}}^2}\left(\boldsymbol{\Xi}\right)=\frac{1}{N-M}tr\left(\boldsymbol{\Pi}_{\mathbf{B}(\boldsymbol{\Xi})}^{\perp}\widehat{\mathbf{R}}_{\mathbf{x}}\right)\\ \widehat{\mathbf{R}}_{\boldsymbol{\sigma}}\left(\boldsymbol{\Xi}\right)=\mathbf{B}^{\dagger}\left(\boldsymbol{\Xi}\right)\left(\widehat{\mathbf{R}}_{\mathbf{x}}-\widehat{\sigma_{\mathbf{n}}^2}\left(\boldsymbol{\Xi}\right)\mathbf{I}_N\right)\mathbf{B}^{\dagger}\left(\boldsymbol{\Xi}\right)^H\\ \widehat{\boldsymbol{\Xi}}_S=\arg\min_{\boldsymbol{\Xi}}\left\{\left|\mathbf{B}\left(\boldsymbol{\Xi}\right)\widehat{\mathbf{R}}_{\boldsymbol{\sigma}}\left(\boldsymbol{\Xi}\right)\mathbf{B}^H\left(\boldsymbol{\Xi}\right)+\widehat{\sigma_{\mathbf{n}}^2}\left(\boldsymbol{\Xi}\right)\mathbf{I}_N\right|\right\}\end{cases} \tag{130}$$

Equivalence des critères du MVS et du MVD à fort RSB

Les auteurs de [RFBL07] ont démontré l'équivalence des critères du MVS et du MVD à fort RSB en remarquant que si $C_S\left(\boldsymbol{\Xi}\right)$ et $C_D\left(\boldsymbol{\Xi}\right)$ sont les critères à minimiser pour obtenir l'EMVD (128) et l'EMVS (130) :

$$\begin{aligned}C_D\left(\boldsymbol{\Xi}\right)&=tr\left(\boldsymbol{\Pi}_{\mathbf{B}(\boldsymbol{\Xi})}^{\perp}\widehat{\mathbf{R}}_{\mathbf{x}}\right)\\ C_S\left(\boldsymbol{\Xi}\right)&=\left|\mathbf{B}\left(\boldsymbol{\Xi}\right)\widehat{\mathbf{R}}_{\boldsymbol{\sigma}}\left(\boldsymbol{\Xi}\right)\mathbf{B}^H\left(\boldsymbol{\Xi}\right)+\widehat{\sigma_{\mathbf{n}}^2}\left(\boldsymbol{\Xi}\right)\mathbf{I}_N\right|=\left|\boldsymbol{\Pi}_{\mathbf{B}(\boldsymbol{\Xi})}\widehat{\mathbf{R}}_{\mathbf{x}}\left(\boldsymbol{\Xi}\right)\boldsymbol{\Pi}_{\mathbf{B}(\boldsymbol{\Xi})}+\frac{C_{MVD}\left(\boldsymbol{\varepsilon}\right)}{N-M}\boldsymbol{\Pi}_{\mathbf{B}(\boldsymbol{\Xi})}^{\perp}\right|\end{aligned}$$

alors puisque $\mathbf{I}_N=\boldsymbol{\Pi}_{\mathbf{B}}^{\perp}+\boldsymbol{\Pi}_{\mathbf{B}}$:

$$C_S\left(\boldsymbol{\Xi}\right)=\left|\boldsymbol{\Pi}_{\mathbf{B}(\boldsymbol{\Xi})}\widehat{\mathbf{R}}_{\mathbf{x}}\left(\boldsymbol{\Xi}\right)\boldsymbol{\Pi}_{\mathbf{B}(\boldsymbol{\Xi})}\right|\left(\frac{C_D\left(\boldsymbol{\Xi}\right)}{N-M}\right)^{N-M}$$

$$\Downarrow$$

$$C_S\left(\boldsymbol{\Xi}\right)\quad\overset{RSB\to\infty}{\longrightarrow}\quad\left|\mathbf{B}\left(\boldsymbol{\Xi}\right)\widehat{\mathbf{R}}_{\boldsymbol{\sigma}}\left(\boldsymbol{\Xi}\right)\mathbf{B}\left(\boldsymbol{\Xi}\right)^H\right|\left(\frac{C_D\left(\boldsymbol{\Xi}\right)}{N-M}\right)^{N-M}$$

$$\Downarrow$$

$$\min\left\{C_S\left(\boldsymbol{\Xi}\right)\right\}\quad\overset{RSB\to\infty}{\Longleftrightarrow}\quad\min\left\{C_D\left(\boldsymbol{\Xi}\right)\right\}$$

Ce résultat permet une unification des algorithmes de recherche de l'EMV pour les modèles d'observation stochastique et déterministe.

2) Performances asymptotiques des EMV:

Rappelons que la BCR repose sur l'hypothèse d'un estimateur $\widehat{\boldsymbol{\theta}}$ de $\boldsymbol{\theta}$ localement sans biais (59) :

$$E_{\boldsymbol{\theta}^0+d\boldsymbol{\theta}}\left[\widehat{\boldsymbol{\theta}^0}\left(\mathbf{X}\right)\right]=\boldsymbol{\theta}^0+d\boldsymbol{\theta}+\mathbf{o}_{\boldsymbol{\theta}^0}\left(\|d\boldsymbol{\theta}\|\right)$$

et localement le meilleur en $\boldsymbol{\theta}^0$ (60) :

$$\mathbf{G}_{\boldsymbol{\theta}^0}\left(\widehat{\boldsymbol{\theta}^0}\left(\mathbf{X}\right)-\boldsymbol{\theta}^0\right)\geq\mathbf{BCR}_{\boldsymbol{\theta}|\boldsymbol{\theta}}\left(\boldsymbol{\theta}^0\right),\quad\mathbf{BCR}_{\boldsymbol{\theta}|\boldsymbol{\theta}}\left(\boldsymbol{\theta}\right)=E_{\boldsymbol{\theta}^0}\left[\frac{\partial\ln p\left(\mathbf{X};\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\left(\frac{\partial\ln p\left(\mathbf{X};\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right)^T\right]^{-1}.$$

Plus spécifiquement, la BCR "Stochastique" pour les paramètres $\boldsymbol{\Xi}$ [OVSN93] s'écrit :

$$\left(\mathbf{BCRS}_{\boldsymbol{\Xi}|\boldsymbol{\theta}}^{-1}\left(\boldsymbol{\theta}\right)\right)_{i,j}=\frac{2T}{\sigma_{\mathbf{n}}^2}\operatorname{Re}\left\{tr\left(\frac{\partial\mathbf{B}\left(\boldsymbol{\Xi}\right)}{\partial\Xi_j}^H\boldsymbol{\Pi}_{\mathbf{B}(\boldsymbol{\Xi})}^{\perp}\frac{\partial\mathbf{B}\left(\boldsymbol{\Xi}\right)}{\partial\Xi_i}\mathbf{R}_{\boldsymbol{\sigma}}\mathbf{B}\left(\boldsymbol{\Xi}\right)^H\mathbf{R}_{\mathbf{x}}^{-1}\mathbf{B}\left(\boldsymbol{\Xi}\right)\mathbf{R}_{\boldsymbol{\sigma}}\right)\right\} \tag{131}$$

et la BCR "Déterministe" pour les paramètres $\boldsymbol{\Xi}$ [OVSN93] s'écrit :

$$\left(\mathbf{BCRD}_{\boldsymbol{\Xi}|\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta})\right)_{i,j}=\frac{2T}{\sigma_{\mathbf{n}}^2}\operatorname{Re}\left\{tr\left(\frac{\partial\mathbf{B}(\boldsymbol{\Xi})}{\partial\Xi_j}^H\boldsymbol{\Pi}_{\mathbf{B}(\boldsymbol{\Xi})}^{\perp}\frac{\partial\mathbf{B}(\boldsymbol{\Xi})}{\partial\Xi_i}\widehat{\mathbf{R}}_{\boldsymbol{\sigma}}\right)\right\},\quad \widehat{\mathbf{R}}_{\boldsymbol{\sigma}}=\frac{1}{L}\sum_{l=1}^{L}\boldsymbol{\sigma}^l\left(\boldsymbol{\sigma}^l\right)^H \tag{132}$$

Les BCR Stochastique et Déterministe sont donc calculables quelle que soit la forme de $\mathbf{B}(\boldsymbol{\Xi})$ et servent de référence pour évaluer les performances d'un estimateur donnée en terme d'efficacité relative, c'est à dire de capacité à avoir une variance d'estimation égale à sa valeur minimale possible : la BCR.

Asymptotiquement (L / N / RSB $\to\infty$) les EMV peuvent être comparés à la BCR [OVSN93] [Ber99] **[T3]** [FBL04] [RFCL06] [RFBL07] :

• à nombre d'observations L tendant vers l'infini (nombre de capteurs N fini, RSB fini) :

l'EMVS est asymptotiquement gaussien et efficace :

$$\mathbf{G}_{\boldsymbol{\theta}^0}\left(\widehat{\boldsymbol{\Xi}^0}_S-\boldsymbol{\Xi}^0\right)\approx\mathbf{BCRS}_{\boldsymbol{\Xi}|\boldsymbol{\theta}}\left(\boldsymbol{\theta}^0\right)$$

l'EMVD est asymptotiquement gaussien mais non efficace ($\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Xi})$ est non consistant (nombre d'inconnues croissant avec L) mais asymptotiquement sans biais). De plus, dans le cas d'un seul paramètre inconnu par source $\left(\boldsymbol{\varepsilon}\triangleq\varepsilon,\boldsymbol{\Xi}=(\varepsilon_1,\ldots,\varepsilon_M)^T\right)$ alors :

$$\mathbf{G}_{\boldsymbol{\theta}^0}\left(\widehat{\boldsymbol{\Xi}^0}_D-\boldsymbol{\Xi}^0\right)\approx\mathbf{BCRD}_{\boldsymbol{\Xi}|\boldsymbol{\theta}}\left(\boldsymbol{\theta}^0\right)+ 2T\mathbf{BCRD}_{\boldsymbol{\Xi}|\boldsymbol{\theta}}\left(\boldsymbol{\theta}^0\right)\operatorname{Re}\left\{\frac{\partial\mathbf{B}\left(\boldsymbol{\Xi}^0\right)}{\partial\varepsilon}^H\boldsymbol{\Pi}_{\mathbf{B}(\boldsymbol{\Xi}^0)}^{\perp}\frac{\partial\mathbf{B}\left(\boldsymbol{\Xi}^0\right)}{\partial\varepsilon}\odot\left(\mathbf{B}\left(\boldsymbol{\Xi}^0\right)^T\mathbf{B}\left(\boldsymbol{\Xi}^0\right)^*\right)^{-1}\right\}\mathbf{BCRD}_{\boldsymbol{\Xi}|\boldsymbol{\theta}}\left(\boldsymbol{\theta}^0\right)$$

• à nombre d'observations L et nombre de capteurs N tendant vers l'infini (RSB fini) :

l'EMVD est asymptotiquement gaussien et efficace :

$$\mathbf{G}_{\boldsymbol{\theta}^0}\left(\widehat{\boldsymbol{\Xi}^0}_D-\boldsymbol{\Xi}^0\right)\approx\mathbf{BCRD}_{\boldsymbol{\Xi}|\boldsymbol{\theta}}\left(\boldsymbol{\theta}^0\right)$$

• à RSB tendant vers l'infini (nombre d'observations L fini, nombre de capteurs N fini) :

l'EMVS est asymptotiquement non gaussien et non efficace

l'EMVD est asymptotiquement gaussien et efficace :

$$\mathbf{G}_{\boldsymbol{\theta}^0}\left(\widehat{\boldsymbol{\Xi}^0}_D-\boldsymbol{\Xi}^0\right)\approx\mathbf{BCRD}_{\boldsymbol{\Xi}|\boldsymbol{\theta}}\left(\boldsymbol{\theta}^0\right)$$

3) Mise en oeuvre de l'EMVD :

La principale difficulté dans la mise en oeuvre de l'EMVD provient de la forme de sa solution (128) :

$$\left\{\widehat{\boldsymbol{\Sigma}},\widehat{\boldsymbol{\Xi}}\right\}=\underset{\{\boldsymbol{\Sigma},\boldsymbol{\Xi}\}}{\arg\min}\left\{C(\boldsymbol{\Sigma},\boldsymbol{\Xi})\triangleq\sum_{l=1}^{L}\left\|\mathbf{x}^l-\mathbf{B}(\boldsymbol{\Xi})\boldsymbol{\sigma}^l\right\|^2\right\}\Rightarrow\left\{\begin{array}{l}\widehat{\boldsymbol{\sigma}^l}=\left(\mathbf{B}\left(\widehat{\boldsymbol{\Xi}}\right)^H\mathbf{B}\left(\widehat{\boldsymbol{\Xi}}\right)\right)^{-1}\mathbf{B}\left(\widehat{\boldsymbol{\Xi}}\right)^H\mathbf{x}^l\\ \widehat{\boldsymbol{\Xi}}=\arg\max\limits_{\boldsymbol{\Xi}}\left\{tr\left(\boldsymbol{\Pi}_{\mathbf{B}(\boldsymbol{\Xi})}\widehat{\mathbf{R}}_{\mathbf{x}}\right)\right\}\\ \widehat{\boldsymbol{\Xi}}=\arg\min\limits_{\boldsymbol{\Xi}}\left\{tr\left(\boldsymbol{\Pi}_{\mathbf{B}(\boldsymbol{\Xi})}^{\perp}\widehat{\mathbf{R}}_{\mathbf{x}}\right)\right\}\end{array}\right. \tag{133}$$

qui requiert la maximisation d'une fonction dépendant de $M\times P$ paramètres, procédure généralement irréalisable d'un point de vue du temps de calcul lorsque $M\times P>>3$.

Cette limitation est à l'origine de la famille des approches "sous-espace" [OVSN93] dont les plus connues sont : MUSIC, WSF, ESPRIT. Cette famille d'algorithmes exploite l'hypothèse de non corrélation des sources de signal ($\widehat{\mathbf{R}}_{\boldsymbol{\sigma}}$ de rang plein) et constitue un cas particulier d'approximation de l'EMVD. En effet, si nous considérons la décomposition en valeurs/vecteurs propres de $\widehat{\mathbf{R}}_{\mathbf{x}}$, $\widehat{\mathbf{R}}_{\mathbf{x}} = \widehat{\mathbf{U}}\widehat{\boldsymbol{\lambda}}\widehat{\mathbf{U}}$ avec $\widehat{\boldsymbol{\lambda}} = \mathbf{Diag}\left(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_N\right)$ et $\widehat{\mathbf{U}}^H\widehat{\mathbf{U}} = \mathbf{I}$, alors

$$tr\left(\mathbf{\Pi}_{\mathbf{B}}^{\perp}\widehat{\mathbf{R}}_{\mathbf{x}}\right) = tr\left(\mathbf{\Pi}_{\mathbf{B}}^{\perp}\widehat{\mathbf{U}}\widehat{\boldsymbol{\lambda}}\widehat{\mathbf{U}}^H\right) = tr\left(\widehat{\boldsymbol{\lambda}}\widehat{\mathbf{U}}^H\mathbf{\Pi}_{\mathbf{B}}^{\perp}\widehat{\mathbf{U}}\right) = \sum_{n=1}^{N}\widehat{\lambda}_n\left\|\mathbf{\Pi}_{\mathbf{B}}^{\perp}\widehat{\mathbf{u}}_n\right\|^2$$

et

$$\arg\min_{\boldsymbol{\Xi}}\left\{tr\left(\mathbf{\Pi}_{\mathbf{B}(\boldsymbol{\Xi})}^{\perp}\widehat{\mathbf{R}}_{\mathbf{x}}\right)\right\} = \arg\min_{\boldsymbol{\Xi}}\left\{\sum_{n=1}^{N}\widehat{\lambda}_n\left\|\mathbf{\Pi}_{\mathbf{B}(\boldsymbol{\Xi})}^{\perp}\widehat{\mathbf{u}}_n\right\|^2\right\}.$$

Par conséquent $\widehat{\boldsymbol{\Xi}}_{MV}$ est la valeur de $\boldsymbol{\Xi}$ pour laquelle $\mathbf{\Pi}_{\mathbf{B}(\boldsymbol{\Xi})}^{\perp}$ s'identifie au projecteur sur le sous espace vectoriel (s.e.v.) engendré par les $N-M$ vecteurs propres associés aux $N-M$ plus petites v.p. $\widehat{\lambda}_n$, encore appelé sous espace "bruit". On démontre alors que si $\widehat{\mathbf{R}}_{\boldsymbol{\sigma}}$ est de rang plein et $\mathbf{B}(\boldsymbol{\Xi}) = [\mathbf{b}(\boldsymbol{\varepsilon}_1), \mathbf{b}(\boldsymbol{\varepsilon}_2), \cdots, \mathbf{b}(\boldsymbol{\varepsilon}_N)]$ est de rang plein $\forall \boldsymbol{\varepsilon}_1 \neq \boldsymbol{\varepsilon}_2 \neq \cdots \neq \boldsymbol{\varepsilon}_N$ (condition de fonction de transfert vectorielle non ambiguë, aussi appelée condition d'identifiabilité des paramètres [OVSN93]), la minimisation de $C(\boldsymbol{\Sigma}, \boldsymbol{\Xi})$ dépendant de $M \times P$ paramètres se réduit en un problème de projection orthogonale sur l'espace "bruit" (ou sur son espace orthogonal, l'espace "signal") dépendant de P paramètres.
Or l'hypothèse de non corrélation des sources n'est pas toujours réaliste en pratique. D'autre part, même si la matrice de corrélation des sources $\widehat{\mathbf{R}}_{\boldsymbol{\sigma}}$ est non corrélée, pour que son estimée soit également non corrélée, il faut au moins que $L \geq M$ (idéalement $L \geq 3M$), ce qui n'est pas toujours réalisable d'un point de vue pratique à mesure que M augmente.
Une autre approche destinée à réduire le temps de calcul pour la mise en oeuvre de l'EMVD est basée sur les techniques de "descente" itérative monoparamètre (relativement aux composantes de $\boldsymbol{\Xi}$) appliquées à la fonction $C_D(\boldsymbol{\Sigma}, \boldsymbol{\Xi})$ (133) comme par exemple :
• les algorithmes Clean [TS88], Clean Relax (version améliorée de Clean) [LS96], basés sur la minimisation itérative de $C(\boldsymbol{\Sigma}, \boldsymbol{\Xi})$ relativement aux paramètres d'une source $\{\widehat{\boldsymbol{\varepsilon}}_m, \sigma_{m,1}, \ldots, \sigma_{m,L}\}$:

$$\left.\begin{array}{l}\left\{\frac{\partial C(\boldsymbol{\Sigma},\boldsymbol{\Xi})}{\partial \sigma_{m,l}^*} = 0\right\}_{l\in[1,L]} \\ \frac{\partial C(\boldsymbol{\Sigma},\boldsymbol{\Xi})}{\partial \boldsymbol{\varepsilon}_m} = \mathbf{0}\end{array}\right\} \Leftrightarrow \left\{\begin{array}{l}\widehat{\sigma}_{m,l} = \frac{\mathbf{b}_m^H\left(\mathbf{x}^l - \bar{\mathbf{B}}_m\bar{\boldsymbol{\sigma}}_{m,l}\right)}{\mathbf{b}_m^H\mathbf{b}_m} \\ \widehat{\boldsymbol{\varepsilon}}_m = \arg\min_{\boldsymbol{\varepsilon}_m}\left\{\sum_{l=1}^{L}\left\|\mathbf{\Pi}_{\mathbf{b}(\boldsymbol{\varepsilon}_m)}^{\perp}\left(\mathbf{x}^l - \bar{\mathbf{B}}_m\bar{\boldsymbol{\sigma}}_{m,l}\right)\right\|^2\right\} \\ \widehat{\boldsymbol{\varepsilon}}_m = \arg\max_{\boldsymbol{\varepsilon}_m}\left\{\sum_{l=1}^{L}\left\|\mathbf{\Pi}_{\mathbf{b}(\boldsymbol{\varepsilon}_m)}\left(\mathbf{x}^l - \bar{\mathbf{B}}_m\bar{\boldsymbol{\sigma}}_{m,l}\right)\right\|^2\right\}\end{array}\right.,$$

$$\bar{\boldsymbol{\sigma}}_{m,l} = (\sigma_{l,l})_{l\in[1,M],l\neq m}, \quad \mathbf{B} = \mathbf{B}(\boldsymbol{\Xi}) = \left[(\mathbf{b}_m)_{m\in[1,M]}\right], \quad \bar{\mathbf{B}}_m = \left[(\mathbf{b}_l)_{l\in[1,M],l\neq m}\right], \quad \mathbf{b}_m = \mathbf{b}(\boldsymbol{\varepsilon}_m)$$

• ou l'algorithme Alternating Projection [ZW88], basé sur la minimisation itérative de $C(\boldsymbol{\Sigma}, \boldsymbol{\Xi})$ relativement aux paramètres $\{\widehat{\boldsymbol{\varepsilon}}_m\}$ d'une seule source et aux amplitudes de toutes les sources $\{\boldsymbol{\Sigma}\}$:

$$\left.\begin{array}{l}\left\{\frac{\partial C(\boldsymbol{\Sigma},\boldsymbol{\Xi})}{\partial \sigma_{m,l}^*} = 0\right\}_{m\in[1,M],l\in[1,L]} \\ \frac{\partial C(\boldsymbol{\Sigma},\boldsymbol{\Xi})}{\partial \boldsymbol{\varepsilon}_m} = \mathbf{0}\end{array}\right\} \Leftrightarrow \left\{\begin{array}{l}\widehat{\boldsymbol{\sigma}}_l = \left(\mathbf{B}^H\mathbf{B}\right)^{-1}\mathbf{B}^H\mathbf{x}^l \\ \widehat{\boldsymbol{\varepsilon}}_m = \arg\min_{\boldsymbol{\varepsilon}_m}\left\{\sum_{l=1}^{L}\left\|\mathbf{\Pi}_{\mathbf{B}}^{\perp}\mathbf{x}^l\right\|^2\right\} \\ \widehat{\boldsymbol{\varepsilon}}_m = \arg\max_{\boldsymbol{\varepsilon}_m}\left\{\sum_{l=1}^{L}\left\|\mathbf{\Pi}_{\left(\mathbf{\Pi}_{\bar{\mathbf{B}}_m}^{\perp}\mathbf{b}(\boldsymbol{\varepsilon}_m)\right)}\mathbf{x}^l\right\|^2\right\} \\ \mathbf{B} = \mathbf{B}(\boldsymbol{\Xi}) = \left[(\mathbf{b}_m)_{m\in[1,M]}\right], \quad \bar{\mathbf{B}}_m = \left[(\mathbf{b}_l)_{l\in[1,M],l\neq m}\right], \quad \mathbf{b}_m = \mathbf{b}(\boldsymbol{\varepsilon}_m)\end{array}\right.$$

Les principaux avantages de cette approche est sa généralité (aucune hypothèse particulière sur la corrélation des sources), sa facilité de mise en oeuvre et sa vitesse de convergence (minimisation monoparamètre à chaque itération). Sa principale limitation est la convergence itérative vers des minima locaux qui ne convergent pas nécessairement vers le minimum global. Cependant, il est possible d'améliorer la probabilité de convergence vers le minimum global, en généralisant les deux techniques précédentes à une descente itérative multiparamètres (avec une augmentation

du temps de convergence en contrepartie).
Enfin notons que par construction, la valeur du critère $C(\boldsymbol{\Sigma}, \boldsymbol{\Xi})$ obtenue après utilisation de l'algorithme Alternating Projection à partir d'un couple $(\boldsymbol{\Sigma}, \boldsymbol{\Xi})$ donné est toujours $\leq$ à la valeur du critère $C(\boldsymbol{\Sigma}, \boldsymbol{\Xi})$ obtenue après utilisation des algorithmes Clean/Clean Relax à partir du même couple $(\boldsymbol{\Sigma}, \boldsymbol{\Xi})$. Par contre ce résultat ne peut être extrapolé au minimum local obtenu au final, car les chemins de descente suivis par les 2 types d'algorithme ne sont a priori pas les mêmes.

Cas particulier à 2 sources et 1 paramètre inconnu par source

Lorsque $M = 2$ et $P = 1$, alors [ZW88] :

$$\left\|\boldsymbol{\Pi}_{\mathbf{B}}\mathbf{x}^l\right\|^2 = \left\|\boldsymbol{\Pi}_{\mathbf{b}_1}\mathbf{x}^l\right\|^2 + \left\|\boldsymbol{\Pi}_{\left(\boldsymbol{\Pi}^{\perp}_{\mathbf{b}_1}\mathbf{b}_2\right)}\mathbf{x}^l\right\|^2, \quad \mathbf{B} \triangleq \mathbf{B}(\boldsymbol{\Xi}) = [\mathbf{b}_1, \mathbf{b}_2], \; \mathbf{b}_1 = \mathbf{b}(\varepsilon_1), \; \mathbf{b}_2 = \mathbf{b}(\varepsilon_2)$$

soit encore :

$$tr\left(\boldsymbol{\Pi}_{\mathbf{B}}\widehat{\mathbf{R}}_{\mathbf{x}}\right) = tr\left(\boldsymbol{\Pi}_{\mathbf{b}_1}\widehat{\mathbf{R}}_{\mathbf{x}}\right) + tr\left(\boldsymbol{\Pi}_{\left(\boldsymbol{\Pi}^{\perp}_{\mathbf{b}_1}\mathbf{b}_2\right)}\widehat{\mathbf{R}}_{\mathbf{x}}\right)$$

Si de plus $\mathbf{b}^H(\varepsilon_1)\,\mathbf{b}(\varepsilon_2) = c(\Delta\varepsilon)\,, \Delta\varepsilon = \varepsilon_2 - \varepsilon_1$ alors :

$$tr\left(\boldsymbol{\Pi}_{\mathbf{B}}\widehat{\mathbf{R}}_{\mathbf{x}}\right) = tr\left(\boldsymbol{\Pi}_{\mathbf{b}_1}\widehat{\mathbf{R}}_{\mathbf{x}}\right) + C\left(\varepsilon_1, \Delta\varepsilon, \widehat{\mathbf{R}}_{\mathbf{x}}\right), \quad C\left(\varepsilon_1, \Delta\varepsilon, \widehat{\mathbf{R}}_{\mathbf{x}}\right) = tr\left(\boldsymbol{\Pi}_{\left(\boldsymbol{\Pi}^{\perp}_{\mathbf{b}_1}\mathbf{b}_2\right)}\widehat{\mathbf{R}}_{\mathbf{x}}\right)$$

ce qui permet de reparamétrer la recherche de $\widehat{\boldsymbol{\Xi}}$:

$$\widehat{\boldsymbol{\Xi}} = \arg\max_{\boldsymbol{\Xi}}\left\{tr\left(\boldsymbol{\Pi}_{\mathbf{B}(\boldsymbol{\Xi})}\widehat{\mathbf{R}}_{\mathbf{x}}\right)\right\} \quad \Leftrightarrow \quad \left(\widehat{\varepsilon_1, \Delta\varepsilon}\right) = \arg\max_{(\varepsilon_1, \Delta\varepsilon)}\left\{tr\left(\boldsymbol{\Pi}_{\mathbf{b}(\varepsilon_1)}\widehat{\mathbf{R}}_{\mathbf{x}}\right) + C\left(\varepsilon_1, \Delta\varepsilon, \widehat{\mathbf{R}}_{\mathbf{x}}\right)\right\}$$

soit encore :

$$\left(\widehat{\varepsilon_1, \Delta\varepsilon}\right) = \arg\max_{\varepsilon_1}\left\{tr\left(\boldsymbol{\Pi}_{\mathbf{b}(\varepsilon_1)}\widehat{\mathbf{R}}_{\mathbf{x}}\right) + \max_{\Delta\varepsilon}\left\{C\left(\varepsilon_1, \Delta\varepsilon, \widehat{\mathbf{R}}_{\mathbf{x}}\right)\right\}\right\}$$

En général, si $\Delta\varepsilon \approx 0$ alors on sait résoudre (par d.l. au 1er ordre) :

$$\widehat{\Delta\varepsilon} = \max_{\Delta\varepsilon}\left\{C\left(\varepsilon_1, \Delta\varepsilon, \widehat{\mathbf{R}}_{\mathbf{x}}\right)\right\} \approx \widehat{\Delta\varepsilon}\left(\varepsilon_1, \widehat{\mathbf{R}}_{\mathbf{x}}\right)$$

et ramener la recherche sur $\left(\widehat{\varepsilon_1, \Delta\varepsilon}\right)$ à la recherche sur $\widehat{\varepsilon_1}$:

$$\widehat{\varepsilon_1} = \arg\max_{\varepsilon_1}\left\{tr\left(\boldsymbol{\Pi}_{\mathbf{b}(\varepsilon_1)}\widehat{\mathbf{R}}_{\mathbf{x}}\right) + C\left(\varepsilon_1, \widehat{\Delta\varepsilon}\left(\varepsilon_1, \widehat{\mathbf{R}}_{\mathbf{x}}\right), \widehat{\mathbf{R}}_{\mathbf{x}}\right)\right\},$$

ce qui permet d'accélérer notablement la procédure de recherche de l'EMV $\widehat{\boldsymbol{\Xi}}$ comme nous le proposons dans [J14] [C29].

C. Eléments de la théorie de la détection. Test d'hypothèses binaire.

Nous considérons le cadre général du test d'hypothèses binaire dont un cas particulier est celui qui nous intéresse, à savoir le cas modélisant un signal d'intérêt intermittent qui n'est pas toujours présent dans les observations $\mathbf{x}$:

$$\begin{array}{rcl} H_0 \text{ (environnement seul)} & : & \mathbf{x} = \mathbf{n} \\ H_1 \text{ (environnement et signal)} & : & \mathbf{x} = \mathbf{n} + \mathbf{s} \end{array}$$

Il s'agit alors de savoir si le signal observé contient (H_1) ou non (H_0) le signal d'intérêt $\mathbf{s}$ superposé au signal $\mathbf{n}$ provenant de l'environnement permanent.

1) Règles optimales de détection:

Le problème de la détermination de l'hypothèse observée provient du fait que généralement les ensembles $\Omega_{H_0} = \{\mathbf{x} = \mathbf{n}\}$ et $\Omega_{H_1} = \{\mathbf{x} = \mathbf{n} + \mathbf{s}\}$ ne sont pas disjoints $(\Omega_{H_0} \cap \Omega_{H_1} \neq \varnothing)$, mais confondus $\Omega_{H_0} = \Omega_{H_1} = \Omega = \mathbb{C}^M$. Dans ce cas, puisque toute valeur observée $\mathbf{x}$ peut provenir de l'un où l'autre des modèles d'observation, il devient nécessaire de se doter d'une règle de décision. Cette règle consiste à partitionner l'ensemble des observations Ω en deux sous-ensembles $\mathcal{D}_0$ et $\mathcal{D}_1$ complémentaires $(\mathcal{D}_0 \cap \mathcal{D}_1 = \varnothing)$ associés respectivement à une hypothèse de modèle d'observation : $\mathcal{D}_0 \to H_0$ et $\mathcal{D}_1 \to H_1$. La pertinence de la règle de décision (choix de $\mathcal{D}_0$ et de $\mathcal{D}_1$) dépend de l'information probabiliste connue pour chaque modèle d'observation, notamment l'expression analytique des densités de probabilité conditionnelles : $p(\mathbf{x} \mid H_0)$ et $p(\mathbf{x} \mid H_1)$. En effet dans ce cas, certaines stratégies de décision (règles de décision) peuvent être mises en oeuvre afin de minimiser le risque d'erreur dans le choix du modèle (hypothèse) observé. Par exemple, si la probabilité que chaque modèle d'observation se réalise $P(H_0)$ et $P(H_1)$ (probabilités *a priori*) est également connue, la probabilité de commettre une erreur de décision induite par une règle donnée ($\mathcal{D}_0$ et $\mathcal{D}_1$) - critère de Bayes [VT68, p. 30] - est calculable :

$$P_{erreur} = P((\mathbf{x} \in \mathcal{D}_0) \cap H_1) + P((\mathbf{x} \in \mathcal{D}_1) \cap H_0) = \int_{\mathcal{D}_0} p(\mathbf{x} \mid H_1)\, d\mathbf{x} P(H_1) + \int_{\mathcal{D}_1} p(\mathbf{x} \mid H_0)\, d\mathbf{x} P(H_0) \quad (134)$$

et sa minimisation [VT68, p. 30] est obtenue par la "règle de Bayes" :

$$\frac{p(\mathbf{x} \mid H_1)}{p(\mathbf{x} \mid H_0)} \underset{H_1}{\overset{H_0}{\lessgtr}} \frac{P(H_0)}{P(H_1)} \quad (135)$$

Malheureusement, dans la plupart des cas pratiques, la présence du signal d'intérêt n'est pas prévisible statistiquement : $P(H_0)$ et $P(H_1)$ ne sont pas a priori accessibles et la règle de Bayes (135) est inapplicable. Néanmoins, il est possible de définir un nouveau critère sous-optimal basé sur une analyse opérationnelle. Dans ce cadre, la présence d'un signal d'intérêt déclenche la mobilisation d'une ressource, cette ressource étant disponible en quantité limitée. Une gestion parcimonieuse de la ressource doit veiller à éviter son "gaspillage" sur des erreurs de décision sur environnement seul (H_0) - fausses alarmes - et rendre la *probabilité de fausses alarmes :*

$$P_{FA} = P(\mathbf{x} \in \mathcal{D}_1 \mid H_0) = \int_{\mathcal{D}_1} p(\mathbf{x} \mid H_0)\, d\mathbf{x} \quad (136)$$

aussi petite que possible. Une solution triviale est $P_{FA} = 0 \Leftrightarrow \mathcal{D}_1 = \varnothing$. Le problème d'une telle solution est qu'elle ne permet pas non plus de détecter la présence du signal d'intérêt lorsqu'il est présent (H_1), car alors la *probabilité de détection :*

$$P_{\mathcal{D}} = P(\mathbf{x} \in \mathcal{D}_1 \mid H_1) = \int_{\mathcal{D}_1} p(\mathbf{x} \mid H_1)\, d\mathbf{x} \quad (137)$$

est également nulle. Il s'agit donc de trouver le meilleur compromis en terme de perte de ressource acceptable sur fausse alarme (P_{FA}) et d'allocation optimale de ressource lorsque le signal d'intérêt est présent (P_D), compromis formulé à travers le critère de Neyman-Pearson [VT68, p. 33] :

$$\max\{P_{\mathcal{D}}\} \text{ pour une } P_{FA} = (P_{FA})_0 \quad (138)$$

dont la minimisation [VT68, p. 30] est obtenue par la "règle de Neyman-Pearson" :

$$\frac{p(\mathbf{x} \mid H_1)}{p(\mathbf{x} \mid H_0)} \underset{H_1}{\overset{H_0}{\lessgtr}} \lambda \quad (139)$$

Finalement, dans le cas du test d'hypothèses binaires, les règles de détection optimale (135) ou sous-optimale (139) se présente sous la forme commune suivante :

$$RV(\mathbf{x}) = \frac{p(\mathbf{x} \mid H_1)}{p(\mathbf{x} \mid H_0)} \underset{H_1}{\overset{H_0}{\lessgtr}} T \quad (140)$$

où $RV(\mathbf{x})$ est le rapport de vraisemblance (RV), T représente un seuil de décision (ou de détection) et (140) est le test du rapport de vraisemblance (TRV).

2) Test d'hypothèses composites:

En général, les différentes hypothèses de modèle d'observation dépendent de paramètres déterministes inconnus :

$$\begin{array}{lcll} H_0 \text{ (environnement seul)} & : & \mathbf{x} = \mathbf{x}(\boldsymbol{\theta}_{H_0}) = \mathbf{n}(\boldsymbol{\theta}_{\mathbf{n}}), & \boldsymbol{\theta}_{H_0} \equiv \boldsymbol{\theta}_{\mathbf{n}} \\ H_1 \text{ (environnement et signal)} & : & \mathbf{x} = \mathbf{x}(\boldsymbol{\theta}_{H_1}) = \mathbf{n}(\boldsymbol{\theta}_{\mathbf{n}}) + \mathbf{s}(\boldsymbol{\theta}_{\mathbf{s}}), & \boldsymbol{\theta}_{H_1}^T \equiv \left(\boldsymbol{\theta}_{\mathbf{s}}^T, \boldsymbol{\theta}_{\mathbf{n}}^T\right)^T \end{array}$$

et la règle de décision optimale (140) est a priori fonction du vecteur de paramètres inconnus $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_{H_0}^T, \boldsymbol{\theta}_{H1}^T\right)^T$:

$$RV(\mathbf{x};\boldsymbol{\theta}) = \frac{p(\mathbf{x} \mid H_1; \boldsymbol{\theta}_{H_1})}{p(\mathbf{x} \mid H_0; \boldsymbol{\theta}_{H_0})}, \tag{141}$$

ce qui définit le problème du test d'hypothèses composites (Composite Hypothesis Testing Problem [VT68]). En effet, si les paramètres inconnus $\boldsymbol{\theta}$ ne peuvent être éliminé du RV (141), alors la règle de décision optimale est clairvoyante; ce qui est le cas en général. La transformation de la règle clairvoyante (141) en règle réalisable est obtenue en remplaçant, sous chacune des hypothèses, les paramètres inconnus par un estimateur, ce qui conduit à un estimateur du RV (141) :

$$\widehat{RV(\mathbf{x};\boldsymbol{\theta})}(\mathbf{x}) = \frac{p\left(\mathbf{x} \mid H_1; \widehat{\boldsymbol{\theta}_{H_1}}(\mathbf{x})\right)}{p\left(\mathbf{x} \mid H_0; \widehat{\boldsymbol{\theta}_{H_0}}(\mathbf{x})\right)} \tag{142}$$

Il semble intuitivement assez évident de chercher, sous chaque hypothèse, des estimateurs $\widehat{\boldsymbol{\theta}_{H_0}}(\mathbf{x})$ et $\widehat{\boldsymbol{\theta}_{H_1}}(\mathbf{x})$ aussi précis que possible afin que le test réalisable ainsi construit définisse une règle de décision aussi proche que possible de la règle optimale désirée :

$$\widehat{RV(\mathbf{x};\boldsymbol{\theta})}(\mathbf{x}) = \frac{p\left(\mathbf{x} \mid H_1; \widehat{\boldsymbol{\theta}_{H_1}}(\mathbf{x})\right)}{p\left(\mathbf{x} \mid H_0; \widehat{\boldsymbol{\theta}_{H_0}}(\mathbf{x})\right)} \approx \frac{p(\mathbf{x} \mid H_1; \boldsymbol{\theta}_{H_1})}{p(\mathbf{x} \mid H_0; \boldsymbol{\theta}_{H_0})} = RV(\mathbf{x};\boldsymbol{\theta})$$

Malheureusement, l'inexistence d'une stratégie générale d'estimation optimale conduisant à des estimateurs réalisables (cf. § IV-B précédent) rend le choix de l'estimateur du paramètre inconnu $\boldsymbol{\theta}$ sous chaque hypothèse a priori arbitraire, si ce n'est à prendre en considération certaines propriétés asymptotiques (en le nombre d'observations ou en le rapport signal à bruit) en général non vérifiées par le problème courant. C'est le cas de la détermination de $\widehat{\boldsymbol{\theta}_{H_0}}(\mathbf{x})$ et $\widehat{\boldsymbol{\theta}_{H_1}}(\mathbf{x})$ au sens du maximum de vraisemblance, le RV devenant alors le rapport de vraisemblance généralisé (RVG, "Generalized Likelihood Ratio - GLR") conduisant au test du rapport de vraisemblance généralisé (TRVG, "GLRT") :

$$\frac{p\left(\mathbf{x} \mid H_1; \widehat{\boldsymbol{\theta}_{H_1 MV}}(\mathbf{x})\right)}{p\left(\mathbf{x} \mid H_0; \widehat{\boldsymbol{\theta}_{H_0 MV}}(\mathbf{x})\right)} \underset{H_1}{\overset{H_0}{\lessgtr}} T \tag{143}$$

D. Notations

- a, les lettres en italiques représentent une quantité scalaire.
- $\mathbf{a}$, les lettres minuscules en gras représentent une quantité vectorielle (vecteur colonne).
- $\mathbf{A}$, les lettre majuscules en gras représentent une quantité matricielle.
- a_n ou $(\mathbf{a})_n$ est l'élément correspondant à la $n^{ème}$ coordonnée du vecteur $\mathbf{a}$.
- $A_{n,m}$ ou $(\mathbf{A})_{n,m}$ est l'élément correspondant à la $n^{ème}$ ligne et la $m^{ème}$ colonne de la matrice $\mathbf{A}$.
- $\mathrm{Re}\{a\}$ est la partie réelle de a.
- $\mathrm{Im}\{a\}$ est la partie imaginaire de a.
- $\mathbf{A}^T$, le symbole ^{T} indique l'opérateur transposé.
- $\mathbf{A}^*$, le symbole * indique l'opérateur de conjugaison.
- $\mathbf{A}^H$, le symbole ^{H} indique l'opérateur Hermitien (transposé conjugué).
- $|\mathbf{A}|$ est le déterminant de la matrice $\mathbf{A}$.
- $tr(\mathbf{A})$ est la trace de la matrice carrée $\mathbf{A}$.

- $\mathbf{Diag}\,(\mathbf{a})$ représente la matrice diagonale telle que $(\mathbf{Diag}\,(\mathbf{a}))_{ii} = (\mathbf{a})_i$.
- $[\mathbf{A}, \mathbf{B}]$ représente la matrice résultante de la concaténation horizontale des deux matrices $\mathbf{A}$ et $\mathbf{B}$.
- $\mathbf{I}_N$ représente la matrice identité de taille $N \times N$.
- $\mathbf{1}_N$ est le vecteur de taille $N \times 1$ défini par : $\mathbf{1}_N = (1, \ldots, 1)^T$.
- $\mathbf{vec}\,(\mathbf{A})$ est l'opérateur de vectorisation de la matrice $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_N]$: $\mathbf{vec}\,(\mathbf{A})^T = \left(\mathbf{a}_1^T, \ldots, \mathbf{a}_N^T\right)$.
- $im\,\{\mathbf{A}\}$ est l'image de la matrice $\mathbf{A}$, i.e. le sous-espace vectoriel engendré par les colonnes de la matrice $\mathbf{A}$.
- $\ker\{\mathbf{A}\}$ est le noyau de la matrice $\mathbf{A}$, i.e. le sous-espace vectoriel engendré par les vecteurs $\mathbf{x}$ verifiant $\mathbf{Ax} = \mathbf{0}$.
- $S^{\perp}$ est le sous-espace vectoriel orthogonal au sous-espace vectoriel S.
- $\mathbf{A} \geq \mathbf{B}$, signifie que la matrice $\mathbf{A} - \mathbf{B}$ est non-négative.
- $E\,[\cdot]$ représente l'espérance mathématique.
- $\langle\ \mid\ \rangle$ et $\|.\|$ représentent respectivement un produit scalaire et sa norme associée.
- Si $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_P)^T$, alors: $\frac{\partial}{\partial \boldsymbol{\theta}} = \left(\frac{\partial}{\partial \theta_1}, \frac{\partial}{\partial \theta_2}, \ldots, \frac{\partial}{\partial \theta_P}\right)^T$, $\frac{\partial}{\partial \boldsymbol{\theta}^T} = \left(\frac{\partial}{\partial \theta_1}, \frac{\partial}{\partial \theta_2}, \ldots, \frac{\partial}{\partial \theta_P}\right)$.
- $\frac{\partial \mathbf{f}(\theta_0)}{\partial \theta} = \left.\frac{\partial \mathbf{f}(\theta)}{\partial \theta}\right|_{\theta=\theta_0}$, $\frac{\partial f(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = \left.\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$, $\frac{\partial \mathbf{f}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^T} = \left.\frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$
- $\mathcal{M}_{\mathbb{R}}\,(N, P)$ représente l'ensemble des matrices réelles à N lignes et P colonnes.
- $\mathcal{M}_{\mathbb{C}}\,(N, P)$ représente l'ensemble des matrices complexes à N lignes et P colonnes.
- $\odot$ désigne le produit d'Hadamard.
- $\otimes$ désigne le produit de Kronecker.
- $1\,(\mathbf{x})$ représente la fonction du vecteur $\mathbf{x}$ constante et égale à 1.
- $im\,\{\mathbf{f}\,(\mathbf{x})\}$ est l'image de la fonction vectorielle de la variable vectorielle $\mathbf{x}$.
- $\underline{\mathbf{u}}$ représente un des trois cas suivants :

$$\underline{\mathbf{u}} : \begin{cases} \underline{\mathbf{u}} = \mathbf{u} & \text{si } \mathbf{u} \in \mathbb{R}^I \\ \underline{\mathbf{u}} = \left(\mathbf{u}^T, \mathbf{u}^H\right)^T & \text{si } \mathbf{u} \in \mathbb{C}^I \text{ et } \mathbf{u} \notin \mathbb{R}^I \\ \underline{\mathbf{u}} = \left(\mathbf{u}_c^T, \mathbf{u}_c^H, \mathbf{u}_r^T\right)^T & \text{si } \mathbf{u} = \left(\mathbf{u}_c^T, \mathbf{u}_r^T\right)^T, \quad \mathbf{u}_c \in \mathbb{C}^I \text{ et } \mathbf{u}_c \notin \mathbb{R}^I, \quad \mathbf{u}_r \in \mathbb{R}^{I'} \end{cases} \tag{144}$$

Nous adoptons la définition du produit hermitien utilisée dans les livres de mathématiques [Bou74] [Haz02] [HJ99] où une forme sesquilinéaire est une fonction de deux variables de l'espace vectoriel complexe $\mathbb{U}$, qui est linéaire en la première variable et semi-linéaire en la deuxième :

$$\langle\ \mid\ \rangle : \mathbb{U} \times \mathbb{U} \to \mathbb{C} \ / \ \forall \mathbf{u}, \mathbf{c} \in \mathbb{U},\ \forall \lambda \in \mathbb{C},\ \ \langle \lambda \mathbf{u} \mid \mathbf{c} \rangle = \lambda \langle \mathbf{u} \mid \mathbf{c} \rangle,\ \ \langle \mathbf{u} \mid \lambda \mathbf{c} \rangle = \lambda^* \langle \mathbf{u} \mid \mathbf{c} \rangle \tag{145}$$

Cette notation permet de définir la matrice de Gram associée à deux familles de vecteurs de l'espace vectoriel $\mathbb{U}$, $\{\mathbf{u}_q\}_1^Q = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_Q\}$ et $\{\mathbf{c}_i\}_1^I = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_I\}$ [HJ99] :

$$\mathbf{G}\left(\{\mathbf{u}_q\}_1^Q, \{\mathbf{c}_i\}_1^I\right) = \begin{bmatrix} \langle \mathbf{u}_1 \mid \mathbf{c}_1 \rangle & \ldots & \langle \mathbf{u}_Q \mid \mathbf{c}_1 \rangle \\ \vdots & \ldots & \vdots \\ \langle \mathbf{u}_1 \mid \mathbf{c}_I \rangle & \ldots & \langle \mathbf{u}_Q \mid \mathbf{c}_I \rangle \end{bmatrix} \in \mathcal{M}_{\mathbb{C}}\,(I, Q) \tag{146}$$

$$\left(\mathbf{G}\left(\{\mathbf{u}_q\}_1^Q, \{\mathbf{c}_i\}_1^I\right)\right)_{i,q} = \langle \mathbf{u}_q \mid \mathbf{c}_i \rangle, \qquad \mathbf{G}\left(\{\mathbf{c}_i\}_1^I, \{\mathbf{u}_q\}_1^Q\right) = \mathbf{G}\left(\{\mathbf{u}_q\}_1^Q, \{\mathbf{c}_i\}_1^I\right)^H$$

définition conduisant à :

$$\left\langle \sum_{q=1}^{Q} x_q \mathbf{u}_q \mid \sum_{i=1}^{I} y_i \mathbf{c}_i \right\rangle = \mathbf{y}^H \mathbf{G}\left(\{\mathbf{u}_q\}_1^Q, \{\mathbf{c}_i\}_1^I\right) \mathbf{x}, \quad \mathbf{x} = (x_1, \ldots, x_Q)^T, \mathbf{y} = (y_1, \ldots, y_I)^T. \tag{147}$$

Pour simplifier la notation, nous écrivons :

$$\mathbf{G}\left(\{\mathbf{u}_q\}_1^Q\right) \triangleq \mathbf{G}\left(\{\mathbf{u}_q\}_1^Q, \{\mathbf{u}_q\}_1^Q\right) \tag{148}$$

Dans le cas où $\mathbb{U}$ est un espace vectoriel réel, la forme sesquilinéaire (145) devient une forme bilinéaire symétrique et les définitions (146)(147) restent inchangées : il suffit de remplacer l'opérateur Hermitien ^{H} par l'opérateur transposé ^{T}.

E. On the influence of a detection step on lower bounds for deterministic parameter estimation (IEEE TSP)

IEEE Trans. on SP, 53(11): 4080-4090, 2005.

On the Influence of a Detection Step on Lower Bounds for Deterministic Parameter Estimation.

Eric Chaumette, Pascal Larzabal and Philippe Forster

Abstract

A wide variety of actual processing requires a detection step, whose main effect is to restrict the set of observations available for parameter estimation. Therefore, as a contribution to the theoretical formulation of the joint detection and estimation problem, we address the derivation of lower bounds for deterministic parameters conditioned by a binary hypothesis testing problem. The main result is the introduction of a general scheme - detailed in the particular case of CRB - enabling the derivation of conditional deterministic MSE lower bounds. To prove that it is meaningful, we also show, with the help of a fundamental application, that the problem of lower bound tightness at low SNR may arise from an incorrect lower bound formulation that does not take into account the true nature of the problem under investigation: a joint detection-estimation problem.

Index Terms: Deterministic parameter estimation, joint detection and estimation problem, MSE conditional lower bounds, array processing, Monopulse DOA estimation

I. Notation

$P()$ denotes a probability
$P(A \mid B)$ denotes a conditional probability
$f_\theta(x)$ denotes a probability density function depending on parameter θ
$f_\theta(x \mid y)$ denotes a conditional probability density function
$\frac{\partial f_{\theta_0}(x)}{\partial \theta} = \left[\frac{\partial f_\theta(x)}{\partial \theta}\right]_{\theta=\theta_0}$
If $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_L)^T$, then: $\frac{\partial}{\partial \boldsymbol{\theta}} = \left(\frac{\partial}{\partial \theta_1}, \frac{\partial}{\partial \theta_2}, \ldots, \frac{\partial}{\partial \theta_L}\right)^T, \quad \frac{\partial}{\partial \boldsymbol{\theta}^T} = \left(\frac{\partial}{\partial \theta_1}, \frac{\partial}{\partial \theta_2}, \ldots, \frac{\partial}{\partial \theta_L}\right)$
$\widehat{r}$ denotes an estimator of r
$\mathbf{Id}_I$ denotes Identity matrix with dimensions (I,I)
If $z = x + jy$ is a complex variable, then: $\int h(z)\, dz = \iint h(x,y)\, dxdy$
If $\mathbf{x}$ is a K-dimensional vector, then: $\int_D h(\mathbf{x})\, d\mathbf{x} = \int_D \cdots \int h(x_1, \ldots, x_K)\, dx_1 \ldots dx_K$
$o(x^n) = x^n \varepsilon(x)$ where $\lim_{x \to 0} \varepsilon(x) = 0$
δ_l^k: is 1 if $k = l$, 0 if $k \neq l$

II. Introduction

Lower bounds on the minimum mean square error (MSE) in estimating a set of parameters from noisy observations provide the best performance of any estimators in terms of the MSE. Originally they were introduced [1, p. 473] to investigate fundamental limits of a parameter estimation problem or to assess the relative performance of a specific estimator. They also have been widely used since as a mean to assess the exact MSE of Maximum Likelihood Estimators (MLE) for problems where it is difficult to evaluate. They can be divided in two families [2, p. 52]. The first family treats the set of parameters as an unknown deterministic quantity, and provides bounds on the MSE in estimating any selected values of the parameters ("locally" best estimators). The second family assumes that the

E. Chaumette is with Thales Naval France, Bagneux, France. E-mail: eric.chaumette@noos.fr.
P. Larzabal is with SATIE/CNRS, Ecole Normale Superieure de Cachan, Cachan, France. E-mail: larzabal@satie.ens-cachan.fr.
P. Forster is with GEA, Pôle scientifique et technologique de Paris 10, Ville d'Avray, France. E-mail: philippe.forster@cva.u-paris10.fr.

parameters are random variables with known *a priori* distributions. In this paper, we will focus on the first family, i.e. deterministic parameter estimation.

We are primarily interested in lower bounds for signal parameter estimation problems in which the source signal is a function of unknown deterministic parameters and is embedded in nuisance signals that are also functions of unknown deterministic parameters. In these problems, the Cramer-Rao bound (CRB) [1, p. 475] is the most popular bound, because it is generally the simplest to compute, and in many cases it can be achieved asymptotically by the MLE of parameters, where asymptotically means either with respect to the number of independent snapshots [1, p. 500] and/or with respect to the signal-to-noise ratio (SNR) [3]. Another attractive feature is to be the lowest lower bound on MSE of unbiased estimators, since it derives from the weakest formulation of unbiasness at the vicinity of any selected value of the parameters [4][5]. At the opposite, the highest lower bound on MSE of unbiased estimators has been derived by Barankin [4][6][7] who introduced and studied the strongest formulation of unbiasness, that is to say unbiasness over an interval of parameter values including the selected value. Additionally, numerous works ([4]-[8] and references in [5] and [8]) devoted to the placing of bounds on MSE have shown that CRB and Barankin bound (BB) can be regarded as key representative of two general classes of bounds, respectively the Small-Error bounds and the Large-Error bounds. Indeed, in non-linear estimation problems three distinct regions of operation can be observed. In the asymptotic region, the MSE is small and, in many cases, close to the Small-Error bounds. In the *a priori* performance region where the number of independent snapshots and/or the SNR are very low, the observations provide little information and the MSE is close to that obtained from the prior knowledge about the problem [5][8][10]. Between these two extremes, there is an additional ambiguity region, also called the transition region. In this region, the MSE of MLEs usually deteriorates rapidly with respect to Small-Error bounds and exhibits a threshold behavior corresponding to a "performance breakdown" highlighted by Large-Error bounds [5][8][9][10]. The nature of this phenomenon is specified by a complicated non-smooth behaviour of the likelihood function in the "threshold" area where it tends to generate outliers [4][11].

However in nearly all fields of science and engineering, a wide variety of processing requires a binary detection step designed to decide if a signal is present or not in noise. Intuitively, such a detection step is expected to improve the lower bounds tightness by selecting instances with relatively high signal energy - sufficient to exceed the detection threshold - and disregarding instances belonging to the *a priori* region that deteriorate the MSE. Additionally, as a detection step restricts the set of observations available for parameter estimation, any accurate MSE lower bound should take this statistical conditioning into account, which may appear as a major difficulty. Indeed two problems have to be solved: calculation of the conditional p.d.f. and derivation of conditional lower bound. Fortunately, combining previous results from [7] and [10], derivation of all existing bounds with statistical conditioning is possible for detection step involved in actual applications (see §III-C). The present paper introduces thus a new contribution to the theoretical characterization of the joint detection and estimation problem, since this problem is seldom covered in the open literature - including reference books [2][12][13][14] - where detection and estimation performance are treated as separable problems.

The paper is organized as follows. Section III is tutorial in nature. It presents a general scheme for conditional lower bounds derivation based on a didactic approach of wide scope that is illustrated in CRB case for single or multiple real parameters. Section IV introduces a fundamental application - estimation of the steering vector of a 2 sensors array - that allows a preliminary analysis to be presented on the expected effect of observation selection on lower bounds tightness.

III. Derivation of Conditional Lower Bounds

The present contribution to characterization of the joint detection and estimation problem for deterministic parameters, although relatively simple using the approach developed in [7][10], is new, as far as we know.

A. Background on lower bounds as a norm minimization problem

Let $\mathbf{x}$ be the random observations vector and Ω be the observation space. Denote by $f_{\theta}(\mathbf{x})$ the p.d.f. of observations depending on an unknown deterministic real parameter θ to be estimated. Let $\mathcal{F}_{\Omega}$ be the real vector space of square integrable functions over Ω. The MSE of a particular estimator $\widehat{\theta_0}(\mathbf{x})$ of θ_0, where θ_0 is a selected value of the parameter θ, is defined by:

$$MSE_{\theta_0} = E_{\theta_0}\left[\left(\widehat{\theta_0}(\mathbf{x}) - \theta_0\right)^2\right] = \int\limits_{\Omega} \left(\widehat{\theta_0}(\mathbf{x}) - \theta_0\right)^2 f_{\theta_0}(\mathbf{x})\, d\mathbf{x} \tag{1}$$

and can be rewritten as [7]:

$$MSE_{\theta_0} = \left\|\widehat{\theta_0}(\mathbf{x}) - \theta_0\right\|^2_{\theta_0} \tag{2}$$

if we consider the following scalar product defined over $\mathcal{F}_\Omega$:

$$\langle g \mid h\rangle_{\theta_0} = E_{\theta_0}\left[g(\mathbf{x})\, h(\mathbf{x})\right] = \int\limits_{\Omega} \left[g(\mathbf{x})\, h(\mathbf{x})\right] f_{\theta_0}(\mathbf{x})\, d\mathbf{x} \tag{3}$$

Therefore the search for a lower bound on the MSE is a norm minimization problem over a vector space of square integrable functions (same approach in [10] with a slightly different scalar product). To avoid the trivial solution $MSE_{\theta_0} = 0 \Longleftrightarrow \widehat{\theta_0}(\mathbf{x}) = \theta_0$, some constraints must be added. The simplest are linear constraints, for which the following result exists:

Minimization Lemma Let $\mathcal{F}$ be a vectorial space of any dimension (finite or infinite) on the field of complex numbers $\mathbb{C}$. Let there be a Hermitian product (positive bilinear Hermitian form) defined on $\mathcal{F}$ and denoted $\langle \mathbf{u} \mid \mathbf{v}\rangle$, if $\mathbf{u}$ and $\mathbf{v}$ are two vectors of $\mathcal{F}$. Let $(\mathbf{c}_1, \ldots, \mathbf{c}_K)$ be a family of K independent vectors of $\mathcal{F}$ and $\mathbf{f} = (f_1, \ldots, f_K)^T$ a vector of K complex values. Then a straightforward exercise in Hermitian product establishes that the solution of minimization of $\langle \mathbf{u} \mid \mathbf{u}\rangle$ under the P linear constraints $\langle \mathbf{u} \mid \mathbf{c}_k\rangle = f_k$, $k \in [1, K]$ is:

$$\min\left\{\langle \mathbf{u} \mid \mathbf{u}\rangle\right\} = \mathbf{f}^H \mathbf{G}^{-1}\mathbf{f} \quad \text{for} \quad \mathbf{u}_{opt} = \sum_{k=1}^{K} \alpha_k \mathbf{c}_k \tag{4a}$$

$$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)^T = \mathbf{G}^{-1}\mathbf{f}, \quad \mathbf{G}_{k,l} = \langle \mathbf{c}_l \mid \mathbf{c}_k\rangle, \quad k, l \in [1, K] \tag{4b}$$

where $\mathbf{f}^H$ is the transpose conjugate of complex vector $\mathbf{f}$. Note that expressions (4a-b) are valid for vectorial space on the field of real numbers $\mathbb{R}$, provided $\langle \mathbf{u} \mid \mathbf{v}\rangle$ stands for a scalar product (positive bilinear symmetric form).

Applied to MSE minimization, this result allows the introduction of a didactic approach highlighting the assumptions associated with different bounds. Indeed, [7] and [10] have shown in the real single parameter case, that all known bounds on the MSE – Cramer-Rao, Bhattacharya, Barankin, Hammersley-Chapman-Robbins and Abel bounds – are different solutions of the same norm minimization problem (2) under sets of appropriate linear constraints (possibly infinite but countable).

As an example of lower bound derivation using this remarkable approach, we propose in next section a derivation of CRB that provides an unpublished introduction to the necessity of constraints and to the various unbiasness criterions. Additionally, it also provides an example of its extension to the multiple parameters case that is required for the application investigated in section IV.

B. Application to CRB derivation

1) The single real parameter case: The problem under investigation is the search for an estimator $\widehat{\theta_0}(\mathbf{x})_{opt} \in \mathcal{F}_\Omega$ of θ_0 minimizing the MSE (1). The equality

$$E_{\theta_0}\left[\left(\widehat{\theta_0}(\mathbf{x}) - \theta_0\right)^2\right] = Var_{\theta_0}\left[\widehat{\theta_0}(\mathbf{x})\right] + \left(E_{\theta_0}\left[\widehat{\theta_0}(\mathbf{x})\right] - \theta_0\right)^2 \tag{5}$$

suggests the introduction of the constraint:

$$E_{\theta_0}\left[\widehat{\theta_0}(\mathbf{x})\right] = \theta_0 \tag{6}$$

Indeed, the variance of a random variable (for example $\widehat{\theta_0}(\mathbf{x})$) can always be reduced by averaging I independent observations, in contrast to its expected value that, remaining constant, imposes:

$$E_{\theta_0}\left[\left(\widehat{\theta_0}(\mathbf{x}) - \theta_0\right)^2\right] \geq \left(E_{\theta_0}\left[\widehat{\theta_0}(\mathbf{x})\right] - \theta_0\right)^2$$

Under the constraint (6), the minimization of (1) is now equivalent to minimize:

$$\min\left\{E_{\theta_0}\left[\phi\left(\mathbf{x}\right)^2\right]\right\} \text{ with } E_{\theta_0}\left[\phi\left(\mathbf{x}\right)\right]=0 \tag{7}$$

substituting $\phi(\mathbf{x}) = \widehat{\theta_0}(\mathbf{x}) - \theta_0$. The set Φ of $\phi(\mathbf{x}) \in \mathcal{F}_\Omega$ satisfying $E_{\theta_0}[\phi(\mathbf{x})] = 0$ forms a vectorial subspace of $\mathcal{F}_\Omega$. In this vectorial subspace, the minimization lemma (4a-b) is applicable. With no other constraint, it leads to the trivial solution $\phi(\mathbf{x}) = 0 \Longleftrightarrow \widehat{\theta_0}(\mathbf{x})_{opt} = \theta_0$, an unacceptable solution, because it not only requires *a priori* knowledge of the parameter to be estimated, but is also independent of the observations. To render the optimum solution dependent on the observations, all that is required is to define a constraint that is not satisfiable by the trivial solution, for example:

$$E_{\theta_0+d\theta}\left[\phi\left(\mathbf{x}\right)\right]=h\left(d\theta\right)\ , \tag{8}$$

but compatible with $E_{\theta_0}[\phi(\mathbf{x})] = 0$, which requires $h(0) = 0$.

This general problem has been characterized by Barankin [6] who derived a necessary and sufficient condition for existence of a solution to (8) when $d\theta$ belongs to a real interval $[-a, b]$, $a, b > 0$. He also provided a general scheme to compute the best possible lower bound (BB) on the variance of such solution [4][7]. Unfortunately, whatever function $h()$, (8) is a quite demanding requirement since it defines a continuous continuum of linear constraints. As a result, there is no general closed form of the solution [10] neither of the BB [4][7]. This obstacle to practical use of Barankin results can be overcome by substituting a countable - possibly infinite - set of linear constraints for the initial continuous continuum [7][10]. If $h(d\theta) = d\theta$ then (8) defines an unbiased estimate constraint in the Barankin sense. A first possible discretization of $h(d\theta) = d\theta$ is a first-order extension, in the neighbourhood of θ_0, of the constraint (6):

$$E_{\theta_0+d\theta}\left[\widehat{\theta_0}\left(\mathbf{x}\right)\right]=\left(\theta_0+d\theta\right)+o\left(d\theta\right) \Longleftrightarrow \frac{\partial E_{\theta_0}\left[\widehat{\theta_0}\left(\mathbf{x}\right)\right]}{\partial\theta}=1 \tag{9}$$

which is the "Cramer-Rao" local approximation of unbiasness in the Barankin sense. It is generally referred as the "locally unbiasness" constraint. As it is equivalent to:

$$\int_\Omega \frac{\partial f_{\theta_0}\left(\mathbf{x}\right)}{\partial\theta}\phi\left(\mathbf{x}\right)d\mathbf{x}=1=\left\langle\frac{\partial \ln f_{\theta_0}}{\partial\theta}\mid\phi\right\rangle_{\theta_0} \tag{10}$$

the solution "locally best at θ_0" and "locally unbiased at θ_0" is (4a-b):

$$f=1,\quad G=F\left(\theta_0\right) \tag{11a}$$

$$\left(\widehat{\theta_0}\left(\mathbf{x}\right)-\theta_0\right)_{opt}=F\left(\theta_0\right)^{-1}\frac{\partial\ln f_{\theta_0}\left(\mathbf{x}\right)}{\partial\theta} \tag{11b}$$

$$\min\left\{MSE_{\theta_0}\right\}=f^TG^{-1}f=\frac{1}{F\left(\theta_0\right)} \tag{11c}$$

where $F(\theta)$ is the Fisher information at θ defined by:

$$F\left(\theta\right)=\left\langle\frac{\partial\ln f_\theta}{\partial\theta}\mid\frac{\partial\ln f_\theta}{\partial\theta}\right\rangle_\theta=E_\theta\left[\left(\frac{\partial\ln f_\theta\left(\mathbf{x}\right)}{\partial\theta}\right)^2\right] \tag{12}$$

Actually a more general family of local approximations of unbiasness in the Barankin sense can be defined by:

$$E_{\theta_0+d\theta}\left[\widehat{\theta_0}\left(\mathbf{x}\right)\right]=\left(\theta_0+d\theta\right)+o\left(d\theta^N\right)$$

It generates the Bhattacharya bounds family (see [7][10] for expressions) that encompasses the CRB ($N = 1$ case).

2) The multiple real parameters case: We now consider the case where the p.d.f. of the observations $f_{\boldsymbol{\theta}}(\mathbf{x})$ depends on a vector of K parameters $\boldsymbol{\theta}$=$(\theta_1,\ldots,\theta_K)$ belonging to $\mathbb{R}^K$, the canonical basis of which we denote $(\mathbf{e}_1\ldots\mathbf{e}_K)$. Let $\boldsymbol{\theta}_0$ be a particular value of $\boldsymbol{\theta}$, and $\widehat{\theta_{0,k}}(\mathbf{x})$ an estimator of $\theta_{0,k}$, the k^{th} coordinate of $\boldsymbol{\theta}_0$.
Now the "locally unbiased" (9) criterion must be extended to the multiple parameters context. It seems quite natural to consider that $\widehat{\theta_{0,k}}(\mathbf{x})$ is a "locally unbiased" estimator of $\theta_{0,k}$ if:

$$\begin{aligned} E_{\boldsymbol{\theta}_0+d\theta\mathbf{e}_k}\left[\widehat{\theta_{0,k}}(\mathbf{x})\right] &= \theta_{0,k}+d\theta+o(d\theta) \\ E_{\boldsymbol{\theta}_0+d\theta\mathbf{e}_l}\left[\widehat{\theta_{0,k}}(\mathbf{x})\right] &= \theta_{0,k}+o(d\theta), \quad \forall l\neq k \end{aligned}$$

which means that, up to the first order and in the neighbourhood of $\boldsymbol{\theta}_0$, $\widehat{\theta_{0,k}}(\mathbf{x})$ remains an unbiased estimator of $\theta_{0,k}$ independently of a - small - variation of the other parameters. These K equations are equivalent to the following $(K+1)$ linear constraints:

$$E_{\boldsymbol{\theta}_0}\left[\widehat{\theta_{0,k}}(\mathbf{x})\right]=\theta_{0,k} \tag{13}$$

$$\frac{\partial E_{\boldsymbol{\theta}_0}\left[\widehat{\theta_{0,k}}(\mathbf{x})\right]}{\partial\theta_l}=\delta_l^k \tag{14}$$

Substituting $\phi(\mathbf{x})=\widehat{\theta_{0,k}}(\mathbf{x})-\theta_{0,k}$, the K constraints (14) become:

$$\int_\Omega \frac{\partial f_{\boldsymbol{\theta}_0}(\mathbf{x})}{\partial\theta_l}\phi(\mathbf{x})\,d\mathbf{x}=\delta_l^k=\left\langle\frac{\partial\ln f_{\boldsymbol{\theta}_0}}{\partial\theta_l}\mid\phi\right\rangle_{\theta_0} \tag{15}$$

and the minimization of $MSE_{\theta_{0,k}}$ under (13) and (14) is then equivalent to minimization problem (7) under (15). Consequently the solution is (4a-b):

$$\mathbf{f}=\mathbf{e}_k,\quad \mathbf{G}=\mathbf{F}(\boldsymbol{\theta}_0),\quad \boldsymbol{\alpha}=\mathbf{F}(\boldsymbol{\theta}_0)^{-1}\mathbf{e}_k \tag{16a}$$

$$\left(\widehat{\theta_{0,k}}(\mathbf{x})-\theta_{0,k}\right)_{opt}=\sum_{l=1}^{l=K}\alpha_l\frac{\partial\ln f_{\boldsymbol{\theta}_0}(\mathbf{x})}{\partial\theta_l} \tag{16b}$$

$$\min\left\{MSE_{\theta_{0,k}}\right\}=\mathbf{f}^T\mathbf{G}^{-1}\mathbf{f}=\mathbf{e}_k^T\mathbf{F}(\boldsymbol{\theta}_0)^{-1}\mathbf{e}_k \tag{16c}$$

where $\mathbf{F}(\boldsymbol{\theta}_0)$ is the Fisher information matrix (FIM) defined by:

$$\mathbf{F}(\boldsymbol{\theta})_{k,l}=\left\langle\frac{\partial\ln f_{\boldsymbol{\theta}}}{\partial\theta_k}\mid\frac{\partial\ln f_{\boldsymbol{\theta}}}{\partial\theta_l}\right\rangle_{\boldsymbol{\theta}}=E_{\boldsymbol{\theta}}\left[\frac{\partial\ln f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial\theta_k}\frac{\partial\ln f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial\theta_l}\right] \tag{17}$$

Moreover, the above derivation can be extended to the estimation of multiple functions depending on multiples parameters, leading thus to the matrix form of CRB inequality [15].

C. Extension to Conditional Lower Bounds

Since most appropriate linear constraints involved in Cramer-Rao, Bhattacharya or Abel bounds derivation depend on implicit interchange of derivatives and integrals - see (9) and (10), (14) and (15), [7][10] -, the mathematical property required in their original form - without conditioning - is the uniform integrability condition on the derivatives of the p.d.f.. If the observations set is restricted to a subset D of Ω, for example by a detection step, then the p.d.f. of observations $f_\theta(\mathbf{x})$ becomes a conditional p.d.f.:

$$f_\theta(\mathbf{x}\mid D)=\frac{f_\theta(\mathbf{x})}{\int_D f_\theta(\mathbf{x})\,d\mathbf{x}} \tag{18}$$

and scalar product (3) becomes:

$$\langle g\mid h\rangle_{\theta_0|D}=E_{\theta_0}\left[g(\mathbf{x})h(\mathbf{x})\mid D\right]=\int_D\left[g(\mathbf{x})h(\mathbf{x})\right]f_{\theta_0}(\mathbf{x}\mid D)\,d\mathbf{x} \tag{19}$$

where $\int_D f_\theta(\mathbf{x})\,d\mathbf{x} = P(D) = P_D(\theta)$ is the probability of conditioning event D. It is obvious that if subset D does not depend on parameter θ, scalar product definitions (3) and (19) are of the same form: they have the same properties and requirement regarding computation of their derivatives. Consequently, from a formal point of view, whatever bound is considered its conditional formulation will be obtained by substituting D and $f_\theta(\mathbf{x} \mid D)$ for Ω and $f_\theta(\mathbf{x})$ in the various expressions.

Contrarily, if subset D does depend on parameter θ generalization of lower bounds expression is not an elementary exercise since it involves calculation of integral derivatives with respect to its domain. Although it is certainly an interesting mathematical problem, this case is of little interest for actual applications where realizable detection test - defining the conditioning set D - can not depend on the unknown value θ. Let us recall that optimal decision rules are based on the exact statistics of the observations [2, p. 23]. Their expressions require knowledge of the p.d.f. of observations under each hypothesis and the *a priori* probability of each hypothesis, if known (Bayes criterion). If no *a priori* probability of hypotheses is available, then in the particular case of binary hypothesis testing the criterion used most often is the likelihood ratio test (LRT) derived by Neyman-Pearson [2, p. 33]. Unfortunately these optimal detection tests are generally not realizable since they often depend on certain of unknown parameters. There are intended for providing the best attainable performance of any decision rule for a given problem. Therefore, a common approach to design realizable tests is to replace the unknown parameters by estimates function of observations, the detection problem becoming a composite hypothesis testing problem (CHTP) [2, p. 86]. Although not necessarily optimal for detection performance, the estimates are generally chosen in the maximum likelihood sense (MLEs), so obtaining the generalized likelihood ratio test (GLRT) [2, p. 92]. In such approach, either analytical expressions of unknown parameter estimators are available and the detection test becomes an analytical expression of observations $\mathbf{x}$ only, or one must resort to numerical maximization methods, which simplest implementation is a discrete search on a fixed set $\Theta = \{\theta_0, \theta_1, \dots \theta_N\}$ of parameters possible values (Joint Detection and Estimation algorithm). In this case, the detection test still does not depend on the true value of the parameter, but only on $\mathbf{x}$ and set Θ.

As a result, for actual applications, whatever lower bound is considered its conditional formulation as defined above can be applied. For example, one can derive the two useful following expressions (see Appendix VI-B) of the Conditional Fisher Information Matrix (CFIM)

$$\mathbf{F}(\boldsymbol{\theta} \mid D)_{k,l} = E_{\boldsymbol{\theta}}\left[\frac{\partial \ln f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k}\frac{\partial \ln f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_l} \mid D\right] - \frac{\partial \ln P_D(\boldsymbol{\theta})}{\partial \theta_k}\frac{\partial \ln P_D(\boldsymbol{\theta})}{\partial \theta_l} \tag{20}$$

$$\mathbf{F}(\boldsymbol{\theta} \mid D)_{k,l} = -E_{\boldsymbol{\theta}}\left[\frac{\partial^2 \ln f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k \partial \theta_l} \mid D\right] + \frac{\partial^2 \ln P_D(\boldsymbol{\theta})}{\partial \theta_k \partial \theta_l} \tag{21}$$

that encompass usual unconditional FIM expressions (17).

IV. Application to array processing

This section deals with estimation of the steering vector of a 2 sensors array. Why is this application fundamental? Firstly, because estimation of the direction of arrival (DOA) of a signal source by means of a 2 sensors array is one of the oldest and most widely used high-resolution techniques, even nowadays, in most operational tracking systems (because of its performance to price ratio). Secondly, as in nearly all fields of science and engineering, its processing requires a detection step. Last but not least, a complete statistical prediction can be computed analytically for a Rayleigh signal source, including GLRT performance, conditional bias, MSE of MLE and conditional CRB. It is the perfect example to illustrate the main theoretical points ensuing from conditioning: conditional bound expressions, "conditional" efficiency, importance of bias at low SNR; in short, that most studies of lower bound tightness at low SNR should take into account the true nature of the problem under investigation: a joint detection-estimation problem.

A. DOA estimation with a 2 sensors array

Assume that a signal source situated at an angle θ (deviation angle from array boresight) is received on a 2 sensors (Σ and Δ) array in the presence of a circular, zero mean, white (both temporally and spatially), complex

Gaussian thermal noise. A common model of the observation equation dedicated to this problem - after Hilbert Filtering - is the following receiver signal vector:

$$\mathbf{v}(t) = \begin{pmatrix} \Sigma(t) \\ \Delta(t) \end{pmatrix} = \alpha(t) \begin{pmatrix} g_\Sigma \\ g_\Delta \end{pmatrix} + \begin{pmatrix} n_\Sigma(t) \\ n_\Delta(t) \end{pmatrix} = \beta(t)\,\mathbf{x} + \mathbf{n}(t) \tag{22}$$

where $\beta(t) = \alpha(t)\, g_\Sigma$, $\mathbf{x} = (1, r)^T$, $r = \frac{g_\Delta}{g_\Sigma}$, $n_\Sigma(t)$ and $n_\Delta(t)$ represent Gaussian receiver noise, g_Σ and g_Δ represent the one-way complex sensor voltage pattern at angle θ and $\alpha(t)$ represents the complex amplitude of the source (including power budget equation, signal processing gains).

In the particular case of a 2 sensors array, the angular information is contained in the ratio $r(\theta) = \frac{g_\Delta(\theta)}{g_\Sigma(\theta)}$, provided the function $\theta \longrightarrow r(\theta)$ is invertible. In actual 2 sensors arrays beamwidth/resolution constraint generally prevents this assumption from being verified for any θ in $\left[\frac{-\pi}{2}, \frac{\pi}{2}\right]$. Nevertheless with appropriate sensors patterns design - uniform sum excitation for Σ and linear odd difference excitation for Δ - and collocated sensors [16, p. 290], the property can hold for $r_x(\theta) = \mathrm{Re}\left\{\frac{g_\Delta(\theta)}{g_\Sigma(\theta)}\right\}$ for θ belonging to Σ main beam, i.e. between the first pattern nulls. Such 2 sensors array are generally called monopulse antennas where $r_x(\theta)$ is the monopulse ratio and $\theta = r_x^{-1}\left(\frac{g_\Delta}{g_\Sigma}\right)$ is the deviation angle function. If a linear relation $r_x = k\theta$ is assumed - which is true at the vicinity of boresight [16, p. 294] - then statistical prediction of $\widehat{\theta} = \frac{\widehat{r}_x}{k}$ can be easily derived from statistical prediction of $\widehat{r}_x$. It is the reason why in open literature the deviation angle function is generally reduced to a linear function characterized by a Monopulse Slope and most DOA statistical performance analysis are related to $\widehat{r}_x$. We will consider this approximation in the following.

B. Background on detection theory applied to a 2 sensors array

Based on one array snapshot $\mathbf{v}(t)$ when the amplitude fluctuation law of the signal source is of Rayleigh type (circular zero mean, temporally white, complex Gaussian discrete random process), we want to decide whether to accept the null hypothesis (noise only) H_0:

$$H_0 : \mathbf{v}(t) = \mathbf{n}(t), \qquad f(\mathbf{v} \mid H_0) = \frac{e^{-\frac{\mathbf{v}^H\mathbf{v}}{\sigma_n^2}}}{\pi^2 \left(\sigma_n^2\right)^2} \tag{23}$$

or to accept the alternate hypothesis (signal plus noise) H_1:

$$H_1 : \mathbf{v}(t) = \beta(t)\,\mathbf{x} + \mathbf{n}(t), \quad f(\mathbf{v} \mid H_1) = \frac{e^{-\mathbf{v}^H\mathbf{C}^{-1}\mathbf{v}}}{\pi^2 |\mathbf{C}|}, \quad \mathbf{C} = \mathbf{C}_{\mathbf{v}(t)} = \sigma_\beta^2 \mathbf{x}\mathbf{x}^H + \sigma_n^2 \mathbf{Id}_2 \tag{24}$$

If the noise power σ_n^2 is known, the unknown parameters of model (22) becomes $\sigma_\beta^2 = E\left[|\beta(t)|^2\right]$ and r. Authors in [17] have proved that the GLRT related to this problem has the simple form:

$$|\Delta|^2 + |\Sigma|^2 \underset{H_0}{\overset{H_1}{\gtrless}} T, \qquad \widehat{r} = \frac{\Delta}{\Sigma} \tag{25}$$

leading to the following probability of false alarm P_{FA} and probability of detection P_D:

$$P_{FA} = P(D \mid H_0) = e^{\frac{-T}{\sigma_n^2}} e_1\left(\frac{T}{\sigma_n^2}\right) \tag{26}$$

$$P_D = P(D \mid H_1) = \iint_{x+t \geq T} f_{\chi_2^1}\left(x, |\gamma|^2 t, \sigma^2\right) f_{\chi_2^1}(t, 0, \mathbf{C}_{11})\, dx dt \tag{27}$$

where $D = \left\{|\Delta|^2 + |\Sigma|^2 \geq T\right\}$, $\widehat{r}$ is the MLE of r, T is the detection threshold and:

$\gamma = \frac{\mathbf{C}_{12}^H}{\mathbf{C}_{11}}, \quad \sigma^2 = \frac{\det(\mathbf{C})}{\mathbf{C}_{11}}$

$f_{\chi_2^I}(t, 0, \sigma^2) = \frac{e^{-\frac{t}{\sigma^2}}}{\sigma^2 (I-1)!}\left(\frac{t}{\sigma^2}\right)^{(I-1)}, \quad f_{\chi_2^I}(t, \mu^2, \sigma^2) = \frac{e^{-\frac{(t+\mu^2)}{\sigma^2}}}{\sigma^2} I_{I-1}\left(\frac{2\mu\sqrt{t}}{\sigma^2}\right)\left(\frac{\sqrt{t}}{\mu}\right)^{(I-1)}$

$I_I(t) = \frac{1}{2\pi} \int\limits_0^{2\pi} e^{t\cos(\theta)} \cos(I\theta)\, d\theta, \quad e_N(T) = \sum\limits_{n=0}^{N} \frac{T^n}{n!}$

Moreover, they also succeeded in deriving analytical expressions of the mean and variance of $\widehat{r}_x = \mathrm{Re}\{\widehat{r}\}$ conditioned by the event D of the GLRT, i.e. $E[\widehat{r}_x \mid D]$ and $Var[\widehat{r}_x \mid D]$, for any SNR, using the following identities:

$$\begin{aligned} \mathrm{Re}\{\widehat{r}\}^2 &= \frac{1}{2}\left[|\widehat{r}|^2 + \mathrm{Re}\left\{\widehat{r}^2\right\}\right] \\ Var[\mathrm{Re}\{\widehat{r}\} \mid D] &= E\left[\mathrm{Re}\{\widehat{r}\}^2 \mid D\right] - \mathrm{Re}\{E[\widehat{r} \mid D]\}^2 \end{aligned}$$

and [17]:

$$E[\widehat{r} \mid D] = \frac{\gamma}{P_D} \iint\limits_{x+t \geq T} f_{\chi_2^2}\left(x, |\gamma|^2 t, \sigma^2\right) f_{\chi_2^1}(t, 0, \mathbf{C}_{11})\, dxdt \quad (28)$$

$$E\left[\widehat{r}^2 \mid D\right] = \frac{\gamma^2}{P_D} \iint\limits_{x+t \geq T} f_{\chi_2^3}\left(x, |\gamma|^2 t, \sigma^2\right) f_{\chi_2^1}(t, 0, \mathbf{C}_{11})\, dxdt \quad (29)$$

$$\begin{aligned} E\left[|\widehat{r}|^2 \mid D\right] = \frac{\sigma^2}{P_D} &\iint\limits_{x+t \geq T} f_{\chi_2^2}\left(x, |\gamma|^2 t, \sigma^2\right) \frac{f_{\chi_2^1}(t, 0, \mathbf{C}_{11})}{t} dxdt \quad (30) \\ +\frac{|\gamma|^2}{P_D} &\iint\limits_{x+t \geq T} f_{\chi_2^3}\left(x, |\gamma|^2 t, \sigma^2\right) f_{\chi_2^1}(t, 0, \mathbf{C}_{11})\, dxdt \end{aligned}$$

C. Derivation of CRB of monopulse ratio r_x

The different expressions - conditional or not - of the CRB of monopulse ratio r_x can be established using real parameters formalism but are simpler to derive with the help of complex formalism. As lower bounds may be derived with or without event conditioning, reference to conditioning event is omitted for legibility. The extension to complex parameters of [7][10] approach is almost straightforward using Minimization Lemma (4a-b), a norm definition (2) based on the Hermitian product

$$\langle g \mid h \rangle_{\theta_0} = E_{\theta_0}\left[g(\mathbf{x})\, h(\mathbf{x})^H\right] = \int\limits_{\Omega} \left[g(\mathbf{x})\, h(\mathbf{x})^H\right] f_{\theta_0}(\mathbf{x})\, d\mathbf{x} \,,$$

and some complex derivative identities developed in [18].

For sake of simplicity we sketch the main stages of the derivation for the problem at hand, i.e. the mixture of a single real and a single complex parameter (and its conjugate). In this case, the p.d.f. of the observations $f_{\boldsymbol{\theta}}(\mathbf{x})$ depends on a vector of three real parameters $\boldsymbol{\theta} = \left(\sigma_\beta^2, \mathrm{Re}\{r\}, \mathrm{Im}\{r\}\right)^T = (\theta_1, \theta_2, \theta_3)^T$, and we are interested in the estimation of a function $\rho(\boldsymbol{\theta})$ - possibly complex - for a particular value $\boldsymbol{\theta}_0$ of $\boldsymbol{\theta}$. According to §III-B2, $\widehat{\rho(\boldsymbol{\theta}_0)}(\mathbf{x})$ is a "locally unbiased" estimator of $\rho(\boldsymbol{\theta}_0)$ if:

$$E_{\boldsymbol{\theta}_0 + d\boldsymbol{\theta}}\left[\widehat{\rho(\boldsymbol{\theta}_0)}(\mathbf{x})\right] = \rho(\boldsymbol{\theta}_0 + d\boldsymbol{\theta}) + o(\|d\boldsymbol{\theta}\|) = \rho(\boldsymbol{\theta}_0) + \left(\frac{\partial \rho(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^T}\right) d\boldsymbol{\theta} + o(\|d\boldsymbol{\theta}\|) \quad (31)$$

The problem can also be reparametrized substituting $\boldsymbol{\omega} = \left(\sigma_\beta^2, r, r^H\right)^T$ for $\boldsymbol{\theta}$. If $\rho(\boldsymbol{\omega})$ and $f_{\boldsymbol{\omega}}(\mathbf{x})$ are analytic with respect to r and r^H, then complex and real derivatives verify [18]:

$$\begin{pmatrix} \frac{\partial}{\partial r} \\ \frac{\partial}{\partial r^H} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{-j}{2} \\ \frac{1}{2} & \frac{j}{2} \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial \mathrm{Re}\{r\}} \\ \frac{\partial}{\partial \mathrm{Im}\{r\}} \end{pmatrix}, \quad \frac{\partial}{\partial r^H} = \left(\frac{\partial}{\partial r}\right)^H \quad (32)$$

and (31) can be rewritten as:

$$E_{\boldsymbol{\omega}_0 + d\boldsymbol{\omega}}\left[\widehat{\rho(\boldsymbol{\omega}_0)}(\mathbf{x})\right] = \rho(\boldsymbol{\omega}_0) + \left(\frac{\partial \rho(\boldsymbol{\omega}_0)}{\partial \boldsymbol{\omega}^T}\right) d\boldsymbol{\omega} + o(\|d\boldsymbol{\omega}\|)$$

which, expressed in terms of constraints (14), leads to (4a-b):

$$\min\left\{MSE_{\rho(\boldsymbol{\omega}_0)}\right\} = \left(\frac{\partial\rho\left(\boldsymbol{\omega}_0\right)}{\partial\boldsymbol{\omega}^T}\right)\mathbf{F}\left(\boldsymbol{\omega}_0\right)^{-1}\left(\frac{\partial\rho\left(\boldsymbol{\omega}_0\right)}{\partial\boldsymbol{\omega}^T}\right)^H \tag{33}$$

with:

$$\mathbf{F}\left(\boldsymbol{\omega}_0\right)_{k,l} = E_{\boldsymbol{\omega}_0}\left[\left(\frac{\partial\ln f_{\boldsymbol{\omega}_0}\left(\mathbf{x}\right)}{\partial\omega_k}\right)^H\frac{\partial\ln f_{\boldsymbol{\omega}_0}\left(\mathbf{x}\right)}{\partial\omega_l}\right] \tag{34}$$

The calculation of CRB and CFIM according to (33) and (34) for any values of σ_β^2 and r are detailed in Appendix VI-C. We are more particularly interested in the case $r_x = r = 0$, which corresponds to a source signal located along the main axis of a monopulse antenna ($\theta = 0$). This is a reference case in the study of the performance of such receiving system, since it corresponds to the peak received energy.

D. Results

The most significant results are provided by figures (1), (2) and (3). They display the values of the MSE of MLE of r_x and its related CRBs for unbiased estimates as a function of SNR on sum channel (Σ) for a given P_{FA}. Indeed, when a detection step is taken into account, the P_{FA} is the information of practical interest since it represents the proportion - exactly under hypothesis H_0 and approximately under H_1 - of noise samples that will be transmitted to the estimation process. At each P_{FA} $\left(0.9999, 10^{-4}, 10^{-10}\right)$ corresponds a linear detection threshold T $(0.01, 5.9, 13.2)$ according to (26). In figures legend "CRB" stands for "classical" CRB, whereas "CONDITIONAL CRB" takes into account the detection test.

In the problem at hand, we are in the particular case where the transition region is smooth - see figure (1) - when the detection threshold effect is negligible ($P_{FA} = 0.9999$). Indeed the p.d.f. of $\widehat{r}_x$ follows a Student distribution [19] with a constant mean value 0 (28) and a smoothly increasing variance (no outliers). Nevertheless some general considerations can still be drawn from this particular case. As intuitively expected the detection step modify MSE behaviour mainly in the transition region. It plays a crucial role in selecting instances with relatively high signal energy - sufficient to exceed the detection threshold - and disregarding instances belonging to the *a priori* region that deteriorate the MSE. This is perfectly depicted by figures (2) and (3) where the increase of threshold detection decreases MSE values in the transition region. Another expected result is the proof of the necessity of using conditional form of a given lower bound to be able to keep at least its lower bound property. This is highlighted in the SNR region $[-10, 10]$ dB on figures (2) and (3) where the conditional CRB is obviously the only meaningful expression of the bound.

On the other hand, a surprising result is the tightness of the conditional CRB in most of the transition region. Indeed, a particular property of $\widehat{r}_x$ is not to be efficient when SNR tends to infinity, which originates from its Student p.d.f. [20]. Actually this phenomenon reveals that conditioning of the observations by an event D may significantly modify the conditions required to attain the CRB and thus to obtain an efficient estimator. This phenomenon is highlighted as well by figure (4) which depicts, for a given SNR in the transition region (0 dB), the convergence of $\widehat{r}_x$ towards an efficient estimator as the detection threshold T increases. Theoretical investigation of this phenomenon is a topic for further study as it is a quite complex problem that requires first to find $f_\theta\left(\mathbf{x}\right)$ solution of (11a-c):

$$\widehat{\theta_0}\left(\mathbf{x}\right)_{opt} - \theta_0 = F\left(\theta_0 \mid D\right)^{-1}\left[\frac{\partial\ln f_{\theta_0}\left(\mathbf{x}\right)}{\partial\theta} - \frac{\partial\ln P_D(\theta_0)}{\partial\theta}\right], \quad \forall\mathbf{x}\in D\subset\Omega$$

Additionally, the results also show that there is a limit to the pertinence of the information delivered by the CRB for unbiased estimates at very low SNR, even when conditioning is included, since threshold detection increase has almost no effect on the tightness of the bound in the *a priori* region. The main reason is that a locally unbiased estimator of source signal parameters generally does not exist asymptotically as the SNR decreases to 0. To overcome this limitation, one can resort to biased CRB. In that case (9) becomes:

$$E_{\theta_0+d\theta}\left[\widehat{\theta_0}\left(\mathbf{x}\right)\right] = \left(\theta_0 + d\theta\right) + b\left(\theta_0 + d\theta\right) + o\left(d\theta\right) \Longleftrightarrow \frac{\partial E_{\theta_0}\left[\widehat{\theta_0}\left(\mathbf{x}\right)\right]}{\partial\theta} = 1 + \frac{\partial b\left(\theta_0\right)}{\partial\theta}$$

where $b\left(\theta\right)$ is the bias function, which leads - according to (5) and (10) - to:

$$\min\left\{MSE_{\theta_0}\right\} \geq F(\theta_0)^{-1}\left[1+\frac{\partial b\left(\theta_0\right)}{\partial\theta}\right]^2 + b\left(\theta_0\right)^2$$

It is an attractive theoretical refinement if analytical expression of the bias is available (see §VI-D) as it is shown in figure (5) where introduction of bias ("CONDITIONAL BIASED CRB" curve) has restored Cramer-Rao lower bound property in all regions of operation. Unfortunately the bias depends on the specific estimator and furthermore is hardly ever known in practice. This pessimistic observation must however be balanced against practical considerations. Indeed, it is doubtful whether this limit raises a genuine practical problem in our application, since it appears in an SNR region $(SNR < -10dB)$ where the source signal is simply considered to be absent from an operational point of view $\left(P_D \leq 10^{-3}\right)$. As a consequence, performance assessment in this operating area is of no interest from a contractual point of view.

V. Conclusion

Despite they have been derived for a Small-Error bound representative (CRB) and for a particular case of MLE behaviour (smooth transition region), the preliminary results introduced in the present paper shows that the problem of lower bound tightness at low SNR $(P_D < 1)$ might be overestimated for practical application involving a binary detection test. Indeed, in a joint detection and estimation problem, the detection step determines the operating area of interest where it is worth assessing estimation performance, including lower bounds computation. It naturally raises the question of the practical importance of MSE prediction in the *a priori* region. From this point of view, improvement of CRB behaviour in the transition region resulting from introduction of the detection step conditioning is promising. Therefore, despite the fact that in most practical applications the conditional p.d.f. must be difficult to calculate, this refinement in MSE lower bounds derivation is worth investigating, including Large-Error bound representatives, which is a topic for further study. Indeed, with respect to the classic approaches, an additional merit of the present paper is the presentation of a general didactic scheme enabling the derivation of any kind - conditional/unconditional, biased/unbiased - of deterministic MSE lower bounds for real/complex parameters. As a consequence, this paper offers the signal processing community the tools required to look again at comparisons of bounds tightness including detection conditioning when it appears to be present in the actual implementation.

References

[1] H. Cramer, "Mathematical Methods of Statistics", Princeton, Princeton University Press 1946

[2] H.L. Van Trees, "Detection, Estimation and Modulation Theory, Part 1", New York, Wiley, 1968

[3] N. Bertaux, P. Larzabal, C. Adnet, E. Chaumette, "A Parameterized Maximum Likelihood Method For Multipaths Channels Estimation", SPAWC Conf., p 391-393, 1999

[4] R. McAulay, E.M. Hofstetter, "Barankin Bounds on parameter estimation", IEEE Trans., IT vol 17, pp 669-676, Nov 1971

[5] J.S. Abel, "A bound on mean-square-estimate error", IEEE Trans., IT vol 39, pp. 1675-1680, Sept. 1993

[6] E.W. Barankin, "Locally best unbiased estimates", Ann. Math. Stat., vol. 20, pp 477-501, 1949

[7] F.E. Glave, "A new look at the Barankin Lower Bound", IEEE Trans., IT vol 18, p 349-356, May 1972

[8] J. Tabrikian, J.L. Kromik, "Barankin bounds for source localization in an uncertain ocean environment", IEEE Trans., SP vol 47, pp 2917-2927, Nov. 1999

[9] L. Knockaert, "The Barankin bound and threshold behaviour in frequency estimation", IEEE Trans, SP vol 45, p 2398–2401, Sep. 1997

[10] P. Forster, P. Larzabal, "On lower Bounds For Deterministic Parameter Estimation", ICASSP Conf. 2002, p1137-1140

[11] D.C. Rife, R.R. Boorstyn, "Single tone parameter esrimation from discrete-time observations", IEEE Trans., IT vol 20., p 591-598, Sept. 1974

[12] E.L. Lehmann, "Theory of point estimation", 2nd Ed., New York, Springer 1998

[13] H.L. Van Trees, "Detection, Estimation and Modulation Theory, Part IV: Optimum Array Processing", Wiley-Interscience, Wiley 2002

[14] L.L. Scharf, "Detection, Estimation and Time Series Analysis", Adison-Wesley, 1991

[15] E. Chaumette, "Contribution to the performance characterization of the joint Detection-Estimation problem", PhD Thesis, Ecole Normale Supérieure de Cachan, France, Dec. 2004

[16] N. Levanon, "Radar Principles", Wiley-Interscience, Wiley 1988

[17] E. Chaumette, P. Larzabal, "Optimal Detection Theory Applied to Monopulse Antenna", ICASSP Conf. 2004, Vol II p. 1065-1068

[18] D.H. Brandwood, "A complex gradient operator and its application in adaptive array theory", IEE Proc. F and H, vol 130, p 11-16, 1983

[19] B-E. Tullsson, "Monopulse tracking of Rayleigh targets, a simple approach", IEEE Trans. AES, vol. AES-27, p 520-531, 1991

[20] A. Renaux, P. Forster, E. Boyer, P. Larzabal, "Non Efficiency and Non Gaussianity of a Maximum Likelihood Estimator at High Signal to Noise Ratio and Finite Number of Samples", ICASSP Conf. 2004, Vol II p. 121-124

[21] A. van den Bos, "A Cramer-Rao Lower Bound for Complex Parameters", IEEE Trans., SP vol 42, p 2859, Oct. 1994

[22] E. Chaumette, "Conditionnal CRB of the Monopulse Ratio for Rayleigh Targets", Memo TNF/BRS/EPR - 261/03, Thales Naval France, Bagneux, France, 2003

VI. APPENDIX

A. Acronyms

MSE	mean square error
MLE(s)	maximum likelihood estimator(s)
CRB	Cramer-Rao bound
BB	Barankin bound
SNR	signal-to-noise ratio
FIM	Fisher information matrix
CFIM	conditional Fisher information matrix
LRT	likelihood ratio test
GLRT	generalized likelihood ratio test
CHTP	composite hypothesis testing problem
DOA	direction of arrival
P_{FA}	probability of false alarm
P_D	probability of detection

B. Derivation of CFIM expressions

The first useful form of CFIM (20) can be derived directly from CFIM definition :

$$\begin{aligned}\mathbf{F}(\boldsymbol{\theta} \mid D)_{k,l} &= \int_D \left[\frac{\partial \ln f_{\boldsymbol{\theta}}(\mathbf{x} \mid D)}{\partial \theta_k}\frac{\partial \ln f_{\boldsymbol{\theta}}(\mathbf{x} \mid D)}{\partial \theta_l}\right] f_{\boldsymbol{\theta}}(\mathbf{x} \mid D)\, d\mathbf{x} \\ &= E_{\boldsymbol{\theta}}\left[\frac{\partial \ln f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k}\frac{\partial \ln f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_l} \mid D\right] + \frac{\partial \ln P_D(\boldsymbol{\theta})}{\partial \theta_k}\frac{\partial \ln P_D(\boldsymbol{\theta})}{\partial \theta_l} \\ &\quad -E_{\boldsymbol{\theta}}\left[\frac{\partial \ln f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k} \mid D\right]\frac{\partial \ln P_D(\boldsymbol{\theta})}{\partial \theta_l} - E_{\boldsymbol{\theta}}\left[\frac{\partial \ln f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_l} \mid D\right]\frac{\partial \ln P_D(\boldsymbol{\theta})}{\partial \theta_k}\end{aligned}$$

where:

$$\begin{aligned}\frac{\partial \ln P_D(\boldsymbol{\theta})}{\partial \theta_k} &= \frac{1}{P_D(\boldsymbol{\theta})}\frac{\partial P_D(\boldsymbol{\theta})}{\partial \theta_k} = \frac{1}{P_D(\boldsymbol{\theta})}\int_D \frac{\partial f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k} d\mathbf{x} = \int_D \left(\frac{\partial f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k}\frac{1}{f_{\boldsymbol{\theta}}(\mathbf{x})}\right)\left(\frac{f_{\boldsymbol{\theta}}(\mathbf{x})}{P_D(\boldsymbol{\theta})}\right) d\mathbf{x} \\ &= E_{\boldsymbol{\theta}}\left[\frac{\partial \ln f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k} \mid D\right]\end{aligned}$$

Finally:

$$\mathbf{F}(\boldsymbol{\theta} \mid D)_{k,l} = E_{\boldsymbol{\theta}}\left[\frac{\partial \ln f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k}\frac{\partial \ln f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_l} \mid D\right] - \frac{\partial \ln P_D(\boldsymbol{\theta})}{\partial \theta_k}\frac{\partial \ln P_D(\boldsymbol{\theta})}{\partial \theta_l}$$

The second useful form of CFIM (21) can be derived from the following identity:

$$\int_D f_{\boldsymbol{\theta}}(\mathbf{x} \mid D)\, d\mathbf{x} = 1 \Rightarrow \int_D \frac{\partial f_{\boldsymbol{\theta}}(\mathbf{x} \mid D)}{\partial \theta_k} d\mathbf{x} = 0 = \int_D \frac{\partial \ln f_{\boldsymbol{\theta}}(\mathbf{x} \mid D)}{\partial \theta_k} f_{\boldsymbol{\theta}}(\mathbf{x} \mid D)\, d\mathbf{x}, \quad \forall k \in [1, K]$$

leading to the straightforward result:

$$E_{\boldsymbol{\theta}}\left[\frac{\partial \ln f_{\boldsymbol{\theta}}(\mathbf{x} \mid D)}{\partial \theta_k}\frac{\partial \ln f_{\boldsymbol{\theta}}(\mathbf{x} \mid D)}{\partial \theta_l} \mid D\right] = -E_{\boldsymbol{\theta}}\left[\frac{\partial^2 \ln f_{\boldsymbol{\theta}}(\mathbf{x} \mid D)}{\partial \theta_k \partial \theta_l} \mid D\right]$$

Therefore:

$$\mathbf{F}(\boldsymbol{\theta} \mid D)_{k,l} = -E_{\boldsymbol{\theta}}\left[\frac{\partial^2 \ln f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k \partial \theta_l} \mid D\right] + \frac{\partial^2 \ln P_D(\boldsymbol{\theta})}{\partial \theta_k \partial \theta_l}$$

C. Calculation of CRB of Monopulse Ratio

The general calculation of the CFIM according to (34) can be done in two steps (21): the calculation of $E_{\boldsymbol{\omega}}\left[\frac{\partial^2 \ln f_{\boldsymbol{\omega}}(\mathbf{v})}{\partial \omega_k^H \partial \omega_l} \mid D\right]$ then of $\frac{\partial^2 - \ln P_D(\boldsymbol{\omega})}{\partial \omega_k^H \partial \omega_l}$ with $f_{\boldsymbol{\omega}}(\mathbf{v}) = f(\mathbf{v} \mid H_1)$ (24), $\boldsymbol{\omega} = \left(\sigma_\beta^2, r^H, r\right)^T$. Furthermore, the symmetry properties of the CFIM [21] can be used to show that it depends only on terms $\frac{\partial^2}{\partial \sigma_\beta^2 \partial \sigma_\beta^2}$, $\frac{\partial^2}{\partial r^H \partial \sigma_\beta^2}$,$\frac{\partial^2}{\partial r \partial r^H}$, $\frac{\partial^2}{\partial r^H \partial r^H}$.

1) Computation of $E_{\boldsymbol{\omega}}\left[\frac{\partial^2 \ln f_{\boldsymbol{\omega}}(\mathbf{v})}{\partial \omega_k^H \partial \omega_l} \mid D\right]$*:* In the problem at hand, the calculation of the $\frac{\partial^2 \ln f_{\boldsymbol{\omega}}(\mathbf{v})}{\partial \omega_k^H \partial \omega_l}$ terms is simplified by introducing the log likelihood ratio:

$$\ln\left[\frac{f(\mathbf{v} \mid H_1)}{f(\mathbf{v} \mid H_0)}\right] = s\frac{\mathbf{x}^H \widehat{\mathbf{R}} \mathbf{x}}{1 + s\mathbf{x}^H \mathbf{x}} - \ln\left(1 + s\mathbf{x}^H \mathbf{x}\right)$$

where: $\widehat{\mathbf{R}} = \mathbf{v}\mathbf{v}^H$, σ_n^2 is the known noise power and $s = \frac{\sigma_\beta^2}{\sigma_n^2}$ is the SNR on the sum channel Σ, the unknown parameters vector becoming $\boldsymbol{\omega} = \left(s, r^H, r\right)^T$. To derive required conditional expectations, the following scheme has been used:

1) Computation of first order derivatives:

$$\frac{\partial \ln f_{\boldsymbol{\omega}}(\mathbf{v})}{\partial s} = \frac{n_1\left(\widehat{\mathbf{R}}\right)}{\left(1 + s\mathbf{x}^H \mathbf{x}\right)^2}, \quad \frac{\partial \ln f_{\boldsymbol{\omega}}(\mathbf{v})}{\partial r^H} = \frac{n_2\left(\widehat{\mathbf{R}}\right)s}{\left(1 + s\mathbf{x}^H \mathbf{x}\right)^2}$$

where:

$$\begin{aligned} n_1\left(\widehat{\mathbf{R}}\right) &= \mathbf{x}^H \widehat{\mathbf{R}} \mathbf{x} - \mathbf{x}^H \mathbf{x}\left(1 + s\mathbf{x}^H \mathbf{x}\right) \\ n_2\left(\widehat{\mathbf{R}}\right) &= \mathbf{e}_2^H \widehat{\mathbf{R}} \mathbf{x}\left(1 + s\mathbf{x}^H \mathbf{x}\right) - r\left(s\mathbf{x}^H \widehat{\mathbf{R}} \mathbf{x} + 1 + s\mathbf{x}^H \mathbf{x}\right) \end{aligned}$$

2) Computation of conditional expectations of second order derivatives:

$$\begin{aligned} E_{\boldsymbol{\omega}}\left[\frac{\partial^2 \ln f_{\boldsymbol{\omega}}(\mathbf{v})}{\partial^2 s} \mid D\right] &= \frac{-\mathbf{x}^H \mathbf{x}}{\left(1 + s\mathbf{x}^H \mathbf{x}\right)^2}\left(\mathbf{x}^H \mathbf{x} + 2\frac{E_{\boldsymbol{\omega}}\left[n_1\left(\widehat{\mathbf{R}}\right) \mid D\right]}{1 + s\mathbf{x}^H \mathbf{x}}\right) \\ E_{\boldsymbol{\omega}}\left[\frac{\partial^2 \ln f_{\boldsymbol{\omega}}(\mathbf{v})}{\partial r^H \partial s} \mid D\right] &= \frac{1}{\left(1 + s\mathbf{x}^H \mathbf{x}\right)^2}\left(\mathbf{e}_2^H \widehat{\mathbf{R}} \mathbf{x} - r\left(1 + 2s\mathbf{x}^H \mathbf{x}\right) - 2rs\frac{E_{\boldsymbol{\omega}}\left[n_1\left(\widehat{\mathbf{R}}\right) \mid D\right]}{1 + s\mathbf{x}^H \mathbf{x}}\right) \\ E_{\boldsymbol{\omega}}\left[\frac{\partial^2 \ln f_{\boldsymbol{\omega}}(\mathbf{v})}{\partial^2 r^H} \mid D\right] &= \frac{-s^2 r}{\left(1 + s\mathbf{x}^H \mathbf{x}\right)^2}\left(r - 2\frac{E_{\boldsymbol{\omega}}\left[n_2\left(\widehat{\mathbf{R}}\right) \mid D\right]}{1 + s\mathbf{x}^H \mathbf{x}}\right) \end{aligned}$$

$$\begin{aligned} E_{\boldsymbol{\omega}}\left[\frac{\partial^2 \ln f_{\boldsymbol{\omega}}(\mathbf{v})}{\partial r \partial r^H} \mid D\right] = &-\frac{s}{\left(1 + s\mathbf{x}^H \mathbf{x}\right)^2} 2r^H s\frac{E_{\boldsymbol{\omega}}\left[n_2\left(\widehat{\mathbf{R}}\right) \mid D\right]}{1 + s\mathbf{x}^H \mathbf{x}} \\ &- \frac{s}{\left(1 + s\mathbf{x}^H \mathbf{x}\right)^2}\left[\left(s\mathbf{x}^H E_{\boldsymbol{\omega}}\left[\widehat{\mathbf{R}} \mid D\right]\mathbf{x} + 1 + s\mathbf{x}^H \mathbf{x}\right) + r\left(s\mathbf{x}^H E_{\boldsymbol{\omega}}\left[\widehat{\mathbf{R}} \mid D\right]\mathbf{e}_2 + sr^H\right)\right] \\ &+ \frac{s}{\left(1 + s\mathbf{x}^H \mathbf{x}\right)^2}\left[\left(1 + s\mathbf{x}^H \mathbf{x}\right)\mathbf{e}_2^H E_{\boldsymbol{\omega}}\left[\widehat{\mathbf{R}} \mid D\right]\mathbf{e}_2 + sr^H \mathbf{e}_2^H E_{\boldsymbol{\omega}}\left[\widehat{\mathbf{R}} \mid D\right]\mathbf{x}\right] \end{aligned}$$

where:

$$\begin{aligned}
E_{\boldsymbol{\omega}}\left[n_1\left(\widehat{\mathbf{R}}\right) \mid D\right] &= \mathbf{x}^H E_{\boldsymbol{\omega}}\left[\widehat{\mathbf{R}} \mid D\right]\mathbf{x} - \mathbf{x}^H\mathbf{x}\left(1+s\mathbf{x}^H\mathbf{x}\right)\\
E_{\boldsymbol{\omega}}\left[n_2\left(\widehat{\mathbf{R}}\right) \mid D\right] &= \mathbf{e}_2^H E_{\boldsymbol{\omega}}\left[\widehat{\mathbf{R}} \mid D\right]\mathbf{x}\left(1+s\mathbf{x}^H\mathbf{x}\right) - r\left(s\mathbf{x}^H E_{\boldsymbol{\omega}}\left[\widehat{\mathbf{R}} \mid D\right]\mathbf{x} + 1 + s\mathbf{x}^H\mathbf{x}\right)\\
\mathbf{e}_2 &= (0,1)^T\\
E_{\boldsymbol{\omega}}\left[\widehat{\mathbf{R}} \mid D\right] &= E_{\boldsymbol{\omega}}\left[\mathbf{v}\mathbf{v}^H \mid D\right] = E_{\boldsymbol{\omega}}\left[\begin{pmatrix} \Sigma\Sigma^H & \Sigma\Delta^H \\ \Delta\Sigma^H & \Delta\Delta^H \end{pmatrix} \mid D\right]\\
D &= \left\{|\Delta|^2 + |\Sigma|^2 \geq T\right\}
\end{aligned}$$

and [22]:

$$\begin{aligned}
E_{\boldsymbol{\omega}}\left(\Sigma\Sigma^H \mid D\right) &= \mathbf{C}_{11}\left(\frac{1}{P_D}\iint\limits_{x+t\geq T} f_{\chi_2^1}\left(x, \left|\frac{\mathbf{C}_{12}}{\mathbf{C}_{11}}\right|^2 t, \frac{\det(\mathbf{C})}{\mathbf{C}_{11}}\right) f_{\chi_2^2}\left(t,0,\mathbf{C}_{11}\right) dx dt\right)\\
E_{\boldsymbol{\omega}}\left(\Delta\Delta^H \mid D\right) &= \mathbf{C}_{22}\left(\frac{1}{P_D}\iint\limits_{x+t\geq T} f_{\chi_2^1}\left(t, \left|\frac{\mathbf{C}_{12}}{\mathbf{C}_{22}}\right|^2 x, \frac{\det(\mathbf{C})}{\mathbf{C}_{22}}\right) f_{\chi_2^2}\left(t,0,\mathbf{C}_{22}\right) dx dt\right)\\
E_{\boldsymbol{\omega}}\left(\Sigma\Delta^H \mid D\right) &= \mathbf{C}_{12}\left(\frac{1}{P_D}\iint\limits_{x+t\geq T} f_{\chi_2^2}\left(x, \left|\frac{\mathbf{C}_{12}}{\mathbf{C}_{11}}\right|^2 t, \frac{\det(\mathbf{C})}{\mathbf{C}_{11}}\right) f_{\chi_2^2}\left(t,0,\mathbf{C}_{11}\right) dx dt\right)
\end{aligned}$$

Note that if $D = \Omega$, then:

$$\begin{aligned}
E_{\boldsymbol{\omega}}\left[n_1\left(\widehat{\mathbf{R}}\right)\right] &= E_{\boldsymbol{\omega}}\left[n_2\left(\widehat{\mathbf{R}}\right)\right] = 0\\
E_{\boldsymbol{\omega}}\left[\frac{\partial^2 \ln f_{\boldsymbol{\omega}}(\mathbf{x})}{\partial^2 s}\right] &= \frac{-\left(\mathbf{x}^H\mathbf{x}\right)^2}{\left(1+s\mathbf{x}^H\mathbf{x}\right)^2}, \qquad E_{\boldsymbol{\omega}}\left[\frac{\partial^2 \ln f_{\boldsymbol{\omega}}(\mathbf{x})}{\partial r^H \partial s}\right] = \frac{-\mathbf{x}^H\mathbf{x} rs}{\left(1+s\mathbf{x}^H\mathbf{x}\right)^2}\\
E_{\boldsymbol{\omega}}\left[\frac{\partial^2 \ln f_{\boldsymbol{\omega}}(\mathbf{x})}{\partial^2 r^H}\right] &= \frac{-\mathbf{x}^H\mathbf{x}\left(1+s\right)s^2}{\left(1+s\mathbf{x}^H\mathbf{x}\right)^2}, \quad E_{\boldsymbol{\omega}}\left[\frac{\partial^2 \ln f_{\boldsymbol{\omega}}(\mathbf{x})}{\partial r \partial r^H}\right] = \frac{-s^2 r^2}{\left(1+s\mathbf{x}^H\mathbf{x}\right)^2}
\end{aligned}$$

2) Computation of $\frac{\partial^2 - \ln P_D(\boldsymbol{\omega})}{\partial \omega_k^H \partial \omega_l}$: An other form of $P_D = \int\limits_{\|\mathbf{v}\|^2 \geq T} f_{\boldsymbol{\theta}}(\mathbf{v})\, d\mathbf{v}$ (27) can be derived noticing that:

$$\mathbf{C} = \sigma_n^2\left(s\mathbf{x}\mathbf{x}^H + \mathbf{Id}_2\right) = \sigma_n^2\left[s\|\mathbf{x}\|^2\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right)\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right)^H + \mathbf{Id}_2\right], \quad s = \frac{\sigma_\beta^2}{\sigma_n^2}$$

$\mathbf{C}$ can be diagonalized on the orthonormal basis of eigenvectors $\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}, \left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right)^\perp\right) = (\mathbf{u}_1, \mathbf{u}_2)$ with eigenvalues $\lambda_1 = \sigma_n^2\left(s\|\mathbf{x}\|^2 + 1\right)$ and $\lambda_2 = \sigma_n^2$. Let us consider the change of variable $\mathbf{w} = [\mathbf{u}_1\mathbf{u}_2]^H\mathbf{v}$, then [22]:

$$P_D = \int\limits_{\|\mathbf{w}\|^2\geq T} f_{\boldsymbol{\theta}}(\mathbf{w})\, d\mathbf{w} = \iint\limits_{x+t\geq T} f_{\chi_2^1}\left(x,0,\sigma_n^2\right) f_{\chi_2^1}\left(t,0,\sigma_n^2\left(s\mathbf{x}^H\mathbf{x}+1\right)\right) dx dt \tag{35}$$

To compute the required derivations, the following scheme has been used:

1) Reformulation of $(-\ln P_D(\boldsymbol{\omega}))$ using (35):

$$\begin{aligned}
-\ln P_D(\boldsymbol{\omega}) &= g\left(\lambda(\boldsymbol{\omega})\right)\\
g(\lambda) &= \ln\left(\iint\limits_{x+t\geq T} f_{\chi_2^1}\left(x,0,\sigma_n^2\right) f_{\chi_2^1}\left(t,0,\lambda\right) dx dt\right), \qquad \lambda(\boldsymbol{\omega}) = \sigma_n^2\left(s\mathbf{x}^H\mathbf{x}+1\right)
\end{aligned}$$

2) Reformulation of $\frac{\partial^2 - \ln P_D(\boldsymbol{\omega})}{\partial \omega_k^H \partial \omega_l}$ as a function of derivatives of $g(\lambda)$ and $\lambda(\boldsymbol{\omega})$:

$$\frac{\partial^2 - \ln P_D(\boldsymbol{\omega})}{\partial \omega_k^H \partial \omega_l} = \frac{\partial g(\lambda(\boldsymbol{\omega}))}{\partial \lambda} \frac{\partial^2 \lambda(\boldsymbol{\omega})}{\partial \omega_k^H \partial \omega_l} + \frac{\partial^2 g(\lambda(\boldsymbol{\omega}))}{\partial^2 \lambda} \frac{\partial \lambda(\boldsymbol{\omega})}{\partial \omega_k^H} \frac{\partial \lambda(\boldsymbol{\omega})}{\partial \omega_l}$$

3) Derivation of $\frac{\partial g(\lambda)}{\partial \lambda}$ and $\frac{\partial^2 g(\lambda)}{\partial^2 \lambda}$:
From:

$$\frac{\partial f_{\chi_2^I}(t,0,\lambda)}{\partial \lambda} = \frac{I}{\lambda}\left[f_{\chi_2^{I+1}}(t,0,\lambda) - f_{\chi_2^I}(t,0,\lambda)\right]$$

it is straightforward to show that:

$$\frac{\partial g(\lambda)}{\partial \lambda} = \frac{1}{\lambda}\left[1 - \frac{1}{P_D}\iint_{x+t\geq T} f_{\chi_2^1}\left(x,0,\sigma_n^2\right) f_{\chi_2^2}(t,0,\lambda)\, dxdt\right]$$

$$\begin{aligned}\frac{\partial^2 g(\lambda)}{\partial^2 \lambda} = \frac{-1}{\lambda}\frac{\partial g(\lambda)}{\partial \lambda} &- \frac{2}{\lambda^2 P_D}\iint_{x+t\geq T} f_{\chi_2^1}\left(x,0,\sigma_n^2\right) f_{\chi_2^3}(t,0,\lambda)\, dxdt \\ &+ \frac{1}{\lambda^2 P_D}\iint_{x+t\geq T} f_{\chi_2^1}\left(x,0,\sigma_n^2\right) f_{\chi_2^2}(t,0,\lambda)\, dxdt + \frac{1}{\lambda^2 P_D}\left(\iint_{x+t\geq T} f_{\chi_2^1}\left(x,0,\sigma_n^2\right) f_{\chi_2^2}(t,0,\lambda)\, dxdt\right)^2\end{aligned}$$

4) Derivation of $\frac{\partial \lambda(\boldsymbol{\omega})}{\partial \omega_l}$ and $\frac{\partial^2 \lambda(\boldsymbol{\omega})}{\partial \omega_k^H \partial \omega_l}$:

$$\frac{\partial \lambda(\boldsymbol{\omega})}{\partial s} = \sigma_n^2 \mathbf{x}^H \mathbf{x}, \qquad \frac{\partial \lambda(\boldsymbol{\omega})}{\partial r^H} = \sigma_n^2 s r, \qquad \frac{\partial \lambda(\boldsymbol{\omega})}{\partial r} = \sigma_n^2 s r^H$$

$$\frac{\partial^2 \lambda(\boldsymbol{\omega})}{\partial^2 s} = 0, \qquad \frac{\partial^2 \lambda(\boldsymbol{\omega})}{\partial r \partial s} = \sigma_n^2 r^H, \qquad \frac{\partial^2 \lambda(\boldsymbol{\omega})}{\partial r \partial r^H} = s\sigma_n^2, \qquad \frac{\partial^2 \lambda(\boldsymbol{\omega})}{\partial r^H \partial r^H} = 0$$

5) Final combination of results:

$$\begin{aligned}\frac{\partial^2 P_D(\boldsymbol{\omega})}{\partial^2 s} &= \frac{\partial^2 g(\lambda(\boldsymbol{\omega}))}{\partial^2 \lambda}\left(\sigma_n^2 \mathbf{x}^H \mathbf{x}\right)^2 \\ \frac{\partial^2 P_D(\boldsymbol{\omega})}{\partial r \partial s} &= \frac{\partial g(\lambda(\boldsymbol{\omega}))}{\partial \lambda}\sigma_n^2 r^H + \frac{\partial^2 g(\lambda(\boldsymbol{\omega}))}{\partial^2 \lambda}\left(\sigma_n^2\right)^2 \mathbf{x}^H \mathbf{x} s r^H \\ \frac{\partial^2 P_D(\boldsymbol{\omega})}{\partial r \partial r^H} &= \frac{\partial g(\lambda(\boldsymbol{\omega}))}{\partial \lambda}\sigma_n^2 s + \frac{\partial^2 g(\lambda(\boldsymbol{\omega}))}{\partial^2 \lambda}\left(\sigma_n^2\right)^2 s r s r^H \\ \frac{\partial^2 P_D(\boldsymbol{\omega})}{\partial r^H \partial r^H} &= \frac{\partial^2 g(\lambda(\boldsymbol{\omega}))}{\partial^2 \lambda}\left(\sigma_n^2 s r^H\right)^2\end{aligned}$$

Note that if $D = \Omega$, then: $\frac{\partial^2 - \ln P_D(\boldsymbol{\omega})}{\partial \omega_k^H \partial \omega_l} = 0$

D. Computation of bias derivatives

Let us recall that $\boldsymbol{\omega} = \left(s, r^H, r\right)^T$. We are interested in the calculation of the MSE (33) of estimator $\widehat{\mathrm{Re}\{r\}}(\mathbf{x})$ of $\mathrm{Re}\{r\}$ such that [17]:

$$E_{\boldsymbol{\omega}}\left[\widehat{\mathrm{Re}\{r\}}(\mathbf{x}) \mid D\right] = \rho(\boldsymbol{\omega}) = \mathrm{Re}\left\{\frac{\mathbf{C}_{12}^H}{\mathbf{C}_{11}}\right\}\frac{P(\boldsymbol{\omega})}{P_D(\boldsymbol{\omega})} = \mathrm{Re}\{r\}\frac{s}{1+s}\frac{P(\boldsymbol{\omega})}{P_D(\boldsymbol{\omega})} \tag{36}$$

where:

$$P(\boldsymbol{\omega}) = \iint_{x+t\geq T} f_{\chi_2^2}\left(x, \left|\frac{\mathbf{C}_{12}}{\mathbf{C}_{11}}\right|^2 t, \frac{\det(\mathbf{C})}{\mathbf{C}_{11}}\right) f_{\chi_2^1}(t,0,\mathbf{C}_{11})\, dxdt$$

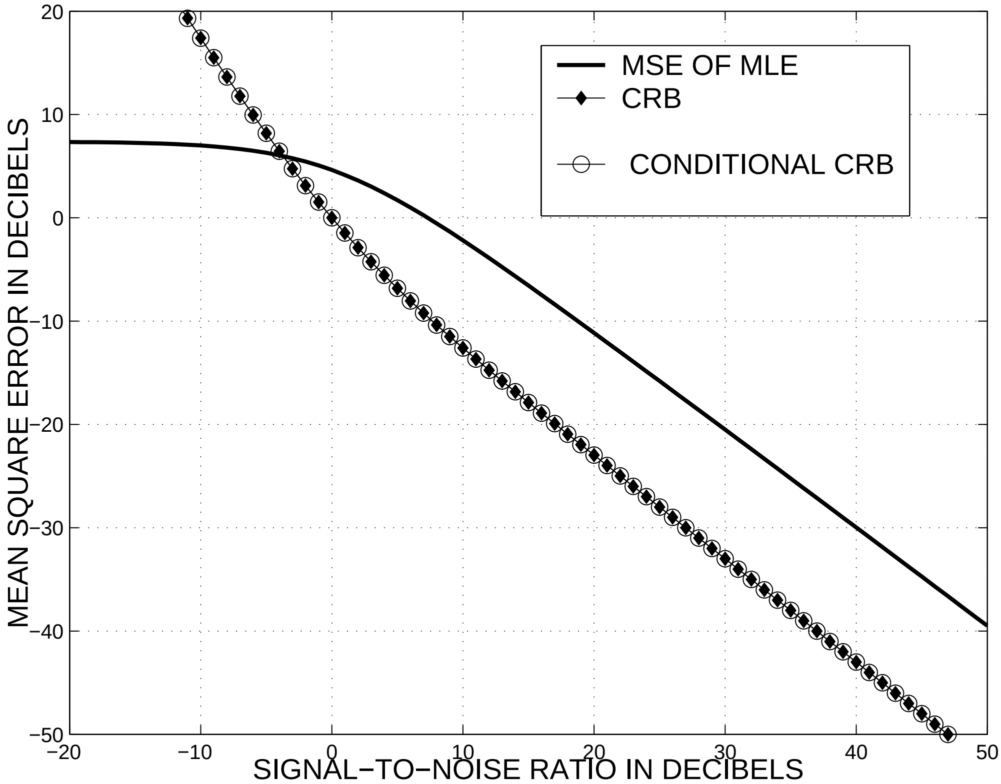


Fig. 1. MSE of MLE and CRB of $\mathrm{Re}\{\widehat{r}\}$ versus Signal-to-Noise Ratio, $P_{FA} = 0.9999$

The calculation of the CRB according to (33) then requires the evaluation of:

$$\frac{\partial \rho(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} = \frac{P(\boldsymbol{\omega})}{P_D(\boldsymbol{\omega})} \frac{\partial \left(\frac{s}{1+s} \frac{r+r^H}{2} \right)}{\partial \boldsymbol{\omega}} + \frac{\partial \left(\frac{P(\boldsymbol{\omega})}{P_D(\boldsymbol{\omega})} \right)}{\partial \boldsymbol{\omega}} \left(\frac{s}{1+s} \frac{r+r^H}{2} \right)$$

If $r = 0$ (see §IV) then $\boldsymbol{\omega} = (s, 0, 0)$ and:

$$\frac{\partial \rho(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} = \frac{P(\boldsymbol{\omega})}{P_D(\boldsymbol{\omega})} \left(\frac{\mathrm{Re}\{r\}}{(1+s)^2}, \frac{s}{1+s}\frac{1}{2}, \frac{s}{1+s}\frac{1}{2} \right)^T = \frac{P(s,0,0)}{P_D(s,0,0)} \left(0, \frac{s}{1+s}\frac{1}{2}, \frac{s}{1+s}\frac{1}{2} \right)^T$$

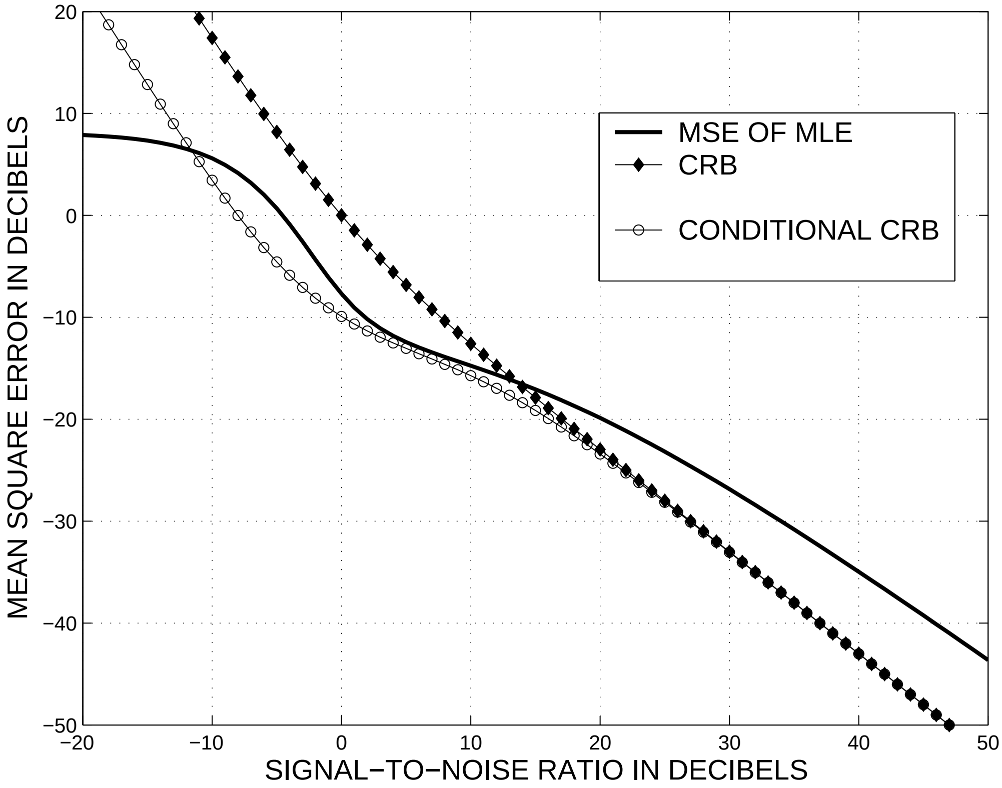


Fig. 2. MSE of MLE and CRB of $\mathrm{Re}\{\widehat{r}\}$ versus Signal-to-Noise Ratio , $P_{FA} = 10^{-4}$

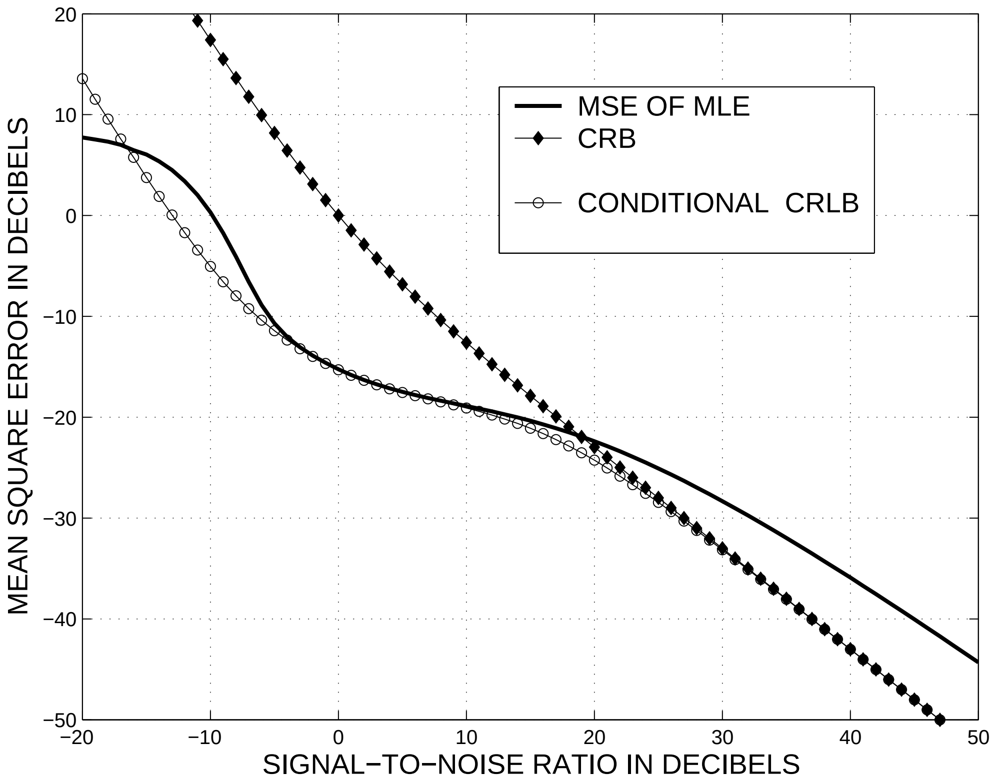


Fig. 3. MSE of MLE and CRB of $\mathrm{Re}\{\widehat{r}\}$ versus Signal-to-Noise Ratio , $P_{FA} = 10^{-10}$

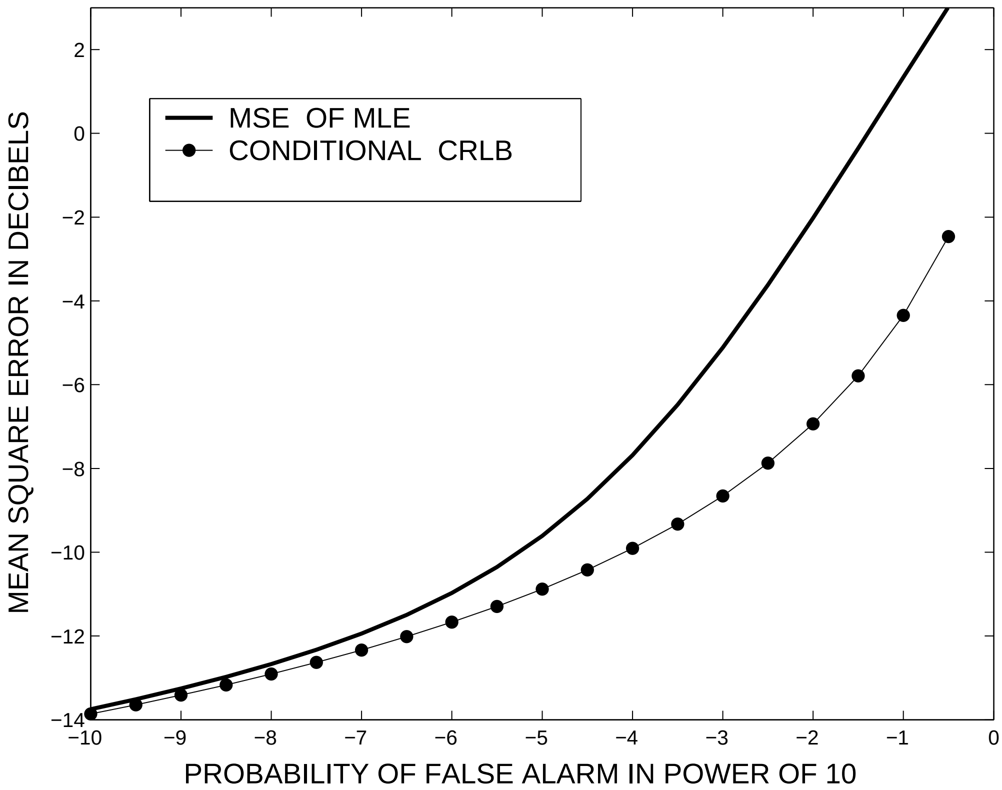


Fig. 4. Relative Efficiency of MLE of $\mathrm{Re}\{r\}$ versus Probability of False Alarm, SNR = 0 dB

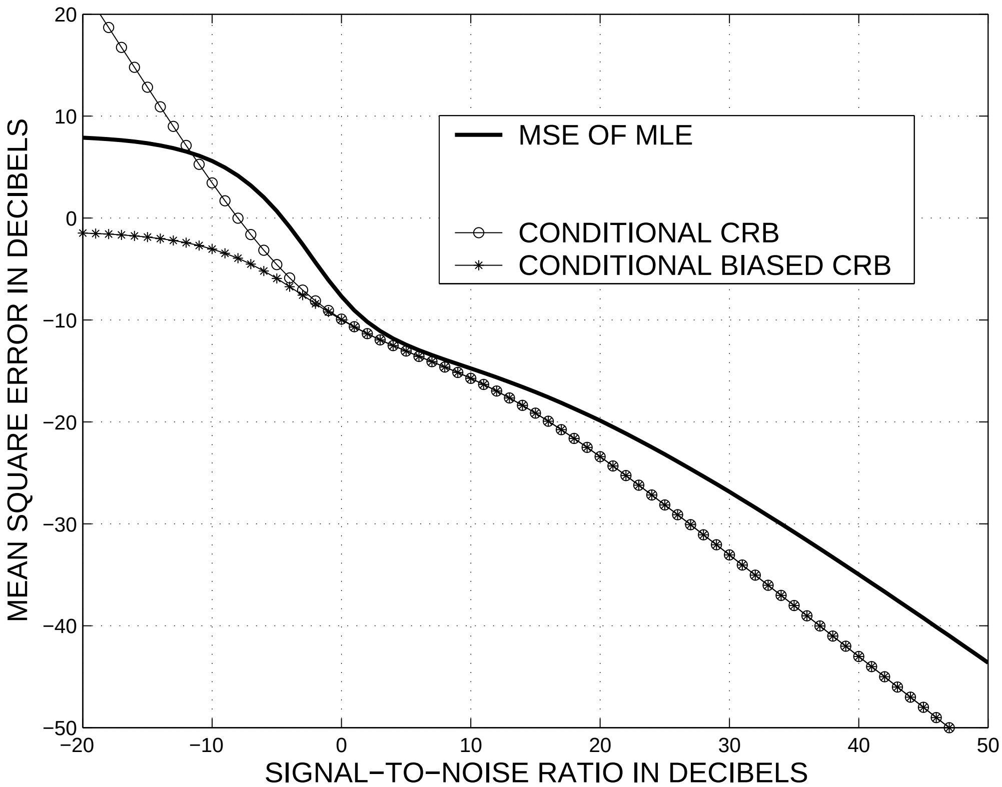


Fig. 5. MSE of MLE and CRB of $\mathrm{Re}\{\widehat{r}\}$ versus Signal-to-Noise Ratio, $P_{FA} = 10^{-4}$

F. A New Barankin Bound Approximation for the Prediction of the Threshold Region Performance of Maximum Likelihood Estimators (IEEE TSP)

IEEE Trans. on SP, 56(11): 5319-5333, 2008.

A New Barankin Bound Approximation for the Prediction of the Threshold Region Performance of Maximum-Likelihood Estimators

Eric Chaumette, Jerome Galy, Angela Quinlan, Pascal Larzabal

Abstract

It is well known that the ML estimator exhibits a threshold effect, i.e. a rapid deterioration of estimation accuracy below a certain SNR or number of snapshots. This effect is caused by outliers and is not captured by standard tools such as the Cramér-Rao bound (CRB). The search of the SNR threshold value (where the CRB becomes unreliable for prediction of ML estimator variance) can be achieved with the help of the Barankin bound (BB), as proposed by many authors. The major drawback of the BB, in comparison with the CRB, is the absence of a general analytical formula, which compels to resort to a discrete form, usually the Mcaulay-Seidman Bound (MSB), requesting the search of an optimum over a set of test points. In this paper, we propose a new practical BB discrete form which provides, for a given set of test points, an improved SNR threshold prediction in comparison with existing approximations (MSB, Abel bound, Mcaulay-Hofstetter bound), at the expense of the computational complexity increased by a factor $\leq (P+1)^3$ where P is the number of unknown parameters. We have derived its expression for the general Gaussian observation model to be used in place of existing approximations.

Index Terms: Deterministic parameter estimation, MSE lower bounds

EDICS: SAS-STAT

I. Introduction

Minimal performance bounds allow for calculation of the best performance that may be achieved in the Mean Square Error (MSE) sense, when estimating a parameter of a signal corrupted by noise. There are two main categories of lower bounds [1]. Those that evaluate the "locally best" behaviour of the estimator and those that consider the "globally best" performance. In the first case, the parameters being estimated are considered to be deterministic, whereas the second category considers the parameters as random variables with an *a priori* probability. This paper is concerned with the first category of bounds concerning deterministic parameters. Historically the first MSE lower bound for deterministic parameters to be derived was the Cramér-Rao Bound (CRB) [2][3][4], which has been the most widely used since. Its popularity is largely due to its simplicity of calculation, the fact that in many cases it can be achieved asymptotically (high SNR [5] and/or large number of snapshots [2]) by Maximum Likelihood Estimators (MLE) [1][6], and last but not least, its noticeable property of being the lowest bound on the MSE of unbiased estimators, since it is derived from the *weakest* formulation of unbiasedness at the vicinity of any selected value of the parameters [7][8][9]. This initial characterization of locally unbiased estimators has been improved first by Bhattacharyya's works [1][7][10] which refined the characterization of local unbiasedness,

Eric Chaumette is with ONERA, the French Aerospace Lab, DEMR/TSI, Chemin de la Hunière 91120 Palaiseau, France. E-mail: eric.chaumette@onera.fr, Phone: +33 1 69 93 64 30, Fax: +33 1 69 93.62 69

J. Galy is with LIRMM (Laboratoire d'Informatique, de Robotique et de Micro-Electronique de Montpellier) 161 rue ada 34392 Montpellier Cedex 5, France. E-mail: galy@lirmm.fr, Phone: +33 4 67 41 86 43, Fax: +33 4 67 41 85 00.

Angela Quinlan is Media Interaction Group, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8568, Japan, E-mail: angela-quinlan@aist.go.jp

P. Larzabal is with SATIE, ENS Cachan, CNRS, Universud, 61 av. du Président Wilson - 94235 Cachan Cedex, France. E-mail: pascal.larzabal@satie.ens-cachan.fr, Phone: +33 1 47 40 27 09.

This work was presented in part at the IEEE Internationnal Conference on Acoustics Speech and Signal Processing, Toulouse, France, May 2006,

This work has been partly funded by the European Network of excellence NEWCOM++ under the number 216715

and significantly generalized by Barankin works [7], who established the general form of the greatest lower bound of any absolute moment of an unbiased estimator. In the particular case of MSE - absolute moment of order two - his work allows the derivation of the highest lower bound on MSE (BB) since it takes into account the *strongest* formulation of unbiasedness, that is to say unbiasedness over an interval of parameter values including the selected value (see section II). Unfortunately the BB is the solution of an integral equation [11][12] with a generally incomputable analytic solution. Therefore, since then, numerous works [8][9][13][14][15][16][17][18] have been devoted to deriving computable approximations of the BB and have shown that the CRB and the BB can be regarded as key representatives of two general classes of bounds, respectively the Small-Error bounds and the Large-Error bounds. These works have also shown that in non-linear estimation problems three distinct regions of operation can be observed. In the asymptotic region, *i.e.* at a high number of independent snapshots and/or high SNR, the MSE is small and, in many cases, close to the Small-Error bounds. In the *a priori* performance region where the number of independent snapshots and/or the SNR are very low, the observations provide little information and the MSE is close to that obtained from the prior knowledge about the problem. Between these two extremes, there is an additional ambiguity region, also called the transition region. In this region, the MSE of MLEs usually deteriorates rapidly with respect to existing MSE lower bounds (Large or Small) and exhibits a threshold behaviour corresponding to a ”performance breakdown”. The nature of this phenomenon originates in the behaviour of the likelihood function in the ”threshold” area where it tends to generate outliers [9][19]. Small-Error bounds are not able to handle the threshold phenomena (*i.e.* when the performance breakdown appears), whereas it is revealed by Large-Error bounds (BB) that can be used to predict the threshold value.

Another possible approach to handle the threshold phenomena is to approximate the MSE of an estimator by a model of operation, as in the method of interval errors (MIE) [1] which assumes that $MSE \simeq (1 - P_{outlier})\, MSE_{asymp}$ $P_{outlier} MSE_{apriori}$, where MSE_{asymp} is the asymptotic MSE, $MSE_{apriori}$ is the MSE in the a priori region and $P_{outlier}$ is the probability of apparition of an outlier. Although MIE has been shown to provide accurate MSE estimation over the three regions [19][20][21][22], obtaining the components necessary to apply this approach to a specific algorithm can be quite nontrivial. Indeed, the successful application of MIE requires good approximations of two quantities: interval error probabilities ($P_{outlier}$) and the asymptotic MSE performance (MSE_{asymp}). And both of these quantities are algorithm dependent. Therefore methods such as MIE do not allow for investigation of fundamental limits of a given estimation problem as it can be done with lower bounds.

On the other hand, if deterministic lower bounds does not suffer from a lack of generality as they are not algorithm dependent, they still have a few drawbacks. A major one is that they do not take into account the support of the parameters. Consequently, when the parameters have a finite support, they cannot give the fundamental limits of an estimator in terms of MSE over all the three regions (see section VI). To fill this lack, when the parameters are assumed to be random, other bounds have been derived: the so-called Bayesian bounds. These bounds take into account the support of the parameters throughout an a priori probability density function and can give the fundamental limits in terms of MSE over all the three regions [23][24][25]. However they are pointless when the parameters are deterministic. A second drawback is that the computation of deterministic lower bounds is relevant only for unbiased estimators. Indeed if any known bias can be taken into account in their formulation (see section II), the bias depends on the specific estimator and furthermore is hardly ever known in practice. Therefore, one must keep in mind that the achievable performance obtained with deterministic lower bounds becomes less informative as the SNR decreases, since most realizable estimators (including MLEs) can not remain unbiased in that region of operation [15]. Nevertheless, as the MLE can be still considered as unbiased at the start of the transition region [15], Large-Error bounds has been shown to handle the threshold phenomena and to provide an estimation of the threshold value, at the expense of some computational cost. Indeed, the BB approximations request the search of an optimum over a set of test points and their tightness depends on the chosen set of test points (see section II). Therefore, for a given set of test points, the search for an easily computable and tight approximation of the BB is still a subject worth investigation for two reasons. First, from a theoretical standpoint, it provides a more accurate estimation of the achievable performance of the ”locally best” unbiased estimator. Second, a more accurate knowledge of the BB should allow a better prediction of the SNR threshold value and avoid misleading conclusions - too optimistic - being drawn from the computation of the CRB as the SNR decreases.

As a contribution to this research effort, we derive a general class of BB approximations based on a generalization of the method introduced by [11] and [14], which has the advantage of a clear interpretation of the hypotheses associated with the different lower bounds (see section III). This generalization includes all previously derived bounds, and provides a meaningful way to classify them (see section III-A). Moreover it suggests a new practical BB approximation which is tighter than existing approximations [8][9][10][13][15], for a given set of test points, at the expense of the computational complexity increased by a factor $\leq (P+1)^3$ where P is the number of unknown parameters (see sections III-B and IV). As a consequence, the new practical BB approximation is expected to provide an improved SNR threshold prediction as well. Although we have derived its expression for the general Gaussian observation model (see Appendix), we focus on the two most used cases : the complex circular deterministic and stochastic signal models (see sections V-B and V-A). Then, with the help of a reference estimation problem (single tone estimation, see section VI), we provide an illustration of the benefit of using the new practical BB approximation for SNR threshold prediction.

Last, for sake of legibility, we have chosen to focus on the SNR threshold effect; nevertheless the same phenomena exist with respect to the number of independent snapshots and the proposed approximation of the Barankin bound can also be used to predict the threshold value in that case.

II. An overview of Barankin Bound literature

For the sake of simplicity we will focus on the estimation of a single real function $g(\theta)$ of a single unknown real deterministic parameter θ. In the following, unless otherwise stated, $\mathbf{x}$ denotes the random observation vector[1] of dimension M, Ω the observations space, and $p(\mathbf{x};\theta)$ the probability density function (p.d.f.) of $\mathbf{x}$ depending on $\theta \in \Theta$, where Θ denotes the parameter space. Let $L^2(\Omega)$ be the real Hilbert space of square integrable functions over Ω.

A. On lower bounds and norm minimization

A fundamental property of the MSE of any estimator $\widehat{g(\theta^0)}(\mathbf{x}) \in L^2(\Omega)$ of $g(\theta^0)$, where θ^0 is a selected value of the parameter θ, is that it is a norm $\|.\|$ associated with a particular scalar product $\langle \mid \rangle_\theta$:

$$\begin{aligned} MSE_{\theta^0}\left[\widehat{g(\theta^0)}(\mathbf{x})\right] &= \left\|\widehat{g(\theta^0)}(\mathbf{x}) - g(\theta^0)\right\|^2_{\theta^0} \\ \langle g(\mathbf{x}) \mid h(\mathbf{x})\rangle_{\theta^0} &= E_{\theta^0}[g(\mathbf{x})h(\mathbf{x})] = \int g(\mathbf{x})h(\mathbf{x})p(\mathbf{x};\theta^0)\,d\mathbf{x} \end{aligned} \tag{1}$$

In the search for a lower bound on the MSE, this property allows the use of two equivalent fundamental results: the generalization of the Cauchy-Schwartz inequality to Gram matrices (generally referred to as the "covariance inequality" [15][26]) and the minimization of a norm under linear constraints [11][14][27]. Nevertheless, we shall prefer the "norm minimization" form as its use provides a better understanding of the hypotheses associated with the different lower bounds on the MSE. Then, let $\mathbb{U}$ be an Euclidean vector space of any dimension (finite or infinite) on the body of real numbers $\mathbb{R}$ which has a scalar product $\langle \mid \rangle$. Let $\mathbf{c}^{[1,K]} = (\mathbf{c}^1, \ldots, \mathbf{c}^K)$ be a free family of K vectors of $\mathbb{U}$ and $\mathbf{v} = (v_1, \ldots, v_K)^T$ a vector of $\mathbb{R}^K$. The problem of the minimization of $\|\mathbf{u}\|^2$ under the K linear constraints $\langle \mathbf{u} \mid \mathbf{c}^k \rangle = v_k$, $k \in [1,K]$ then has the solution [11][14][27]:

$$\begin{aligned} \min\left\{\|\mathbf{u}\|^2\right\} &= \mathbf{v}^T\mathbf{G}_\mathbf{c}^{-1}\mathbf{v} \quad \text{for } \mathbf{u}_{opt} = \sum_{k=1}^{K}\eta_k\mathbf{c}^k \\ (\eta_1, \ldots, \eta_K)^T &= \boldsymbol{\eta} = \mathbf{G}_\mathbf{c}^{-1}\mathbf{v}, \quad [\mathbf{G}_\mathbf{c}]_{k',k} = \left\langle \mathbf{c}^k \mid \mathbf{c}^{k'} \right\rangle \end{aligned} \tag{2}$$

[1] The notational convention adopted is as follows: italic indicates a scalar quantity, as in a; lower case boldface indicates a vector quantity, as in $\mathbf{a}$; upper case boldface indicates a matrix quantity, as in $\mathbf{A}$. The n-th row and m-th column element of the matrix $\mathbf{A}$ will be denoted by $a_{n,m}$ or $[\mathbf{A}]_{n,m}$. The n-th coordinate of the vector $\mathbf{a}$ will be denoted by a_n or $[\mathbf{a}]_n$. $\mathrm{Re}\{A\}$ is the real part of A and $\mathrm{Im}\{A\}$ is the imaginary part of A. The complex conjugaison of a quantity is indicated by a superscript * as in $\mathbf{A}^*$. The matrix transpose is indicated by a superscript ^{T} as in $\mathbf{A}^T$, and the complex conjugate plus matrix transpose is indicated by a superscript ^{H} as in $\mathbf{A}^H$. $|\mathbf{A}|$ is the determinant of the square matrix $\mathbf{A}$. $tr(\mathbf{A})$ is the trace of the square matrix $\mathbf{A}$. $[\mathbf{A} \mid \mathbf{B}]$ denotes the matrix resulting from the horizontal concatenation of matrices $\mathbf{A}$ and $\mathbf{B}$. $\mathbf{I}_M$ is the identity matrix of order M. If $\mathbf{a}$ is a column vector, then $\overrightarrow{\mathbf{a}} = (\mathbf{a}^T, \mathbf{a}^H)^T$ denotes the column vector resulting from the vertical concatenation of vectors $\mathbf{a}$ and $\mathbf{a}^*$. $\mathcal{M}(N,P)$ denotes the vector space of matrices with N rows and P columns.

B. Application to Barankin Bound derivation

As introduced by Barankin, the ultimate constraint that an unbiased estimator $\widehat{g\left(\theta^0\right)}(\mathbf{x})$ of $g\left(\theta^0\right)$ should verify is to be unbiased for all admissible values of the unknown parameter θ:

$$E_\theta\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\right]=g(\theta),\ \forall\theta\in\Theta. \tag{3}$$

In this case the problem of interest becomes:

$$\min\left\{MSE_{\theta^0}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\right]\right\}\text{ under }E_\theta\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\right]=g(\theta),\ \forall\theta\in\Theta \tag{4}$$

and corresponds to the search for the locally-best (at θ^0) unbiased estimator. This problem can be solved by applying the work of Barankin [7] that has been supported by many other studies [9][11][14][28] aimed not only at expressing the principal results of Barankin's (mathematical) theory in a form accessible to most engineers, but also at obtaining "computable" lower bounds approximating the BB. In the following we provide a didactic synthesis of these studies, highlighting all their key results (5) (7) (8) (9) (10a-10d). If $\widehat{g\left(\theta^0\right)}(\mathbf{x})$ is an unbiased estimator of $g\left(\theta^0\right)$ in the Barankin sense (3), then for any test point $\theta^n\in\Theta$:

$$E_{\theta^n}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\right]=g\left(\theta^n\right)=\int\widehat{g\left(\theta^0\right)}(\mathbf{x})\,p\left(\mathbf{x};\theta^n\right)d\mathbf{x}.$$

Consequently, $\forall\mathbf{w}=[w_1,\ldots,w_N]\in\mathbb{R}^N$:

$$E_{\theta^0}\left[\left(\widehat{g\left(\theta^0\right)}(\mathbf{x})-g\left(\theta^0\right)\right)\left(\sum_{n=1}^N w_n\frac{p\left(\mathbf{x};\theta^n\right)}{p\left(\mathbf{x};\theta^0\right)}\right)\right]=\sum_{n=1}^N w_n\left(g\left(\theta^n\right)-g\left(\theta^0\right)\right)$$

Therefore, according to (2), the minimization of $MSE_{\theta^0}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\right]$ under the constraint as above - valid for any subset of test points $\{\theta^n\}^{[1,N]}$ of Θ and $\mathbf{w}\in\mathbb{R}^N$ - implies [7]:

$$MSE_{\theta^0}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\right]\geq\lim_{N\to\infty}\sup_{\mathbf{w},\{\theta^n\}^{[1,N]}}\frac{\left[\sum\limits_{n=1}^N w_n\left(g\left(\theta^n\right)-g\left(\theta^0\right)\right)\right]^2}{E_{\theta^0}\left[\left(\sum\limits_{n=1}^N w_n\frac{p(\mathbf{x};\theta^n)}{p(\mathbf{x};\theta^0)}\right)^2\right]} \tag{5}$$

which is the original form of the BB on MSE. A more concise form can be derived by noting that [9]:

$$\frac{\left[\sum\limits_{n=1}^N w_n\left(g\left(\theta^n\right)-g\left(\theta^0\right)\right)\right]^2}{E_{\theta^0}\left[\left(\sum\limits_{n=1}^N w_n\frac{p(\mathbf{x};\theta^n)}{p(\mathbf{x};\theta^0)}\right)^2\right]}=\frac{\left(\mathbf{w}^T\boldsymbol{\delta}\mathbf{g}\right)^2}{\mathbf{w}^T\mathbf{R}\mathbf{w}}\text{ where }\left\{\begin{array}{c}[\mathbf{R}]_{n,n'}=E_{\theta^0}\left[\frac{p(\mathbf{x};\theta^n)}{p(\mathbf{x};\theta^0)}\frac{p\left(\mathbf{x};\theta^{n'}\right)}{p(\mathbf{x};\theta^0)}\right]\\ \delta g_n=g\left(\theta^n\right)-g\left(\theta^0\right)\end{array}\right. \tag{6}$$

Indeed, since $\frac{(\mathbf{w}^T\boldsymbol{\delta}\mathbf{g})^2}{\mathbf{w}^T\mathbf{R}\mathbf{w}}\leq\boldsymbol{\delta}\mathbf{g}^T\mathbf{R}^{-1}\boldsymbol{\delta}\mathbf{g}$ and reaches its maximum value for $\mathbf{w}=\lambda\mathbf{R}^{-1}\boldsymbol{\delta}\mathbf{g}$, $\lambda\in\mathbb{R}$, then the "reduced" form of the BB [9] is:

$$MSE_{\theta^0}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\right]\geq\lim_{N\to\infty}\sup_{\{\theta^n\}^{[1,N]}}\left\{\boldsymbol{\delta}\mathbf{g}^T\mathbf{R}^{-1}\boldsymbol{\delta}\mathbf{g}\right\} \tag{7}$$

It is then worth noting that (7) is also the solution of:

$$\min\left\{MSE_{\theta^0}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\right]\right\}\text{ under }E_{\theta^n}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\right]=g\left(\theta^n\right),\ \{\theta^n\}^{[1,N]}\in\Theta \tag{8}$$

which corresponds to the fact that the greatest lower bound on the MSE for a finite number of test points $\{\theta^n\}^{[1,N]}$ is obtained by simply expressing the "unbiased" constraint at the test points [9][14]. Therefore any additional weighting $(\mathbf{w})$ leads to a lower bound of inferior or equal value. Consequently the unbiased and locally best estimator (or Locally Minimum Variance Unbiased estimator) $\widehat{g\left(\theta^0\right)}_{lmvu}(\mathbf{x})$ satisfies [14]:

$$\lim_{N\longrightarrow\infty}\begin{cases} \mathbf{R}\left(\frac{\mathbf{w}}{\lambda}\right)=\boldsymbol{\delta}\mathbf{g} \\ \widehat{g\left(\theta^0\right)}_{lmvu}(\mathbf{x})-g\left(\theta^0\right)=\sum_{n=1}^{N}\frac{w_n}{\lambda}\frac{p(\mathbf{x};\theta^n)}{p(\mathbf{x};\theta^0)} \\ MSE_{\theta^0}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\right]\geq MSE_{\theta^0}\left[\widehat{g\left(\theta^0\right)}_{lmvu}(\mathbf{x})\right]=\boldsymbol{\delta}\mathbf{g}^T\mathbf{R}^{-1}\boldsymbol{\delta}\mathbf{g}=\boldsymbol{\delta}\mathbf{g}^T\left(\frac{\mathbf{w}}{\lambda}\right)\end{cases} \tag{9}$$

that leads to, defining $\frac{1}{\lambda}=d\theta=\theta^{n+1}-\theta^n$, [11]:

$$\int K\left(\theta,\theta'\right)w\left(\theta'\right)d\theta'=g\left(\theta\right)-g\left(\theta^0\right) \tag{10a}$$

$$K\left(\theta,\theta'\right)=\int\frac{p\left(\mathbf{x};\theta\right)p\left(\mathbf{x};\theta'\right)}{p\left(\mathbf{x};\theta^0\right)}d\mathbf{x} \tag{10b}$$

$$\widehat{g\left(\theta^0\right)}_{lmvu}(\mathbf{x})-g\left(\theta^0\right)=\int\frac{p\left(\mathbf{x};\theta\right)}{p\left(\mathbf{x};\theta^0\right)}w\left(\theta\right)d\theta \tag{10c}$$

$$MSE_{\theta^0}\left[\widehat{g\left(\theta^0\right)}(\mathbf{x})\right]\geq MSE_{\theta^0}\left[\widehat{g\left(\theta^0\right)}_{lmvu}(\mathbf{x})\right]=\int\left(g\left(\theta\right)-g\left(\theta^0\right)\right)w\left(\theta\right)d\theta \tag{10d}$$

Unfortunately, in most practical cases, it is impossible to find either the limit of (7) or an analytical solution of (10a-b) to obtain an explicit form of $\widehat{g\left(\theta^0\right)}_{lmvu}(\mathbf{x})$ and of the greatest lower bound on the MSE, which somewhat limits its interest. Therefore, the search for an easily computable but tight approximation of the BB is a subject of great theoretical and practical importance.

III. Approximating the Barankin Bound

All known bounds on the MSE have been shown to be different solutions of the same norm minimization problem (1) under sets of appropriate linear constraints (possibly infinite but countable) expressed at the true value (θ^0) only (Small Error bounds) or at a set of test points including the true value $\left\{\boldsymbol{\theta}^0\right\}\cup\left\{\boldsymbol{\theta}^n\right\}^{[1,N]}$ [11][14] (Large Error bounds). Therefore, any BB approximation (Large Error bound) requests the search of an optimum over a set of test points $\left\{\theta^n\right\}^{[1,N]}$ and its tightness depends on the set of used test points (see also [17]). For that reason, for a given set of test points, the search for an easily computable and tighter approximation of the BB is a subject worth investigation. It is worth noting that the nature of the rationale used to derive each Barankin bound approximation have determined its final practical form proposed by each author. For example, in [9] the main goal was to reduce the complexity of use of the Barankin bound by substituting the simplified form (7) for the initial form (5). In [13] and generalized in [15], the rationale is to combine a Small Error bound (Cramér-Rao [13] or Bhattacharyya [15]) with a Large Error bound (McAulay-Seidman [9]) in order to obtain an hybrid bound which accounts for both local and large errors and is able to handle the threshold phenomena. As mentioned by Abel [15] and previously analysed by Glave [14], the expected effect of adding Large Error bound constraints (test points) to Small Error bound constraints (derivatives at θ^0) is to restrict the class of viable estimators and therefore potentially increase the minimum variance. Therefore, according to this approach, a simple extension mentioned by Abel would be to add additional constraints by taking into account higher derivative constraints not just at the true value (θ^0), but at the test points (θ^n) as well. This generalisation can be introduced as the main result of an equivalent but more intuitive rationale that explicitly exploits the dependence of the tightness of the BB approximation on the quality of unbiasedness reached by a given class of estimators. Note that this rationale can also be regarded as another extension (to several test points) of the Bhattacharyya bound constraints (11). The consequence is the form of the novel practical BB approximation naturally suggested (14), as shown hereinafter. Indeed, a simple method for exploration of the unbiasedness assumption, is to consider a general family of local approximations of unbiasedness (3). A natural one is provided by the Taylor series expansion. Therefore, let us consider that both $p\left(\mathbf{x};\theta\right)$ and $g\left(\theta\right)$ can be approximated by piecewise Taylor series expansions of order L^n, that is to say the parameter space Θ can be partitioned in N real sub-intervals I^n over which $p\left(\mathbf{x};\theta\right)$ and $g\left(\theta\right)$ are piecewise continuously differentiable

L^n+1 times with respect to θ. Then, $\forall\theta^n+d\theta\in I^n$:

$$\begin{aligned} g(\theta^n+d\theta) &= g(\theta^n) \quad +\sum_{l=1}^{L^n}\frac{\partial^l g(\theta^n)}{\partial\theta^l}\frac{d\theta^l}{l!}+o\left(d\theta^{L^n}\right) \\ p\left(\mathbf{x};\theta^n+d\theta\right) &= p\left(\mathbf{x};\theta^n\right)+\sum_{l=1}^{L^n}\frac{\partial^l p\left(\mathbf{x};\theta^n\right)}{\partial\theta^l}\frac{d\theta^l}{l!}+o_{\mathbf{x}}\left(d\theta^{L^n}\right) \end{aligned}$$

where $o\left(.\right)$ denotes the usual small oh notation and $o_{\mathbf{x}}\left(d\theta\right)=\varepsilon\left(\mathbf{x},d\theta\right)d\theta$ with $\forall\mathbf{x}\in\Omega$, $\lim\limits_{d\theta\longrightarrow 0}\varepsilon\left(\mathbf{x},d\theta\right)=0$; both being the reminders of the Taylor series expansions with respect to θ. Additionally let us assume that the integrals $\int\left(\frac{\partial^l p(\mathbf{x};\theta)}{\partial\theta^l}\right)^2\frac{1}{p(\mathbf{x};\theta)}d\mathbf{x}$ converge and define piecewise continuous functions of θ on Θ, for all $\theta\in\Theta$, to allow order of integration and differentiation interchange [15]. Then, a few lines of algebra show that, on every sub-interval I^n, a possible local approximation of unbiasedness definition (3) is:

$$E_{\theta^n+d\theta}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right]=g\left(\theta^n+d\theta\right)+o\left(d\theta^{L^n}\right) \tag{11}$$

provided the following (L^n+1) linear constraints are verified:

$$\int\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\frac{\partial^l p\left(\mathbf{x};\theta^n\right)}{\partial\theta^l}d\mathbf{x}=\frac{\partial^l g(\theta^n)}{\partial\theta^l},\quad l\in[0,L^n]$$

or equivalently:

$$E_{\theta^0}\left[\left(\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)-g\left(\theta^0\right)\right)\frac{\frac{\partial^l p(\mathbf{x};\theta^n)}{\partial\theta^l}}{p\left(\mathbf{x};\theta^0\right)}\right]=\left[\frac{\partial^l\left(g(\theta)-g(\theta^0)\right)}{\partial\theta^l}\right]_{\theta^n},\quad l\in[0,L^n] \tag{12}$$

Looked at from that point of view, the local approximation of unbiasedness definition used in (8) is simply equivalent to:

$$E_{\theta^n+d\theta}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right]=g\left(\theta^n+d\theta\right)+o\left(1\right)$$

Then, the set of $\sum\limits_{n=1}^{N}(L^n+1)$ constraints (12) deriving from the N piecewise local approximation of unbiasedness (3) defines a given approximation of the BB denoted by $\mathrm{BB}_{L^1,\ldots,L^N}^{I^1,\ldots,I^N}$ (2):

$$\begin{aligned} \mathrm{BB}_{L^1,\ldots,L^N}^{I^1,\ldots,I^N} &= \mathbf{v}^T\mathbf{G}^{-1}\mathbf{v} \\ \mathbf{v} &= \left[\mathbf{v}_1^T,\ldots,\mathbf{v}_N^T\right]^T,\quad \mathbf{G}=E_{\theta^0}\left[\mathbf{c}\mathbf{c}^T\right] \\ \mathbf{v}_n &= \left[g(\theta^n)-g(\theta^0),\frac{\partial g(\theta^n)}{\partial\theta},\ldots,\frac{\partial^{L^n}g(\theta^n)}{\partial\theta^{L^n}}\right]^T \\ \mathbf{c} &= \left[\mathbf{c}_1^T,\ldots,\mathbf{c}_N^T\right]^T,\quad \mathbf{c}_n=\frac{1}{p\left(\mathbf{x};\theta^0\right)}\left[p\left(\mathbf{x};\theta^n\right),\frac{\partial p\left(\mathbf{x};\theta^n\right)}{\partial\theta},\ldots,\frac{\partial^{L^n}p\left(\mathbf{x};\theta^n\right)}{\partial\theta^{L^n}}\right]^T \end{aligned} \tag{13}$$

As looked-for, if $\min\left\{L^1,\ldots,L^N\right\}$ tends to infinity, a straightforward exercise in mean square convergence establishes that $\mathrm{BB}_{L^1,\ldots,L^N}^{I^1,\ldots,I^N}$ converges in mean-square to the BB [14][29]. An immediate generalization of expression (13) consists of taking its supremum over existing degrees of freedom (sub-interval definitions and series expansion orders). As mentioned above, the merit of the proposed rationale (piecewise approximations of unbiasedness) is to provide a straightforward insight into the unbiasedness assumption from its *weakest* formulation (CRB) to its *strongest* formulation (BB) and, to a certain extent, a measure of the relative tightness of the resulting BB approximations. As a consequence, it provides a meaningful way to classify all existing bounds and suggest an obvious new practical BB approximation (14), as highlighted in the two next sections.

A. A different look at existing BB approximations

Designating the BB approximations as:

• N-piecewise BB approximation of homogeneous order L, if on all sub-intervals I^n the series expansions are of the same order L,

• N-piecewise BB approximation of heterogeneous orders $\left\{L^1,...,L^N\right\}$, if otherwise,

we can provide a new look at previously derived MSE lower bounds:

• the CRB [2] is a 1-piecewise BB approximation of homogeneous order 1, since the constraints are:

$$E_{\theta^0}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right]=g\left(\theta^0\right),\ E_{\theta^0}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\frac{\frac{\partial p\left(\mathbf{x};\theta^0\right)}{\partial\theta}}{p\left(\mathbf{x};\theta^0\right)}\right]=\frac{\partial g\left(\theta^0\right)}{\partial\theta}$$

• the Bhattacharyya bound [1][10] of order L is a 1-piecewise BB approximation of homogeneous order L, since the constraints are:

$$E_{\theta^0}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right]=g\left(\theta^0\right),\ E_{\theta^0}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\frac{\frac{\partial^l p\left(\mathbf{x};\theta^0\right)}{\partial\theta^l}}{p\left(\mathbf{x};\theta^0\right)}\right]=\frac{\partial^l g\left(\theta^0\right)}{\partial\theta^l},\ l\in[1,L]$$

• the Hammersley-Chapman-Robbins bound (HCRB) [8] is the supremum of a 2-piecewise BB approximation of homogeneous order 0, over a set of constraints of type:

$$E_{\theta^0}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right]=g\left(\theta^0\right),\ E_{\theta_1}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right]=g\left(\theta_1\right)$$

• the McAulay-Seidman bound (MSB^N) [9] with $N+1$ test points is an $(N+1)$-piecewise BB approximation of homogeneous order 0, since the constraints are:

$$E_{\theta^0}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right]=g\left(\theta^0\right),\ E_{\theta^n}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right]=g\left(\theta^n\right),\ n\in[1,N]$$

• the Hybrid Barankin-Bhattacharyya bound (HBB$_L^N$) [15] is an $(N+1)$-piecewise BB approximation of heterogeneous order $\{L,0,...,0\}$, since the constraints are:

$$\begin{cases} E_{\theta^0}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right]=g\left(\theta^0\right),\ E_{\theta^0}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\frac{\frac{\partial^l p\left(\mathbf{x};\theta^0\right)}{\partial\theta^l}}{p\left(\mathbf{x};\theta^0\right)}\right]=\frac{\partial^l g\left(\theta^0\right)}{\partial\theta^l},\ l\in[1,L] \\ E_{\theta^n}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right]=g\left(\theta^n\right),\ n\in[1,N] \end{cases}$$

B. A new practical BB approximation

The introduced formalism suggests a very straightforward practical BB approximation - denoted BB$_1^N$ in the following -: the $(N+1)$-piecewise BB approximation of homogeneous order 1 characterized by the set of constraints:

$$\begin{cases} E_{\theta^0}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right]=g\left(\theta^0\right),\ E_{\theta^0}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\frac{\frac{\partial p\left(\mathbf{x};\theta^0\right)}{\partial\theta}}{p\left(\mathbf{x};\theta^0\right)}\right]=\frac{\partial g\left(\theta^0\right)}{\partial\theta} \\ E_{\theta^n}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\right]=g\left(\theta^n\right),\ E_{\theta^n}\left[\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)\frac{\frac{\partial p\left(\mathbf{x};\theta^n\right)}{\partial\theta}}{p\left(\mathbf{x};\theta^n\right)}\right]=\frac{\partial g\left(\theta^n\right)}{\partial\theta} \end{cases},\ n\in[1,N] \tag{14}$$

The proposed bound $\left(\text{BB}_1^N\right)$ appears to be the generalization of the CRB when the parameter space is partitioned in more than one sub-interval, as well as the generalization of the usual BB approximation used in the open literature (the McAulay-Seidman form MSB^N of the BB (7)), and to some extent, an extension of the HBB$_1^N$. The cost is an increase of the computational complexity, as shown in the next section where we generalize the BB$_1^N$ definition to the multiple parameters case. Last, note that the introduced formalism also suggests naturally to resort to BB$_L^N$ - piecewise BB approximation of homogeneous order L - which will be tighter than BB$_1^N$. Nevertheless our aim is to derive a new practical BB approximation for the most general use possible, what generally means as simple as possible to derive analytically . The derivation of BB$_1^N$ for the general Gaussian observation model (see Appendix) clearly shows that any attempt to use higher order not only will lead to a tedious derivation but also will dramatically increases the computational cost.

IV. GENERALIZATION TO SIMULTANEOUS ESTIMATION OF SEVERAL FUNCTIONS OF MULTIPLE PARAMETERS

We consider now the case where the p.d.f. $p(\mathbf{x};\boldsymbol{\theta})$ depends on a vector of P parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_P)$ belonging to $\mathbb{R}^P$. Let $\boldsymbol{\theta}^0$ be a particular value of $\boldsymbol{\theta}$, and $\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})$ an estimator of $\mathbf{g}\left(\boldsymbol{\theta}^0\right)$ where $\mathbf{g}(\boldsymbol{\theta}) = (g_1(\boldsymbol{\theta}), \ldots, g_Q(\boldsymbol{\theta}))$ is a vector of Q real functions of $\boldsymbol{\theta}$. In the multiple parameters context, $\widehat{g_q\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})$ is a "locally unbiased" estimator of $g_q\left(\boldsymbol{\theta}^0\right)$ if:

$$E_{\boldsymbol{\theta}^0+d\boldsymbol{\theta}}\left[\widehat{g_q\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})\right] = g_q\left(\boldsymbol{\theta}^0+d\boldsymbol{\theta}\right) + o(\|d\boldsymbol{\theta}\|) = g_q\left(\boldsymbol{\theta}^0\right) + \frac{\partial g_q\left(\boldsymbol{\theta}^0\right)}{\partial \boldsymbol{\theta}^T} d\boldsymbol{\theta} + o(\|d\boldsymbol{\theta}\|) \tag{15}$$

where $\frac{\partial}{\partial \boldsymbol{\theta}^T} = \left(\frac{\partial}{\partial \theta_1}, \frac{\partial}{\partial \theta_2}, \ldots, \frac{\partial}{\partial \theta_P}\right)$, which means that, up to the first order and in the neighbourhood of $\boldsymbol{\theta}^0$, $\widehat{g_q\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})$ remains an unbiased estimator of $g_q\left(\boldsymbol{\theta}^0 + d\boldsymbol{\theta}\right)$ independently of a small variation $d\boldsymbol{\theta}$ of the parameters. Considering as well that in the neighbourhood of $\boldsymbol{\theta}^0$:

$$p\left(\mathbf{x};\boldsymbol{\theta}^0\right) = p\left(\mathbf{x};\boldsymbol{\theta}^0\right) + \frac{\partial p\left(\mathbf{x};\boldsymbol{\theta}^0\right)}{\partial \boldsymbol{\theta}^T} d\boldsymbol{\theta} + o(\|d\boldsymbol{\theta}\|),$$

the requested locally unbiased property (15) is satisfied for all components of $\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x})$ if the following linear constraints are verified:

$$\left\{\begin{array}{l} E_{\boldsymbol{\theta}^0}\left[\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x}) - \mathbf{g}\left(\boldsymbol{\theta}^0\right)\right)^T\right] = \mathbf{0}^T \\ E_{\boldsymbol{\theta}^0}\left[\left(\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}}\right)\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x}) - \mathbf{g}\left(\boldsymbol{\theta}^0\right)\right)^T\right] = \left(\frac{\partial \mathbf{g}(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}^T}\right)^T \end{array}\right. \tag{16}$$

where $\frac{\partial}{\partial \boldsymbol{\theta}} = \left(\frac{\partial}{\partial \theta_1}, \frac{\partial}{\partial \theta_2}, \ldots, \frac{\partial}{\partial \theta_P}\right)^T$. Applied to any selected value $\boldsymbol{\theta}^n$ of the set of $(N+1)$ test points $\left\{\boldsymbol{\theta}^0\right\} \cup \left\{\boldsymbol{\theta}^n\right\}^{[1,N]}$, (16) becomes:

$$\left\{\begin{array}{l} E_{\boldsymbol{\theta}^0}\left[\frac{p(\mathbf{x};\boldsymbol{\theta}^n)}{p(\mathbf{x};\boldsymbol{\theta}^0)}\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x}) - \mathbf{g}\left(\boldsymbol{\theta}^0\right)\right)^T\right] = \left(\mathbf{g}\left(\boldsymbol{\theta}^n\right) - \mathbf{g}\left(\boldsymbol{\theta}^0\right)\right)^T \\ E_{\boldsymbol{\theta}^0}\left[\left(\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta}^n)}{\partial \boldsymbol{\theta}} \frac{p(\mathbf{x};\boldsymbol{\theta}^n)}{p(\mathbf{x};\boldsymbol{\theta}^0)}\right)\left(\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}(\mathbf{x}) - \mathbf{g}\left(\boldsymbol{\theta}^0\right)\right)^T\right] = \left(\frac{\partial \mathbf{g}(\boldsymbol{\theta}^n)}{\partial \boldsymbol{\theta}^T}\right)^T \end{array}\right. \tag{17}$$

Moreover lemma (2) can be generalized to linear combinations of a family of Q vectors $\mathbf{u}^{[1,Q]} = \left(\mathbf{u}^1, \ldots, \mathbf{u}^Q\right)$ where the minimization problem becomes:

$$\min\left\{\left\|\sum_{q=1}^{Q} \lambda_q \mathbf{u}^q\right\|^2 = \boldsymbol{\lambda}^T \mathbf{G}_{\mathbf{u}} \boldsymbol{\lambda}\right\} \text{ under } \left\langle \mathbf{u}^q \mid \mathbf{c}^k \right\rangle = [\mathbf{V}]_{k,q}\ , \quad \boldsymbol{\lambda} \neq \mathbf{0} \in \mathbb{R}^Q \tag{18}$$

and leads to the matrix inequality:

$$\boldsymbol{\lambda}^T \mathbf{G}_{\mathbf{u}} \boldsymbol{\lambda} \geq \boldsymbol{\lambda}^T \left(\mathbf{V}^T \mathbf{G}_{\mathbf{c}}^{-1} \mathbf{V}\right) \boldsymbol{\lambda} \Leftrightarrow \mathbf{G}_{\mathbf{u}} \geq \mathbf{V}^T \mathbf{G}_{\mathbf{c}}^{-1} \mathbf{V} \tag{19}$$

where $[\mathbf{G}_{\mathbf{u}}]_{q',q} = \left\langle \mathbf{u}^q \mid \mathbf{u}^{q'} \right\rangle$, which is an alternative way to introduce the "covariance inequality" [15][26] focusing on the associated constraints. Then, considering $\mathbf{u}^q = \widehat{g_q\left(\boldsymbol{\theta}^0\right)}(\mathbf{x}) - g_q\left(\boldsymbol{\theta}^0\right)$, $q \in [1, Q]$ and the set of $(N+1)$ constraints systems (17), (13) becomes:

$$\mathrm{BB}_1^N = \mathbf{V}^T \mathbf{G}_{\mathbf{c}}^{-1} \mathbf{V} \text{ where} \left\{\begin{array}{l} \mathbf{G}_{\mathbf{c}} = \begin{bmatrix} \mathbf{MS} & \mathbf{H}^T \\ \mathbf{H} & \mathbf{EFI} \end{bmatrix},\ \mathbf{V} = \left[\boldsymbol{\Delta}\mathbf{G} \mid \frac{\partial \mathbf{g}(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}^T} \mid \ldots \mid \frac{\partial \mathbf{g}(\boldsymbol{\theta}^N)}{\partial \boldsymbol{\theta}^T}\right]^T \\ \boldsymbol{\Delta}\mathbf{G} = \left[\mathbf{g}\left(\boldsymbol{\theta}^0\right) - \mathbf{g}\left(\boldsymbol{\theta}^0\right) \mid \ldots \mid \mathbf{g}\left(\boldsymbol{\theta}^N\right) - \mathbf{g}\left(\boldsymbol{\theta}^0\right)\right] \end{array}\right. \tag{20}$$

and:

$$\mathbf{MS} = E_{\boldsymbol{\theta}^0}\left[\begin{pmatrix} \frac{p(\mathbf{x};\boldsymbol{\theta}^0)}{p(\mathbf{x};\boldsymbol{\theta}^0)} \\ \vdots \\ \frac{p(\mathbf{x};\boldsymbol{\theta}^N)}{p(\mathbf{x};\boldsymbol{\theta}^0)} \end{pmatrix}\begin{pmatrix} \frac{p(\mathbf{x};\boldsymbol{\theta}^0)}{p(\mathbf{x};\boldsymbol{\theta}^0)} \\ \vdots \\ \frac{p(\mathbf{x};\boldsymbol{\theta}^N)}{p(\mathbf{x};\boldsymbol{\theta}^0)} \end{pmatrix}^T\right] \tag{21a}$$

$$\mathbf{EFI} = E_{\boldsymbol{\theta}^0}\left[\begin{pmatrix} \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}} \\ \vdots \\ \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta}^N)}{\partial \boldsymbol{\theta}}\frac{p(\mathbf{x};\boldsymbol{\theta}^N)}{p(\mathbf{x};\boldsymbol{\theta}^0)} \end{pmatrix}\begin{pmatrix} \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}} \\ \vdots \\ \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta}^N)}{\partial \boldsymbol{\theta}}\frac{p(\mathbf{x};\boldsymbol{\theta}^N)}{p(\mathbf{x};\boldsymbol{\theta}^0)} \end{pmatrix}^T\right] \tag{21b}$$

$$\mathbf{H} = E_{\boldsymbol{\theta}^0}\left[\begin{pmatrix} \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}} \\ \vdots \\ \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta}^N)}{\partial \boldsymbol{\theta}}\frac{p(\mathbf{x};\boldsymbol{\theta}^N)}{p(\mathbf{x};\boldsymbol{\theta}^0)} \end{pmatrix}\begin{pmatrix} \frac{p(\mathbf{x};\boldsymbol{\theta}^0)}{p(\mathbf{x};\boldsymbol{\theta}^0)} \\ \vdots \\ \frac{p(\mathbf{x};\boldsymbol{\theta}^N)}{p(\mathbf{x};\boldsymbol{\theta}^0)} \end{pmatrix}^T\right] \tag{21c}$$

Thus $\mathbf{G_c} \in \mathcal{M}\left((P+1)\times(N+1),(P+1)\times(N+1)\right)$, $\mathbf{MS} \in \mathcal{M}\left(N+1,N+1\right)$,
$\mathbf{EFI} \in \mathcal{M}\left(P\times(N+1),P\times(N+1)\right)$, $\mathbf{H} \in \mathcal{M}\left(P\times(N+1),N+1\right)$, and $\mathbf{V} \in \mathcal{M}\left((P+1)\times(N+1),Q\right)$. As reminded hereinafter, **MS** is the McAulay-Seidman matrix. **EFI** stands for the Extended Fisher Information matrix, as it reduces to the **FI** (Fisher Information) matrix when the set of test points is reduced to $\boldsymbol{\theta}^0$ only. **H** is a kind of "hybrid" matrix.

If we restrict (17) to the point unbiased constraint terms, *i.e.*:

$$E_{\boldsymbol{\theta}^0}\left[\frac{p(\mathbf{x};\boldsymbol{\theta}^n)}{p(\mathbf{x};\boldsymbol{\theta}^0)}\left(\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}(\mathbf{x})-\mathbf{g}(\boldsymbol{\theta}^0)\right)^T\right] = E_{\boldsymbol{\theta}^n}\left[\left(\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}(\mathbf{x})-\mathbf{g}(\boldsymbol{\theta}^0)\right)^T\right] = \left(\mathbf{g}(\boldsymbol{\theta}^n)-\mathbf{g}(\boldsymbol{\theta}^0)\right)^T$$

(20) becomes:

$$MSE \geq \mathbf{\Delta G}\ \mathbf{MS}^{-1}\mathbf{\Delta G}^T$$

which is the general expression of the McAulay-Seidman bound (MSB^N, usual BB approximation) for simultaneous estimation of several functions of multiple parameters [9, eq. 7].
If we restrict the set of test points to $\boldsymbol{\theta}^0$ only, then (20) becomes:

$$MSE \geq \left[\mathbf{0} \mid \frac{\partial \mathbf{g}(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}^T}\right]\begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{FI}\end{bmatrix}^{-1}\left[\mathbf{0} \mid \frac{\partial \mathbf{g}(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}^T}\right]^T = \frac{\partial \mathbf{g}(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}^T}\mathbf{FI}^{-1}\left(\frac{\partial \mathbf{g}(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}^T}\right)^T$$

where $\mathbf{FI} = E_{\boldsymbol{\theta}^0}\left[\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}}\left(\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}}\right)^T\right]$, which is the expression of the CRB for simultaneous estimation of several functions of multiple parameters [1].
Last, if we restrict (17) to the point unbiased constraint terms for $\boldsymbol{\theta}^n \neq \boldsymbol{\theta}^0$ but keep the locally unbiased constraint at $\boldsymbol{\theta}^0$ (15), then (20) becomes:

$$MSE \geq \left[\mathbf{\Delta G} \mid \frac{\partial \mathbf{g}(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}^T}\right]\begin{bmatrix} \mathbf{MS} & \mathbf{H}^T \\ \mathbf{H} & \mathbf{FI}\end{bmatrix}^{-1}\left[\mathbf{\Delta G} \mid \frac{\partial \mathbf{g}(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}^T}\right]^T$$

which is the expression of the McAulay-Hofstetter bound (MHB^N) for simultaneous estimation of several functions of multiple parameters [13, eq. 7 and 51] (which is also equal to the Hybrid Barankin-Bhattacharyya bound (HBB$_1^N$) [15]).

Finally, the new proposed practical BB approximation (BB$_1^N$ (20)) appears to be, whatever the number of functions to estimate or the number of unknown parameters, the generalization of the CRB, the MSB^N, and the MHB^N-HBB$_1^N$. As a result:
• for any set of test points $\{\boldsymbol{\theta}^0\} \cup \{\boldsymbol{\theta}^n\}^{[1,N]}$, the BB$_1^N$ will provide a tighter approximation of the BB,
• any rationale to optimize the choice of the set of test points to maximize the MSB^N or the MHB^N-HBB$_1^N$ is

applicable to the BB_1^N to increase its tightness (see [17] for an example).
The cost is the inversion of square matrix with dimension $(P+1)\times(N+1)$ instead of $(N+1)$ for the computation of the MSB^N or instead of $(N+P+1)$ for the computation of the HBB_1^N, what leads to an increase of the computational complexity by a factor $\leq (P+1)^3$ in comparison with existing bounds.
Regardless of the computational cost, the only drawback of the BB_1^N in comparison with the MSB^N, is to require the same regularity conditions as the CRB to be computed, *i.e.* the two first derivatives of $p(\mathbf{x};\boldsymbol{\theta})$ must be differentiable with respect to $\boldsymbol{\theta}$ and these derivatives must be absolutely integrable [1, p. 66].

V. Lower Bounds for the Gaussian observation model

For sake of legibility, the analytical expressions of the components $\mathbf{MS}$, $\mathbf{EFI}$ and $\mathbf{H}$ of BB_1^N given by (21a-21c) where $\mathbf{x}$ is an M-dimensional Gaussian real vector characterized by a p.d.f. $p(\mathbf{x};\boldsymbol{\theta})$ depending on a vector of P real parameters $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_P)$ are derived in Appendix. The main interest in addressing the Gaussian real observation model is to encompass all Gaussian observation models, including the usual complex circular deterministic (conditional) or stochastic (unconditional) observation models [30]. Indeed, let us recall that an M-dimensional Gaussian complex vector $\mathbf{z}$ with mean $\mathbf{m_z}=E[\mathbf{z}]$ and covariance matrix $\mathbf{C_z}=E\left[(\mathbf{z}-\mathbf{m_z})(\mathbf{z}-\mathbf{m_z})^H\right]$, is actually a $2M$-dimensional Gaussian real vector $\mathbf{x}\sim\mathcal{N}_{2M}(\mathbf{m_x},\mathbf{C_x})$ where, up to a permutation, $\mathbf{x}=(\mathrm{Re}\{z_1\},...,\mathrm{Re}\{z_M\},$ $\mathrm{Im}\{z_1\},...,\mathrm{Im}\{z_M\})$. Therefore, a few lines of algebra [31] allows to rewrite the Gaussian real vector p.d.f. as a Gaussian complex vector p.d.f.:

$$p(\mathbf{x};\mathbf{m_x},\mathbf{C_x})=\frac{e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m_x})^T\mathbf{C_x}^{-1}(\mathbf{x}-\mathbf{m_x})}}{\sqrt{2\pi}^{2M}\sqrt{|\mathbf{C_x}|}}=p\left(\overrightarrow{\mathbf{z}};\mathbf{m}_{\overrightarrow{\mathbf{z}}},\mathbf{C}_{\overrightarrow{\mathbf{z}}}\right)=\frac{e^{-\frac{1}{2}\left(\overrightarrow{\mathbf{z}}-\mathbf{m}_{\overrightarrow{\mathbf{z}}}\right)^H\mathbf{C}_{\overrightarrow{\mathbf{z}}}^{-1}\left(\overrightarrow{\mathbf{z}}-\mathbf{m}_{\overrightarrow{\mathbf{z}}}\right)}}{\pi^M\sqrt{|\mathbf{C}_{\overrightarrow{\mathbf{z}}}|}} \tag{22}$$

where:

$$\overrightarrow{\mathbf{z}}=\begin{pmatrix}\mathbf{z}\\ \mathbf{z}^*\end{pmatrix}=\mathbf{T}^{-1}\mathbf{x},\quad \mathbf{T}=\frac{1}{2}\begin{bmatrix}\mathbf{I}_M & \mathbf{I}_M\\ -j\mathbf{I}_M & j\mathbf{I}_M\end{bmatrix},\quad |\mathbf{T}|=\left(\frac{\mathbf{1}}{\mathbf{2}}\right)^M,\quad \mathbf{T}^{-1}=2\mathbf{T}^H, \tag{23}$$

and:

$$\mathbf{m}_{\overrightarrow{\mathbf{z}}}=\mathbf{T}^{-1}\mathbf{m_x},\quad \mathbf{C}_{\overrightarrow{\mathbf{z}}}=\begin{bmatrix}\mathbf{C_z} & \mathbf{C}_{\mathbf{z},\mathbf{z}^*}\\ \mathbf{C}_{\mathbf{z},\mathbf{z}^*}^* & \mathbf{C}_{\mathbf{z}}^*\end{bmatrix}=\mathbf{T}^{-1}\mathbf{C_x}\left(\mathbf{T}^{-1}\right)^H,\quad \mathbf{C}_{\mathbf{z},\mathbf{z}^*}=E\left[(\mathbf{z}-\mathbf{m_z})(\mathbf{z}-\mathbf{m_z})^T\right]$$

Additionally, in many practical problems of interest (radar, sonar, communication, ...), the complex observation vector $\mathbf{z}$ consists of a bandpass signal which is the output of an Hilbert filtering leading to an "in-phase" real part associated to a "quadrature" imaginary part [1][32, §13][33], *i.e.* a complex circular vector $\mathbf{z}\sim\mathcal{CN}_M(\mathbf{m_z},\mathbf{C_z})$ verifying $\mathbf{C}_{\mathbf{z},\mathbf{z}^*}=\mathbf{0}$ with a compact p.d.f. expression:

$$p\left(\overrightarrow{\mathbf{z}};\mathbf{m}_{\overrightarrow{\mathbf{z}}},\mathbf{C}_{\overrightarrow{\mathbf{z}}}\right)=p(\mathbf{z};\mathbf{m_z},\mathbf{C_z})=\frac{e^{-(\mathbf{z}-\mathbf{m_z})^H\mathbf{C_z}^{-1}(\mathbf{z}-\mathbf{m_z})}}{\pi^M|\mathbf{C_z}|} \tag{24}$$

Among any other estimation problem based on the Gaussian complex circular observation model, the estimation of the direction of arrival (DOA) of a source of signal in the space domain, or the estimation of the frequency of a cisoid in the time/frequency domain is of first importance, both from theoretical and practical viewpoints. Theses two problems have been merged into the framework of modern array processing [34] where mostly two different signal models are considered: the deterministic (conditional) signal model and the stochastic (unconditional) signal model [30]. The discussed signal models are Gaussian and the angular/frequency dependency is given by parameters which are connected with the expectation value in the deterministic case and with the covariance matrix in the stochastic one.

However, not all practical problems of interest can afford an Hilbert filtering stage and some deals simply with real observation samples [32, §13]. Furthermore, the non circularity of the signal of interest may be a beneficial property, as highlighted lately in communication applications [35]. In such cases, the relevant expressions of the components of BB_1^N are given in Appendix: (34a-34d), (37-39), (40a-40e).

As shown in Appendix, the derivation of $\mathbf{MS}$, $\mathbf{EFI}$ and $\mathbf{H}$ is based on a factorization property of the Gaussian p.d.f. (33) that suggests a breakdown into components (scalar or matrices) depending only on the selected value

$\boldsymbol{\theta}^0$ and a couple of test points $\left\{\boldsymbol{\theta}^i, \boldsymbol{\theta}^j\right\}^{i,j\in[0,N]}$:

$$\mathbf{MS} = \begin{bmatrix} ms\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^0\right) & \cdots & ms\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^N\right) \\ \vdots & ms\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right) & \vdots \\ ms\left(\boldsymbol{\theta}^N,\boldsymbol{\theta}^0\right) & \cdots & ms\left(\boldsymbol{\theta}^N,\boldsymbol{\theta}^N\right) \end{bmatrix}, \quad ms\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right) = [\mathbf{MS}]_{i,j} \in \mathbb{R} \tag{25a}$$

$$\mathbf{EFI} = \begin{bmatrix} \mathbf{EFI}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^0\right) & \cdots & \mathbf{EFI}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^N\right) \\ \vdots & \mathbf{EFI}\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right) & \vdots \\ \mathbf{EFI}\left(\boldsymbol{\theta}^N,\boldsymbol{\theta}^0\right) & \cdots & \mathbf{EFI}\left(\boldsymbol{\theta}^N,\boldsymbol{\theta}^N\right) \end{bmatrix}, \quad \mathbf{EFI}\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right) \in \mathcal{M}\left(P,P\right) \tag{25b}$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^0\right) & \cdots & \mathbf{h}\left(\boldsymbol{\theta}^0,\boldsymbol{\theta}^N\right) \\ \vdots & \mathbf{h}\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right) & \vdots \\ \mathbf{h}\left(\boldsymbol{\theta}^N,\boldsymbol{\theta}^0\right) & \cdots & \mathbf{h}\left(\boldsymbol{\theta}^N,\boldsymbol{\theta}^N\right) \end{bmatrix}, \quad \mathbf{h}\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right) \in \mathcal{M}\left(P,1\right) \tag{25c}$$

detailed for both complex circular deterministic and stochastic observation models hereinafter.

A. Case of complex circular Gaussian stochastic signal model

Under the stochastic signal model assumption: $\mathbf{m_z} = \mathbf{0}$, $\mathbf{C_z} = \mathbf{C_z}\left(\boldsymbol{\theta}\right)$ and:

$$[\mathbf{MS}]_{i,j} = \frac{\left|\mathbf{C}_\mathbf{z}^{ij}\right| \left|\mathbf{C_z}\left(\boldsymbol{\theta}^0\right)\right|}{\left|\mathbf{C_z}\left(\boldsymbol{\theta}^i\right)\right| \left|\mathbf{C_z}\left(\boldsymbol{\theta}^j\right)\right|} \tag{26a}$$

$$\left[\mathbf{h}\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right)\right]_p = [\mathbf{MS}]_{i,j} \operatorname{Re}\left\{ tr\left(\frac{\partial \mathbf{C_z}\left(\boldsymbol{\theta}^i\right)^{-1}}{\partial\theta_p} \left(\mathbf{C_z}\left(\boldsymbol{\theta}^i\right) - \mathbf{C}_\mathbf{z}^{ij}\right)\right)\right\} \tag{26b}$$

$$\begin{aligned}\left[\mathbf{EFI}\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right)\right]_{p,p'} &= [\mathbf{MS}]_{i,j} \left[\operatorname{Re}\left\{ tr\left(\frac{\partial \mathbf{C_z}\left(\boldsymbol{\theta}^i\right)^{-1}}{\partial\theta_p} \mathbf{C}_\mathbf{z}^{ij} \frac{\partial \mathbf{C_z}\left(\boldsymbol{\theta}^j\right)^{-1}}{\partial\theta_{p'}} \mathbf{C}_\mathbf{z}^{ij}\right)\right\}\right. \\ &\left. + \operatorname{Re}\left\{ tr\left(\frac{\partial \mathbf{C_z}\left(\boldsymbol{\theta}^i\right)^{-1}}{\partial\theta_p} \left(\mathbf{C_z}\left(\boldsymbol{\theta}^i\right) - \mathbf{C}_\mathbf{z}^{ij}\right)\right)\right\} \operatorname{Re}\left\{ tr\left(\frac{\partial \mathbf{C_z}\left(\boldsymbol{\theta}^j\right)^{-1}}{\partial\theta_{p'}} \left(\mathbf{C_z}\left(\boldsymbol{\theta}^j\right) - \mathbf{C}_\mathbf{z}^{ij}\right)\right)\right\}\right] \\ \mathbf{C}_\mathbf{z}^{ij} &= \left[\mathbf{C_z}\left(\boldsymbol{\theta}^i\right)^{-1} + \mathbf{C_z}\left(\boldsymbol{\theta}^j\right)^{-1} - \mathbf{C_z}\left(\boldsymbol{\theta}^0\right)^{-1}\right]^{-1}\end{aligned} \tag{26c}$$

where $i, j \in [0, N]$ and $p, p\prime \in [1, P]$.

B. Case of complex circular Gaussian deterministic signal model

Under the deterministic model assumption, generally the signal is supposed to be embedded in a spatially white noise with constant but unknown power σ^2 [30]: $\boldsymbol{\theta} = \left[\boldsymbol{\epsilon}, \sigma^2\right]^T$, $\mathbf{m_z} = \mathbf{m_z}\left(\boldsymbol{\epsilon}\right)$, $\mathbf{C_z}\left(\boldsymbol{\theta}\right) = \sigma_\mathbf{z}^2\mathbf{I}$, where $\boldsymbol{\epsilon} = [\epsilon_1, \ldots, \epsilon_{P-1}]$ are the signal unknown parameters. Then:

$$[\mathbf{MS}]_{i,j} = \left(\frac{c_\mathbf{z}^{ij}\left(\sigma_\mathbf{z}^2\right)^0}{\left(\sigma_\mathbf{z}^2\right)^j\left(\sigma_\mathbf{z}^2\right)^i}\right)^M e^{c_\mathbf{z}^{ij}\|\mathbf{m}_\mathbf{z}^{ij}\|^2 - \delta_\mathbf{z}^{ij}} \tag{27a}$$

$$\left[\mathbf{h}\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right)\right]_p = [\mathbf{MS}]_{i,j}\left[\boldsymbol{\alpha}\left(\boldsymbol{\theta}^i\right)\right]_p \tag{27b}$$

$$\left[\mathbf{EFI}\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right)\right]_{p,p'} = \frac{4\left[\mathbf{MS}\right]_{i,j}}{\left(\sigma_{\mathbf{z}}^2\right)^i\left(\sigma_{\mathbf{z}}^2\right)^j}\left[\frac{c_{\mathbf{z}}^{ij}}{2}\operatorname{Re}\left\{\frac{\partial\mathbf{m_z}\left(\boldsymbol{\epsilon}^i\right)}{\partial\epsilon_p}^H\frac{\partial\mathbf{m_z}\left(\boldsymbol{\epsilon}^j\right)}{\partial\epsilon_{p'}}\right\}\right.$$
$$\left.+\operatorname{Re}\left\{\frac{\partial\mathbf{m_z}\left(\boldsymbol{\epsilon}^i\right)}{\partial\epsilon_p}^H\left(E\left[\mathbf{z}\right]-\mathbf{m_z}\left(\boldsymbol{\epsilon}^i\right)\right)\right\}\operatorname{Re}\left\{\frac{\partial\mathbf{m_z}\left(\boldsymbol{\epsilon}^j\right)}{\partial\epsilon_{p'}}^H\left(E\left[\mathbf{z}\right]-\mathbf{m_z}\left(\boldsymbol{\epsilon}^j\right)\right)\right\}\right] \tag{27c}$$

$$\left[\mathbf{EFI}\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right)\right]_{P,p'} = \left[\mathbf{MS}\right]_{i,j}\left\{-\frac{1}{2}\left[\boldsymbol{\Gamma}\left(\boldsymbol{\theta}^j,\boldsymbol{\theta}^i\right)\right]_{p',P}-\frac{2M}{\left(\sigma_{\mathbf{z}}^2\right)^i}\left[\boldsymbol{\alpha}\left(\boldsymbol{\theta}^j\right)\right]_{p'}\right\} \tag{27d}$$

$$\left[\mathbf{EFI}\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right)\right]_{p,P} = \left[\mathbf{MS}\right]_{i,j}\left\{-\frac{1}{2}\left[\boldsymbol{\Gamma}\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right)\right]_{p,P}-\frac{2M}{\left(\sigma_{\mathbf{z}}^2\right)^j}\left[\boldsymbol{\alpha}\left(\boldsymbol{\theta}^i\right)\right]_{p}\right\} \tag{27e}$$

$$\left[\mathbf{EFI}\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right)\right]_{P,P} = \left[\mathbf{MS}\right]_{i,j}\left\{\frac{\left[\boldsymbol{\Delta}\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right)\right]_{P,P}}{4}-\frac{4M^2}{\left(\sigma_{\mathbf{z}}^2\right)^i\left(\sigma_{\mathbf{z}}^2\right)^j}-\frac{2M\left[\boldsymbol{\alpha}\left(\boldsymbol{\theta}^j\right)\right]_P}{\left(\sigma_{\mathbf{z}}^2\right)^i}-\frac{2M\left[\boldsymbol{\alpha}\left(\boldsymbol{\theta}^i\right)\right]_P}{\left(\sigma_{\mathbf{z}}^2\right)^j}\right\} \tag{27f}$$

where $l, l\prime, p, p\prime \in [1, P-1]$ and:

$$c_{\mathbf{z}}^{ij} = \left(\frac{1}{\left(\sigma_{\mathbf{z}}^2\right)^i}+\frac{1}{\left(\sigma_{\mathbf{z}}^2\right)^j}-\frac{1}{\left(\sigma_{\mathbf{z}}^2\right)^0}\right)^{-1}, \qquad \mathbf{m}_{\mathbf{z}}^{ij}=\frac{\mathbf{m_z}\left(\boldsymbol{\epsilon}^i\right)}{\left(\sigma_{\mathbf{z}}^2\right)^i}+\frac{\mathbf{m_z}\left(\boldsymbol{\epsilon}^j\right)}{\left(\sigma_{\mathbf{z}}^2\right)^j}-\frac{\mathbf{m_z}\left(\boldsymbol{\epsilon}^0\right)}{\left(\sigma_{\mathbf{z}}^2\right)^0}$$

$$\delta_{\mathbf{z}}^{ij} = \frac{\left\|\mathbf{m_z}\left(\boldsymbol{\epsilon}^i\right)\right\|^2}{\left(\sigma_{\mathbf{z}}^2\right)^i}+\frac{\left\|\mathbf{m_z}\left(\boldsymbol{\epsilon}^j\right)\right\|^2}{\left(\sigma_{\mathbf{z}}^2\right)^j}-\frac{\left\|\mathbf{m_z}\left(\boldsymbol{\epsilon}^0\right)\right\|^2}{\left(\sigma_{\mathbf{z}}^2\right)^0}, \quad E\left[\mathbf{z}\right]=c_{\mathbf{z}}^{ij}\mathbf{m}_{\mathbf{z}}^{ij}$$

$$\left[\boldsymbol{\alpha}\left(\boldsymbol{\theta}\right)\right]_l=\frac{2}{\sigma_{\mathbf{z}}^2}\operatorname{Re}\left\{\frac{\partial\mathbf{m_z}\left(\boldsymbol{\epsilon}\right)}{\partial\epsilon_l}^H\left(E\left[\mathbf{z}\right]-\mathbf{m_z}\left(\boldsymbol{\epsilon}\right)\right)\right\}, \qquad \left[\boldsymbol{\alpha}\left(\boldsymbol{\theta}\right)\right]_P=\frac{2\left\|E\left[\mathbf{z}\right]-\mathbf{m_z}\left(\boldsymbol{\epsilon}\right)\right\|^2+2M\left(c_{\mathbf{z}}^{ij}-\sigma_{\mathbf{z}}^2\right)}{\sigma_{\mathbf{z}}^4}$$

$$\left[\boldsymbol{\Gamma}\left(\boldsymbol{\theta},\boldsymbol{\theta}'\right)\right]_{l,P} = \frac{-8}{\sigma_{\mathbf{z}}^2\sigma_{\mathbf{z}}'^4}\operatorname{Re}\left\{\left[\left(Mc_{\mathbf{z}}^{ij}+\left\|E\left[\mathbf{z}\right]-\mathbf{m_z}\left(\boldsymbol{\epsilon}'\right)\right\|^2\right)\left(E\left[\mathbf{z}\right]-\mathbf{m_z}\left(\boldsymbol{\epsilon}\right)\right)+c_{\mathbf{z}}^{ij}\left(E\left[\mathbf{z}\right]-\mathbf{m_z}\left(\boldsymbol{\epsilon}'\right)\right)\right]^H\right.$$
$$\left.\frac{\partial\mathbf{m_z}\left(\boldsymbol{\epsilon}\right)}{\partial\epsilon_l}\right\}$$

$$\left[\boldsymbol{\Delta}\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right)\right]_{P,P}=\frac{16}{\left(\sigma_{\mathbf{z}}^4\right)^i\left(\sigma_{\mathbf{z}}^4\right)^j}\left\{M\left(c_{\mathbf{z}}^{ij}\right)^2+2c_{\mathbf{z}}^{ij}\operatorname{Re}\left\{\left(E\left[\mathbf{z}\right]-\mathbf{m_z}\left(\boldsymbol{\epsilon}^i\right)\right)^H\left(E\left[\mathbf{z}\right]-\mathbf{m_z}\left(\boldsymbol{\epsilon}^j\right)\right)\right\}\right.$$
$$\left.+\left(Mc_{\mathbf{z}}^{ij}+\left\|E\left[\mathbf{z}\right]-\mathbf{m_z}\left(\boldsymbol{\epsilon}^i\right)\right\|^2\right)\left(Mc_{\mathbf{z}}^{ij}+\left\|E\left[\mathbf{z}\right]-\mathbf{m_z}\left(\boldsymbol{\epsilon}^j\right)\right\|^2\right)\right\}$$

If we consider the special case where $\sigma_{\mathbf{z}}^2$ is known, then $\boldsymbol{\theta}=\boldsymbol{\epsilon}$ and:

$$\left[\mathbf{MS}\right]_{i,j} = e^{\frac{2}{\sigma_{\mathbf{z}}^2}\operatorname{Re}\left\{\left(\mathbf{m_z}(\boldsymbol{\epsilon}^i)-\mathbf{m_z}(\boldsymbol{\epsilon}^0)\right)^H\left(\mathbf{m_z}(\boldsymbol{\epsilon}^j)-\mathbf{m_z}(\boldsymbol{\epsilon}^0)\right)\right\}} \tag{28a}$$

$$\left[\mathbf{h}\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right)\right]_p = \frac{2\left[\mathbf{MS}\right]_{i,j}}{\sigma_{\mathbf{z}}^2}\operatorname{Re}\left\{\frac{\partial\mathbf{m_z}\left(\boldsymbol{\epsilon}^i\right)}{\partial\epsilon_p}^H\left(\mathbf{m_z}\left(\boldsymbol{\epsilon}^j\right)-\mathbf{m_z}\left(\boldsymbol{\epsilon}^0\right)\right)\right\} \tag{28b}$$

$$\left[\mathbf{EFI}\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right)\right]_{p,p'} = \frac{4\left[\mathbf{MS}\right]_{i,j}}{\sigma_{\mathbf{z}}^2}\left[\frac{\sigma_{\mathbf{z}}^2}{2}\operatorname{Re}\left\{\frac{\partial\mathbf{m_z}\left(\boldsymbol{\epsilon}^i\right)}{\partial\epsilon_p}^H\frac{\partial\mathbf{m_z}\left(\boldsymbol{\epsilon}^j\right)}{\partial\epsilon_{p'}}\right\}\right.$$
$$\left.+\operatorname{Re}\left\{\frac{\partial\mathbf{m_z}\left(\boldsymbol{\epsilon}^i\right)}{\partial\epsilon_p}^H\left(\mathbf{m_z}\left(\boldsymbol{\epsilon}^j\right)-\mathbf{m_z}\left(\boldsymbol{\epsilon}^0\right)\right)\right\}\operatorname{Re}\left\{\frac{\partial\mathbf{m_z}\left(\boldsymbol{\epsilon}^j\right)}{\partial\epsilon_{p'}}^H\left(\mathbf{m_z}\left(\boldsymbol{\epsilon}^i\right)-\mathbf{m_z}\left(\boldsymbol{\epsilon}^0\right)\right)\right\}\right] \tag{28c}$$

where $p, p\prime \in [1, P]$.

C. Multiple independent observations

Let us consider S identical independent observations of a Gaussian complex circular random vector $\mathbf{z} \sim \mathcal{CN}_M(\mathbf{m_z}(\boldsymbol{\theta}), \mathbf{C_z}(\boldsymbol{\theta}))$. This observation model is equivalent to a single MS-dimensional Gaussian random vector $\mathbf{y} = \left[\mathbf{z}_1^T, \ldots, \mathbf{z}_S^T\right]^T \sim \mathcal{CN}_{MS}(\mathbf{m_y}(\boldsymbol{\theta}), \mathbf{C_y}(\boldsymbol{\theta}))$ where:

$$\mathbf{m_y}(\boldsymbol{\theta}) = \mathbf{1}_S \otimes \mathbf{m}(\boldsymbol{\theta}), \qquad \mathbf{C_y}(\boldsymbol{\theta}) = \mathbf{I}_S \otimes \mathbf{C}(\boldsymbol{\theta}), \tag{29}$$

$\mathbf{1}_S = [1, \ldots, 1]^T$ and $\otimes$ is the Kronecker product. Then, as $|\mathbf{I}_S \otimes \mathbf{A}| = |\mathbf{A}|^S$, $tr(\mathbf{I}_S \otimes \mathbf{A}) = Str(\mathbf{A})$, $(\mathbf{1}_S \otimes \mathbf{u})^H (\mathbf{I}_S \otimes \mathbf{A})(\mathbf{1}_S \otimes \mathbf{v}) = S\mathbf{u}^H\mathbf{A}\mathbf{v}$, any expression derived in the single observation case can be used in the multiple observations case provided the following algebraic values are updated as follows: $|\mathbf{A}| \rightarrow |\mathbf{A}|^S$, $\mathbf{u}^H\mathbf{A}\mathbf{v} \rightarrow S\mathbf{u}^H\mathbf{A}\mathbf{v}$ and $tr(\mathbf{A}) \rightarrow Str(\mathbf{A})$. The same result holds for the real observation model (see Appendix), provided that $\mathbf{u}^T\mathbf{A}\mathbf{v} \rightarrow S\mathbf{u}^T\mathbf{A}\mathbf{v}$.

VI. Lower bounds comparison: single tone threshold analysis

Let the M-dimensional complex observation vector $\mathbf{z}$ be modelled by:

$$\mathbf{z} = a\boldsymbol{\psi}\left(\theta^0\right) + \mathbf{n}, \quad \boldsymbol{\psi}(\theta) = \left[1, e^{j2\pi\theta}, ..., e^{j(M-1)2\pi\theta}\right]^T, \quad \theta \in]-0.5, 0.5[\tag{30}$$

where θ is the unknown parameter to estimate, $\sigma_a^2 = E\left[|a|^2\right]$ is the SNR and $\mathbf{n}$ is a complex circular Gaussian noise, with zero mean and known covariance matrix $\mathbf{C_n} = \mathbf{I}_M$. Let $\mathbf{y} = \left[\mathbf{z}_1^T, \ldots, \mathbf{z}_S^T\right]^T$ be the vector resulting from the concatenation of S identical independent observations of $\mathbf{z}$. Then the p.d.f. of $\mathbf{z}$ is given either by (31a) for the deterministic signal model or by (31b) for the stochastic one:

$$p\left(\mathbf{z};\theta^0\right) = \frac{e^{-\left(\mathbf{z}-a\boldsymbol{\psi}(\theta^0)\right)^H \mathbf{C}_\mathbf{z}^{-1}\left(\mathbf{z}-a\boldsymbol{\psi}(\theta^0)\right)}}{\pi^M |\mathbf{C_z}|} = \frac{e^{-\|\mathbf{z}-a\boldsymbol{\psi}(\theta^0)\|^2}}{\pi^M} \tag{31a}$$

$$p\left(\mathbf{z};\theta^0\right) = \frac{e^{-\mathbf{z}^H\mathbf{C}_\mathbf{z}^{-1}\mathbf{z}}}{\pi^M |\mathbf{C_z}|} = \frac{e^{-\mathbf{z}^H\left(\sigma_a^2\boldsymbol{\psi}(\theta^0)\boldsymbol{\psi}(\theta^0)^H + \mathbf{I}_M\right)^{-1}\mathbf{z}}}{\pi^M \left|\sigma_a^2\boldsymbol{\psi}\left(\theta^0\right)\boldsymbol{\psi}\left(\theta^0\right)^H + \mathbf{I}_M\right|} \tag{31b}$$

leading to either the following components of BB_1^N for the deterministic signal model (28a-28c):

$$\begin{aligned}
[\mathbf{MS}]_{i,j} &= e^{2Sa^2 \operatorname{Re}\left\{\left(\boldsymbol{\psi}(\theta^i)-\boldsymbol{\psi}(\theta^0)\right)^H\left(\boldsymbol{\psi}(\theta^j)-\boldsymbol{\psi}(\theta^0)\right)\right\}} \\
\left[\mathbf{h}\left(\theta^i,\theta^j\right)\right]_1 &= 2Sa^2 [\mathbf{MS}]_{i,j} \operatorname{Re}\left\{\frac{\partial\boldsymbol{\psi}\left(\theta^i\right)}{\partial\theta}^H \left(\boldsymbol{\psi}\left(\theta^j\right) - \boldsymbol{\psi}\left(\theta^0\right)\right)\right\} \\
\left[\mathbf{EFI}\left(\theta^i,\theta^j\right)\right]_{1,1} &= 4Sa^2 [\mathbf{MS}]_{i,j} \left[\frac{1}{2}\operatorname{Re}\left\{\frac{\partial\boldsymbol{\psi}\left(\theta^i\right)}{\partial\theta}^H \frac{\partial\boldsymbol{\psi}\left(\theta^j\right)}{\partial\theta}\right\}\right. \\
&\left. + Sa^2 \operatorname{Re}\left\{\frac{\partial\boldsymbol{\psi}\left(\theta^i\right)}{\partial\theta}^H \left(\boldsymbol{\psi}\left(\theta^j\right) - \boldsymbol{\psi}\left(\theta^0\right)\right)\right\} \operatorname{Re}\left\{\frac{\partial\boldsymbol{\psi}\left(\theta^j\right)}{\partial\theta}^H \left(\boldsymbol{\psi}\left(\theta^i\right) - \boldsymbol{\psi}\left(\theta^0\right)\right)\right\}\right]
\end{aligned}$$

or, for the stochastic one (26a-26c), to:

$$\begin{aligned}
\mathbf{C}_{\mathbf{z}}^{ij} &= \left[\mathbf{I}_M - \frac{\sigma_a^2}{1+M\sigma_a^2}\left[\boldsymbol{\psi}\left(\theta^i\right)\boldsymbol{\psi}\left(\theta^i\right)^H + \boldsymbol{\psi}\left(\theta^j\right)\boldsymbol{\psi}\left(\theta^j\right)^H - \boldsymbol{\psi}\left(\theta^0\right)\boldsymbol{\psi}\left(\theta^0\right)^H\right]\right]^{-1} \\
[\mathbf{MS}]_{i,j} &= \left(\frac{\left|\mathbf{C}_{\mathbf{z}}^{ij}\right|}{1+M\sigma_a^2}\right)^S \\
\left[\mathbf{h}\left(\theta^i,\theta^j\right)\right]_1 &= S\,[\mathbf{MS}]_{i,j}\,\mathrm{Re}\left\{tr\left(\frac{\partial \mathbf{C}_{\mathbf{z}}\left(\theta^i\right)^{-1}}{\partial\theta}\left(\mathbf{C}_{\mathbf{z}}\left(\theta^i\right)-\mathbf{C}_{\mathbf{z}}^{ij}\right)\right)\right\} \\
\left[\mathbf{EFI}\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right)\right]_{1,1} &= S\,[\mathbf{MS}]_{i,j}\left[\mathrm{Re}\left\{tr\left(\frac{\partial \mathbf{C}_{\mathbf{z}}\left(\theta^i\right)^{-1}}{\partial\theta}\mathbf{C}_{\mathbf{z}}^{ij}\frac{\partial \mathbf{C}_{\mathbf{z}}\left(\theta^j\right)^{-1}}{\partial\theta}\mathbf{C}_{\mathbf{z}}^{ij}\right)\right\}\right. \\
&\quad \left. +S\,\mathrm{Re}\left\{tr\left(\frac{\partial \mathbf{C}_{\mathbf{z}}\left(\theta^i\right)^{-1}}{\partial\theta}\left(\mathbf{C}_{\mathbf{z}}\left(\theta^i\right)-\mathbf{C}_{\mathbf{z}}^{ij}\right)\right)\right\}\mathrm{Re}\left\{tr\left(\frac{\partial \mathbf{C}_{\mathbf{z}}\left(\theta^j\right)^{-1}}{\partial\theta}\left(\mathbf{C}_{\mathbf{z}}\left(\theta^j\right)-\mathbf{C}_{\mathbf{z}}^{ij}\right)\right)\right\}\right]
\end{aligned}$$

where:

$$\begin{aligned}
\mathbf{C}_{\mathbf{z}}(\theta)^{-1} &= \mathbf{I}_M - \frac{\sigma_a^2\boldsymbol{\psi}(\theta)\boldsymbol{\psi}(\theta)^H}{1+\sigma_a^2\boldsymbol{\psi}(\theta)^H\boldsymbol{\psi}(\theta)} = \mathbf{I}_M - \frac{\sigma_a^2}{1+M\sigma_a^2}\boldsymbol{\psi}(\theta)\boldsymbol{\psi}(\theta)^H \\
\left|\mathbf{C}_{\mathbf{z}}(\theta)\right| &= 1+\sigma_a^2\boldsymbol{\psi}(\theta)^H\boldsymbol{\psi}(\theta) = 1+M\sigma_a^2 \\
\frac{\partial \mathbf{C}_{\mathbf{z}}(\theta)^{-1}}{\partial\theta} &= -\frac{\sigma_a^2}{1+M\sigma_a^2}\left(\frac{\partial\boldsymbol{\psi}(\theta)}{\partial\theta}\boldsymbol{\psi}(\theta)^H + \boldsymbol{\psi}(\theta)\frac{\partial\boldsymbol{\psi}(\theta)}{\partial\theta}^H\right), \quad \frac{\partial\boldsymbol{\psi}(\theta)}{\partial\theta} = \mathbf{D}\boldsymbol{\psi}(\theta), \quad [\mathbf{D}]_{m,m} = j2\pi(m-1)
\end{aligned}$$

The MLE of θ is given by (32a) for the deterministic signal model or by (32b) for the stochastic one [19]:

$$\widehat{\theta}_{DML} = \max_{\theta}\left\{\sum_{s=1}^{S}\mathrm{Re}\left\{\boldsymbol{\psi}(\theta)^H\mathbf{z}_s\right\}\right\} \tag{32a}$$

$$\widehat{\theta}_{SML} = \max_{\theta}\left\{\sum_{s=1}^{S}\left|\boldsymbol{\psi}(\theta)^H\mathbf{z}_s\right|^2\right\} \tag{32b}$$

We consider the reference estimation case where $\theta^0 = 0$. For any set of $N+1$ test points $\left\{\theta^0\right\}\cup\left\{\theta^n\right\}^{[1,N]}$, only the MSB^N and the MHB^N-HBB_1^N are of a complexity comparable with the BB_1^N. Nevertheless, we also include in the comparison the HCRB as it is the simplest and the most used representative of Large Errors bounds. All mentioned lower bounds can be computed from the components of BB_1^N, with rearrangement (see section IV). For the sake of fair comparison with the HCRB, the MSB^N, HBB_1^N and BB_1^N are computed as supremum over the possible values of $\left\{\theta^0\right\}\cup\left\{\theta^n\right\}^{[1,N]}$.

As previously mentioned, our primary motivation was to derive an easily computable and tighter approximation of the BB. Therefore, to be of practical interest, the proposed bound BB_1^N should exhibit an improvement in the prediction of the SNR threshold for a small number of test points. It is the reason why we focus on the cases where the number of test points used (per unknown parameter) is 1 or 2.

If $N = 1$ (one test point), note that, for the deterministic cisoid model (31a), the closed form expression of the HCRB and of the HBB_1^N has been given respectively in [16] and [18]. Then numerous analytical simulations not reported here, have shown that the improvement in the SNR threshold prediction brought by the BB_1^1 in comparison with HBB_1^1 is not significant, at least for the single tone estimation problem.

If $N = 2$ (two test points), the expected improvement is demonstrated by figures (1-6) which show the evolution of the various bounds as a function of SNR in various cases: $S = 10$ (deterministic), $S = 100$ (stochastic) and $M = 2, 10, 32$. The variance of the MLE is also shown in order to compare the threshold behaviour of the bounds. For the sake of simplicity $\left\{\theta^1,\theta^2\right\} = \left\{d\theta, -d\theta\right\}$, $d\theta \in \left]0, 0.5\right[$. The proposed bound BB_1^N clearly results in a tighter BB approximation than the other bounds and allows a better prediction of the SNR threshold value.

Additionally, the present results suggest that the true value of the BB may be significantly underestimated by existing approximations, questioning some previously drawn conclusions on MLE variance prediction by Deterministic Large

Error Bounds.
Last, let us recall that any rationale to optimize the choice of the set of test points to maximize the MSB^N or the HBB_1^N is applicable to the BB_1^N to increase its tightness [17][18].

VII. Conclusion

We have proposed a method for the derivation of a general class of BB approximations which has the advantage of a clear interpretation. This method includes all previously derived bounds, and provides a meaningful way to classify them. Moreover it suggests a new practical BB approximation, the BB_1^N, to be used in place of existing approximations (MSB^N [9], MHB^N-HBB_1^N [13][15]). Indeed, for a given set of test points, the BB_1^N provides by construction a tighter BB approximation at the expense of a bounded increase of the computational complexity. As a consequence the BB_1^N is expected to provide an improved SNR threshold prediction, as shown on a reference estimation problem. We have therefore derived its expression for the general Gaussian observation model and more particularly for the complex circular deterministic and stochastic observation models, to be used in any related estimation problem. Additionally, we also provide with the expression of the MSB^N for the general Gaussian observation model to be used when the BB_1^N is not applicable.

VIII. Appendix: Derivation of the BB_1^N for the Gaussian observation model

We use notations defined in section V. Let us consider an M-dimensional Gaussian real vector with mean $\mathbf{m_x} = \mathbf{m}(\boldsymbol{\theta})$ and covariance matrix $\mathbf{C_x} = \mathbf{C}(\boldsymbol{\theta})$: $\mathbf{x} \sim \mathcal{N}_M(\mathbf{m}(\boldsymbol{\theta}), \mathbf{C}(\boldsymbol{\theta}))$ and $p(\mathbf{x};\boldsymbol{\theta}) = p(\mathbf{x};\mathbf{m}(\boldsymbol{\theta}), \mathbf{C}(\boldsymbol{\theta}))$ (22). The derivation of the components $\mathbf{MS}$, $\mathbf{EFI}$ and $\mathbf{H}$ of BB_1^N (21a-21c) is based on the following factorization property of the Gaussian real p.d.f.:

$$\frac{p(\mathbf{x};\boldsymbol{\theta}^i)\,p(\mathbf{x};\boldsymbol{\theta}^j)}{p(\mathbf{x};\boldsymbol{\theta}^0)} = \frac{p(\mathbf{x};\mathbf{m}(\boldsymbol{\theta}^i),\mathbf{C}(\boldsymbol{\theta}^i))\,p(\mathbf{x};\mathbf{m}(\boldsymbol{\theta}^j),\mathbf{C}(\boldsymbol{\theta}^j))}{p(\mathbf{x};\mathbf{m}(\boldsymbol{\theta}^0),\mathbf{C}(\boldsymbol{\theta}^0))} = [\mathbf{MS}]_{i,j}\; p(\mathbf{x};\mathbf{C}^{ij}\mathbf{m}^{ij},\mathbf{C}^{ij}) \tag{33}$$

where :

$$\mathbf{C}^{ij} = \left[\mathbf{C}(\boldsymbol{\theta}^i)^{-1} + \mathbf{C}(\boldsymbol{\theta}^j)^{-1} - \mathbf{C}(\boldsymbol{\theta}^0)^{-1}\right]^{-1} \tag{34a}$$

$$\mathbf{m}^{ij} = \mathbf{C}(\boldsymbol{\theta}^i)^{-1}\mathbf{m}(\boldsymbol{\theta}^i) + \mathbf{C}(\boldsymbol{\theta}^j)^{-1}\mathbf{m}(\boldsymbol{\theta}^j) - \mathbf{C}(\boldsymbol{\theta}^0)^{-1}\mathbf{m}(\boldsymbol{\theta}^0) \tag{34b}$$

$$\delta^{ij} = \mathbf{m}(\boldsymbol{\theta}^i)^T\mathbf{C}(\boldsymbol{\theta}^i)^{-1}\mathbf{m}(\boldsymbol{\theta}^i) + \mathbf{m}(\boldsymbol{\theta}^j)^T\mathbf{C}(\boldsymbol{\theta}^j)^{-1}\mathbf{m}(\boldsymbol{\theta}^j) - \mathbf{m}(\boldsymbol{\theta}^0)^T\mathbf{C}(\boldsymbol{\theta}^0)^{-1}\mathbf{m}(\boldsymbol{\theta}^0) \tag{34c}$$

$$[\mathbf{MS}]_{i,j} = \sqrt{\frac{|\mathbf{C}^{ij}|\,|\mathbf{C}(\boldsymbol{\theta}^0)|}{|\mathbf{C}(\boldsymbol{\theta}^i)|\,|\mathbf{C}(\boldsymbol{\theta}^j)|}}\,e^{\frac{1}{2}\left[(\mathbf{m}^{ij})^T\mathbf{C}^{ij}\mathbf{m}^{ij} - \delta^{ij}\right]} \tag{34d}$$

which suggests a breakdown into items $\left\{[\mathbf{MS}]_{i,j}, \mathbf{EFI}(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j), \mathbf{h}(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j)\right\}$ depending only on the selected value $\boldsymbol{\theta}^0$ and a couple of test points $\left\{\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right\}^{i,j\in[0,N]}$, as detailed in (25a-25c).
Indeed, denoting $E[g(\mathbf{x})] = \int g(\mathbf{x})\,p(\mathbf{x};\mathbf{C}^{ij}\mathbf{m}^{ij},\mathbf{C}^{ij})\,d\mathbf{x}$, then:

$$[\mathbf{MS}]_{i,j} = E_{\boldsymbol{\theta}^0}\left[\frac{p(\mathbf{x};\boldsymbol{\theta}^i)}{p(\mathbf{x};\boldsymbol{\theta}^0)}\frac{p(\mathbf{x};\boldsymbol{\theta}^j)}{p(\mathbf{x};\boldsymbol{\theta}^0)}\right] = [\mathbf{MS}]_{i,j}\int p(\mathbf{x};\mathbf{C}^{ij}\mathbf{m}^{ij},\mathbf{C}^{ij})\,d\mathbf{x}$$

$$\begin{aligned}\mathbf{EFI}(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j) &= E_{\boldsymbol{\theta}^0}\left[\frac{\partial\ln p(\mathbf{x};\boldsymbol{\theta}^i)}{\partial\boldsymbol{\theta}}\left(\frac{\partial\ln p(\mathbf{x};\boldsymbol{\theta}^j)}{\partial\boldsymbol{\theta}}\right)^T\frac{p(\mathbf{x};\boldsymbol{\theta}^i)}{p(\mathbf{x};\boldsymbol{\theta}^0)}\frac{p(\mathbf{x};\boldsymbol{\theta}^j)}{p(\mathbf{x};\boldsymbol{\theta}^0)}\right] \\ &= [\mathbf{MS}]_{i,j}\;E\left[\frac{\partial\ln p(\mathbf{x};\boldsymbol{\theta}^i)}{\partial\boldsymbol{\theta}}\left(\frac{\partial\ln p(\mathbf{x};\boldsymbol{\theta}^j)}{\partial\boldsymbol{\theta}}\right)^T\right]\end{aligned}$$

$$\mathbf{h}(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j) = E_{\boldsymbol{\theta}^0}\left[\frac{\partial\ln p(\mathbf{x};\boldsymbol{\theta}^i)}{\partial\boldsymbol{\theta}}\frac{p(\mathbf{x};\boldsymbol{\theta}^i)}{p(\mathbf{x};\boldsymbol{\theta}^0)}\frac{p(\mathbf{x};\boldsymbol{\theta}^j)}{p(\mathbf{x};\boldsymbol{\theta}^0)}\right] = [\mathbf{MS}]_{i,j}\;E\left[\frac{\partial\ln p(\mathbf{x};\boldsymbol{\theta}^i)}{\partial\boldsymbol{\theta}}\right]$$

Therefore in the following, $\mathbf{x} \sim \mathcal{N}_M\left(\mathbf{C}^{ij}\mathbf{m}^{ij}, \mathbf{C}^{ij}\right)$ where $\mathbf{C}^{ij}$, $\mathbf{m}^{ij}$ are given by (34a-34d). To compute the missing expectations, let us introduce $\phi(\mathbf{x};\boldsymbol{\theta})$ and $\widehat{\mathbf{C}(\boldsymbol{\theta})}$ defined as:

$$p(\mathbf{x};\boldsymbol{\theta}) = \frac{e^{-\frac{1}{2}\phi(\mathbf{x};\boldsymbol{\theta})}}{\sqrt{2\pi}^M\sqrt{|\mathbf{C}(\boldsymbol{\theta})|}}, \quad \begin{cases} \phi(\mathbf{x};\boldsymbol{\theta}) = (\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}))^T\mathbf{C}(\boldsymbol{\theta})^{-1}(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta})) = tr\left(\mathbf{C}(\boldsymbol{\theta})^{-1}\widehat{\mathbf{C}(\boldsymbol{\theta})}\right) \\ \widehat{\mathbf{C}(\boldsymbol{\theta})} = (\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}))(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}))^T \end{cases}$$

and recall that $\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\omega} = -\mathbf{C}(\boldsymbol{\theta})^{-1}\frac{\partial\mathbf{C}(\boldsymbol{\theta})}{\partial\omega}\mathbf{C}(\boldsymbol{\theta})^{-1}$, $\frac{\partial\ln|\mathbf{C}(\boldsymbol{\theta})|}{\partial\omega} = -tr\left(\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\omega}\mathbf{C}(\boldsymbol{\theta})\right)$. Then:

$$\frac{\partial\phi(\mathbf{x};\boldsymbol{\theta})}{\partial\omega} = -2\frac{\partial\mathbf{m}(\boldsymbol{\theta})}{\partial\omega}^T\mathbf{C}(\boldsymbol{\theta})^{-1}(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta})) + (\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}))^T\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\omega}(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta})) \quad (35)$$

$$\frac{\partial\ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\omega} = \frac{1}{2}tr\left(\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\omega}\left(\mathbf{C}(\boldsymbol{\theta})-\widehat{\mathbf{C}(\boldsymbol{\theta})}\right)\right) + \frac{\partial\mathbf{m}(\boldsymbol{\theta})}{\partial\omega}^T\mathbf{C}(\boldsymbol{\theta})^{-1}(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta})) \quad (36)$$

$$E\left[\widehat{\mathbf{C}(\boldsymbol{\theta})}\right] = \mathbf{C}^{ij} + (E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta}))(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta}))^T$$

- **Computation of** $E\left[\frac{\partial\ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\omega}\right]$. From (36):

$$E\left[\frac{\partial\ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\omega}\right] = \frac{1}{2}tr\left(\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\omega}\left(\mathbf{C}(\boldsymbol{\theta})-E\left[\widehat{\mathbf{C}(\boldsymbol{\theta})}\right]\right)\right) + \frac{\partial\mathbf{m}(\boldsymbol{\theta})}{\partial\omega}^T\mathbf{C}(\boldsymbol{\theta})^{-1}(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta}))$$

- **Computation of** $E\left[\frac{\partial\ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\omega}\frac{\partial\ln p(\mathbf{x};\boldsymbol{\theta}')}{\partial\omega'}\right]$. From (36):

$$\begin{aligned}\frac{\partial\ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\omega}\frac{\partial\ln p(\mathbf{x};\boldsymbol{\theta}')}{\partial\omega'} = {} & \frac{1}{2}tr\left(\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\omega}\mathbf{C}(\boldsymbol{\theta})\right)\frac{\partial\ln p(\mathbf{x};\boldsymbol{\theta}')}{\partial\omega'} + \frac{\partial\ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\omega}\frac{1}{2}tr\left(\frac{\partial\mathbf{C}(\boldsymbol{\theta}')^{-1}}{\partial\omega'}\mathbf{C}(\boldsymbol{\theta}')\right) \\ & -\frac{1}{4}tr\left(\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\omega}\mathbf{C}(\boldsymbol{\theta})\right)tr\left(\frac{\partial\mathbf{C}(\boldsymbol{\theta}')^{-1}}{\partial\omega'}\mathbf{C}(\boldsymbol{\theta}')\right) + \frac{1}{4}\frac{\partial\phi(\mathbf{x};\boldsymbol{\theta})}{\partial\omega}\frac{\partial\phi(\mathbf{x};\boldsymbol{\theta}')}{\partial\omega'}\end{aligned}$$

therefore:

$$\begin{aligned}E\left[\frac{\partial\ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\omega}\frac{\partial\ln p(\mathbf{x};\boldsymbol{\theta}')}{\partial\omega'}\right] = {} & \frac{1}{2}tr\left(\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\omega}\mathbf{C}(\boldsymbol{\theta})\right)E\left[\frac{\partial\ln p(\mathbf{x};\boldsymbol{\theta}')}{\partial\omega'}\right] + \frac{1}{4}E\left[\frac{\partial\phi(\mathbf{x};\boldsymbol{\theta})}{\partial\omega}\frac{\partial\phi(\mathbf{x};\boldsymbol{\theta}')}{\partial\omega'}\right] \\ & + \frac{1}{2}tr\left(\frac{\partial\mathbf{C}(\boldsymbol{\theta}')^{-1}}{\partial\omega'}\mathbf{C}(\boldsymbol{\theta}')\right)E\left[\frac{\partial\ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\omega}\right] - \frac{1}{4}tr\left(\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\omega}\mathbf{C}(\boldsymbol{\theta})\right)tr\left(\frac{\partial\mathbf{C}(\boldsymbol{\theta}')^{-1}}{\partial\omega'}\mathbf{C}(\boldsymbol{\theta}')\right)\end{aligned}$$

where (35):

$$\begin{aligned}E\left[\frac{\partial\phi(\mathbf{x};\boldsymbol{\theta})}{\partial\omega}\frac{\partial\phi(\mathbf{x};\boldsymbol{\theta}')}{\partial\omega'}\right] = {} & 4\frac{\partial\mathbf{m}(\boldsymbol{\theta})}{\partial\omega}^T\mathbf{C}(\boldsymbol{\theta})^{-1}E\left[(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}))(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}'))^T\right]\mathbf{C}(\boldsymbol{\theta}')^{-1}\frac{\partial\mathbf{m}(\boldsymbol{\theta}')}{\partial\omega'} \\ & - 2\frac{\partial\mathbf{m}(\boldsymbol{\theta})}{\partial\omega}^T\mathbf{C}(\boldsymbol{\theta})^{-1}E\left[(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}))(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}'))^T\frac{\partial\mathbf{C}(\boldsymbol{\theta}')^{-1}}{\partial\omega'}(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}'))\right] \\ & - 2\frac{\partial\mathbf{m}(\boldsymbol{\theta}')}{\partial\omega'}^T\mathbf{C}(\boldsymbol{\theta}')^{-1}E\left[(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}'))(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}))^T\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\omega}(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}))\right] \\ & + E\left[(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}))^T\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\omega}(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}))(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}'))^T\frac{\partial\mathbf{C}(\boldsymbol{\theta}')^{-1}}{\partial\omega'}(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}'))\right]\end{aligned}$$

As $E\left[(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}))\left(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}')\right)^T\right]=\mathbf{C}^{ij}+(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta}))\left(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta}')\right)^T$ and [36]:

$$E\left[(\mathbf{x}-\mathbf{a})(\mathbf{x}-\mathbf{b})^T\mathbf{A}(\mathbf{x}-\mathbf{c})\right]=tr\left(\mathbf{A}\mathbf{C}_\mathbf{x}\right)(\mathbf{m}_\mathbf{x}-\mathbf{a})+\mathbf{C}_\mathbf{x}\mathbf{A}^T(\mathbf{m}_\mathbf{x}-\mathbf{b})+\mathbf{C}_\mathbf{x}\mathbf{A}(\mathbf{m}_\mathbf{x}-\mathbf{c})$$
$$+(\mathbf{m}_\mathbf{x}-\mathbf{a})(\mathbf{m}_\mathbf{x}-\mathbf{b})^T\mathbf{A}(\mathbf{m}_\mathbf{x}-\mathbf{c})$$
$$E\left[(\mathbf{x}-\mathbf{a})^T\mathbf{B}(\mathbf{x}-\mathbf{b})(\mathbf{x}-\mathbf{c})^T\mathbf{D}(\mathbf{x}-\mathbf{d})\right]=tr\left(\mathbf{C}_\mathbf{x}\left(\mathbf{D}+\mathbf{D}^T\right)\mathbf{C}_\mathbf{x}\mathbf{B}^T\right)$$
$$+\left((\mathbf{m}_\mathbf{x}-\mathbf{a})^T\mathbf{B}+(\mathbf{m}_\mathbf{x}-\mathbf{b})^T\mathbf{B}^T\right)\mathbf{C}_\mathbf{x}\left(\mathbf{D}(\mathbf{m}_\mathbf{x}-\mathbf{d})+\mathbf{D}^T(\mathbf{m}_\mathbf{x}-\mathbf{c})\right)$$
$$+\left(tr\left(\mathbf{C}_\mathbf{x}\mathbf{B}^T\right)+(\mathbf{m}_\mathbf{x}-\mathbf{a})^T\mathbf{B}(\mathbf{m}_\mathbf{x}-\mathbf{b})\right)\left(tr\left(\mathbf{C}_\mathbf{x}\mathbf{D}^T\right)+(\mathbf{m}_\mathbf{x}-\mathbf{c})^T\mathbf{D}(\mathbf{m}_\mathbf{x}-\mathbf{d})\right)$$

then:

$$E\left[(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}))\left(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}')\right)^T\frac{\partial\mathbf{C}(\boldsymbol{\theta}')^{-1}}{\partial\omega'}\left(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}')\right)\right]=2\mathbf{C}^{ij}\frac{\partial\mathbf{C}(\boldsymbol{\theta}')^{-1}}{\partial\omega'}\left(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta}')\right)$$
$$+tr\left(\frac{\partial\mathbf{C}(\boldsymbol{\theta}')^{-1}}{\partial\omega'}\mathbf{C}^{ij}\right)(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta}))+(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta}))\left(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta}')\right)^T\frac{\partial\mathbf{C}(\boldsymbol{\theta}')^{-1}}{\partial\omega'}\left(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta}')\right)$$

$$E\left[(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}))^T\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\omega}(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}))\left(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}')\right)^T\frac{\partial\mathbf{C}(\boldsymbol{\theta}')^{-1}}{\partial\omega'}\left(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}')\right)\right]=$$
$$2tr\left(\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\omega}\mathbf{C}^{ij}\frac{\partial\mathbf{C}(\boldsymbol{\theta}')^{-1}}{\partial\omega'}\mathbf{C}^{ij}\right)+4(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta}))^T\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\omega}\mathbf{C}^{ij}\frac{\partial\mathbf{C}(\boldsymbol{\theta}')^{-1}}{\partial\omega'}\left(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta}')\right)$$
$$+\left(tr\left(\mathbf{C}^{ij}\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\omega}\right)+(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta}))^T\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\omega}(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta}))\right)\times$$
$$\left(tr\left(\mathbf{C}^{ij}\frac{\partial\mathbf{C}(\boldsymbol{\theta}')^{-1}}{\partial\omega'}\right)+\left(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta}')\right)^T\frac{\partial\mathbf{C}(\boldsymbol{\theta}')^{-1}}{\partial\omega'}\left(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta}')\right)\right)$$

$$E\left[(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}))^T\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\omega}(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}))\left(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}')\right)^T\frac{\partial\mathbf{C}(\boldsymbol{\theta}')^{-1}}{\partial\omega'}\left(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}')\right)\right]=$$
$$2tr\left(\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\omega}\mathbf{C}^{ij}\frac{\partial\mathbf{C}(\boldsymbol{\theta}')^{-1}}{\partial\omega'}\mathbf{C}^{ij}\right)+4(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta}))^T\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\omega}\mathbf{C}^{ij}\frac{\partial\mathbf{C}(\boldsymbol{\theta}')^{-1}}{\partial\omega'}\left(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta}')\right)$$
$$+tr\left(\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\omega}E\left[\widehat{\mathbf{C}(\boldsymbol{\theta})}\right]\right)tr\left(\frac{\partial\mathbf{C}(\boldsymbol{\theta}')^{-1}}{\partial\omega'}E\left[\widehat{\mathbf{C}(\boldsymbol{\theta}')}\right]\right)$$

$$E\left[(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}))\left(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}')\right)^T\frac{\partial\mathbf{C}(\boldsymbol{\theta}')^{-1}}{\partial\omega'}\left(\mathbf{x}-\mathbf{m}(\boldsymbol{\theta}')\right)\right]=tr\left(\frac{\partial\mathbf{C}(\boldsymbol{\theta}')^{-1}}{\partial\omega'}E\left[\widehat{\mathbf{C}(\boldsymbol{\theta}')}\right]\right)(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta}))$$
$$+2\mathbf{C}^{ij}\frac{\partial\mathbf{C}(\boldsymbol{\theta}')^{-1}}{\partial\omega'}\left(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta}')\right)$$

Finally:

$$[\mathbf{MS}]_{i,j}=E_{\boldsymbol{\theta}^0}\left[\frac{p\left(\mathbf{x};\boldsymbol{\theta}^i\right)}{p\left(\mathbf{x};\boldsymbol{\theta}^0\right)}\frac{p\left(\mathbf{x};\boldsymbol{\theta}^j\right)}{p\left(\mathbf{x};\boldsymbol{\theta}^0\right)}\right]=\sqrt{\frac{\left|\mathbf{C}^{ij}\right|\left|\mathbf{C}\left(\boldsymbol{\theta}^0\right)\right|}{\left|\mathbf{C}\left(\boldsymbol{\theta}^i\right)\right|\left|\mathbf{C}\left(\boldsymbol{\theta}^j\right)\right|}}e^{\frac{1}{2}\left[(\mathbf{m}^{ij})^T\mathbf{C}^{ij}\mathbf{m}^{ij}-\delta^{ij}\right]}\tag{37}$$

$$\left[\mathbf{h}\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right)\right]_p=E_{\boldsymbol{\theta}^0}\left[\frac{\partial\ln p\left(\mathbf{x};\boldsymbol{\theta}^i\right)}{\partial\theta_p}\frac{p\left(\mathbf{x};\boldsymbol{\theta}^i\right)p\left(\mathbf{x};\boldsymbol{\theta}^j\right)}{p\left(\mathbf{x};\boldsymbol{\theta}^0\right)^2}\right]=\left[\boldsymbol{\alpha}\left(\boldsymbol{\theta}^i\right)\right]_p[\mathbf{MS}]_{i,j}\tag{38}$$

$$\left[\mathbf{EFI}\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right)\right]_{p,p'} = E_{\boldsymbol{\theta}^0}\left[\frac{\partial \ln p\left(\mathbf{x};\boldsymbol{\theta}^i\right)}{\partial\theta_p}\frac{\partial \ln p\left(\mathbf{x};\boldsymbol{\theta}^j\right)}{\partial\theta_{p'}}\frac{p\left(\mathbf{x};\boldsymbol{\theta}^i\right)p\left(\mathbf{x};\boldsymbol{\theta}^j\right)}{p\left(\mathbf{x};\boldsymbol{\theta}^0\right)^2}\right] \quad (39)$$

$$= [\mathbf{MS}]_{i,j}\left\{\left[\boldsymbol{\Pi}\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right)\right]_{p,p'}+\frac{1}{4}\left[\boldsymbol{\Delta}\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right)\right]_{p,p'}-\frac{1}{2}\left[\boldsymbol{\Gamma}\left(\boldsymbol{\theta}^i,\boldsymbol{\theta}^j\right)\right]_{p,p'}-\frac{1}{2}\left[\boldsymbol{\Gamma}\left(\boldsymbol{\theta}^j,\boldsymbol{\theta}^i\right)\right]_{p',p}\right.$$

$$-\frac{1}{4}tr\left(\frac{\partial\mathbf{C}\left(\boldsymbol{\theta}^i\right)^{-1}}{\partial\theta_p}\mathbf{C}\left(\boldsymbol{\theta}^i\right)\right)tr\left(\frac{\partial\mathbf{C}\left(\boldsymbol{\theta}^j\right)^{-1}}{\partial\theta_{p'}}\mathbf{C}\left(\boldsymbol{\theta}^j\right)\right)$$

$$\left.+\frac{1}{2}tr\left(\frac{\partial\mathbf{C}\left(\boldsymbol{\theta}^i\right)^{-1}}{\partial\theta_p}\mathbf{C}\left(\boldsymbol{\theta}^i\right)\right)\left[\boldsymbol{\alpha}\left(\boldsymbol{\theta}^j\right)\right]_{p'}+\frac{1}{2}tr\left(\frac{\partial\mathbf{C}\left(\boldsymbol{\theta}^j\right)^{-1}}{\partial\theta_{p'}}\mathbf{C}\left(\boldsymbol{\theta}^j\right)\right)\left[\boldsymbol{\alpha}\left(\boldsymbol{\theta}^i\right)\right]_{p}\right\}$$

where $\mathbf{C}^{ij}$, $\mathbf{m}^{ij}$, δ^{ij} are given by (34a-34d) and:

$$E\left[\widehat{\mathbf{C}(\boldsymbol{\theta})}\right]=\mathbf{C}^{ij}+\left(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta})\right)\left(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta})\right)^T, \qquad E[\mathbf{x}]=\mathbf{C}^{ij}\mathbf{m}^{ij} \quad (40a)$$

$$[\boldsymbol{\alpha}(\boldsymbol{\theta})]_l=\frac{1}{2}tr\left(\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\theta_l}\left(\mathbf{C}(\boldsymbol{\theta})-E\left[\widehat{\mathbf{C}(\boldsymbol{\theta})}\right]\right)\right)+\frac{\partial\mathbf{m}(\boldsymbol{\theta})}{\partial\theta_l}^T\mathbf{C}(\boldsymbol{\theta})^{-1}\left(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta})\right) \quad (40b)$$

$$\left[\boldsymbol{\Pi}\left(\boldsymbol{\theta},\boldsymbol{\theta}'\right)\right]_{l,l'}=\frac{\partial\mathbf{m}(\boldsymbol{\theta})}{\partial\theta_l}^T\mathbf{C}(\boldsymbol{\theta})^{-1}\left[\mathbf{C}^{ij}+\left(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta})\right)\left(E[\mathbf{x}]-\mathbf{m}\left(\boldsymbol{\theta}'\right)\right)^T\right]\mathbf{C}\left(\boldsymbol{\theta}'\right)^{-1}\frac{\partial\mathbf{m}\left(\boldsymbol{\theta}'\right)}{\partial\theta_{l'}} \quad (40c)$$

$$\left[\boldsymbol{\Gamma}\left(\boldsymbol{\theta},\boldsymbol{\theta}'\right)\right]_{l,l'}=\frac{\partial\mathbf{m}(\boldsymbol{\theta})}{\partial\theta_l}^T\mathbf{C}(\boldsymbol{\theta})^{-1}\left\{tr\left(\frac{\partial\mathbf{C}\left(\boldsymbol{\theta}'\right)^{-1}}{\partial\theta_{l'}}E\left[\widehat{\mathbf{C}\left(\boldsymbol{\theta}'\right)}\right]\right)\left(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta})\right)\right. \quad (40d)$$

$$\left.+2\mathbf{C}^{ij}\frac{\partial\mathbf{C}\left(\boldsymbol{\theta}'\right)^{-1}}{\partial\theta_{l'}}\left(E[\mathbf{x}]-\mathbf{m}\left(\boldsymbol{\theta}'\right)\right)\right\}$$

$$\left[\boldsymbol{\Delta}\left(\boldsymbol{\theta},\boldsymbol{\theta}'\right)\right]_{l,l'}=tr\left(\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\theta_l}E\left[\widehat{\mathbf{C}(\boldsymbol{\theta})}\right]\right)tr\left(\frac{\partial\mathbf{C}\left(\boldsymbol{\theta}'\right)^{-1}}{\partial\theta_{l'}}E\left[\widehat{\mathbf{C}\left(\boldsymbol{\theta}'\right)}\right]\right) \quad (40e)$$

$$+4\left(E[\mathbf{x}]-\mathbf{m}(\boldsymbol{\theta})\right)^T\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\theta_l}\mathbf{C}^{ij}\frac{\partial\mathbf{C}\left(\boldsymbol{\theta}'\right)^{-1}}{\partial\theta_{l'}}\left(E[\mathbf{x}]-\mathbf{m}\left(\boldsymbol{\theta}'\right)\right)+2tr\left(\frac{\partial\mathbf{C}(\boldsymbol{\theta})^{-1}}{\partial\theta_l}\mathbf{C}^{ij}\frac{\partial\mathbf{C}\left(\boldsymbol{\theta}'\right)^{-1}}{\partial\theta_{l'}}\mathbf{C}^{ij}\right)$$

The M-dimensional complex non circular observation model (22) can be processed directly using above expressions derived for the real observation model provided $\mathbf{m}(\boldsymbol{\theta})=\mathbf{m}_{\overrightarrow{\mathbf{z}}}(\boldsymbol{\theta})$, $\mathbf{C}(\boldsymbol{\theta})=\mathbf{C}_{\overrightarrow{\mathbf{z}}}(\boldsymbol{\theta})$ and the substitution of the vector transpose conjugate $\left(^H\right)$ for the vector transpose $\left(^T\right)$.
This is a straightforward result deriving from the following identities: $\mathbf{m_x}=\mathbf{T}\mathbf{m}_{\overrightarrow{\mathbf{z}}}$, $\mathbf{C_x}=\mathbf{T}\mathbf{C}_{\overrightarrow{\mathbf{z}}}\mathbf{T}^H$, $\mathbf{m_x}^T=\mathbf{m_x}^H=\mathbf{m}_{\overrightarrow{\mathbf{z}}}^H\mathbf{T}^H$, $\left|\mathbf{TAT}^H\right|=\frac{|\mathbf{A}|}{\mathbf{2^M}}$, $tr\left(\left(\mathbf{TAT}^H\right)^{-1}\left(\mathbf{TBT}^H\right)\right)=tr\left(\mathbf{A}^{-1}\mathbf{B}\right)$,
$tr\left(\left(\mathbf{TAT}^H\right)^{-1}\left(\mathbf{TBT}^H\right)\left(\mathbf{TCT}^H\right)^{-1}\left(\mathbf{TDT}^H\right)\right)=tr\left(\mathbf{A}^{-1}\mathbf{B}\mathbf{C}^{-1}\mathbf{D}\right)$, where $\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}\in\mathcal{M}(M,M)$.
Further simplifications, introduced in section V, can be obtained in the special case of the M-dimensional complex circular observation model (24), with the help of a few lines of algebra by noticing that $\forall\mathbf{u},\mathbf{v}\left(\mathbf{T}\overrightarrow{\mathbf{u}}\right)^T\left(\mathbf{T}\overrightarrow{\mathbf{v}}\right)=\text{Re}\left\{\mathbf{u}^H\mathbf{v}\right\}$ and using relation $\mathbf{C}_{\mathbf{z},\mathbf{z}^*}=\mathbf{0}$, since then:

$$\mathbf{C_x}(\boldsymbol{\theta})=\mathbf{T}\mathbf{C}_{\overrightarrow{\mathbf{z}}}(\boldsymbol{\theta})\mathbf{T}^H=\mathbf{T}\begin{bmatrix}\mathbf{C_z}(\boldsymbol{\theta}) & \mathbf{0}\\ \mathbf{0} & \mathbf{C_z}(\boldsymbol{\theta})^*\end{bmatrix}\mathbf{T}^H, \quad \mathbf{m_x}(\boldsymbol{\theta})=\mathbf{T}\mathbf{m}_{\overrightarrow{\mathbf{z}}}(\boldsymbol{\theta})=\mathbf{T}\overrightarrow{\mathbf{m_z}}(\boldsymbol{\theta}),$$

therefore $\frac{\partial\mathbf{m_x}(\boldsymbol{\theta})}{\partial\theta}=\mathbf{T}\overrightarrow{\frac{\partial\mathbf{m_z}(\boldsymbol{\theta})}{\partial\theta}}$ and:

$$\frac{\partial\mathbf{C_x}(\boldsymbol{\theta})}{\partial\theta}=\mathbf{T}\begin{bmatrix}\frac{\partial\mathbf{C_z}(\boldsymbol{\theta})}{\partial\theta} & \mathbf{0}\\ \mathbf{0} & \frac{\partial\mathbf{C_z}(\boldsymbol{\theta})^*}{\partial\theta}\end{bmatrix}\mathbf{T}^H, \quad \frac{\partial\mathbf{C_x}(\boldsymbol{\theta})^{-1}}{\partial\theta}=\left(\mathbf{T}^H\right)^{-1}\begin{bmatrix}\frac{\partial\mathbf{C_z}(\boldsymbol{\theta})^{-1}}{\partial\theta} & \mathbf{0}\\ \mathbf{0} & \left(\frac{\partial\mathbf{C_z}(\boldsymbol{\theta})^{-1}}{\partial\theta}\right)^*\end{bmatrix}\mathbf{T}^{-1}$$

where $\mathbf{x}$ is a $2M$-dimensional real vector.

REFERENCES

[1] H.L. Van Trees, *Detection, Estimation and Modulation Theory, Part 1*. New York: Wiley, 1968.

[2] H. Cramér, *Mathematical Methods of Statistics. Princeton*. NJ: Princeton Univ. Press, 1946.

[3] M. Fréchet, "Sur l'extension de certaines évaluations statistiques au cas de petits échantillons," Rev. Int. Stat, vol. 11, pp. 182-205, 1943.

[4] C.R. Rao, "Information and accuracy attainable in the estimation of statistical parameters," Bull. Calcutta Math. Soc., vol. 37, pp. 81-91, 1945.

[5] A. Renaux, P. Forster, E. Chaumette, and P. Larzabal, "On the high SNR CML estimator full statistical characterization," IEEE Trans. Signal Process., vol. 54, no. 12, pp. 4840-4843, Dec. 2006.

[6] R.A. Fisher, "On an absolute criterion for fitting frequency curves," Mess. of Math., vol. 41, pp. 155-160, 1912.

[7] E.W. Barankin, "Locally best unbiased estimates," Ann. Math. Stat., vol. 20, no. 4, pp. 477-501, 1949.

[8] D.G. Chapman and H. Robbins, "Minimum variance estimation without regularity assumptions," Ann. Math. Stat., vol. 22, no. 4, pp. 581-586, 1951.

[9] R. McAulay and L.P. Seidman, "A useful form of the Barankin lower bound and its application to PPM threshold analysis," IEEE Trans. Inf. Theory, vol. 15, no. 2, pp. 273-279, Mar. 1969.

[10] A. Bhattacharyya, "On some analogues of the amount of information and their use in statistical estimation," Sankhya Indian J. Stat, vol. 8, pp. 1-14, 1946.

[11] P. Forster and P. Larzabal, "On lower bounds for deterministic parameter estimation," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., pp.1137-1140, 2002.

[12] T.L. Marzetta, "Computing the Barankin bound, by solving an unconstrained quadratic optimization problem," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., pp. 3829-3832, 1997.

[13] R. McAulay and E.M. Hofstetter, "Barankin Bounds on parameter estimation," IEEE Trans. Inf. Theory, vol. 17, no. 6, pp. 669-676, Nov. 1971.

[14] F.E. Glave, "A new look at the Barankin lower bound," IEEE Trans. Inf. Theory, vol. 18, no. 3, pp. 349-356, May 1972.

[15] J.S. Abel, "A bound on mean-square-estimate error," IEEE Trans. Inf. Theory, vol. 39, no. 5, pp. 1675-1680, Sep. 1993.

[16] L. Knockaert, "The Barankin bound and threshold behaviour in frequency estimation," IEEE Trans. Signal Process., vol. 45, no. 9, pp. 2398–2401, Sep. 1997.

[17] J. Tabrikian and J.L. Krolik, "Barankin bounds for source localization in an uncertain ocean environnment," IEEE Trans. Signal Process., vol. 47, no. 11, pp. 2917-2927, Nov. 1999.

[18] A. Renaux, L. N. Atallah, P. Forster, and P. Larzabal, "A useful form of the Abel bound and its application to estimator threshold prediction," IEEE Trans. Signal Process, vol. 55, no. 5, pp. 2365-2369, May 2007.

[19] D.C. Rife and R.R. Boorstyn, "Single tone parameter estimation from discrete-time observations," IEEE Trans. Inf. Theory, vol. 20, no.5, pp. 591-598, Sep. 1974.

[20] P. Forster, P. Larzabal, and E. Boyer, "Threshold performance analysis of maximum likelihood DOA estimation," IEEE Trans. Signal Process., vol. 52, no. 11, Nov. 2004.

[21] F. Athley, "Threshold Region Performance of Maximum Likelihood Direction of Arrival Estimators," IEEE Trans., SP vol. 53, pp.1359-1373, apr. 2005.

[22] C. D. Richmond, "Capon algorithm mean squared error threshold SNR prediction and probability of resolution," IEEE Trans. Signal Process., vol. 53, no. 8, pp. 2748-2764, Aug. 2005.

[23] A. J.Weiss and E.Weinstein, "A lower bound on the mean square error in random parameter estimation," IEEE Trans. Inf. Theory, vol. 31, no. 9, pp. 680–682, Sep. 1985.

[24] K. Bell, Y. Steinberg, Y. Ephraim, and H. L. V. Trees, "Extended Ziv Zakai lower bound for vector parameter estimation," IEEE Trans. Signal Process., vol. 43, no. 3, pp. 624-638, Mar. 1997.

[25] A. Renaux, "Weiss-Weinstein bound for data aided carrier estimation," IEEE Signal Process. Letters, vol. 14, no. 4, pp. 283–286, Apr. 2007.

[26] C. R. Blyth, "Necessary and sufficient conditions for inequalities of Cramér-Rao type," Ann. Math. Stat., vol. 2, no. 3, pp. 464–473, 1974.

[27] E. Chaumette, P. Larzabal, and P. Forster, "On the influence of a detection step on lower bounds for deterministic parameters estimation," IEEE Trans. Signal Process., vol. 53, no. 11, pp. 4080-4090, Nov. 2005.

[28] J. Kiefer, "On minimum variance estimators," Ann. Math. Stat., vol. 23, no. 4, pp. 627–629, 1952.

[29] D.A.S. Fraser and I. Guttman, "Bhattacharyya bounds without regularity assumptions," Ann. Math. Stat., vol. 23, no. 4., pp. 629–632, 1952.

[30] P. Stoica and A. Nehorai, "Performances study of conditional and unconditional direction of arrival estimation," IEEE Trans. Signal Process., vol. 38, no. 10, pp. 1783-1795, Oct. 1990.

[31] A. Van Den Bos, "The multivariate complex normal distribution -a generalization," IEEE Trans. Inf. Theory, vol. 41, no. 2, pp 537-539, Mar. 1995.

[32] S.M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1998.

[33] B. Picinbono, "On circularity," IEEE Trans. Signal Process. vol. 42, no. 12, pp. 3473-3482, Dec. 1994.

[34] H.L. Van Trees, *Optimum Array Processing*. New-York: Wiley, 2002.

[35] P. Chevalier and F. Pipon, "New insights into optimal widely linear array receivers for the demodulation of BPSK, MSK, and GMSK signals corrupted by noncircular interferences-application to SAIC," IEEE Trans. Signal Process., vol. 54, no. 3, pp. 870-883, Mar. 2006.

[36] M. Brookes, "The Matrix reference manual," [online] Matrix Reference Manual - 2005.

[37] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* NJ: Prentice-Hall, 1993.

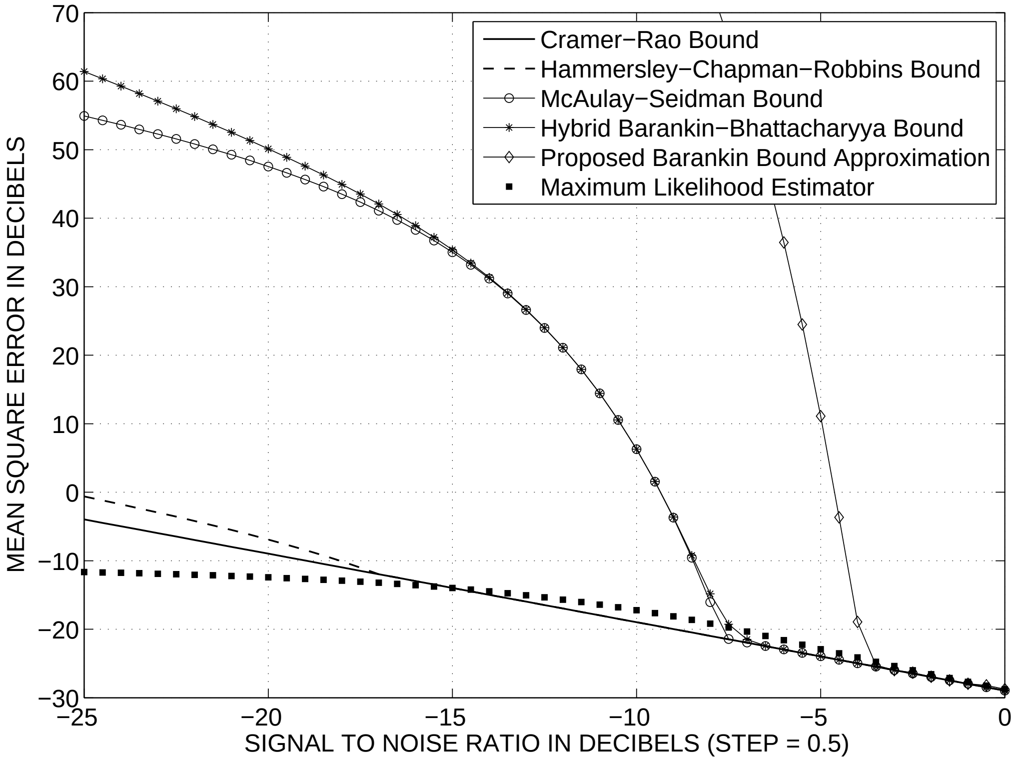


Fig. 1. MSE lower bounds versus SNR for the Deterministic signal model, $M = 2$, $S = 10$ independant observations

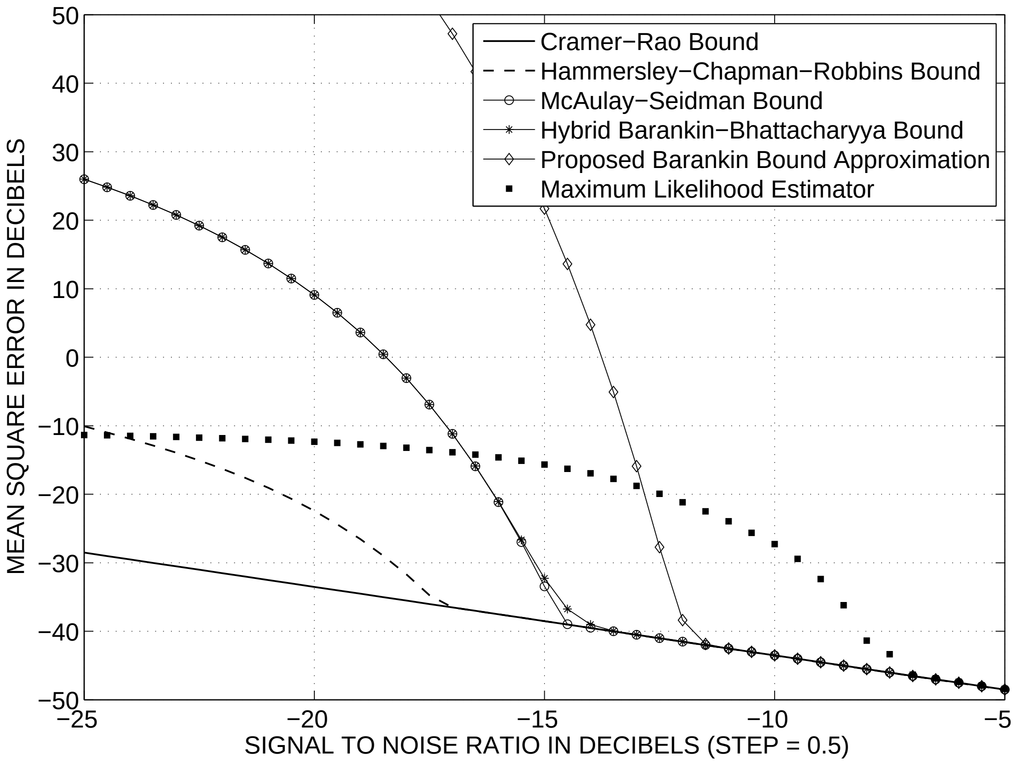


Fig. 2. MSE lower bounds versus SNR for the Deterministic signal model, $M = 10$, $S = 10$ independant observations

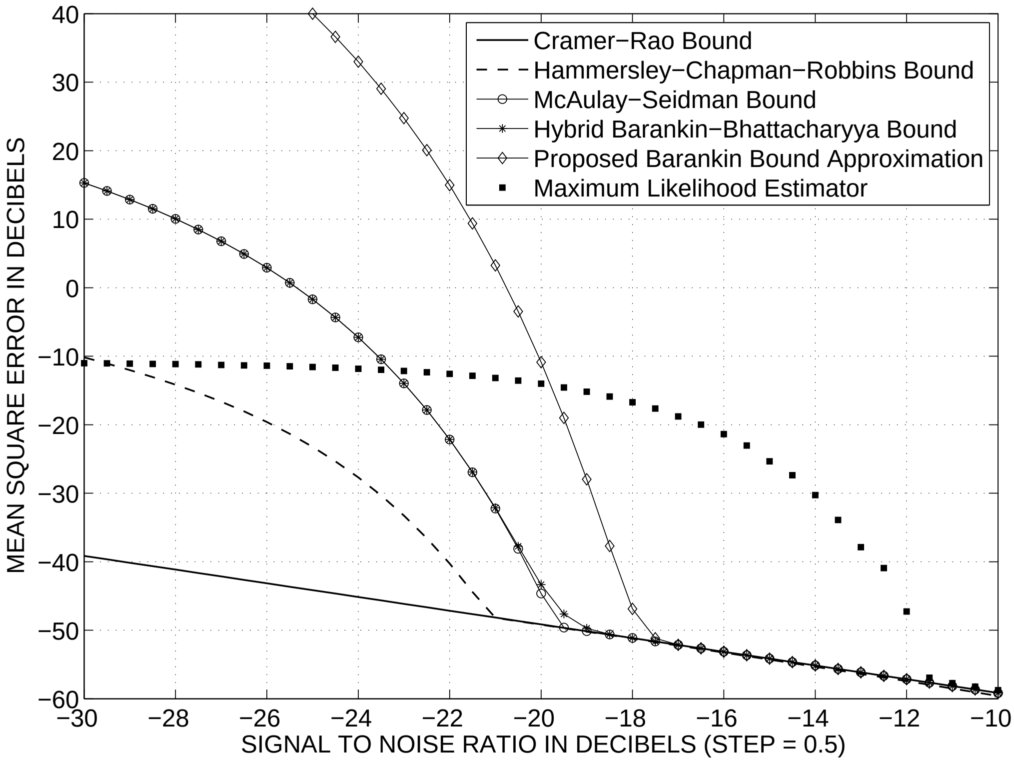


Fig. 3. MSE lower bounds versus SNR for the Deterministic signal model, $M = 32$, $S = 10$ independant observations

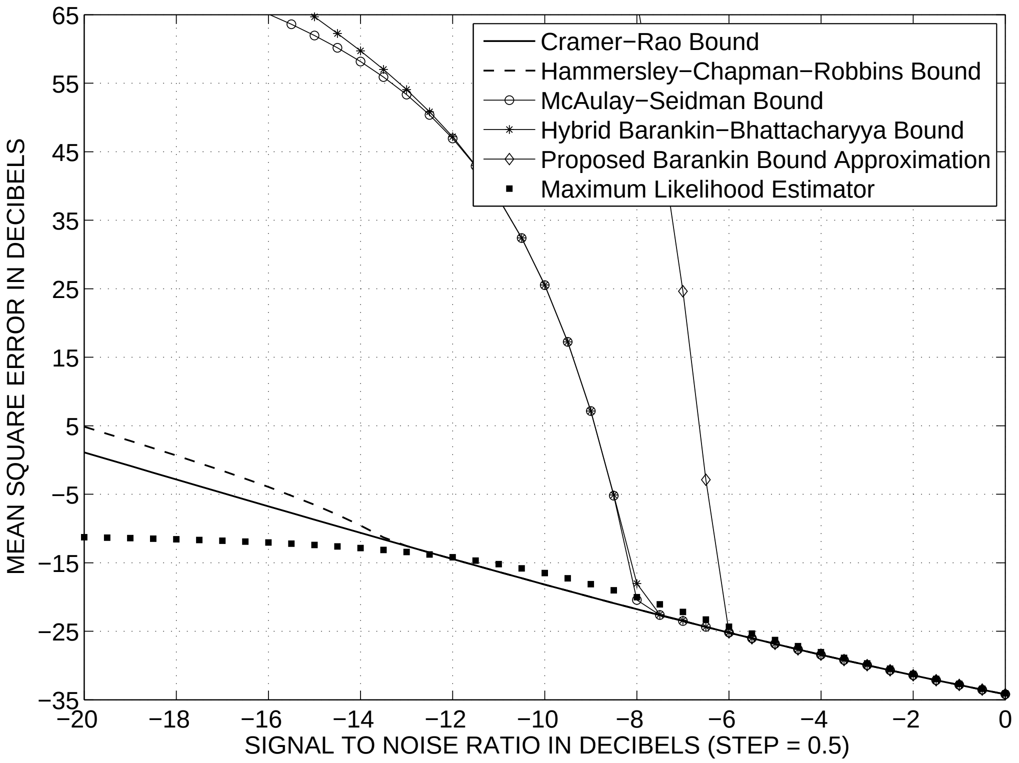


Fig. 4. MSE lower bounds versus SNR for the Stochastic signal model, $M = 2$, $S = 100$ independant observations

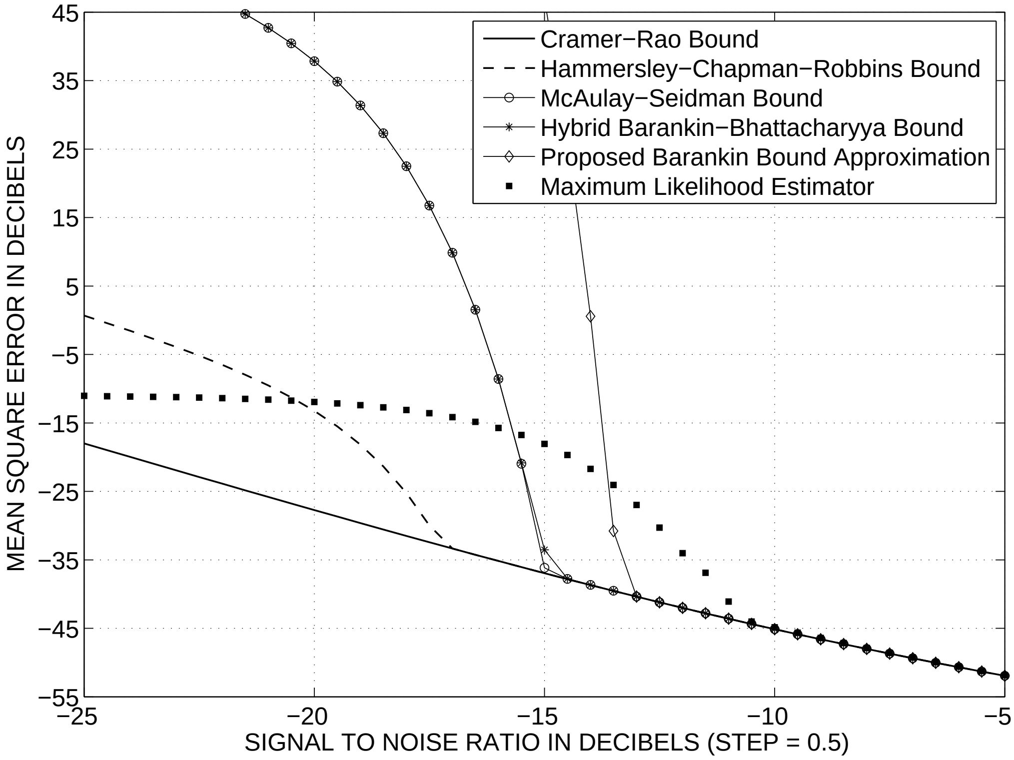


Fig. 5. MSE lower bounds versus SNR for the Stochastic signal model, $M = 10$, $S = 100$ independant observations

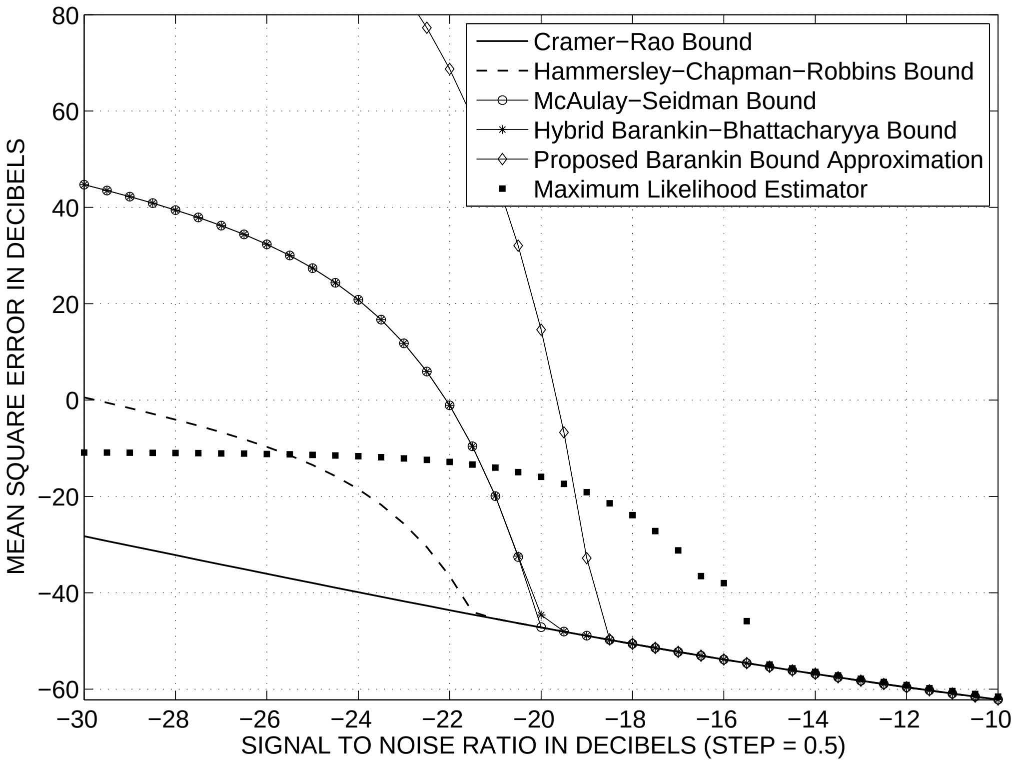


Fig. 6. MSE lower bounds versus SNR for the Stochastic signal model, $M = 32$, $S = 100$ independant observations

G. Monopulse-radar tracking of swerling III-IV targets using multiple observations (IEEE TAES)

IEEE Trans. on AES, 44(2): 520-537, 2008.

Monopulse-Radar Tracking of Swerling 3-4 Targets using Multiple Observations.

Eric Chaumette and Pascal Larzabal

Abstract

This paper proposes a novel statistical prediction of monopulse errors [1] for a radar Swerling 3-4 target embedded in noise or noise jamming where multiple observations are available. First, the study of the Maximum Likelihood Estimator (MLE) of the complex monopulse ratio for a Swerling 3-4 target embedded in spatially white noise allows us to extend the use of the MLE practical approximate form introduced by Mosca [2] for Swerling 0-1-2 cases. Afterward, we derive analytical formulas for both the mean and variance of the MLE in approximate form conditioned by the usual detection step performed on the sum channel of a monopulse antenna. Last, we provide a comparison of target direction of arrival (DOA) estimation performance based on monopulse ratio estimation as a function of the Swerling model in the context of a multifunction radar.

Index Terms

Radar, Direction of Arrival estimation, Monopulse antennas, Maximum likelihood estimation, Variance estimation

I. Notation

$\overrightarrow{X}, \overrightarrow{x}$ denote vectors (complex or real)
$\mathbf{C}, \widehat{\mathbf{R}}, \mathbf{m}$ denote matrices (complex or real) and $|\ |$ denotes a determinant
$\Sigma_i, \Delta_i, \alpha_i, \beta_i, g_\Sigma, n_{\Sigma_i}, g_\Delta, n_{\Delta_i}, r, \ldots$ denote scalar values (complex or real)
$\mathrm{Re}\{\ \}$ and $\mathrm{Im}\{\ \}$ denote respectively the "real part" and the "imaginary part" of a complex value
$f(\ )$ and $f(x \mid y)$ denote respectively a probability density function (p.d.f.) and a conditional p.d.f.
$P(\ )$ and $P(A \mid B)$ denotes respectively a probability and a conditional probability
$\widehat{r}$ denotes an estimator of r
$E[\ ]$ and $Var[\ ]$ denote respectively mean (expectation) and variance
$\overrightarrow{m}_{\overrightarrow{X}} = E\left[\overrightarrow{X}\right]$ denotes the mean of random vector $\overrightarrow{X}$
$\mathbf{C}_{\overrightarrow{X},\overrightarrow{Y}} = E\left[\left(\overrightarrow{X} - \overrightarrow{m}_{\overrightarrow{X}}\right)\left(\overrightarrow{Y} - \overrightarrow{m}_{\overrightarrow{Y}}\right)^H\right]$ denotes the covariance matrix of random vectors $\overrightarrow{X},\overrightarrow{Y}$
$\mathbf{C}_{\overrightarrow{X}} = \mathbf{C}_{\overrightarrow{X},\overrightarrow{X}}$ denotes the covariance matrix of random vector $\overrightarrow{X}$
$\mathbf{Id}_I$ denotes Identity matrix with dimensions (I,I)
$\mathbf{C}_{ij}$ denotes element (i,j) of matrix $\mathbf{C}$
If $z = x + jy$ is a complex variable, then: $\int h(z)\,dz = \iint h(x,y)\,dxdy$
If $\overrightarrow{\alpha}$ is a I-dimensional vector, then: $\int\limits_{\overrightarrow{\alpha}} h(\overrightarrow{\alpha})\,\overrightarrow{d\alpha} = \int\limits_{\alpha_1,\ldots,\alpha_I}\cdots\int h(\alpha_1,\ldots,\alpha_I)\,d\alpha_1\ldots d\alpha_I$
If $\overrightarrow{\alpha}$ is a K-dimensional vector, and $\overrightarrow{\Sigma}$ a I-dimensional vector, then:

E. Chaumette is with Thales Air Defence, Hameau de Roussigny - 91470 Limours, France. E-mail: eric.chaumette@noos.fr, Eric.Chaumette@fr.thalesgroup.com, Phone: 33 1 64 91 69 00, Fax: 33 1 64 91 69 89.

P. Larzabal is with SATIE/CNRS, Ecole Normale Supérieure de Cachan, 61 avenue du Président Wilson - 94235 Cachan Cedex, France. E-mail: pascal.larzabal@satie.ens-cachan.fr, Phone: 33 1 47 40 27 09.

This work has been partially performed in the framework of the European Community Contract no. 507325, NEWCOM

$$\iint h\left(\overrightarrow{\alpha}, \overrightarrow{\Sigma}\right) \overrightarrow{d\alpha} \overrightarrow{d\Sigma} = \underset{\alpha_1,..,\alpha_K,\Sigma_1,..,\Sigma_I}{\int \cdots \int} f\left(\alpha_1, .., \alpha_K, \Sigma_1, .., \Sigma_I, \right) d\alpha_1..d\alpha_K d\Sigma_1..d\Sigma_I$$

If $\overrightarrow{\Sigma} = \left(\Sigma_1, \Sigma_2, \ldots, \Sigma_I\right)^T$, then: $\frac{\partial}{\partial \overrightarrow{\Sigma}} = \left(\frac{\partial}{\partial \Sigma_1}, \ldots, \frac{\partial}{\partial \Sigma_I}\right)^T, \quad \frac{\partial}{\partial \overrightarrow{\Sigma}^T} = \left(\frac{\partial}{\partial \Sigma_1}, \ldots, \frac{\partial}{\partial \Sigma_I}\right)$

$$e_N\left(T\right) = \sum_{n=0}^{N} \frac{T^n}{n!} = \frac{e^T}{N!} \int_T^\infty e^{-t} t^N dt, \quad E_1\left(T\right) = \int_T^\infty \frac{e^{-t}}{t} dt$$

II. Introduction

The estimation of the direction of arrival (DOA) of a target by means of a radar consisting of a two-sensor array is one of the oldest and most widely used high-precision techniques, even nowadays, in most operational tracking systems [1]. Its principle is the following: assume that a target situated at an angle θ (deviation angle from array boresight) is received on a two-sensor (Σ and Δ) array in the presence of a circular, zero mean, white (both temporally and spatially), complex Gaussian thermal noise. A common model of the observation equation dedicated to this problem - after Hilbert Filtering - is the following receiver signal vector [1]:

$$\overrightarrow{v}\left(t\right) = \begin{pmatrix} \Sigma\left(t\right) \\ \Delta\left(t\right) \end{pmatrix} = \alpha(t) \begin{pmatrix} g_\Sigma\left(\theta\right) \\ g_\Delta\left(\theta\right) \end{pmatrix} + \begin{pmatrix} n_\Sigma\left(t\right) \\ n_\Delta\left(t\right) \end{pmatrix} = \alpha(t) \overrightarrow{g}\left(\theta\right) + \overrightarrow{n}(t) \tag{1}$$

where $n_\Sigma\left(t\right)$ and $n_\Delta\left(t\right)$ represent Gaussian receiver noise, $g_\Sigma\left(\theta\right)$ and $g_\Delta\left(\theta\right)$ represent the one-way complex antenna voltage pattern at angle θ and $\alpha(t)$ represents the complex amplitude of the radar target (including power budget equation, signal processing gains). In the particular case of a two-sensor array, the angular information is contained in the ratio $r\left(\theta\right) = \frac{g_\Delta(\theta)}{g_\Sigma(\theta)}$, provided the function $\theta \longrightarrow r\left(\theta\right)$ is invertible. In actual two-sensor arrays the beamwidth constraint generally prevents this assumption from being verified for any θ in $\left[\frac{-\pi}{2}, \frac{\pi}{2}\right]$. Nevertheless with appropriate sensors patterns design - for instance, collocated sensors paired in phase with uniform sum excitation for Σ and linear odd difference excitation for Δ [1, p. 290] - $r\left(\theta\right)$ is real and the property can hold for θ belonging to Σ main lobe, i.e. between the first pattern nulls. Such two-sensor array are generally called monopulse antennas where $r\left(\theta\right)$ is the monopulse ratio and $\theta = r^{-1}\left(\frac{g_\Delta}{g_\Sigma}\right)$ is the deviation angle function. If a linear relation $r = k\theta$ is assumed - which is true at the vicinity of boresight [1, p. 294] (see figure (1)) - then statistical prediction of any estimator $\widehat{r}$ of r provides as well statistical prediction of the corresponding estimator of θ defined by: $\widehat{\theta} = \frac{\widehat{r}}{k}$. It is the reason why in open literature [1, p. 294] the deviation angle function is generally reduced to a linear function characterized by a Monopulse Slope and most DOA statistical performance analyses are related to the monopulse ratio r. We will consider this approximation in the present paper.

In the static situation in which the target does not alter its relative position with respect to the monopulse antenna during I independent measurements at time $t_i, i \in [1, I]$, the Maximum Likelihood Estimator (MLE) $\widehat{r}_{ML}$ of the target monopulse ratio has been derived by Mosca [2] for a non fluctuating target or a fluctuating target with a Rayleigh or unknown a priori amplitude fluctuation law, embedded in spatially white noise with known powers. In the fluctuating case, Mosca also provided an approximate form - valid at high Signal-to-Noise-Ratio (SNR) and target close to boresight -:

$$\widehat{r}_{ML}^{MOSCA} \approx \mathrm{Re}\left\{\frac{\sum_{i=1}^{I} \Sigma\left(t_i\right)^H \Delta\left(t_i\right)}{\sum_{i=1}^{I} \left|\Sigma\left(t_i\right)\right|^2}\right\} \tag{2}$$

that has become since the practical monopulse ratio estimator. Historically, the statistical properties of this approximate MLE (2) have been first treated focusing on statistics averaged over all possible amplitudes of both sum and difference channels for:

- a Swerling 1 ($I = 1$ observation) or Swerling 2 ($I \geq 2$ observations) target [3] (Rayleigh amplitude). This case has been completely characterized - including the computation of the probability density function (p.d.f.) - first by Kanter using the characteristic functions [4][5]. More recently Tullsson [6] has proposed a simpler approach based on conditional Gaussian distribution that allows not only to obtain the same results, but also to take into account the true monopulse estimation problem where the receiver sum-channel is completed by a detection step (see also Seifer [7]), which leads to the computation of conditional mean and conditional variance. Additionally

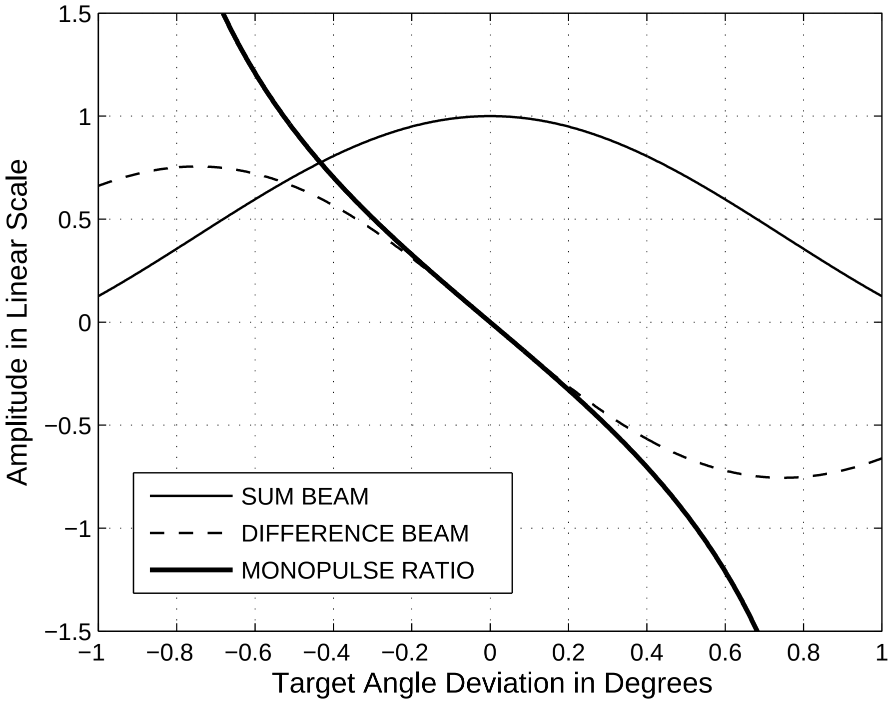


Fig. 1. Beams Pattern and Monopulse Ratio of Monopulse Rectangular Reflector Antenna (1° beamwidth)

this calculation scheme has inspired us a pertinent breakdown of the p.d.f of the observations into the form of a product of polynomials and Gaussian p.d.f.s in order to treat the problem as a calculation of moments of Gaussian random variables (see VII-C).

• a Swerling 0 target [3] (non fluctuating amplitude) which has been characterized by Seifer [7] in terms of conditional mean and variance taken into account as well the sum beam detection step.

Regarding the Swerling 3 ($I = 1$ observation) or Swerling 4 ($I \geq 2$ observations) target case [3], which is a most interesting case from an operational point of view in radar since it is the default fluctuation model for missiles, it has been characterized only partially so far by authors in [8] who have expressed the p.d.f. of the observation model of a monopulse antenna where 2 independent Swerling 3 targets are present [8, eq. 42]. Unfortunately due to the complexity of the p.d.f. obtained they have not been able to obtain analytic expression of MLEs of the 2 unknown monopulse ratios.

Thus, none of these past works has provided yet a relevant statistical characterization of the monopulse ratio MLE for a Swerling 3 or 4 target, which is the originality of the present paper where we demonstrate the validity of Mosca MLE approximate form (2) for a Swerling 3–4 target and derives its statistical prediction in terms of conditional mean and variance taken into account the sum beam detection step.

For completeness, let us recall that the contribution of colored noise (jammers) can be cancelled by resorting to adaptive arrays which are very effective in improving the signal-to-noise-plus-interference ratio (SNIR). However, due to the nulls in the adapted pattern, the shape of the beams can be severely distorted. Therefore, this technique requires corrected adaptive monopulse formulas which performance has been characterized by Nickel [10] for Rayleigh targets and jammers, including the sum beam detection step.

Last, let us mention that an alternative statistical treatment of monopulse parameters dedicated to support Kalman Filter Tracking has been lately introduced in [11]: for each independent monopulse measurement, the complex monopulse ratio is conditioned on the measured amplitude of the sum signal. They have not only derived a complete characterization of their approach (p.d.f., Cramer Rao Bound, variance of Conditional Maximum Likelihood), but also provided new algorithms for the detection of interference from two or more targets in monopulse measurement and for the extraction of separate angle measurements for two closely spaced (unresolved) Rayleigh [12] and Swerling 3 targets.

Throughout the paper, for sake of simplicity in expressions, most references to observation time $t_i, i \in [1, I]$ will be expressed implicitly by use of subscript $_i$ and explicit dependence on θ of $g_\Sigma, g_\Delta, r, \overrightarrow{g}, ...$ will be omitted.

III. Monopulse ratio MLEs

A. Background: Deterministic and Stochastic MLEs of the monopulse ratio

Although not introduced in these terms in the initial work of Mosca [2], Monopulse measurement can nowadays be considered within the framework of modern Direction of Arrival (DOA) estimation analysis [9] as being the particular problem of estimation of the steering vector of a two-sensor array. In this framework mostly two different signal models are considered: the deterministic and the stochastic signal model. The discussed signal models are Gaussian and the angular dependency is given by parameters which are connected with the expectation value in the deterministic case and with the covariance matrix in the stochastic one. In the particular case of a monopulse antenna (1) and I independent measurements, the p.d.f. is of the following general form [2][9]:

$$\begin{aligned} f\left(\overrightarrow{\Sigma}, \overrightarrow{\Delta}\right) &= f\left(\overrightarrow{V}\right) = \int f\left(\overrightarrow{V} \mid \overrightarrow{\alpha'}\right) f\left(\overrightarrow{\alpha'}\right) \overrightarrow{d\alpha'} \\ \overrightarrow{\Sigma} &= \left(\Sigma_1 = \Sigma\left(t_1\right), \ldots, \Sigma_I = \Sigma\left(t_I\right)\right)^T, \quad \overrightarrow{\Delta} = \left(\Delta_1 = \Delta\left(t_1\right), \ldots, \Delta_I = \Delta\left(t_I\right)\right)^T \\ \overrightarrow{V} &= \left(\overrightarrow{v_1}^T, \ldots, \overrightarrow{v_I}^T\right)^T, \quad \overrightarrow{v_i} = \left(\Sigma_i, \Delta_i\right)^T \end{aligned} \tag{3}$$

where [9, §8.5]:

$$\begin{aligned} f\left(\overrightarrow{V} \mid \overrightarrow{\alpha'}\right) &= \prod_{i=1}^{I} \frac{e^{-\left(\overrightarrow{v_i} - \alpha'_i \overrightarrow{g}\right)^H \mathbf{C}_{\overrightarrow{n}}^{-1} \left(\overrightarrow{v_i} - \alpha'_i \overrightarrow{g}\right)}}{\pi^2 \left|\mathbf{C}_{\overrightarrow{n}}\right|} = \frac{e^{-ITr\left\{\mathbf{C}_{\overrightarrow{n}}^{-1} \widehat{\mathbf{C}_{\overrightarrow{v}}}\right\}}}{\left(\pi^2 \left|\mathbf{C}_{\overrightarrow{n}}\right|\right)^I} \\ \widehat{\mathbf{C}_{\overrightarrow{v}}} &= \frac{1}{I} \sum_{i=1}^{I} \left(\overrightarrow{v_i} - \alpha'_i \overrightarrow{g}\right) \left(\overrightarrow{v_i} - \alpha'_i \overrightarrow{g}\right)^H, \quad \overrightarrow{g} = \left(g_\Sigma, g_\Delta\right)^T \end{aligned} \tag{4}$$

The deterministic model is obtained when [2]:

$$f\left(\overrightarrow{\alpha'}\right) = \prod_{i=1}^{I} \delta\left(\alpha'_i - \alpha_i\right) \quad \Rightarrow \quad f\left(\overrightarrow{V}\right) = f\left(\overrightarrow{V} \mid \overrightarrow{\alpha}\right) \tag{5}$$

and the stochastic model when [9, §8.5]:

$$\begin{aligned} f\left(\overrightarrow{\alpha'}\right) &= \frac{e^{-\frac{\overrightarrow{\alpha}^H \overrightarrow{\alpha}}{\sigma_\alpha^2}}}{\left(\pi \sigma_\alpha^2\right)^I} \quad \Rightarrow \quad f\left(\overrightarrow{V}\right) = \prod_{i=1}^{I} \frac{e^{-\overrightarrow{v_i}^H \mathbf{C}_{\overrightarrow{v}}^{-1} \overrightarrow{v_i}}}{\pi^2 \left|\mathbf{C}_{\overrightarrow{v}}\right|} = \frac{e^{-ITr\left\{\mathbf{C}_{\overrightarrow{v}}^{-1} \widehat{\mathbf{C}_{\overrightarrow{v}}}\right\}}}{\left(\pi^2 \left|\mathbf{C}_{\overrightarrow{v}}\right|\right)^I} \\ \widehat{\mathbf{C}_{\overrightarrow{v}}} &= \frac{1}{I} \sum_{i=1}^{I} \overrightarrow{v_i} \overrightarrow{v_i}^H, \quad \mathbf{C}_{\overrightarrow{v}} = \sigma_\alpha^2 \overrightarrow{g} \overrightarrow{g}^H + \mathbf{C}_{\overrightarrow{n}}, \quad \sigma_\alpha^2 = E\left[\left|\alpha_i\right|^2\right] \end{aligned} \tag{6}$$

If we refer to standard Swerling models of fluctuation of a radar target [3], the deterministic model corresponds to the Swerling 0 model if $I = 1$ or $\alpha_i = \alpha, i \in [1, I]$, and the stochastic model corresponds either to Swerling 1 model if $I = 1$, or to Swerling 2 model if $I \geq 2$. Assuming that the only unknown parameters are the target parameters and that the noise is spatially white with known powers $\left(\sigma_{n_\Sigma}^2, \sigma_{n_\Delta}^2\right)$ on both channels, both for the deterministic - when the α_i are not supposed equal - and stochastic model, the MLE estimator $\widehat{r}_{ML}$ of the complex monopulse ratio is given by [13][14]:

$$\widehat{r}_{ML} = \arg\max \left\{ \frac{\overrightarrow{x}^H \widehat{\mathbf{R}_{\overrightarrow{v}}} \overrightarrow{x}}{\overrightarrow{x}^H \overrightarrow{x}} \right\} = \frac{Tr\left(\widehat{\mathbf{R}_{\overrightarrow{v}}}\right) + \sqrt{Tr\left(\widehat{\mathbf{R}_{\overrightarrow{v}}}\right)^2 - 4\left|\widehat{\mathbf{R}_{\overrightarrow{v}}}\right| - 2\left\|\overrightarrow{\Sigma}\right\|^2}}{2\overrightarrow{\Delta}^H \overrightarrow{\Sigma}} \tag{7}$$

where $\overrightarrow{x} = (1, r)^T$ and $\widehat{\mathbf{R}_{\overrightarrow{v}}} = \frac{1}{I} \sum_{i=1}^{I} \overrightarrow{v_i} \overrightarrow{v_i}^H$.

This results is an extension to results previously derived by Mosca [2] under the assumption that the observation of a single target where the monopulse ratio is real. Indeed, in practice, the reception channels Σ and Δ of a monopulse antenna are paired in phase so as to obtain a real monopulse ratio. Nevertheless, Asseo in [15] has shown that it was worth considering the monopulse ratio as a complex value: whereas the real part supplies the angular measurement, the imaginary part provides a simple detection test for the simultaneous presence of several targets [1][5][6][7][15], i.e. for the validity of the angular measurement. In this case $\text{Im}\{r\} \neq 0$, although $\text{Im}\{r\} = 0$ in the case of a single target.

The usual hypothesis of high SNR and target close to boresight $\left(\left\|\overrightarrow{\Sigma}\right\|^2 >> \left\|\overrightarrow{\Delta}\right\|^2\right)$ [2] leads to the well-known following approximation [2][13][14]:

$$\widehat{r}_{ML} \approx \frac{\overrightarrow{\Sigma}^H \overrightarrow{\Delta}}{\overrightarrow{\Sigma}^H \overrightarrow{\Sigma}} = \frac{\sum_{i=1}^{I} \Sigma_i^H \Delta_i}{\sum_{i=1}^{I} |\Sigma_i|^2} \tag{8}$$

Last, to be complete, let us mention that :

- expression (8) is the exact $\widehat{r}_{ML}$ if $I = 1$, since then:

$$\widehat{r}_{ML} = \frac{\Delta_1}{\Sigma_1} \tag{9}$$

- in the deterministic model, if $\alpha_i = \alpha, i \in [1, I]$, then:

$$\widehat{r}_{ML} = \frac{\sum_{i=1}^{I} \Delta_i}{\sum_{i=1}^{I} \Sigma_i} \tag{10}$$

  which is in fact the computation of the MLE given by (9) in the case of a single measurement of a target of amplitude α embedded in white noise, measurement which happens to be the output of the Matched Filter (coherent processing) applied to a target which does not fluctuate during the set of measurements [2]. In the deterministic model with I independent measurements, the choice of using expression (8) or (10) generally depends on the operational context (duration between measurements, waveforms, ...) which determines the validity of the hypothesis of a non fluctuating target.

B. Swerling 3-4 MLE of the monopulse ratio

Swerling 3 ($I = 1$) and Swerling 4 ($I \geq 2$) models of amplitude fluctuation law of a radar target consists of the following target amplitude p.d.f. [3][8]:

$$f(\alpha_i) = \frac{4\,|\alpha_i|^2}{\pi(\sigma_\alpha^2)^2} e^{-\frac{2|\alpha_i|^2}{\sigma_\alpha^2}}, \quad \sigma_\alpha^2 = E\left[|\alpha_i|^2\right]$$

which is not only a most interesting case from an operational point of view in radar, since it is the default statistical fluctuation model for missiles, but also not a classical study case in the DOA / array processing open literature [9]. Indeed, to our best knowledge, the only attempt so far to derive the MLE of the monopulse ratio for Swerling 3-4 models has been performed by authors in [8] who have expressed the p.d.f. of the observation model of a monopulse antenna (1) where 2 independent Swerling 3 (single observation) target are present [8, eq. 42]. Unfortunately due to the complexity of the p.d.f. obtained they have not been able to obtain analytic expression of MLEs of the 2 unknown monopulse ratios and have resorted to a classical search over a parameter-space grid. As a contribution to this research effort, we derive in the next section the analytical expression of the p.d.f. of the observation model of a monopulse antenna (1) where 1 single Swerling 3-4 target (single or multiple observations) is present. But there again, if the complexity of the p.d.f. prevents us from obtaining an exact analytical expression of the MLE, it nevertheless allows us to derive an approximate form which appears to be the standard extended Mosca approximation (8), as shown hereinafter.

1) Calculation of the p.d.f. of a single observation: To simplify the expressions we define $\overrightarrow{v}_i = \overrightarrow{v}$, $\alpha_i = \alpha$ and $\sigma^2 = \frac{\sigma_\alpha^2}{2}$. Then:

$$f\left(\overrightarrow{v}\right) = \int f\left(\overrightarrow{v} \mid \alpha\right) f\left(\alpha\right) d\alpha = \int \frac{|\alpha|^2}{\pi\left(\sigma^2\right)^2} \frac{e^{\frac{-1}{\sigma^2}\left[|\alpha|^2 + \sigma^2\left(\overrightarrow{v} - \alpha\overrightarrow{g}\right)^H \mathbf{C}_{\overrightarrow{n}}^{-1}\left(\overrightarrow{v} - \alpha\overrightarrow{g}\right)\right]}}{\pi^2\left|\mathbf{C}_{\overrightarrow{n}}\right|} d\alpha \tag{11}$$

The above expression of $f\left(\overrightarrow{v}\right)$ can be simplified by noticing that (see details in Appendix VII-A):

$$\begin{aligned} A &= |\alpha|^2 + \sigma^2\left(\overrightarrow{v} - \alpha\overrightarrow{g}\right)^H \mathbf{C}_{\overrightarrow{n}}^{-1}\left(\overrightarrow{v} - \alpha\overrightarrow{g}\right) \\ &= \sigma^2\overrightarrow{v}^H\mathbf{C}^{-1}\overrightarrow{v} + \left(1 + \sigma^2\overrightarrow{g}^H\mathbf{C}_{\overrightarrow{n}}^{-1}\overrightarrow{g}\right)\left|\alpha - \frac{\sigma^2\overrightarrow{g}^H\mathbf{C}_{\overrightarrow{n}}^{-1}\overrightarrow{v}}{1 + \sigma^2\overrightarrow{g}^H\mathbf{C}_{\overrightarrow{n}}^{-1}\overrightarrow{g}}\right|^2 \end{aligned}$$

where:

$$\mathbf{C} = \sigma^2\overrightarrow{g}\,\overrightarrow{g}^H + \mathbf{C}_{\overrightarrow{n}}, \quad |\mathbf{C}| = \left|\mathbf{C}_{\overrightarrow{n}}\right|\left(1 + \sigma^2\overrightarrow{g}^H\mathbf{C}_{\overrightarrow{n}}^{-1}\overrightarrow{g}\right).$$

Consequently:

$$f\left(\overrightarrow{v}\right) = \frac{e^{-\overrightarrow{v}^H\mathbf{C}^{-1}\overrightarrow{v}}}{\sigma^2\pi^2\left|\mathbf{C}_{\overrightarrow{n}}\right|\left(1 + \sigma^2\overrightarrow{g}^H\mathbf{C}_{\overrightarrow{n}}^{-1}\overrightarrow{g}\right)}\left[\int \frac{|\alpha|^2}{\pi\left(\frac{\sigma^2}{1+\sigma^2\overrightarrow{g}^H\mathbf{C}_{\overrightarrow{n}}^{-1}\overrightarrow{g}}\right)} e^{-\frac{1+\sigma^2\overrightarrow{g}^H\mathbf{C}_{\overrightarrow{n}}^{-1}\overrightarrow{g}}{\sigma^2}\left|\alpha - \frac{\sigma^2\overrightarrow{g}^H\mathbf{C}_{\overrightarrow{n}}^{-1}\overrightarrow{v}}{1+\sigma^2\overrightarrow{g}^H\mathbf{C}_{\overrightarrow{n}}^{-1}\overrightarrow{g}}\right|^2} d\alpha\right]$$

and finally:

$$f\left(\overrightarrow{v}\right) = \frac{e^{-\overrightarrow{v}^H\mathbf{C}^{-1}\overrightarrow{v}}}{\pi^2|\mathbf{C}|}\left[\frac{1}{1 + \sigma^2\overrightarrow{g}^H\mathbf{C}_{\overrightarrow{n}}^{-1}\overrightarrow{g}} + \sigma^2\left|\overrightarrow{v}^H\mathbf{C}^{-1}\overrightarrow{g}\right|^2\right], \qquad \mathbf{C} = \frac{\sigma_\alpha^2}{2}\overrightarrow{g}\,\overrightarrow{g}^H + \mathbf{C}_{\overrightarrow{n}} \tag{12}$$

2) Calculation of the p.d.f. of I observations: When I independent observations are available, then the p.d.f. of $f\left(\overrightarrow{V}\right)$ is given by (12):

$$\begin{aligned} f\left(\overrightarrow{V}\right) &= \prod_{i=1}^{I} f\left(\overrightarrow{v_i}\right) = \prod_{i=1}^{I} \frac{e^{-\overrightarrow{v_i}^H\mathbf{C}^{-1}\overrightarrow{v_i}}}{\pi^2|\mathbf{C}|}\left(\frac{1}{1 + \frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\mathbf{C}_{\overrightarrow{n}}^{-1}\overrightarrow{x}} + \frac{\sigma_\beta^2}{2}\left|\overrightarrow{v_i}^H\mathbf{C}^{-1}\overrightarrow{x}\right|^2\right) \\ \mathbf{C} &= \frac{\sigma_\beta^2}{2}\overrightarrow{x}\,\overrightarrow{x}^H + \mathbf{C}_{\overrightarrow{n}}, \quad \overrightarrow{x} = (1, r)^T, \quad \sigma_\beta^2 = \sigma_\alpha^2\left|g_\Sigma\right|^2 \end{aligned} \tag{13}$$

Additionally, it is worth noting that expressions (12) and (13) hold whatever is the common dimension of the $\overrightarrow{v_i}$.

3) Derivation of MLE of complex monopulse ratio: We consider in the following the standard noise characteristics in a monopulse antenna [1][2]: spatially white with known powers $\left(\sigma_{n_\Sigma}^2, \sigma_{n_\Delta}^2\right)$ on both channels and independent from observation to observation. Without any loss of generality and for the sake of simplicity in expressions, we can assume - since the noise powers are known - that the observations have been normalized by $\left(1/\sqrt{\sigma_{n_\Sigma}^2}, 1/\sqrt{\sigma_{n_\Delta}^2}\right)$ factors on each channel in order to obtain a normalized spatially white noise ($\mathbf{C}_{\overrightarrow{n}} = \mathbf{Id}_2$). The MLE of the complex monopulse ratio is obtained for σ_β^2 and $r = r_x + jr_y$ values for which $f\left(\overrightarrow{V}\right)$ (13) is maximum [9], i.e. resorting to both real and complex derivatives [17]:

$$\frac{\partial f\left(\overrightarrow{V}\right)}{\partial\sigma_\beta^2} = 0 \Leftrightarrow \frac{\partial \ln f\left(\overrightarrow{V}\right)}{\partial\sigma_\beta^2} = 0 \tag{14a}$$

$$\frac{\partial f\left(\overrightarrow{V}\right)}{\partial r^*} = 0 \Leftrightarrow \frac{\partial \ln f\left(\overrightarrow{V}\right)}{\partial r^*} = 0 \tag{14b}$$

Unfortunately system of equations (14a-b) leads to a system of 2 implicit equations without any obvious analytical solution neither for $\widehat{\sigma^2_{\beta}}_{ML}$ or $\widehat{r}_{ML}$ (see Appendix VII-B for details). Nevertheless the combination of (14a-b) and some rearrangements allows to show that $\widehat{\sigma^2_{\beta}}_{ML}$ and $\widehat{r}_{ML}$ are the solution of:

$$\sum_{i=1}^{I}\left(1+\frac{1}{1+\mu_i\left(\sigma_\beta^2,r,r^*\right)}\right)\frac{\partial}{\partial r^*}\left[\frac{\vec{x}^H\left(\vec{v_i}\vec{v_i}^H\right)\vec{x}}{\vec{x}^H\vec{x}}\right]=0 \tag{15}$$

$$\mu_i\left(\sigma_\beta^2,r,r^*\right)=\frac{\sigma_\beta^2}{2}\frac{\left|\vec{v_i}^H\vec{x}\right|^2}{1+\frac{\sigma_\beta^2}{2}\vec{x}^H\vec{x}}$$

The main interest of the above maximum condition is its equivalent form at high SNR, since (see Appendix VII-B)

$$\lim_{\sigma_\beta^2\to\infty}P\left(\frac{1}{1+\mu_i\left(\sigma_\beta^2,r,r^*\right)}=0\right)=1,$$

which demonstrates that asymptotically at high SNR (15) is "almost surely" equivalent to:

$$\sum_{i=1}^{I}\frac{\partial}{\partial r^*}\left[\frac{\vec{x}^H\left(\vec{v_i}\vec{v_i}^H\right)\vec{x}}{\vec{x}^H\vec{x}}\right]=0=\frac{\partial}{\partial r^*}\left[\frac{\vec{x}^H\left(\sum\limits_{i=1}^{I}\vec{v_i}\vec{v_i}^H\right)\vec{x}}{\vec{x}^H\vec{x}}\right]=\arg\max\left\{\frac{\vec{x}^H\widehat{\mathbf{R}_{\vec{v}}}\vec{x}}{\vec{x}^H\vec{x}}\right\}$$

that is to say, has "almost surely" the same expression as the MLE of the complex monopulse ratio for the deterministic and stochastic models (7). Additionally, the "almost surely" equivalence criterion becomes an exact equivalence in the Swerling 3 model, i.e., if only a single observation is available. In the same way, the usual approximation of high SNR and target close to boresight leads to the classical practical approximation (8), which complete the proof of the relevance of the use of (8) as an asymptotic approximation of the MLE (or exact value if a single observation) of the complex monopulse ratio whatever the Swerling Fluctuation model.

IV. Conditional mean and variance

In most tracking devices the monopulse antenna is connected to a receiver chain which, after various analogue and/or digital processing, is completed by a threshold detection applied to the receiver sum-channel (Σ) [1]. If an unknown amplitude fluctuation is assumed, then the generalized likelihood ratio test (GLRT) reduces to the energy detector defined by [16, §7.3]:

$$\left\|\vec{\Sigma}\right\|^2\geq T \tag{16}$$

For sake of simplicity in formulas, the event of a threshold detection $\left\{\vec{\Sigma}\mid\left\|\vec{\Sigma}\right\|^2\geq T\right\}$ when the target is present is denoted by D and its probability by P_D - Probability of Detection [1] -. The detection threshold T is determined by the desired Probability of False Alarm [1] - detection on thermal noise only - according to the usual formula:

$$P_{FA}=e^{\frac{-T}{\sigma^2_{n_\Sigma}}}e_I\left(\frac{T}{\sigma^2_{n_\Sigma}}\right). \tag{17}$$

As the monopulse measurement is performed only on detected samples, mean and variance must be computed taking into account this observations selection criterion requesting use of conditional mean $E\left[\mathrm{Re}\left\{\widehat{r}_{ML}\right\}\mid D\right]$ and conditional variance $Var\left[\mathrm{Re}\left\{\widehat{r}_{ML}\right\}\mid D\right]$, where $\widehat{r}_{ML}$ is defined by equation (8).
One way to compute conditional mean and variance is to introduce (see Appendix VII-C) an auxiliary variable $\vec{U}$ and an auxiliary function $s\left(\vec{U}\right)$ defined by:

$$\vec{U}=\left(\frac{\Sigma_1^H\Delta_1}{\vec{\Sigma}^H\vec{\Sigma}},..,\frac{\Sigma_I^H\Delta_I}{\vec{\Sigma}^H\vec{\Sigma}}\right)^T,\qquad s\left(\vec{U}\right)=\sum_{i=1}^{I}\mathrm{Re}\left\{U_i\right\}-\mathrm{Re}\left\{\frac{\mathbf{C}_{12}^H}{\mathbf{C}_{11}}\right\}=\mathrm{Re}\left\{\widehat{r}_{ML}\right\}-\mathrm{Re}\left\{\frac{\mathbf{C}_{12}^H}{\mathbf{C}_{11}}\right\}$$

which allow to express $E\left[\mathrm{Re}\left\{\widehat{r}_{ML}\right\} \mid D\right]$ and $Var\left[\mathrm{Re}\left\{\widehat{r}_{ML}\right\} \mid D\right]$ as:

$$E\left[\mathrm{Re}\left\{\widehat{r}_{ML}\right\} \mid D\right] = \mathrm{Re}\left\{\frac{\mathbf{C}_{12}^H}{\mathbf{C}_{11}}\right\} + E\left[s\left(\overrightarrow{U}\right) \mid D\right] \quad (18)$$

$$Var\left[\mathrm{Re}\left\{\widehat{r}_{ML}\right\} \mid D\right] = Var\left[s\left(\overrightarrow{U}\right) \mid D\right] = E\left[s\left(\overrightarrow{U}\right)^2 \mid D\right] - E\left[s\left(\overrightarrow{U}\right) \mid D\right]^2 \quad (19)$$

After some fastidious calculus detailed in appendix VII-C, $E\left[s\left(\overrightarrow{U}\right) \mid D\right]$ and $E\left[s\left(\overrightarrow{U}\right)^2 \mid D\right]$ are given by:

- if $I = 1$:

$$E\left[s\left(\overrightarrow{U}\right) \mid D\right] = \frac{\sigma_\alpha^2}{2}\left|k_2\right|^2 \mathrm{Re}\left\{n\right\}\left|\mathbf{C}\right| \frac{1}{\mathbf{C}_{11} + \frac{T}{\mathbf{C}_{11}}\frac{\sigma_\alpha^2}{2}\left|g_\Sigma\right|^2} \quad (20a)$$

$$E\left[s\left(\overrightarrow{U}\right)^2 \mid D\right] = \frac{\left|\mathbf{C}\right|}{2\left(\mathbf{C}_{11}\right)^2}\left[\frac{\frac{\sigma_\alpha^2}{2}\frac{\left|g_\Sigma\right|^2}{\mathbf{C}_{11}} + \left(1 + \frac{\sigma_\alpha^2}{2}\left|k_2\right|^2\left|\mathbf{C}\right|\right)\frac{e^{\frac{T}{\mathbf{C}_{11}}}}{\mathbf{C}_{11}}E_1\left(\frac{T}{\mathbf{C}_{11}}\right)}{\mathbf{1} + \frac{T}{\mathbf{C}_{11}}\frac{\sigma_\alpha^2}{2}\frac{\left|g_\Sigma\right|^2}{\mathbf{C}_{11}}}\right] \quad (20b)$$

- if $I \geq 2$:

$$E\left[s\left(\overrightarrow{U}\right) \mid D\right] = \frac{I\sigma_\alpha^2 A_4}{2A_0}\left|k_2\right|^2 \mathrm{Re}\left\{n\right\}\left|\mathbf{C}\right| \quad (21a)$$

$$E\left[s\left(\overrightarrow{U}\right)^2 \mid D\right] = \frac{\left|\mathbf{C}\right| A_1}{2A_0\left(\mathbf{C}_{11}\right)^2} + \frac{I\sigma_\alpha^2\left|k_2\right|^2\left|\mathbf{C}\right|^2 A_2}{4A_0\left(\mathbf{C}_{11}\right)^2} + \frac{I\left(I-1\right)A_3}{A_0}\left[\frac{\sigma_\alpha^2}{2}\left|k_2\right|^2 \mathrm{Re}\left\{n\right\}\left|\mathbf{C}\right|\right]^2 \quad (21b)$$

and P_D is given - whatever I - by:

$$P_D = \frac{e^{-\frac{T}{\mathbf{C}_{11}}}}{\left(\mathbf{C}_{11}\right)^I}A_0 \quad (22)$$

where:

$\overrightarrow{k} = \begin{pmatrix} k_1 \\ k_2 \end{pmatrix} = \mathbf{C}^{-1}\overrightarrow{g}$, $n = \left(\frac{k_1}{k_2} + \frac{\mathbf{C}_{12}}{\mathbf{C}_{11}}\right)^H$

$A_0 = \sum_{m=0}^{I} p_I^m \left(\mathbf{C}_{11}\right)^m e_{I-1+m}\left(\frac{T}{\mathbf{C}_{11}}\right)(I-1+m)!$

$A_1 = \sum_{m=0}^{I} p_I^m \left(\mathbf{C}_{11}\right)^m e_{I-2+m}\left(\frac{T}{\mathbf{C}_{11}}\right)(I-2+m)!$

$A_2 = \sum_{m=0}^{I-1} h_{I-1}^m \left(\mathbf{C}_{11}\right)^m e_{I-2+m}\left(\frac{T}{\mathbf{C}_{11}}\right)(I-2+m)!$

$A_3 = \sum_{m=0}^{I-2} h_{I-2}^m \left(\mathbf{C}_{11}\right)^m e_{I-1+m}\left(\frac{T}{\mathbf{C}_{11}}\right)\frac{(I-1+m)!}{(I+m)(I+m+1)}$

$A_4 = \sum_{m=0}^{I-1} h_{I-1}^m \left(\mathbf{C}_{11}\right)^m e_{I-1+m}\left(\frac{T}{\mathbf{C}_{11}}\right)(I-1+m)!$

p_I^m and h_I^m are coefficients of the following recursive polynomials $P_I\left(t\right)$ and $H_I\left(t\right)$ defined by:

$$\begin{cases} P_I\left(t\right) = \left(\sum_{m=0}^{I} p_I^m t^m\right)t^{I-1} = Q_{I-1}\left(t\right) + bH_{I-1}\left(t\right) \\ H_I\left(t\right) = \left(\sum_{m=0}^{I} h_I^m t^m\right)t^{I+1} = \int_0^t Q_I\left(z\right)dz \\ Q_I\left(t\right) = \int_0^t P_I\left(z\right)dz \end{cases} \quad \text{if } I \geq 1 \text{ where } \begin{cases} b = \frac{\sigma_\alpha^2}{2}\frac{\left|g_\Sigma\right|^2}{\mathbf{C}_{11}} \\ Q_0(t) = 1 \\ H_0\left(t\right) = t \end{cases}$$

It is worth noting that the above statistical prediction has been derived for any $\mathbf{C}_{\overrightarrow{n}}$ and encompasses therefore the thermal noise only or thermal noise plus jammers cases.
Last, for sake of simplicity in formulas and derivation, both sum-channel and difference-channel are assumed to be normalized by $\left(1/\sqrt{\sigma^2_{n_\Sigma}} = 1/\sqrt{(\mathbf{C}_{\overrightarrow{n}})_{11}},\, 1/\sqrt{\sigma^2_{n_\Delta}} = 1/\sqrt{(\mathbf{C}_{\overrightarrow{n}})_{22}}\right)$ factors. Therefore the normalized values of g_Σ, g_Δ and $\mathbf{C}_{\overrightarrow{n}}$ must be used in the analytical formulas above.

V. Performance Analysis

As already mentioned, the estimation of target DOA by computation of a monopulse ratio is still the most widely used high-precision technique in operational tracking systems, particularly in modern multifunction Radars which must operate both surveillance and tracking mode [1]. Therefore, among all the various types of analysis on target DOA estimation performance that could be drawn from analytical formulas (18)(19)(20a-b)(21a-b) - including calibration faults, presence of Gaussian stand-off jammers, thermal noise correlation, ... - we will focus on the statistical prediction of target DOA as a function of its Swerling model where the target is embedded in a temporally and spatially white thermal noise. Indeed, it is an usual customer requirement that a radar designer or manufacturer must comply with. This requirement is generally expressed in terms of DOA estimation precision for target angle deviation values belonging to an interval centered around boresight - generally the sum beam $-3dB$ beamwidth - and a target amplitude leading to $P_D = 0.9$ for a given P_{FA} when target is at boresight.
As an example, we consider a modern multifunction radar with a sum beam (Σ) having a 1° beamwidth at $-3dB$ and an associated difference beam (Δ) as shown in figure (1). The monopulse beams are supposed to be perfectly paired in phase and equivalent - in the sum beam main lobe - to a plane reflector illuminated by a uniform current distribution for the sum channel and an odd linear distribution for the difference channel, which generates the following voltage patterns [1]:

$$g_\Sigma(\theta) = \frac{\sqrt{4\pi}}{\lambda}\sqrt{L}\left[\frac{\sin(h(\theta))}{h(\theta)}\right], \quad g_\Delta(\theta) = \frac{\sqrt{4\pi}}{\lambda}\sqrt{3L}\left[\frac{\sin(h(\theta))}{h(\theta)^2} - \frac{\cos(h(\theta))}{h(\theta)}\right]$$

where $h(\theta) = \pi\frac{L}{\lambda}\sin(\theta)$, L is the dimension of the antenna and λ is the wavelength.
In a multifunction radar:
• the most likely number of independent observations to compute the monopulse ratio is 1 or 2, since due to time budget constraints its tracking mode can not generally afford more than two successive observations per tracked target,
• a typical P_{FA} in tracking mode is 10^{-4}.
To perform the comparison of performance in DOA estimation as a function of the Swerling model, we have resorted to theoretical formulas derived in [7] for the Swerling 0 case, in [6] for the Swerling 1-2 case, and in the present paper (18)(19)(20a-b)(21a-b) for the Swerling 3-4 case, completed by Monte-Carlo simulations (5.10^5 independent trials) in the Swerling 3-4 case.
If a performance comparison only requires the assessment of the Probability of Detection (see figures (2) and (3)) and of the Conditional Root Mean Square Error (RMSE, see figures (4) and (5)) of the target angle deviation MLE $\widehat{\theta}_{ML} = \frac{\mathrm{Re}\{\widehat{r}_{ML}\}}{k}$, we have additionally provided the Conditional Mean (see figures (6) and (7)) and the Conditional Standard Deviation (STD, see figures (8) and (9)) for a more complete overview of estimation performance factors. All the introduced figures has been restricted to positive values of target angle deviation θ for straightforward symmetry reasons since, in our example, the monopulse ratio $r(\theta) = \frac{g_\Delta(\theta)}{g_\Sigma(\theta)}$ is an odd function (see figure (1)). In all legends, "Theo" stands for Theoretical results and "Simu" for Simulation results (Monte-Carlo).

First of all, a perfect agreement between simulations and theoretical formulas can be observed in the swerling 3-4 case in all figures, which proves the validity of the new results (18)(19)(20a-b)(21a-b) derived in the present paper.

Then, let us recall that the Swerling 3-4 target model has been designed [3] to be an intermediate case, from the amplitude p.d.f. spread standpoint, between the Dirac p.d.f. (Swerling 0 target) and the Rayleigh p.d.f. (Swerling 1-2 target), leading to intermediate probability of detection as illustrated by figures (2) and (3). Although not introduced in the present paper, we have also conducted the same simulations (1 or 2 observations) for the additional $P_{FA} = 10^{-3}$, $P_{FA} = 10^{-5}$, $P_{FA} = 10^{-6}$ values encompassing thus the usual P_{FA} values operated in a multifunction radar. From all these simulations, the only noticeable result is the extension of the intermediate behavior of the Swerling 3-4 fluctuation law to conditional RMSE and STD performance in the monopulse antenna beamwidth at $-3dB$ (see figures (4)(5) and (8) (9)). On the other hand, such property is not applicable to the bias, and no particular conclusion should be drawn from figures (6) and (7).

VI. Conclusion

The present paper provides analytical and efficient formulas to compute performance of monopulse ratio estimation when a monopulse radar faces a Swerling 3-4 fluctuating target (missile for example). Indeed, despite their apparent complexity, formulas (20a-b)(21a-b) are very simple to program as they mainly consists of polynomials. Therefore, this work, by completing results previously derived [4][5][6][7], allows any monopulse radar designer to have a comprehensive set of analytical formulas to investigate the influence of his design on target DOA estimation performance for all the reference target amplitude fluctuation laws in radar.

VII. Appendix

A. Derivation of the p.d.f. of observations for a Swerling 3 target :

We start from expression (11) of $f(\overrightarrow{v})$. This expression can be simplified by noticing that:

$$\begin{aligned} A &= |\alpha|^2 + \sigma^2 (\overrightarrow{v} - \alpha \overrightarrow{g})^H \mathbf{C}_{\overrightarrow{n}}^{-1} (\overrightarrow{v} - \alpha \overrightarrow{g}) \\ &= \sigma^2 \overrightarrow{v}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{v} + |\alpha|^2 \left(1 + \sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g}\right) - 2\sigma^2 \operatorname{Re}\left\{\overrightarrow{v}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g} \alpha\right\} \\ &= \sigma^2 \overrightarrow{v}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{v} + \left(1 + \sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g}\right) \left[\left|\alpha - \frac{\sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{v}}{1 + \sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g}}\right|^2 - \left|\frac{\sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{v}}{1 + \sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g}}\right|^2\right] \\ &= \sigma^2 \left[\overrightarrow{v}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{v} - \frac{\sigma^2 \left|\overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{v}\right|^2}{1 + \sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g}}\right] + \left(1 + \sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g}\right) \left|\alpha - \frac{\sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{v}}{1 + \sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g}}\right|^2 \end{aligned}$$

$$\begin{aligned} A &= \sigma^2 \overrightarrow{v}^H \left[\mathbf{C}_{\overrightarrow{n}}^{-1} - \frac{\sigma^2 \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g} \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1}}{1 + \sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g}}\right] \overrightarrow{v} + \left(1 + \sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g}\right) \left|\alpha - \frac{\sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{v}}{1 + \sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g}}\right|^2 \\ &= \sigma^2 \overrightarrow{v} \mathbf{C}^{-1} \overrightarrow{v} + \left(1 + \sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g}\right) \left|\alpha - \frac{\sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{v}}{1 + \sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g}}\right|^2 \end{aligned}$$

where:

$$\mathbf{C} = \sigma^2 \overrightarrow{g} \overrightarrow{g}^H + \mathbf{C}_{\overrightarrow{n}} = \frac{\sigma_\alpha^2}{2} \overrightarrow{g} \overrightarrow{g}^H + \mathbf{C}_{\overrightarrow{n}}, \quad |\mathbf{C}| = |\mathbf{C}_{\overrightarrow{n}}| \left(1 + \sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g}\right)$$

Consequently:

$$\begin{aligned} f(\overrightarrow{v}) &= \frac{e^{-\overrightarrow{v}^H \mathbf{C}^{-1} \overrightarrow{v}}}{\pi^2 |\mathbf{C}_{\overrightarrow{n}}|} \int \frac{|\alpha|^2}{\pi (\sigma^2)^2} e^{-\frac{1 + \sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g}}{\sigma^2} \left|\alpha - \frac{\sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{v}}{1 + \sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g}}\right|^2} d\alpha \\ &= \frac{e^{-\overrightarrow{v}^H \mathbf{C}^{-1} \overrightarrow{v}}}{\sigma^2 \pi^2 |\mathbf{C}_{\overrightarrow{n}}| \left(1 + \sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g}\right)} \left[\int \frac{|\alpha|^2}{\pi \left(\frac{\sigma^2}{1 + \sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g}}\right)} e^{-\frac{1 + \sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g}}{\sigma^2} \left|\alpha - \frac{\sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{v}}{1 + \sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g}}\right|^2} d\alpha\right] \\ &= \frac{e^{-\overrightarrow{v}^H \mathbf{C}^{-1} \overrightarrow{v}}}{\sigma^2 \pi^2 |\mathbf{C}|} E\left[|\alpha|^2\right] \end{aligned}$$

where:

$$E\left[|\alpha|^2\right] = Var(\alpha) + |E[\alpha]|^2 = \frac{\sigma^2}{1 + \sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g}} + \left|\frac{\sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{v}}{1 + \sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g}}\right|^2$$

therefore:

$$f(\overrightarrow{v}) = \frac{e^{-\overrightarrow{v}^H \mathbf{C}^{-1} \overrightarrow{v}}}{\sigma^2 \pi^2 |\mathbf{C}|} \left[\frac{\sigma^2}{1 + \sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g}} + \left|\frac{\sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{v}}{1 + \sigma^2 \overrightarrow{g}^H \mathbf{C}_{\overrightarrow{n}}^{-1} \overrightarrow{g}}\right|^2\right]$$

Moreover, since:

$$\mathbf{C}^{-1}\overrightarrow{g}=\frac{\mathbf{C}_{\overrightarrow{n}}^{-1}\overrightarrow{g}}{1+\sigma^2\overrightarrow{g}^H\mathbf{C}_{\overrightarrow{n}}^{-1}\overrightarrow{g}},$$

finally:

$$f\left(\overrightarrow{v}\right)=\frac{e^{-\overrightarrow{v}^H\mathbf{C}^{-1}\overrightarrow{v}}}{\pi^2\left|\mathbf{C}\right|}\left[\frac{1}{1+\sigma^2\overrightarrow{g}^H\mathbf{C}_{\overrightarrow{n}}^{-1}\overrightarrow{g}}+\sigma^2\left|\overrightarrow{v}^H\mathbf{C}^{-1}\overrightarrow{g}\right|^2\right],\quad \mathbf{C}=\frac{\sigma_\alpha^2}{2}\overrightarrow{g}\,\overrightarrow{g}^H+\mathbf{C}_{\overrightarrow{n}}$$

B. Derivation of the MLE of complex monopulse ratio for a Swerling 3-4 target

We start from equations (14a-b). From (13) where $\mathbf{C}_{\overrightarrow{n}}=\mathbf{Id}_2$, it comes:

$$\begin{aligned}\ln f\left(\overrightarrow{V}\right)&=Cst+\sum_{i=1}^{I}\left[\mu_i\left(\sigma_\beta^2,r,r^*\right)+\ln\left(1+\mu_i\left(\sigma_\beta^2,r,r^*\right)\right)-2\ln\left(1+\frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\overrightarrow{x}\right)\right]\\ \mu_i\left(\sigma_\beta^2,r,r^*\right)&=\frac{\sigma_\beta^2}{2}\frac{\left|\overrightarrow{v_i}^H\overrightarrow{x}\right|^2}{1+\frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\overrightarrow{x}}\end{aligned}$$

Therefore:

$$\frac{\partial\ln f\left(\overrightarrow{V}\right)}{\partial\sigma_\beta^2}=\sum_{i=1}^{I}\left[\frac{\partial\mu_i\left(\sigma_\beta^2,r,r^*\right)}{\partial\sigma_\beta^2}\left(1+\frac{1}{1+\mu_i\left(\sigma_\beta^2,r,r^*\right)}\right)-\frac{\overrightarrow{x}^H\overrightarrow{x}}{1+\frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\overrightarrow{x}}\right]$$

where:

$$\frac{\partial\mu_i\left(\sigma_\beta^2,r,r^*\right)}{\partial\sigma_\beta^2}=\frac{\left|\overrightarrow{v_i}^H\overrightarrow{x}\right|^2}{2\left(1+\frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\overrightarrow{x}\right)^2}=\frac{\mu_i\left(\sigma_\beta^2,r,r^*\right)}{\sigma_\beta^2\left(1+\frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\overrightarrow{x}\right)}$$

which leads to:

$$\frac{\partial\ln f\left(\overrightarrow{V}\right)}{\partial\sigma_\beta^2}=\frac{\sum_{i=1}^{I}\left[\mu_i\left(\sigma_\beta^2,r,r^*\right)\left(1+\frac{1}{1+\mu_i\left(\sigma_\beta^2,r,r^*\right)}\right)-\sigma_\beta^2\overrightarrow{x}^H\overrightarrow{x}\right]}{\sigma_\beta^2\left(1+\frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\overrightarrow{x}\right)}$$

Finally:

$$\frac{\partial\ln f\left(\overrightarrow{V}\right)}{\partial\sigma_\beta^2}=0\Leftrightarrow\sum_{i=1}^{I}\mu_i\left(\sigma_\beta^2,r,r^*\right)\left(1+\frac{1}{1+\mu_i\left(\sigma_\beta^2,r,r^*\right)}\right)=2I\frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\overrightarrow{x}\tag{23}$$

In the same way, using complex derivation [17]:

$$\frac{\partial\ln f\left(\overrightarrow{V}\right)}{\partial r^*}=\sum_{i=1}^{I}\left[\frac{\partial\mu_i\left(\sigma_\beta^2,r,r^*\right)}{\partial r^*}\left(1+\frac{1}{1+\mu_i\left(\sigma_\beta^2,r,r^*\right)}\right)-\frac{\sigma_\beta^2 r}{1+\frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\overrightarrow{x}}\right]$$

where:

$$\begin{aligned}\frac{\partial\mu_i\left(\sigma_\beta^2,r,r^*\right)}{\partial r^*}&=\frac{\sigma_\beta^2}{2}\left[\frac{\frac{\partial\left(\left|\overrightarrow{v_i}^H\overrightarrow{x}\right|^2\right)}{\partial r^*}}{1+\frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\overrightarrow{x}}-\frac{\left|\overrightarrow{v_i}^H\overrightarrow{x}\right|^2\left(\frac{\sigma_\beta^2}{2}r\right)}{\left(1+\frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\overrightarrow{x}\right)^2}\right]\\&=\frac{\sigma_\beta^2}{2\left(1+\frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\overrightarrow{x}\right)}\left[\frac{\partial\left(\left|\overrightarrow{v_i}^H\overrightarrow{x}\right|^2\right)}{\partial r^*}-r\mu_i\left(\sigma_\beta^2,r,r^*\right)\right]\end{aligned}$$

which leads to:

$$\frac{\partial \ln f\left(\overrightarrow{V}\right)}{\partial r^*}=\left(\frac{\sigma_\beta^2}{2}\right)\frac{\sum_{i=1}^{I}\left(1+\frac{1}{1+\mu_i\left(\sigma_\beta^2,r,r^*\right)}\right)\frac{\partial\left(\left|\overrightarrow{v_i}^H\overrightarrow{x}\right|^2\right)}{\partial r^*}-r\left[\sum_{i=1}^{I}\left(1+\frac{1}{1+\mu_i\left(\sigma_\beta^2,r,r^*\right)}\right)\mu_i\left(\sigma_\beta^2,r,r^*\right)\right]-2Ir}{\left(1+\frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\overrightarrow{x}\right)}$$

or equivalently - taking into account (23) -:

$$\frac{\partial \ln f\left(\overrightarrow{V}\right)}{\partial r^*}=\left(\frac{\sigma_\beta^2}{2}\right)\frac{\sum_{i=1}^{I}\left(1+\frac{1}{1+\mu_i\left(\sigma_\beta^2,r,r^*\right)}\right)\frac{\partial\left(\left|\overrightarrow{v_i}^H\overrightarrow{x}\right|^2\right)}{\partial r^*}-r2I\left(1+\frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\overrightarrow{x}\right)}{\left(1+\frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\overrightarrow{x}\right)}$$

Finally, as (23) is also equivalent to:

$$\sum_{i=1}^{I}\left(1+\frac{1}{1+\mu_i\left(\sigma_\beta^2,r,r^*\right)}\right)\frac{\left|\overrightarrow{v_i}^H\overrightarrow{x}\right|^2}{\overrightarrow{x}^H\overrightarrow{x}}=2I\left(1+\frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\overrightarrow{x}\right)$$

it comes:

$$\frac{\partial \ln f\left(\overrightarrow{V}\right)}{\partial r^*}=\frac{\frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\overrightarrow{x}}{1+\frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\overrightarrow{x}}\left[\sum_{i=1}^{I}\left(1+\frac{1}{1+\mu_i\left(\sigma_\beta^2,r,r^*\right)}\right)\left(\frac{\frac{\partial\left(\left|\overrightarrow{v_i}^H\overrightarrow{x}\right|^2\right)}{\partial r^*}}{\overrightarrow{x}^H\overrightarrow{x}}-\frac{r\left|\overrightarrow{v_i}^H\overrightarrow{x}\right|^2}{\left(\overrightarrow{x}^H\overrightarrow{x}\right)^2}\right)\right]$$

Thus, equations (14a-b) leads to the necessary MLE condition :

$$\frac{\partial \ln f\left(\overrightarrow{V}\right)}{\partial \sigma_\beta^2}=0 \text{ and } \frac{\partial \ln f\left(\overrightarrow{V}\right)}{\partial r^*}=0\Rightarrow\sum_{i=1}^{I}\left(1+\frac{1}{1+\mu_i\left(\sigma_\beta^2,r,r^*\right)}\right)\frac{\partial\left(\frac{\left|\overrightarrow{v_i}^H\overrightarrow{x}\right|^2}{\overrightarrow{x}^H\overrightarrow{x}}\right)}{\partial r^*}=0 \tag{24}$$

Moreover, let us consider the following probability:

$$P\left(\left|\nu_i\right|^2\geq\varepsilon\right),\quad \nu_i=\frac{\overrightarrow{x}^H}{\left\|\overrightarrow{x}\right\|}\overrightarrow{v_i}=\beta_i\left\|\overrightarrow{x}\right\|+\overrightarrow{n_i}^H\frac{\overrightarrow{x}}{\left\|\overrightarrow{x}\right\|},\quad \varepsilon\geq 0. \tag{25}$$

Then $P\left(\left|\nu_i\right|^2\geq\varepsilon\right)$ is simply the probability of detection (16) where $\overrightarrow{\Sigma}=(\nu_i)\,,T=\varepsilon,I=1$; and according to (22):

$$P\left(\left|\nu_i\right|^2\geq\varepsilon\right)=\frac{e^{-\frac{\varepsilon}{\mathbf{C}_{11}}}}{\mathbf{C}_{11}}A_0,\quad A_0=1+b\mathbf{C}_{11}e_1\left(\frac{\varepsilon}{\mathbf{C}_{11}}\right),\quad \mathbf{C}_{11}=\frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\overrightarrow{x}+1,\quad b=\frac{\frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\overrightarrow{x}}{\mathbf{C}_{11}},$$

i.e.:

$$P\left(\left|\nu_i\right|^2\geq\varepsilon\right)=e^{-\varepsilon'}\left(\frac{1+\frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\overrightarrow{x}e_1\left(\varepsilon'\right)}{\frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\overrightarrow{x}+1}\right),\quad \varepsilon'=\frac{\varepsilon}{\frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\overrightarrow{x}+1}.$$

Therefore:

$$\forall\varepsilon\geq 0,\quad \lim_{\sigma_\beta^2\to\infty}P\left(\left|\nu_i\right|^2\geq\varepsilon\right)=1$$

and consequently - as $\mu_i\left(\sigma_\beta^2,r,r^*\right)=\left(1-\frac{1}{1+\frac{\sigma_\beta^2}{2}\overrightarrow{x}^H\overrightarrow{x}}\right)\left|\nu_i\right|^2$ -:

$$\forall\lambda\in\left]0,1\right],\quad \lim_{\sigma_\beta^2\to\infty}P\left(\frac{1}{1+\mu_i\left(\sigma_\beta^2,r,r^*\right)}\leq\lambda\right)=1,$$

that is to say:

$$\lim_{\sigma_\beta^2 \to \infty} P\left(\frac{1}{1+\mu_i\left(\sigma_\beta^2, r, r^*\right)} = 0\right) = 1,$$

which proves that asymptotically at high SNR (24) is ”almost surely” equivalent to:

$$\sum_{i=1}^{I} \frac{\partial}{\partial r^*}\left[\frac{\overrightarrow{x}^H\left(\overrightarrow{v_i}\,\overrightarrow{v_i}^H\right)\overrightarrow{x}}{\overrightarrow{x}^H\overrightarrow{x}}\right] = 0 = \frac{\partial}{\partial r^*}\left[\frac{\overrightarrow{x}^H\left(\sum\limits_{i=1}^{I}\overrightarrow{v_i}\,\overrightarrow{v_i}^H\right)\overrightarrow{x}}{\overrightarrow{x}^H\overrightarrow{x}}\right] = \text{arg max}\left\{\frac{\overrightarrow{x}^H\widehat{\mathbf{R}_{\overrightarrow{v}}}\overrightarrow{x}}{\overrightarrow{x}^H\overrightarrow{x}}\right\}$$

C. Computation of conditional mean $E\left[s\left(\overrightarrow{U}\right) \mid D\right]$ *and conditional moment of order 2* $E\left[s\left(\overrightarrow{U}\right)^2 \mid D\right]$

1) Change of variable, calculation of the Jacobian and expression of the new p.d.f.: The form (13) of $f\left(\overrightarrow{V}\right)$ does not allow application of the calculation scheme introduced for Rayleigh targets [6] relying on the properties of Gaussian random variables. Nevertheless, this calculation scheme has inspired us an alternative method consisting in breaking $f\left(\overrightarrow{V}\right)$ down into the form of a product of polynomials and Gaussian p.d.f.s in order to treat the problem as a calculation of moments of Gaussian random variables. To do this, just consider the change of variable:

$$\left(U_i = \frac{\Sigma_i^*\Delta_i}{\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}},\ W_i = \Sigma_i\right)_{i\in[1,I]}, \quad \widehat{r} = \sum_{i=1}^{I} U_i \tag{26}$$

The Jacobian of the transformation is then written:

$$J_{\overrightarrow{\Sigma},\overrightarrow{\Delta}}\left(\overrightarrow{W},\overrightarrow{U}\right) = \left|\frac{\partial\left(\overrightarrow{\Sigma_r}^T,\overrightarrow{\Sigma_j}^T,\overrightarrow{\Delta_r}^T,\overrightarrow{\Delta_j}^T\right)}{\partial\left(\overrightarrow{W_r}^T,\overrightarrow{W_j}^T,\overrightarrow{U_r}^T,\overrightarrow{U_j}^T\right)^T}\right|, \quad \left\{\begin{array}{ll}\overrightarrow{\Sigma} = \overrightarrow{\Sigma_r} + j\overrightarrow{\Sigma_j}, & \overrightarrow{\Delta} = \overrightarrow{\Delta_r} + j\overrightarrow{\Delta_j}\\ \overrightarrow{U} = \overrightarrow{U_r} + j\overrightarrow{U_j}, & \overrightarrow{W} = \overrightarrow{W_r} + j\overrightarrow{W_j}\end{array}\right.$$

with:

$$\left(\Delta_i = \left(\overrightarrow{W}^H\overrightarrow{W}\right)\frac{U_iW_i}{|W_i|^2},\ \Sigma_i = W_i\right)_{i\in[1,I]}$$

Thus:

$$\begin{aligned} J_{\overrightarrow{\Sigma},\overrightarrow{\Delta}}\left(\overrightarrow{W},\overrightarrow{U}\right) &= \left|\begin{array}{ccc}\mathbf{Id}_I & \mathbf{0} & \frac{\partial\left(\overrightarrow{\Delta_r}^T,\overrightarrow{\Delta_j}^T\right)}{\partial\left(\overrightarrow{W_r}^T,\overrightarrow{W_j}^T\right)^T}\\ \mathbf{0} & \mathbf{Id}_I & \\ \mathbf{0} & \mathbf{0} & \frac{\partial\left(\overrightarrow{\Delta_r}^T,\overrightarrow{\Delta_j}^T\right)}{\partial\left(\overrightarrow{U_r}^T,\overrightarrow{U_j}^T\right)^T}\\ \mathbf{0} & \mathbf{0} & \end{array}\right| = \left|\begin{array}{cc}\frac{\partial\overrightarrow{\Delta_r}^T}{\partial\overrightarrow{U_r}} & \frac{\partial\overrightarrow{\Delta_j}^T}{\partial\overrightarrow{U_r}}\\ \frac{\partial\overrightarrow{\Delta_r}^T}{\partial\overrightarrow{U_j}} & \frac{\partial\overrightarrow{\Delta_j}^T}{\partial\overrightarrow{U_j}}\end{array}\right| \\ &= \left|\frac{\partial\overrightarrow{\Delta_r}^T}{\partial\overrightarrow{U_r}}\right|\left|\frac{\partial\overrightarrow{\Delta_j}^T}{\partial\overrightarrow{U_j}} - \frac{\partial\overrightarrow{\Delta_r}^T}{\partial\overrightarrow{U_j}}\left(\frac{\partial\overrightarrow{\Delta_r}^T}{\partial\overrightarrow{U_r}}\right)^{-1}\frac{\partial\overrightarrow{\Delta_j}^T}{\partial\overrightarrow{U_r}}\right| \text{ [18, p. 21]}\end{aligned}$$

with:

$$\frac{\partial\overrightarrow{\Delta_r}^T}{\partial\overrightarrow{U_r}} = \overrightarrow{W}^H\overrightarrow{W}\left[\begin{array}{ccc}\frac{\mathrm{Re}\{W_1\}}{|W_1|^2} & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \frac{\mathrm{Re}\{W_I\}}{|W_I|^2}\end{array}\right], \quad \frac{\partial\overrightarrow{\Delta_r}^T}{\partial\overrightarrow{U_j}} = \overrightarrow{W}^H\overrightarrow{W}\left[\begin{array}{ccc}\frac{-\mathrm{Im}\{W_1\}}{|W_1|^2} & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \frac{-\mathrm{Im}\{W_I\}}{|W_1|^2}\end{array}\right]$$

$$\frac{\partial\overrightarrow{\Delta_j}^T}{\partial\overrightarrow{U_r}} = \overrightarrow{W}^H\overrightarrow{W}\left[\begin{array}{ccc}\frac{\mathrm{Im}\{W_1\}}{|W_1|^2} & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \frac{\mathrm{Im}\{W_I\}}{|W_I|^2}\end{array}\right], \quad \frac{\partial\overrightarrow{\Delta_j}^T}{\partial\overrightarrow{U_j}} = \overrightarrow{W}^H\overrightarrow{W}\left[\begin{array}{ccc}\frac{\mathrm{Re}\{W_1\}}{|W_1|^2} & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \frac{\mathrm{Re}\{W_I\}}{|W_1|^2}\end{array}\right]$$

Finally:

$$J_{\overrightarrow{\Sigma},\overrightarrow{\Delta}}\left(\overrightarrow{W},\overrightarrow{U}\right)=\left(\overrightarrow{W}^H\overrightarrow{W}\right)^{2I}\left(\prod_{i=1}^{I}\frac{\mathrm{Re}\left\{W_i\right\}}{\left|W_i\right|^2}\right)\begin{vmatrix}\frac{1}{\mathrm{Re}\{W_1\}} & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \frac{1}{\mathrm{Re}\{W_I\}}\end{vmatrix}=\prod_{i=1}^{I}\frac{\left(\overrightarrow{W}^H\overrightarrow{W}\right)^2}{\left|W_i\right|^2}$$

leading to:

$$\begin{aligned}f\left(\overrightarrow{W},\overrightarrow{U}\right) &= f_{\overrightarrow{\Sigma},\overrightarrow{\Delta}}\left(\overrightarrow{\Sigma}\left(\overrightarrow{W},\overrightarrow{U}\right),\overrightarrow{\Delta}\left(\overrightarrow{W},\overrightarrow{U}\right)\right)\left|J_{\overrightarrow{\Sigma},\overrightarrow{\Delta}}\left(\overrightarrow{W},\overrightarrow{U}\right)\right|\\ &= f_{\overrightarrow{\Sigma},\overrightarrow{\Delta}}\left(\overrightarrow{W},\overrightarrow{W}^H\overrightarrow{W}\left(\frac{U_1}{W_1^H},..,\frac{U_I}{W_I^H}\right)^T\right)\left(\prod_{i=1}^{I}\frac{\left(\overrightarrow{W}^H\overrightarrow{W}\right)^2}{\left|W_i\right|^2}\right)\end{aligned}$$

which can be broken down into the form - as $\overrightarrow{W}=\overrightarrow{\Sigma}$ -:

$$f\left(\overrightarrow{\Sigma},\overrightarrow{U}\right)=\mathcal{N}\left(\overrightarrow{\Sigma},\overrightarrow{0},\mathbf{C}_{\overrightarrow{\Sigma}}^{eq}\right)\ \left[\mathcal{N}\left(\overrightarrow{U},\overrightarrow{m}_{\overrightarrow{U}}^{eq},\mathbf{C}_{\overrightarrow{U}}^{eq}\right)\ B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right] \tag{27}$$

where:

$$\mathcal{N}\left(\overrightarrow{X},\overrightarrow{m}_{\overrightarrow{X}},\mathbf{C}_{\overrightarrow{X}}\right)=\frac{e^{-\left(\overrightarrow{X}-\overrightarrow{m}_{\overrightarrow{X}}\right)^H\mathbf{C}_{\overrightarrow{X}}^{-1}\left(\overrightarrow{X}-\overrightarrow{m}_{\overrightarrow{X}}\right)}}{\pi^I\left|\mathbf{C}_{\overrightarrow{X}}\right|}$$

$$\mathbf{C}_{\overrightarrow{\Sigma}}^{eq}=\mathbf{C}_{11}\mathbf{Id}_I$$

$$\mathbf{C}_{\overrightarrow{U}}^{eq}=\frac{|\mathbf{C}|}{\mathbf{C}_{11}}\frac{1}{\left(\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}\right)^2}\begin{bmatrix}\left|\Sigma_1\right|^2 & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \left|\Sigma_I\right|^2\end{bmatrix},\quad \overrightarrow{m}_{\overrightarrow{U}}^{eq}=\frac{\mathbf{C}_{12}^H}{\mathbf{C}_{11}}\frac{1}{\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}}\begin{pmatrix}\left|\Sigma_1\right|^2\\ \vdots\\ \left|\Sigma_I\right|^2\end{pmatrix}$$

$$B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)=\prod_{i=1}^{I}b_{\overrightarrow{\Sigma}}^i\left(U_i\right),\quad b_{\overrightarrow{\Sigma}}^i\left(U_i\right)=h+\frac{\sigma_\alpha^2}{2}\left|k_2\right|^2\frac{|\mathbf{C}|}{\mathbf{C}_{11}}\frac{1}{\left(\mathbf{C}_{\overrightarrow{U}}^{eq}\right)_{ii}}\left|\left(U_i-m_{U_i}^{eq}\right)+n\frac{\left|\Sigma_i\right|^2}{\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}}\right|^2$$

$$\overrightarrow{g}=\begin{pmatrix}g_\Sigma\\ g_\Delta\end{pmatrix},\quad \overrightarrow{k}=\begin{pmatrix}k_1\\ k_2\end{pmatrix}=\mathbf{C}^{-1}\overrightarrow{g},\quad n=\left(\frac{k_1}{k_2}+\frac{\mathbf{C}_{12}}{\mathbf{C}_{11}}\right)^H,\quad h=\frac{1}{1+\frac{\sigma_\alpha^2}{2}\overrightarrow{g}^H\mathbf{C}_{\overrightarrow{n}}^{-1}\overrightarrow{g}}$$

The guiding idea is to use the change of variable defined above to convert the conditional expectation calculations into calculations of expectations of polynomials of Gaussian circular complex random variables. Indeed, for any function $s\left(\overrightarrow{U}\right)$, according to (27) :

$$E\left[s\left(\overrightarrow{U}\right)\mid D\right]=\iint_{\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}\geq T}s\left(\overrightarrow{U}\right)\frac{f\left(\overrightarrow{\Sigma},\overrightarrow{U}\right)}{P_D}d\overrightarrow{\Sigma}d\overrightarrow{U}\ =\int_{\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}\geq T}E\left[s\left(\overrightarrow{U}\right)B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right]\frac{\mathcal{N}\left(\overrightarrow{\Sigma},\overrightarrow{0},\mathbf{C}_{\overrightarrow{\Sigma}}^{eq}\right)}{P_D}d\overrightarrow{\Sigma} \tag{28}$$

where:

$$E\left[s\left(\overrightarrow{U}\right)B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right]=\int\left[s\left(\overrightarrow{U}\right)B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right]\mathcal{N}\left(\overrightarrow{U},\overrightarrow{m}_{\overrightarrow{U}}^{eq},\mathbf{C}_{\overrightarrow{U}}^{eq}\right)d\overrightarrow{U} \tag{29}$$

Therefore, if we consider the function $s\left(\overrightarrow{U}\right)$ defined by:

$$s\left(\overrightarrow{U}\right)=\sum_{i=1}^{I}\mathrm{Re}\left\{U_i\right\}-\mathrm{Re}\left\{\frac{\mathbf{C}_{12}^H}{\mathbf{C}_{11}}\right\}=\sum_{i=1}^{I}\mathrm{Re}\left\{U_i-m_{U_i}^{eq}\right\}=\mathrm{Re}\left\{\widehat{r}\right\}-\mathrm{Re}\left\{\frac{\mathbf{C}_{12}^H}{\mathbf{C}_{11}}\right\} \tag{30}$$

then:

$$E\left[s\left(\overrightarrow{U}\right)\mid D\right]=E\left[\mathrm{Re}\left\{\widehat{r}_{ML}\right\}\mid D\right]-\mathrm{Re}\left\{\frac{\mathbf{C}_{12}^H}{\mathbf{C}_{11}}\right\}$$

and:

$$E\left[s\left(\overrightarrow{U}\right)^2 \mid D\right] = Var\left[s\left(\overrightarrow{U}\right) \mid D\right] + E\left[s\left(\overrightarrow{U}\right) \mid D\right]^2 = Var\left[\mathrm{Re}\left\{\widehat{r}_{ML}\right\} \mid D\right] + E\left[s\left(\overrightarrow{U}\right) \mid D\right]^2$$

which allows to compute both $E\left[\mathrm{Re}\left\{\widehat{r}_{ML}\right\} \mid D\right]$ and $Var\left[\mathrm{Re}\left\{\widehat{r}_{ML}\right\} \mid D\right]$.

2) Conditional mean, first intermediate formulation: According to (28), to get $E\left[s\left(\overrightarrow{U}\right) \mid D\right]$, we must compute $E\left[s\left(\overrightarrow{U}\right) B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right]$ first. This computation can be down using the following breakdown:

$$\begin{aligned} E\left[s\left(\overrightarrow{U}\right) B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right] &= \sum_{i=1}^{I} E\left[\mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\} b_{\overrightarrow{\Sigma}}^{i}\left(U_i\right)\right] \prod_{j\neq i} E\left[b_{\overrightarrow{\Sigma}}^{j}\left(U_j\right)\right] \\ &= \sum_{i=1}^{I} E\left[\mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\} b_{\overrightarrow{\Sigma}}^{i}\left(U_i\right)\right] \frac{E\left[B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right]}{E\left[b_{\overrightarrow{\Sigma}}^{i}\left(U_i\right)\right]} \end{aligned}$$

since, for a given $\overrightarrow{\Sigma}$ value, the set of random variables $\left(U_i - m_{U_i}^{eq}\right)_{i\in[i,I]}$ are circular, independent, complex Gaussian variables [19] with covariance matrix $\mathbf{C}_{\overrightarrow{U}}^{eq}$. Therefore, as

$$b_{\overrightarrow{\Sigma}}^{i}\left(U_i\right) = \left(h + \frac{\sigma_\alpha^2}{2}\left|nk_2\right|^2\left|\Sigma_i\right|^2\right) + \frac{\sigma_\alpha^2\left|k_2\right|^2\left|\mathbf{C}\right|\left|U_i - m_{U_i}^{eq}\right|^2}{2\mathbf{C}_{11}\left(\mathbf{C}_{\overrightarrow{U}}^{eq}\right)_{ii}} + \overrightarrow{\Sigma}^H\overrightarrow{\Sigma}\,\mathrm{Re}\left\{\left(U_i - m_{U_i}^{eq}\right)n^H\right\}\sigma_\alpha^2\left|k_2\right|^2$$

then:

$$E\left[b_{\overrightarrow{\Sigma}}^{i}\left(U_i\right)\right] = \left(h + \frac{\sigma_\alpha^2}{2}\left|nk_2\right|^2\left|\Sigma_i\right|^2\right) + \frac{\sigma_\alpha^2\left|k_2\right|^2\left|\mathbf{C}\right| E\left[\left|U_i - m_{U_i}^{eq}\right|^2\right]}{2\mathbf{C}_{11}\left(\mathbf{C}_{\overrightarrow{U}}^{eq}\right)_{ii}} + \overrightarrow{\Sigma}^H\overrightarrow{\Sigma}\,\mathrm{Re}\left\{E\left[U_i - m_{U_i}^{eq}\right]n^H\right\}\sigma_\alpha^2\left|k_2\right|^2$$

and:

$$\begin{aligned} E\left[\left(U_i - m_{U_i}^{eq}\right) b_{\overrightarrow{\Sigma}}^{i}\left(U_i\right)\right] = &\left(h + \frac{\sigma_\alpha^2}{2}\left|nk_2\right|^2\left|\Sigma_i\right|^2\right) E\left[U_i - m_{U_i}^{eq}\right] + \frac{\sigma_\alpha^2\left|k_2\right|^2\left|\mathbf{C}\right|}{2\mathbf{C}_{11}\left(\mathbf{C}_{\overrightarrow{U}}^{eq}\right)_{ii}} E\left[\left(U_i - m_{U_i}^{eq}\right)^2\left(U_i - m_{U_i}^{eq}\right)^H\right] \\ &+ \overrightarrow{\Sigma}^H\overrightarrow{\Sigma}\frac{\sigma_\alpha^2\left|k_2\right|^2}{2} E\left[\left(U_i - m_{U_i}^{eq}\right)^2\right] n^H + \overrightarrow{\Sigma}^H\overrightarrow{\Sigma}\frac{\sigma_\alpha^2\left|k_2\right|^2}{2} E\left[\left|U_i - m_{U_i}^{eq}\right|^2\right] n \end{aligned}$$

where [19]: $E\left[U_i - m_{U_i}^{eq}\right] = 0, \quad E\left[\left(U_i - m_{U_i}^{eq}\right)^2\right] = 0, \quad E\left[\left(U_i - m_{U_i}^{eq}\right)^2\left(U_i - m_{U_i}^{eq}\right)^H\right] = 0$.

Thus:

$$\begin{aligned} E\left[b_{\overrightarrow{\Sigma}}^{i}\left(U_i\right)\right] &= \left(h + \frac{\sigma_\alpha^2}{2}\left|nk_2\right|^2\left|\Sigma_i\right|^2\right) + \frac{\sigma_\alpha^2\left|k_2\right|^2\left|\mathbf{C}\right|}{2\mathbf{C}_{11}}\frac{\left(\mathbf{C}_{\overrightarrow{U}}^{eq}\right)_{ii}}{\left(\mathbf{C}_{\overrightarrow{U}}^{eq}\right)_{ii}} \\ E\left[\left(U_i - m_{U_i}^{eq}\right) b_{\overrightarrow{\Sigma}}^{i}\left(U_i\right)\right] &= \overrightarrow{\Sigma}^H\overrightarrow{\Sigma}\frac{\sigma_\alpha^2\left|k_2\right|^2}{2} n\left(\mathbf{C}_{\overrightarrow{U}}^{eq}\right)_{ii} = \frac{\sigma_\alpha^2\left|k_2\right|^2\left|\mathbf{C}\right| n}{2\mathbf{C}_{11}}\frac{\left|\Sigma_i\right|^2}{\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}} \end{aligned}$$

Finally, as $h + \frac{\sigma_\alpha^2}{2}\left|k_2\right|^2\frac{\left|\mathbf{C}\right|}{\mathbf{C}_{11}} = \frac{\mathbf{1}}{\mathbf{C}_{11}}, \quad \left|nk_2\right|^2 = \left(\frac{\left|g_\Sigma\right|}{\mathbf{C}_{11}}\right)^2$:

$$E\left[b_{\overrightarrow{\Sigma}}^{i}\left(U_i\right)\right] = \frac{1}{\mathbf{C}_{11}} + \frac{\sigma_\alpha^2}{2}\frac{\left|g_\Sigma\right|^2}{\mathbf{C}_{11}}\frac{\left|\Sigma_i\right|^2}{\mathbf{C}_{11}} \tag{31}$$

$$E\left[\mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\} b_{\overrightarrow{\Sigma}}^{i}\left(U_i\right)\right] = \frac{\sigma_\alpha^2\left|k_2\right|^2\left|\mathbf{C}\right|\mathrm{Re}\left\{n\right\}}{2\mathbf{C}_{11}}\frac{\left|\Sigma_i\right|^2}{\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}} \tag{32}$$

$$E\left[s\left(\overrightarrow{U}\right) B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right] = \left(\frac{\sigma_\alpha^2}{2}\left|k_2\right|^2\frac{\left|\mathbf{C}\right|}{\mathbf{C}_{11}}\mathrm{Re}\left\{n\right\}\right)\sum_{i=1}^{I}\frac{\left|\Sigma_i\right|^2}{\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}}\frac{E\left[B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right]}{E\left[b_{\overrightarrow{\Sigma}}^{i}\left(U_i\right)\right]} \tag{33}$$

3) Conditional moment of order 2, first intermediate formulation: According to (28), to get $E\left[s\left(\overrightarrow{U}\right)^2 \mid D\right]$, we must compute $E\left[s\left(\overrightarrow{U}\right)^2 B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right]$ first. From (30):

$$s\left(\overrightarrow{U}\right)^2 = \sum_{i=1}^{I} \mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\}^2 + 2\sum_{i<j} \mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\} \mathrm{Re}\left\{U_j - m_{U_j}^{eq}\right\}$$

then:

$$E\left[s\left(\overrightarrow{U}\right)^2 B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right] = \sum_{i=1}^{I} E\left[\mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\}^2 B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right] + 2\sum_{i<j} E\left[\mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\} \mathrm{Re}\left\{U_j - m_{U_j}^{eq}\right\} B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right]$$

where:

$$E\left[\mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\}^2 B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right] = E\left[\mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\}^2 b_{\overrightarrow{\Sigma}}^{i}\left(U_i\right)\right] \frac{E\left[B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right]}{E\left[b_{\overrightarrow{\Sigma}}^{i}\left(U_i\right)\right]} \tag{34}$$

$$E\left[\mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\} \mathrm{Re}\left\{U_j - m_{U_j}^{eq}\right\} B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right] =$$
$$E\left[\mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\} b_{\overrightarrow{\Sigma}}^{i}\left(U_i\right)\right] E\left[\mathrm{Re}\left\{U_j - m_{U_j}^{eq}\right\} b_{\overrightarrow{\Sigma}}^{j}\left(U_j\right)\right] \frac{E\left[B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right]}{E\left[b_{\overrightarrow{\Sigma}}^{i}\left(U_i\right)\right] E\left[b_{\overrightarrow{\Sigma}}^{j}\left(U_j\right)\right]} \tag{35}$$

Most factors requested to compute (34) and (35) have already been computed in previous section (Appendix VII-C2) except:

$$\begin{aligned} E\left[\mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\}^2 b_{\overrightarrow{\Sigma}}^{i}\left(U_i\right)\right] = & \left(h + \frac{\sigma_\alpha^2}{2}\left|nk_2\right|^2\left|\Sigma_i\right|^2\right) E\left[\mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\}^2\right] \\ & + \frac{\sigma_\alpha^2\left|k_2\right|^2\left|\mathbf{C}\right|}{2\mathbf{C}_{11}\left(\mathbf{C}_{\overrightarrow{U}}^{eq}\right)_{ii}} E\left[\mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\}^4\right] \\ & + \frac{\sigma_\alpha^2\left|k_2\right|^2\left|\mathbf{C}\right|}{2\mathbf{C}_{11}\left(\mathbf{C}_{\overrightarrow{U}}^{eq}\right)_{ii}} E\left[\mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\}^2 \mathrm{Im}\left\{U_i - m_{U_i}^{eq}\right\}^2\right] \\ & + \overrightarrow{\Sigma}^H\overrightarrow{\Sigma}\frac{\sigma_\alpha^2\left|k_2\right|^2}{2} 2\,\mathrm{Re}\left\{E\left[\mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\}^2\left(U_i - m_{U_i}^{eq}\right)\right] n^H\right\} \end{aligned}$$

where [19]:

$E\left[\mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\}^2\right] = E\left[\mathrm{Im}\left\{U_i - m_{U_i}^{eq}\right\}^2\right] = \frac{1}{2}\left(\mathbf{C}_{\overrightarrow{U}}^{eq}\right)_{ii}$

$E\left[\mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\}^2 \mathrm{Im}\left\{U_i - m_{U_i}^{eq}\right\}^2\right] = E\left[\mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\}^2\right] E\left[\mathrm{Im}\left\{U_i - m_{U_i}^{eq}\right\}^2\right] = \frac{1}{4}\left(\mathbf{C}_{\overrightarrow{U}}^{eq}\right)_{ii}^2$

$E\left[\mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\}^2\left(U_i - m_{U_i}^{eq}\right)\right] = 0$

$E\left[\mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\}^4\right] = 3E\left[\mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\}^2\right]^2 = \frac{3}{4}\left(\mathbf{C}_{\overrightarrow{U}}^{eq}\right)_{ii}^2$.

Therefore (34) and (35) become:

$$E\left[\mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\}^2 B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right] = \frac{\left(\mathbf{C}_{\overrightarrow{U}}^{eq}\right)_{ii}}{2} E\left[B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right]\left\{1 + \frac{\sigma_\alpha^2}{2}\left|k_2\right|^2\frac{\left|\mathbf{C}\right|}{\mathbf{C}_{11}}\frac{1}{E\left[b_{\overrightarrow{\Sigma}}^{i}\left(U_i\right)\right]}\right\}$$

$$E\left[\mathrm{Re}\left\{U_i - m_{U_i}^{eq}\right\} \mathrm{Re}\left\{U_j - m_{U_j}^{eq}\right\} B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right] = \frac{\left(\frac{\sigma_\alpha^2}{2}\left|k_2\right|^2\frac{\left|\mathbf{C}\right|}{\mathbf{C}_{11}}\mathrm{Re}\left\{n\right\}\right)^2\left|\Sigma_i\right|^2\left|\Sigma_j\right|^2 E\left[B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right]}{\left(\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}\right)^2 E\left[b_{\overrightarrow{\Sigma}}^{i}\left(U_i\right)\right] E\left[b_{\overrightarrow{\Sigma}}^{j}\left(U_j\right)\right]}$$

and finally:

$$E\left[s\left(\overrightarrow{U}\right)^2 B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right] = \frac{|\mathbf{C}|}{2\mathbf{C}_{11}}\frac{E\left[B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right]}{\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}} + \frac{\sigma_\alpha^2|k_2|^2|\mathbf{C}|^2}{4\left(\mathbf{C}_{11}\right)^2}\sum_{i=1}^{I}\frac{E\left[B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right]}{E\left[b^i_{\overrightarrow{\Sigma}}\left(U_i\right)\right]}\frac{|\Sigma_i|^2}{\left(\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}\right)^2} + 2\left(\frac{\sigma_\alpha^2}{2}|k_2|^2\frac{|\mathbf{C}|}{\mathbf{C}_{11}}\mathrm{Re}\left\{n\right\}\right)^2\sum_{i<j}\frac{|\Sigma_i|^2|\Sigma_j|^2 E\left[B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right]}{\left(\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}\right)^2 E\left[b^i_{\overrightarrow{\Sigma}}\left(U_i\right)\right]E\left[b^j_{\overrightarrow{\Sigma}}\left(U_j\right)\right]} \quad (36)$$

4) Conditional mean and moment of order 2, 2nd intermediate formulations: Finally, by introducing (33) in (28), it comes:

$$E\left[s\left(\overrightarrow{U}\right)\mid D\right] = \left(\frac{\sigma_\alpha^2}{2}|k_2|^2\frac{|\mathbf{C}|}{\mathbf{C}_{11}}\mathrm{Re}\left\{n\right\}\right)\sum_{i=1}^{I}\int_{\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}\geq T}\frac{|\Sigma_i|^2}{\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}}\frac{E\left[B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right]}{E\left[b^i_{\overrightarrow{\Sigma}}\left(U_i\right)\right]}\frac{\mathcal{N}\left(\overrightarrow{\Sigma},\overrightarrow{0},\mathbf{C}^{eq}_{\overrightarrow{\Sigma}}\right)}{P_D}d\overrightarrow{\Sigma} \quad (37)$$

It is now particularly clever to notice that:

$$f\left(\overrightarrow{\Sigma}\right) = \mathcal{N}\left(\overrightarrow{\Sigma},\overrightarrow{0},\mathbf{C}^{eq}_{\overrightarrow{\Sigma}}\right)\int B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\mathcal{N}\left(\overrightarrow{U},\overrightarrow{m}^{eq}_{\overrightarrow{U}},\mathbf{C}^{eq}_{\overrightarrow{U}}\right)d\overrightarrow{U} = E\left[B_{\overrightarrow{\Sigma}}\left(\overrightarrow{U}\right)\right]\mathcal{N}\left(\overrightarrow{\Sigma},\overrightarrow{0},\mathbf{C}^{eq}_{\overrightarrow{\Sigma}}\right) \quad (38)$$

since it allows to express (37) as:

$$E\left[s\left(\overrightarrow{U}\right)\mid D\right] = \left(\frac{\sigma_\alpha^2}{2}|k_2|^2\frac{|\mathbf{C}|}{\mathbf{C}_{11}}\mathrm{Re}\left\{n\right\}\right)\sum_{i=1}^{I}E\left[\frac{|\Sigma_i|^2}{\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}}\frac{1}{E\left[b^i_{\overrightarrow{\Sigma}}\left(U_i\right)\right]}\mid D\right] \quad (39)$$

In the same way, introducing (36) in (28) and taken into account (38) leads to:

$$E\left[s\left(\overrightarrow{U}\right)^2\mid D\right] = \frac{|\mathbf{C}|}{2\mathbf{C}_{11}}E\left[\frac{1}{\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}}\mid D\right] + \frac{\sigma_\alpha^2|k_2|^2}{4}\frac{|\mathbf{C}|^2}{\left(\mathbf{C}_{11}\right)^2}\sum_{i=1}^{I}E\left[\frac{|\Sigma_i|^2}{\left(\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}\right)^2E\left[b^i_{\overrightarrow{\Sigma}}\left(U_i\right)\right]}\mid D\right] + 2\left(\frac{\sigma_\alpha^2}{2}|k_2|^2\frac{|\mathbf{C}|}{\mathbf{C}_{11}}\mathrm{Re}\left\{n\right\}\right)^2\sum_{i<j}E\left[\frac{1}{\left(\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}\right)^2}\frac{|\Sigma_i|^2|\Sigma_j|^2}{E\left[b^i_{\overrightarrow{\Sigma}}\left(U_i\right)\right]E\left[b^j_{\overrightarrow{\Sigma}}\left(U_j\right)\right]}\mid D\right] \quad (40)$$

To compute the remaining unknown conditional expectations in (39)(40), it is helpful to notice first that:

$$P_D = \int_T^\infty f\left(\varpi_I\right)d\varpi_I,\quad E\left[\frac{1}{\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}}\mid D\right] = \int_T^\infty\frac{f\left(\varpi_I\right)}{P_D\varpi_I}d\varpi_I,\quad E\left[\frac{1}{\left(\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}\right)^2}\mid D\right] = \int_T^\infty\frac{f\left(\varpi_I\right)}{P_D\varpi_I^2}d\varpi_I \quad (41)$$

if we consider the change of variable $\varpi_I = \overrightarrow{\Sigma}^H\overrightarrow{\Sigma} = \sum_{i=1}^{I}|\Sigma_i|^2$. Indeed, one can also establishes that:

$$E\left[\frac{1}{\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}}\mid D\right] = \frac{1}{P_D\mathbf{C}_{11}}\int_T^\infty\frac{f\left(x_I\right)}{x_I}dx_I + \frac{b}{\mathbf{C}_{11}}E\left[\frac{1}{\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}}\frac{|\Sigma_i|^2}{E\left[b^i_{\overrightarrow{\Sigma}}\left(U_i\right)\right]}\mid D\right] \quad (42)$$

$$E\left[\frac{1}{\left(\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}\right)^2}\mid D\right] = \frac{1}{P_D\mathbf{C}_{11}}\int_T^\infty\frac{f\left(x_I\right)}{x_I^2}dx_I + \frac{b}{\mathbf{C}_{11}}E\left[\frac{1}{\left(\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}\right)^2}\frac{|\Sigma_i|^2}{E\left[b^i_{\overrightarrow{\Sigma}}\left(U_i\right)\right]}\mid D\right] \quad (43)$$

$$E\left[\frac{1}{\left(\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}\right)^2}\mid D\right] = \frac{2}{P_D\mathbf{C}_{11}}\int_T^\infty \frac{f(x_I)}{x_I^2}dx_I - \frac{1}{P_D(\mathbf{C}_{11})^2}\int_T^\infty \frac{f(y_I)}{y_I^2}dy_I + \frac{b^2}{(\mathbf{C}_{11})^2}E\left[\frac{1}{\left(\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}\right)^2}\frac{|\Sigma_i|^2|\Sigma_j|^2}{E\left[b^i_{\overrightarrow{\Sigma}}(U_i)\right]E\left[b^j_{\overrightarrow{\Sigma}}(U_j)\right]}\mid D\right] \quad (44)$$

where $b=\frac{\sigma_\alpha^2}{2}\frac{|g_\Sigma|^2}{\mathbf{C}_{11}}$ and x_I, y_I are random variables defined by:

$$x_I=\begin{cases} x_1=u_1 \\ x_I=\varpi_{I-1}+u_I\ ,\ I\geq 2\end{cases},\quad y_I=\begin{cases} y_2=u_1+u_2 \\ y_I=\varpi_{I-2}+u_{I-1}+u_I,\ I\geq 3\end{cases},$$

$(u_i)_{i\in[1,I]}$ being Rayleigh independent with p.d.f. $f(u_i)=\frac{e^{-\frac{u_i}{\mathbf{C}_{11}}}}{\mathbf{C}_{11}}$.

Although not immediately straightforward, the above changes of unknowns (41)(42)(43)(44) allows to derive the remaining unknown conditional expectations since all integrals involving random variables ϖ_I, x_I and y_I are computable.

5) Conditional mean and moment of order 2, final analytical formulas: Indeed $f(\varpi_I)$, $f(x_I)$ and $f(y_I)$ are easily obtained using recursion. From (12):

$$f(\Sigma_i)=\frac{e^{-\frac{\Sigma_i^H\Sigma_i}{\mathbf{C}_{11}}}}{\pi\mathbf{C}_{11}}\left[h+\frac{\sigma_\alpha^2}{2}|k_2|^2\frac{|\mathbf{C}|}{\mathbf{C}_{11}}+\frac{\sigma_\alpha^2}{2}|nk_2|^2|\Sigma_i|^2\right]=\frac{e^{-\frac{\Sigma_i^H\Sigma_i}{\mathbf{C}_{11}}}}{\pi\mathbf{C}_{11}}\left[\frac{1}{\mathbf{C}_{11}}+\frac{\sigma_\alpha^2}{2}\left(\frac{|g_\Sigma|}{\mathbf{C}_{11}}\right)^2|\Sigma_i|^2\right]$$

and, denoting $\Sigma_i=\sqrt{z_i}e^{j\theta}\Leftrightarrow z_i=|\Sigma_i|^2$:

$$f(z_i)=\int_0^{2\pi}\frac{e^{-\frac{z_i}{\mathbf{C}_{11}}}}{\pi(\mathbf{C}_{11})^2}[1+bz_i]\frac{d\theta}{2}=\frac{e^{-\frac{z_i}{\mathbf{C}_{11}}}}{(\mathbf{C}_{11})^2}[1+bz_i]$$

Therefore:

$$f(\varpi_1)=\frac{e^{-\frac{\varpi_1}{\mathbf{C}_{11}}}}{(\mathbf{C}_{11})^2}P_1(\varpi_1),\quad P_1(\varpi_1)=1+b\varpi_1$$

Now, let us assume that:

$$f(\varpi_{I-1})=\frac{e^{-\frac{\varpi_{I-1}}{\mathbf{C}_{11}}}}{(\mathbf{C}_{11})^{2(I-1)}}P_{I-1}(\varpi_{I-1})$$

then the change of variables $\varpi_I=\varpi_{I-1}+|\Sigma_I|^2,\ t=\varpi_{I-1}$ leads to:

$$f(\varpi_I)=\int_0^{\varpi_I}\frac{e^{-\frac{t}{\mathbf{C}_{11}}}}{(\mathbf{C}_{11})^{2(I-1)}}P_{I-1}(t)\frac{e^{-\frac{(\varpi_I-t)}{\mathbf{C}_{11}}}}{(\mathbf{C}_{11})^2}[1+b(\varpi_I-t)]dt=\frac{e^{-\frac{\varpi_I}{\mathbf{C}_{11}}}}{(\mathbf{C}_{11})^{2I}}\int_0^{\varpi_I}P_{I-1}(t)[1+b(\varpi_I-t)]dt\ ,$$

that is to say:

$$f(\varpi_I)=\frac{e^{-\frac{\varpi_I}{\mathbf{C}_{11}}}}{(\mathbf{C}_{11})^{2I}}P_I(\varpi_I),\quad P_I(t)=Q_{I-1}(t)+bH_{I-1}(t),\quad Q_I(t)=\int_0^t P_I(z)\,dz,\quad H_I(t)=\int_0^t Q_I(z)\,dz$$

As $Q_0(t)=1$ and $H_0(t)=t$, then $P_I(t), Q_I(t)$ and $H_I(t)$ are polynomials of the form:

$$P_I(t)=\left[\sum_{m=0}^I p_I^m t^m\right]t^{I-1},\quad Q_I(t)=\left[\sum_{m=0}^I q_I^m t^m\right]t^I,\quad H_I(t)=\left[\sum_{m=0}^I h_I^m t^m\right]t^{I+1}$$

In the same way, some straightforward computations establish that:

$$f\left(x_I\right)=\frac{e^{-\frac{x_I}{\mathbf{C}_{11}}}}{\left(\mathbf{C}_{11}\right)^{2I-1}}Q_{I-1}\left(x_I\right),\quad f\left(y_I\right)=\frac{e^{-\frac{y_I}{\mathbf{C}_{11}}}}{\left(\mathbf{C}_{11}\right)^{2I-2}}H_{I-2}\left(y_I\right)$$

Thus (41) can be rewritten as:

$$P_D = \int_T^{\infty}\frac{e^{-\frac{t}{\mathbf{C}_{11}}}}{\left(\mathbf{C}_{11}\right)^{2I}}P_I(t)dt$$

$$E\left[\frac{1}{\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}}\mid D\right] = \frac{1}{P_D}\int_T^{\infty}\frac{e^{-\frac{t}{\mathbf{C}_{11}}}}{(\mathbf{C}_{11})^{2I}}\frac{P_I\left(t\right)}{t}dt,\quad E\left[\frac{1}{\left(\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}\right)^2}\mid D\right]=\frac{1}{P_D}\int_T^{\infty}\frac{e^{-\frac{t}{\mathbf{C}_{11}}}}{(\mathbf{C}_{11})^{2I}}\frac{P_I\left(t\right)}{t^2}dt$$

which, introduced in (42)(43)(44), allows to get the remaining unknown conditional expectations:

$$E\left[\frac{1}{\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}}\frac{\left|\Sigma_i\right|^2}{E\left[b^i_{\overrightarrow{\Sigma}}\left(U_i\right)\right]}\mid D\right]=\frac{1}{P_D}\int_T^{\infty}\frac{e^{-\frac{t}{\mathbf{C}_{11}}}}{(\mathbf{C}_{11})^{2I-1}}\frac{H_{I-1}\left(t\right)}{t}dt$$

$$E\left[\frac{1}{\left(\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}\right)^2}\frac{\left|\Sigma_i\right|^2}{E\left[b^i_{\overrightarrow{\Sigma}}\left(U_i\right)\right]}\mid D\right]=\frac{1}{P_D}\int_T^{\infty}\frac{e^{-\frac{t}{\mathbf{C}_{11}}}}{(\mathbf{C}_{11})^{2I-1}}\frac{H_{I-1}\left(t\right)}{t^2}dt$$

$$E\left[\frac{1}{\left(\overrightarrow{\Sigma}^H\overrightarrow{\Sigma}\right)^2}\frac{\left|\Sigma_i\right|^2\left|\Sigma_j\right|^2}{E\left[b^i_{\overrightarrow{\Sigma}}\left(U_i\right)\right]E\left[b^j_{\overrightarrow{\Sigma}}\left(U_j\right)\right]}\mid D\right]=\frac{1}{P_D}\int_T^{\infty}\frac{e^{-\frac{t}{\mathbf{C}_{11}}}}{(\mathbf{C}_{11})^{2I-2}}\frac{1}{t^2}\left(\int_0^t\left[\int_0^z H_{I-2}\left(u\right)du\right]dz\right)dt$$

Last, the following identity:

$$\int_T^{\infty}t^N e^{-t}dt=e^{-T}e_N\left(T\right)N!,\ N\geq 0,\quad e_N\left(T\right)=\sum_{n=0}^{N}\frac{T^n}{n!}$$

allows to complete the derivation and to obtain formulas (20a-b)(21a-b)(22). For example:

$$P_D=\int_T^{\infty}\frac{e^{-\frac{t}{\mathbf{C}_{11}}}}{\left(\mathbf{C}_{11}\right)^{2I}}P_I(t)dt=\int_T^{\infty}\frac{e^{-\frac{t}{\mathbf{C}_{11}}}}{\left(\mathbf{C}_{11}\right)^{2I}}\left[\sum_{m=0}^{I}p_I^m t^m\right]t^{I-1}dt=\frac{1}{\left(\mathbf{C}_{11}\right)^I}\sum_{m=0}^{I}p_I^m\left(\mathbf{C}_{11}\right)^m\int_T^{\infty}\frac{e^{-\frac{t}{\mathbf{C}_{11}}}}{\mathbf{C}_{11}}\left(\frac{t}{\mathbf{C}_{11}}\right)^{I-1+m}dt$$

i.e.:

$$P_D=\frac{e^{-\frac{T}{\mathbf{C}_{11}}}}{\left(\mathbf{C}_{11}\right)^I}A_0,\quad A_0=\sum_{m=0}^{I}p_I^m\left(\mathbf{C}_{11}\right)^m e_{I-1+m}\left(\frac{T}{\mathbf{C}_{11}}\right)(I-1+m)!$$

VIII. Acknowledgments

Among the comments on the original manuscript, the search of a derivation of the MLE of complex monopulse ratio for a Swerling 3-4 target (section III) was suggested by one reviewer and helped to provide a more thorough characterization of the problem at hand.

REFERENCES

[1] N. Levanon, ”Radar Principles”, Wiley-Interscience, Wiley 1988
[2] E. Mosca, ”Angle Estimation in Amplitude Comparison Monopulse Systems”, IEEE Trans., AES-5 (Mar 1969), 205-212
[3] P. Swerling, ”Probability of detection for fluctuating targets”, IEEE Trans., IT-6 (Apr 1960), 269- 308
[4] I. Kanter, ”The Ratios of Functions of Random Variables”, IEEE Trans., AES-13 (Nov 1977), 624-630
[5] I. Kanter, ”Multiple Gaussian targets, the track-on_jam problem”, IEEE Trans., AES-13 (Nov 1977), 620-623
[6] B.E. Tullsson, ”Monopulse tracking of Rayleigh targets, a simple approach”, IEEE Trans., AES-27 (May 1991), 520-531
[7] A.D. Seifer, ”Monopulse-radar angle tracking in noise or noise jamming”, IEEE Trans., AES-28 (Jul 1992), 622-637
[8] A. Sinha, T. Kirubarajan, Y. Bar-Shalom, ”Maximum Likelihood Angle Extractor for Two Closely Spaced Target”, IEEE Trans., AES-38 (Jan 2002), 183-201
[9] H.L. Van Trees,”Optimum Array Processing”, New-York, Wiley-Interscience, 2002
[10] U. Nickel, ”Performance of corrected monopulse estimation”, IEE Proc. Radar Sonar Navig.-146 (1) (Feb 1999), 17-24
[11] W.D. Blair, M. Brandt-Pierce, ”Statistical description of monopulse parameters for Tracking Rayleigh targets”, IEEE Trans., AES-34 (Apr 1998), 597–610
[12] W.D. Blair, M. Brandt-Pierce, ”Monopulse DOA estimation of two unresolved Rayleigh targets”, IEEE Trans., AES-37 (Apr 2001), 452-469
[13] E. Chaumette, P. Larzabal, ”Optimal Detection Theory Applied to Monopulse Antenna”, ICASSP Conf. 2004, Vol II, 1065-1068
[14] E. Chaumette, P. Larzabal, ”Optimal Monopulse Tracking of Signal Source Of Unknown Amplitude”, EUSIPCO Conf. 2004, 1445-1448
[15] S. Asseo, ”Detection of Target Multiplicity using Monopulse Quadrature Angle”, IEEE Trans., AES-17 (Mar 1981), 271-280
[16] S.M. Kay, ”Fundamentals of Statistical Signal Processing: detection theory”, Prentice-Hall, 1998
[17] D.H. Brandwood, ”A complex gradient operator and its application in adaptive array theory”, IEE Proc. F and H - 130 (1983), 11-16
[18] R.A. Horn, C.R. Johnson, ”Matrix Analysis”, Cambridge University Press, 1999
[19] A. Papoulis, ”Probability, Random Variables and Stochastic Processes”, 3rd Ed., New York, McGraw-Hill, 1991

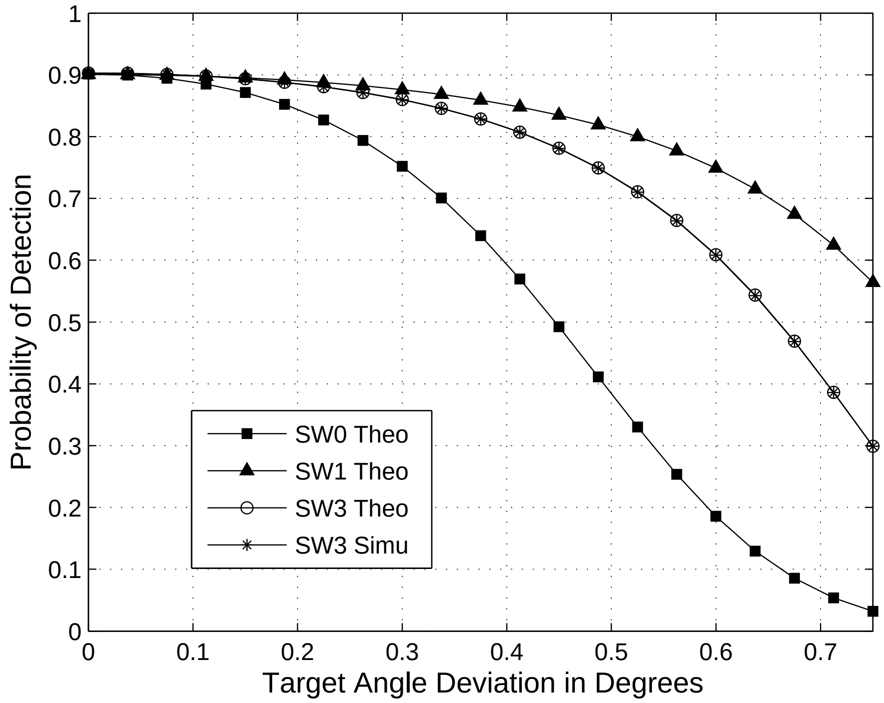


Fig. 2. Probability of Detection versus Target angle deviation - 1 observation

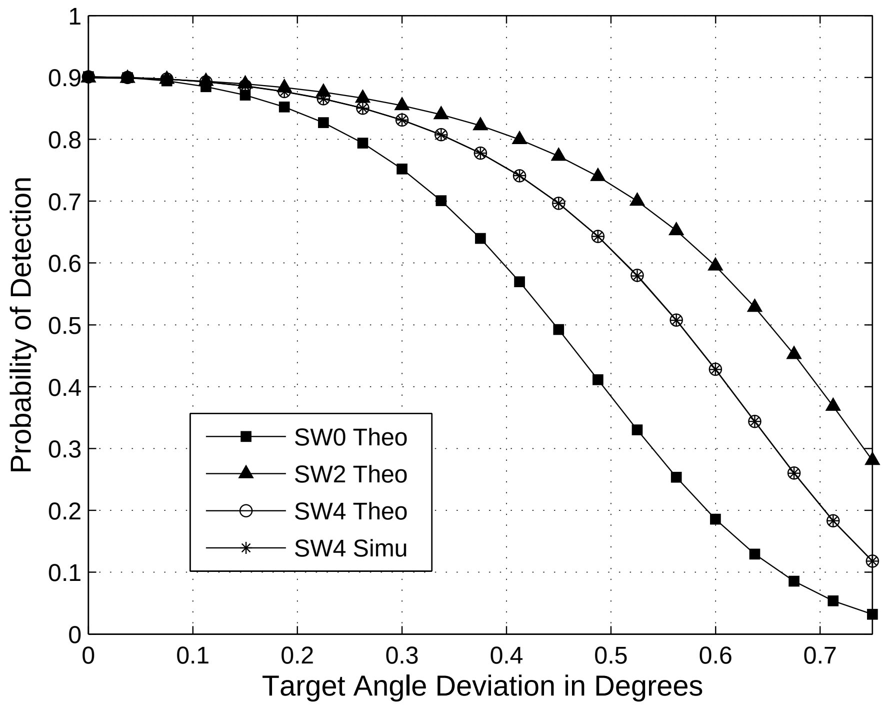


Fig. 3. Probability of Detection versus Target angle deviation - 2 observations

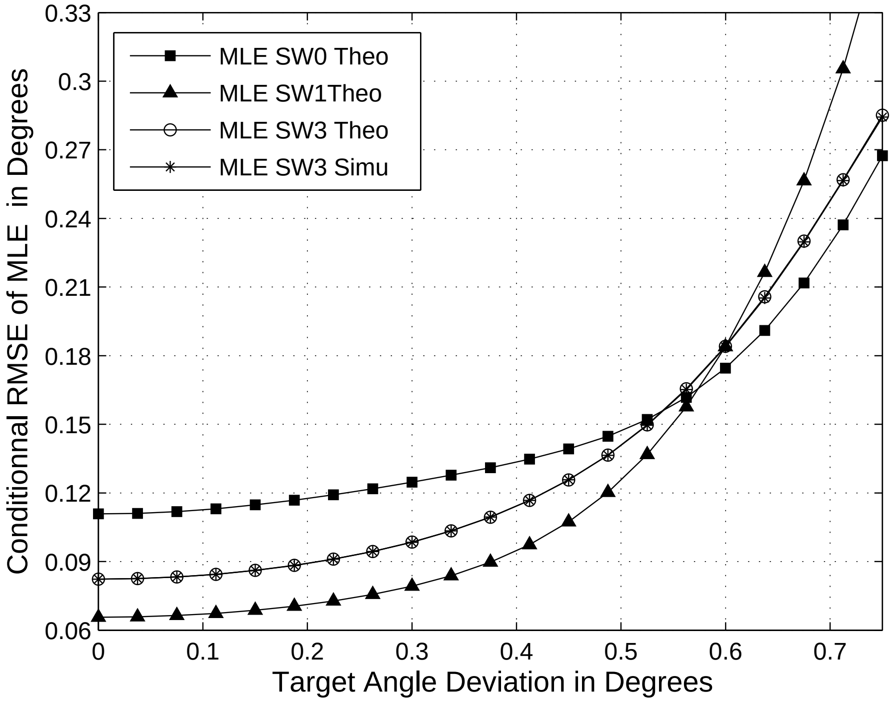


Fig. 4. Conditonal RMSE of $\widehat{\theta}_{ML} = \frac{\mathrm{Re}\{\widehat{r}_{ML}\}}{k}$ versus Target angle deviation - 1 observation

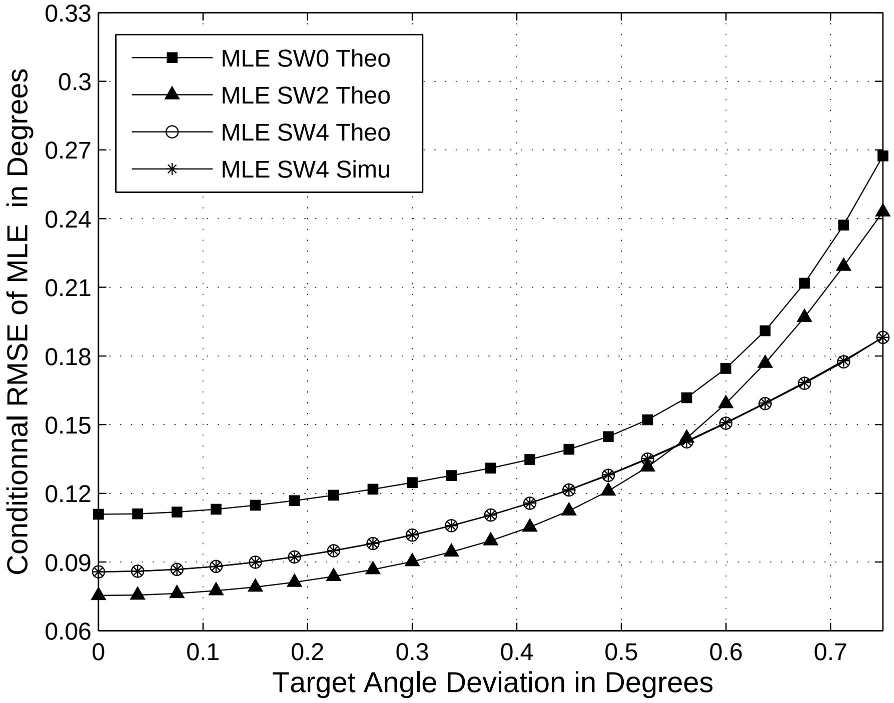


Fig. 5. Conditonal RMSE of $\widehat{\theta}_{ML} = \frac{\mathrm{Re}\{\widehat{r}_{ML}\}}{k}$ versus Target angle deviation - 2 observations

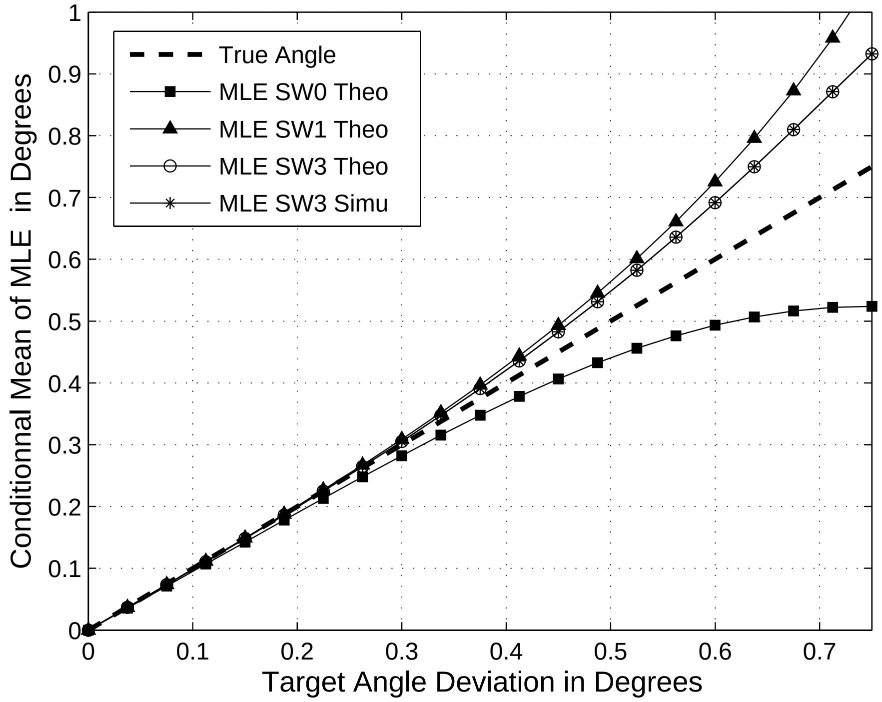


Fig. 6. Conditonal Mean of $\widehat{\theta}_{ML} = \frac{\mathrm{Re}\{\widehat{r}_{ML}\}}{k}$ versus Target angle deviation - 1 observation

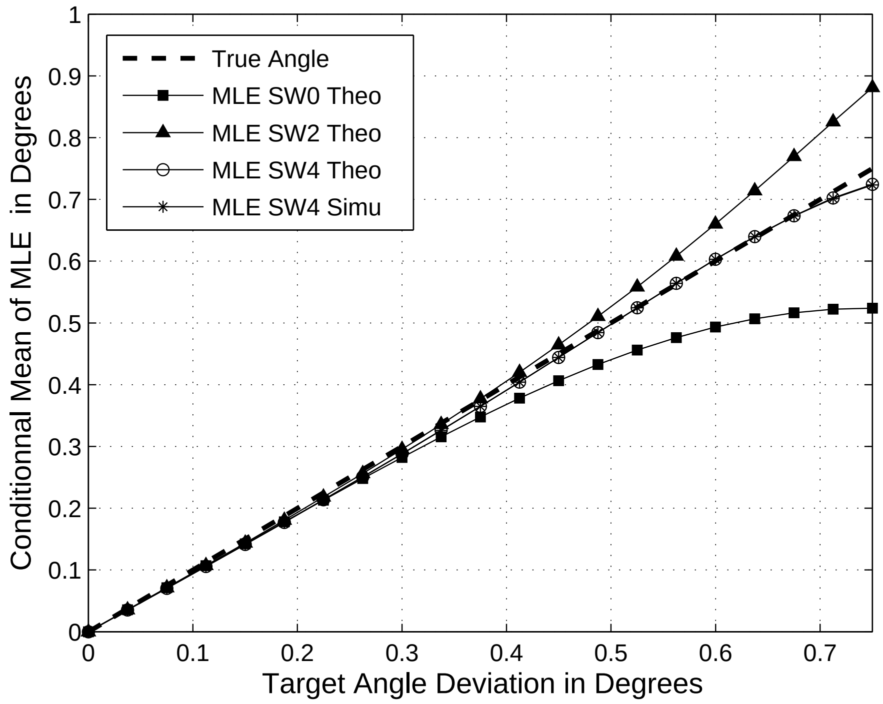


Fig. 7. Conditonal Mean of $\widehat{\theta}_{ML} = \frac{\mathrm{Re}\{\widehat{r}_{ML}\}}{k}$ versus Target angle deviation - 2 observations

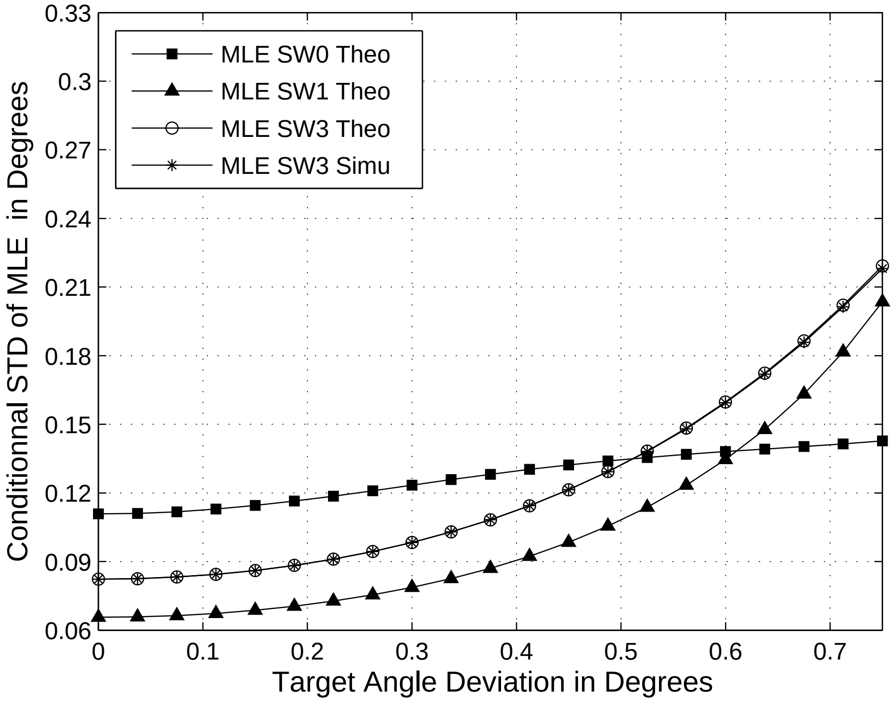


Fig. 8. Conditonal STD of $\widehat{\theta}_{ML} = \frac{\mathrm{Re}\{\widehat{r}_{ML}\}}{k}$ versus Target angle deviation - 1 observation

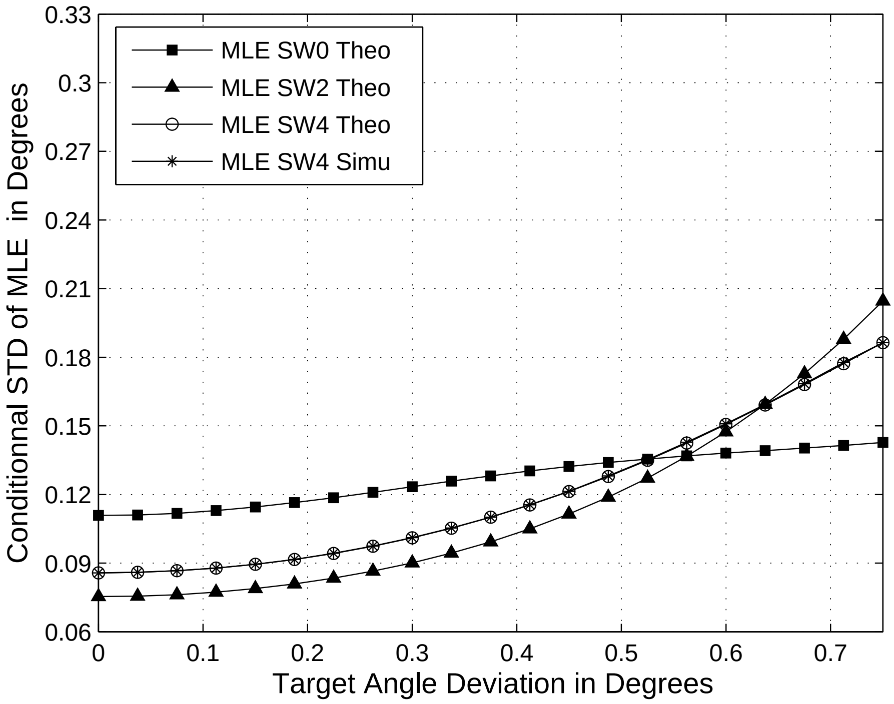


Fig. 9. Conditonal STD of $\widehat{\theta}_{ML} = \frac{\mathrm{Re}\{\widehat{r}_{ML}\}}{k}$ versus Target angle deviation - 2 observations

H. New Results on Deterministic Cramér–Rao Bounds for Real and Complex Parameters (IEEE TSP)

IEEE Trans. on SP, 60(3): 1032-1049, 2012.

New results on Deterministic Cramér-Rao bounds for real and complex parameters

Tarek Menni, Eric Chaumette, Pascal Larzabal, Jean Pierre Barbot

Abstract

The CRB is a lower bound of great interest for system analysis and design in the asymptotic region (high SNR and/or large number of snapshots), as it is simple to calculate and it is usually possible to obtain closed form expressions. The first part of the paper is a generalization to complex parameters of the Barankin rationale for deriving MSE lower bounds, that is the minimization of a norm under a set of linear constraints. With the norm minimization approach the study of FIM singularity, constrained CRB and regularity conditions become straightforward corollaries of the derivation. The second part provides new results useful for system analysis and design: a general reparameterization inequality, the equivalence between reparameterization and equality constraints and an explicit relationship between parameters unidentifiability and FIM singularity.

Index Terms: Deterministic parameter estimation, Cramér–Rao bound, Fisher information matrix

EDICS: SSP-PERF

I. Introduction

Minimal performance bounds allow for calculation of the best performance that can be achieved in the Mean Square Error (MSE) sense, when estimating parameters of a signal corrupted by noise. There are two main categories of lower bounds [1]. Those that evaluate the "locally best" behaviour of the estimator and those that consider the "globally best" performance. In the first case, the parameters being estimated are considered to be deterministic and to be embedded in a noise signal whose parameters are also considered deterministic. The second category considers the parameters as random variables with an *a priori* probability. This paper is concerned with the first category of bounds concerning deterministic parameters.

Historically the first MSE lower bound for deterministic parameters to be derived was the Cramér-Rao Bound (CRB), which was introduced to investigate fundamental limits of a parameter estimation problem or to assess the relative performance of a specific estimator (efficiency) [2][3][4]. It has since become the most popular lower bound due to its simplicity of calculation, the fact that in many cases it can be achieved asymptotically (high SNR [5] and/or large number of snapshots [2]) by Maximum Likelihood Estimators (MLE) [1][6], and last but not least, its noticeable property of being the lowest bound on the MSE of unbiased estimators, since it derives from the *weakest* formulation of unbiasedness at the vicinity of any selected value of the parameters [7][8][9] (see §III-B). This initial characterization of locally unbiased estimators has been significantly generalized by Barankin work [7], who established the general form of the greatest lower bound of any absolute moment of an unbiased estimator. In the particular case of MSE - absolute moment of order two - his work allows the derivation of the highest lower bound on MSE (BB) since it takes into account the *strongest* formulation of unbiasedness, that is to say uniform unbiasedness (unbiasedness over an interval of parameter values including the selected value). Unfortunately the BB is the solution of an integral equation [10][11][12] with a generally incomputable analytic solution. Therefore,

Tarek Menni is a Phd Student at SATIE/CNRS, Ecole Normale Supérieure de Cachan, 61 avenue du Président Wilson - 94235 Cachan Cedex, France. E-mail: tarek.menni@satie.ens-cachan.fr, Phone: +33 1 47 40 55 83 .

Eric Chaumette is with the ONERA, the French Aerospace Lab, DEMR/TSI, Chemin de la Hunière 91120 Palaiseau, France. E-mail: eric.chaumette@onera.fr, Phone: +33 1 80 38 62 96, Fax: +33 1 80 38 62 69

P. Larzabal is with SATIE/CNRS, Ecole Normale Supérieure de Cachan, 61 avenue du Président Wilson - 94235 Cachan Cedex, France. E-mail: pascal.larzabal@satie.ens-cachan.fr, Phone: +33 1 47 40 27 09.

J.P. Barbot is with SATIE/CNRS, Ecole Normale Supérieure de Cachan, 61 avenue du Président Wilson - 94235 Cachan Cedex, France. E-mail: barbot@satie.ens-cachan.fr, Phone: +33 1 47 40 53 20.

This work has been partly funded by the European Network of excellence NEWCOM++ under the number 216715

since then, numerous works detailed in [10][12] have been devoted to deriving computable approximations of the BB and have shown that the CRB and the BB can be regarded as key representatives of two general classes of bounds, respectively the Small-Error bounds and the Large-Error bounds. These works have also shown that in non-linear estimation problems three distinct regions of operation can be observed. In the asymptotic region, the MSE is small and, in many cases, close to the Small-Error bounds. In the *a priori* performance region where the number of independent snapshots and/or the SNR are very low, the observations provide little information and the MSE is close to that obtained from the prior knowledge about the problem. Between these two extremes, there is an additional ambiguity region, also called the transition region. In this region, the MSE of MLEs usually deteriorates rapidly with respect to Small-Error bounds and exhibits a threshold behaviour corresponding to a "performance breakdown". The nature of this phenomenon is specified by a complicated non-smooth behaviour of the likelihood function in the "threshold" area where it tends to generate outliers [9][13]. Small-Error bound such as the CRB are not able to handle the threshold phenomena, whereas it is revealed by Large-Error bounds that can be used to predict the threshold value. Unfortunately, Large-Error bounds closed form expressions hardly ever exist and, even when they exist, each Large-Error bound requests the search of an optimum over a set of test points, leading to a computational cost prohibitive in most applications when the number of unknown parameters increases. Therefore, provided that one keeps in mind the CRB limitations, that is, to become an excessively optimistic lower bound when the observation conditions degrade (low SNR and/or low number of snapshots), the CRB is still a lower bound of great interest for system analysis and design in the asymptotic region, as it is simple to calculate and it is usually possible to obtain closed form expressions.

The first part of the paper - Section (III) - is a generalization to complex parameters of the Barankin rationale for deriving MSE lower bounds, that is the minimization of a norm under a set of linear constraints, as presented in [14, II.B] for real parameters. When the lower bound is the CRB, the set of linear constraints involved reduces to a set of derivative constraints. This property allows a unique simple derivation, whatever the nature (real or complex) of the unknown parameters, that avoids sophisticated matrix manipulations generally used with complex parameters [15][16]. With the norm minimization approach, the case of multiple unknown parameters is easily handled and the study of FIM singularity, constrained CRB, regularity conditions become straightforward corollaries of the derivation, condensing to a few lines previous works [15][17][18][19][20][21] based on the use of the covariance inequality (the alternative approach to derive MSE lower bounds) [22]. The second part of the paper is dedicated to provide new technical results useful for system analysis and design in the asymptotic region. First, for system design and optimization it is worth knowing the general reparameterization inequality provided in Section IV (new to the best of our knowledge). Indeed, a way to improve the estimation of a subset of unknown parameters (parameters of interest for example) can be to introduce, by design choices, either a parameterization change or equality constraints among the other parameters (nuisance parameters for example). Additionally the reparamaterization inequality may allow simple derivations of useful theoretical results as shown in Section V. Last, for system design it is also worth knowing if the parameters of interest are identifiable, that is, if they can be estimated (with or without bias) whatever their values. And if they are not identifiable, at least on which subset of the parameter space. This is addressed in Section VI where we explicitly formulate the relation between unidentifiability and FIM singularity.

An outline of the paper is as follows. Section (II) introduces the algebraic notations used in the paper. In Section (III) a simple rationale for deriving CRB for real and complex parameters is given, and the study of FIM singularity, equality constraints, regularity conditions are revisited in the light of this rationale. Section (IV) establishes the equivalence between reparameterization and equality constraints and provides a useful reparameterization inequality. Section V shows two theoretical applications of the reparameterization inequality. Last, section VI explicitly formulates the relation between unidentifiability and FIM singularity.

II. Notations

The notational convention adopted is as follows: italic indicates a scalar quantity, as in a; lower case boldface indicates a column vector quantity, as in $\mathbf{a}$; upper case boldface indicates a matrix quantity, as in $\mathbf{A}$. The n-th row and m-th column element of the matrix $\mathbf{A}$ will be denoted by $a_{n,m}$ or $(\mathbf{A})_{n,m}$. The n-th coordinate of the column vector $\mathbf{a}$ will be denoted by a_n or $(\mathbf{a})_n$. $\mathrm{Re}\{A\}$ is the real part of A and $\mathrm{Im}\{A\}$ is the imaginary part of A. The matrix/vector transpose is indicated by a superscript ^{T} as in $\mathbf{A}^T$. The matrix/vector conjugate is indicated by a superscript * as in $\mathbf{A}^*$. The matrix/vector transpose conjugate is indicated by a superscript as in $\mathbf{A}^H$. $|\mathbf{A}|$ is the determinant of the square matrix $\mathbf{A}$. $[\mathbf{A}, \mathbf{B}]$ denotes the matrix resulting from the horizontal concatenation of matrices

$\mathbf{A}$ and $\mathbf{B}$. $\left(\mathbf{a}^T, \mathbf{b}^T\right)$ denotes the row vector resulting from the horizontal concatenation of row vectors $\mathbf{a}^T$ and $\mathbf{b}^T$. $\mathbf{I}_M$ is the identity matrix of order M. $\mathbf{vec}(\mathbf{A})$ is a column vector obtained from matrix $\mathbf{A}$ by stacking its column vectors one below another. $S = span\{\mathbf{A}\}$ where $\mathbf{A}$ is a matrix denotes the linear span of the set of its column vectors. $S^\perp$ denotes the orthogonal complement of the subspace S. For two matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A} \geq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive semi-definite. $E[\cdot]$ denotes the expectation operator and $\|.\|$ denotes a norm. $o(.)$ and $O(.)$ denotes respectively the small oh and big Oh notation. If $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_P)^T$, then: $\frac{\partial}{\partial \boldsymbol{\theta}} = \left(\frac{\partial}{\partial \theta_1}, \frac{\partial}{\partial \theta_2}, \ldots, \frac{\partial}{\partial \theta_P}\right)^T$, $\frac{\partial}{\partial \boldsymbol{\theta}^T} = \left(\frac{\partial}{\partial \theta_1}, \frac{\partial}{\partial \theta_2}, \ldots, \frac{\partial}{\partial \theta_P}\right)$. $\mathcal{M}_{\mathbb{R}}(N, P)$ denotes the vector space of real matrices with N rows and P columns. $\mathcal{M}_{\mathbb{C}}(N, P)$ denotes the vector space of complex matrices with N rows and P columns. $\odot$ denotes the Hadamard product. $\otimes$ denotes the Kronecker product. $1(\mathbf{x})$ denotes the constant real-valued function with value equal to 1. $\underline{\mathbf{x}}$ denotes the following notation:

$$\underline{\mathbf{x}} : \begin{cases} \underline{\mathbf{x}} = \mathbf{x} & \text{if } \mathbf{x} \in \mathbb{R}^Q \\ \underline{\mathbf{x}} = \left(\mathbf{x}^T, \mathbf{x}^H\right)^T & \text{if } \mathbf{x} \in \mathbb{C}^Q \text{ and } \mathbf{x} \notin \mathbb{R}^Q \\ \underline{\mathbf{x}} = \left(\mathbf{x}_c^T, \mathbf{x}_c^H, \mathbf{x}_r^T\right)^T & \text{if } \mathbf{x} = \left(\mathbf{x}_c^T, \mathbf{x}_r^T\right)^T, \quad \mathbf{x}_c \in \mathbb{C}^Q \text{ and } \mathbf{x}_c \notin \mathbb{R}^Q, \quad \mathbf{x}_r \in \mathbb{R}^{Q'} \end{cases} \tag{1}$$

Additionally, regarding the definition of Hermitian product, we adopt the convention used in books of mathematics including [23][24][25][26], where a sesquilinear form is a function in two variables on a complex vector space $\mathbb{U}$ which is linear in the first variable and semi-linear in the second:

$$\langle \mid \rangle : \mathbb{U} \times \mathbb{U} \to \mathbb{C} \ / \ \forall \mathbf{u}, \mathbf{c} \in \mathbb{U}, \ \forall \lambda \in \mathbb{C}, \ \langle \lambda \mathbf{u} \mid \mathbf{c} \rangle = \lambda \langle \mathbf{u} \mid \mathbf{c} \rangle, \ \langle \mathbf{u} \mid \lambda \mathbf{c} \rangle = \lambda^* \langle \mathbf{u} \mid \mathbf{c} \rangle \tag{2}$$

This convention allows to define the Gram matrix associated to 2 families of vectors of $\mathbb{U}$, $\{\mathbf{u}\}_{[1,Q]} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_Q\}$ and $\{\mathbf{c}\}_{[1,P]} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_P\}$ as [25]:

$$\mathbf{G}\left(\{\mathbf{u}\}_{[1,Q]}, \{\mathbf{c}\}_{[1,P]}\right) = \begin{bmatrix} \langle \mathbf{u}_1 \mid \mathbf{c}_1 \rangle & \ldots & \langle \mathbf{u}_Q \mid \mathbf{c}_1 \rangle \\ \vdots & \ldots & \vdots \\ \langle \mathbf{u}_1 \mid \mathbf{c}_P \rangle & \ldots & \langle \mathbf{u}_Q \mid \mathbf{c}_P \rangle \end{bmatrix} \in \mathcal{M}_{\mathbb{C}}(P, Q) \tag{3}$$

$$\left(\mathbf{G}\left(\{\mathbf{u}\}_{[1,Q]}, \{\mathbf{c}\}_{[1,P]}\right)\right)_{p,q} = \langle \mathbf{u}_q \mid \mathbf{c}_p \rangle, \qquad \mathbf{G}\left(\{\mathbf{c}\}_{[1,P]}, \{\mathbf{u}\}_{[1,Q]}\right) = \mathbf{G}\left(\{\mathbf{u}\}_{[1,Q]}, \{\mathbf{c}\}_{[1,P]}\right)^H$$

leading to:

$$\left\langle \sum_{q=1}^{Q} x_q \mathbf{u}_q \mid \sum_{p=1}^{P} y_p \mathbf{c}_p \right\rangle = \mathbf{y}^H \mathbf{G}\left(\mathbf{u}_{[1,Q]}, \mathbf{c}_{[1,P]}\right) \mathbf{x}, \quad \mathbf{x} = (x_1, \ldots, x_Q)^T, \mathbf{y} = (y_1, \ldots, y_P)^T \tag{4}$$

For notational convenience:

$$\mathbf{G}\left(\{\mathbf{u}\}_{[1,Q]}\right) = \mathbf{G}\left(\{\mathbf{u}\}_{[1,Q]}, \{\mathbf{u}\}_{[1,Q]}\right) \tag{5}$$

Beware that most reference signal processing books including [1, p1343][27][28] adopt the opposite convention for sesquilinear form, that is to be semi-linear in the first variable and linear in the second. As a consequence, the equivalent form in ”signal processing notation” of any inequality derived in the present paper is obtained by transposing inequality terms. Thanks to the adopted convention, even in the case of complex parameters, the Fisher Information Matrix (see hereinafter) appears to be both Gram matrix and correlation matrix derived from the canonical definition of the MSE, i.e., a norm associated to an Hermitian product depending on the p.d.f. of the observation.

III. A simple and insightful Cramér-Rao Bound derivation

This section is a generalization to complex parameters of the Barankin rationale for deriving MSE lower bounds, that is the minimization of a norm under a set of linear constraints presented in [14, II.B] for real parameters. When the lower bound is the CRB, the set of linear constraints involved reduces to a set of derivative constraints. This property allows a unique simple derivation, whatever the nature (real or complex) of the unknown parameters that:

- avoids sophisticated matrix manipulations generally used with complex parameters [15][16],
- corrects previously incomplete derivation [21],
- allows to condense to a few lines previous works [15][17][18][19][20][21] (based on the use of the covariance inequality [22]) on FIM singularity and constrained CRB,
- allows to clarify standard regularity conditions so far needlessly too restrictive.

A. Differentiability on real or complex field

The sets of complex ($\mathbb{C}$) and real ($\mathbb{R}$) numbers being two fields, the differentiability[1] of a vector of functions $\mathbf{f}(\boldsymbol{\theta}) : (\Bbbk')^P \to \Bbbk^Q$ where $\Bbbk' \equiv \mathbb{C}$ or $\mathbb{R}$ and $\Bbbk \equiv \mathbb{C}$ or $\mathbb{R}$ can be characterized by the following property [24]:

$$\mathbf{f}(\boldsymbol{\theta} + d\boldsymbol{\theta}) = \mathbf{f}(\boldsymbol{\theta}) + \frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} d\boldsymbol{\theta} + \mathbf{o}(d\boldsymbol{\theta}), \quad \mathbf{o}(d\boldsymbol{\theta}) = d\boldsymbol{\theta} \odot \boldsymbol{\varepsilon}(d\boldsymbol{\theta}), \quad \lim_{\|d\boldsymbol{\theta}\| \to 0} \|\boldsymbol{\varepsilon}(d\boldsymbol{\theta})\| = 0 \tag{6}$$

where $\| \ \|$ is the canonical Hermitian (or Euclidean) norm on $(\Bbbk')^P$ and $\Bbbk^Q$: $\|\mathbf{x}\| = \sqrt{\sum_i |x_i|^2}$. As (6) still holds if $\boldsymbol{\theta}$ have mixed components (complex and real), any function of $\boldsymbol{\theta} = \left(\operatorname{Re}\{\boldsymbol{\theta}_c\}^T, \operatorname{Im}\{\boldsymbol{\theta}_c\}^T, \boldsymbol{\theta}_r^T\right)^T \in \mathbb{R}^{2P_c+P_r}, \boldsymbol{\theta}_c \in \mathbb{C}^{P_c}, \boldsymbol{\theta}_r \in \mathbb{R}^{P_r}$, can be written in a dual form:

$$\begin{cases} \mathbf{f}(\boldsymbol{\theta}) : \mathbb{R}^{2P_c+P_r} \to \Bbbk^Q \\ \mathbf{f}(\boldsymbol{\theta}) = \widetilde{\mathbf{f}}(\underline{\boldsymbol{\theta}}), \quad \widetilde{\mathbf{f}}(\mathbf{z}) = \mathbf{f}\left(\begin{pmatrix} \frac{1}{2}(\mathbf{z}_1 + \mathbf{z}_2) \\ \frac{1}{2j}(\mathbf{z}_1 - \mathbf{z}_2) \\ \mathbf{x}_r \end{pmatrix}\right), \ \mathbf{z} = \begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \mathbf{x}_r \end{pmatrix} : \mathbb{C}^{P_c} \times \mathbb{C}^{P_c} \times \mathbb{R}^{P_r} \to \Bbbk^Q \end{cases}$$

where $\underline{\boldsymbol{\theta}} = \left(\boldsymbol{\theta}_c^T, (\boldsymbol{\theta}_c^*)^T, \boldsymbol{\theta}_r^T\right)^T \in \mathbb{C}^{2P_c} \times \mathbb{R}^{P_r}$. Then, if $\mathbf{f}$ and $\widetilde{\mathbf{f}}$ are differentiable (6), we obtain that:

$$\frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} d\boldsymbol{\theta} = \left. \frac{\partial \widetilde{\mathbf{f}}(\mathbf{z})}{\partial \mathbf{z}_1^T} d\mathbf{z}_1 + \frac{\partial \widetilde{\mathbf{f}}(\mathbf{z})}{\partial \mathbf{z}_2^T} d\mathbf{z}_2 + \frac{\partial \widetilde{\mathbf{f}}(\mathbf{z})}{\partial \mathbf{x}_r^T} d\mathbf{x}_r \right|_{\mathbf{z}=\underline{\boldsymbol{\theta}}}$$

which can be rewritten as:

$$\frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} d\boldsymbol{\theta} = \frac{\partial \widetilde{\mathbf{f}}(\underline{\boldsymbol{\theta}})}{\partial (\boldsymbol{\theta}_c)^T} d\boldsymbol{\theta}_c + \frac{\partial \widetilde{\mathbf{f}}(\underline{\boldsymbol{\theta}})}{\partial (\boldsymbol{\theta}_c^*)^T} d\boldsymbol{\theta}_c^* + \frac{\partial \widetilde{\mathbf{f}}(\underline{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}_r^T} d\boldsymbol{\theta}_r, \quad d\boldsymbol{\theta}_c^* = (d\boldsymbol{\theta}_c)^*$$

provided that $\boldsymbol{\theta}_c$ and $\boldsymbol{\theta}_c^*$ are formally considered as independent variables for derivation. Using notation (1) the identity above becomes:

$$\frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} d\boldsymbol{\theta} = \frac{\partial \widetilde{\mathbf{f}}(\underline{\boldsymbol{\theta}})}{\partial \underline{\boldsymbol{\theta}}^T} d\underline{\boldsymbol{\theta}}, \quad d\boldsymbol{\theta}_c^* = (d\boldsymbol{\theta}_c)^*, \tag{7}$$

identity that still holds if $P_r = 0$ or $P_c = 0$. Well known properties straightforwardly derive from (7):

- if $\mathbf{f}(\ )$ is real-valued and $\boldsymbol{\theta} = \left(\operatorname{Re}\{\boldsymbol{\theta}_c\}^T, \operatorname{Im}\{\boldsymbol{\theta}_c\}^T\right)^T$ $(P_r = 0)$ then [29]:

$$\begin{cases} \frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \operatorname{Re}\{\boldsymbol{\theta}_c\}^T} d\operatorname{Re}\{\boldsymbol{\theta}_c\} = \frac{\partial \widetilde{\mathbf{f}}(\underline{\boldsymbol{\theta}})}{\partial (\boldsymbol{\theta}_c)^T} d\operatorname{Re}\{\boldsymbol{\theta}_c\} + \frac{\partial \widetilde{\mathbf{f}}(\underline{\boldsymbol{\theta}})}{\partial (\boldsymbol{\theta}_c^*)^T} d\operatorname{Re}\{\boldsymbol{\theta}_c\} \\ \frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \operatorname{Im}\{\boldsymbol{\theta}_c\}^T} d\operatorname{Im}\{\boldsymbol{\theta}_c\} = j\frac{\partial \widetilde{\mathbf{f}}(\underline{\boldsymbol{\theta}})}{\partial (\boldsymbol{\theta}_c)^T} d\operatorname{Im}\{\boldsymbol{\theta}_c\} - j\frac{\partial \widetilde{\mathbf{f}}(\underline{\boldsymbol{\theta}})}{\partial (\boldsymbol{\theta}_c^*)^T} d\operatorname{Im}\{\boldsymbol{\theta}_c\} \end{cases}$$

$$\Updownarrow$$

$$\begin{cases} \frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \operatorname{Re}\{\boldsymbol{\theta}_c\}^T} = \frac{\partial \widetilde{\mathbf{f}}(\underline{\boldsymbol{\theta}})}{\partial (\boldsymbol{\theta}_c)^T} + \frac{\partial \widetilde{\mathbf{f}}(\underline{\boldsymbol{\theta}})}{\partial (\boldsymbol{\theta}_c^*)^T} \\ \frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \operatorname{Im}\{\boldsymbol{\theta}_c\}^T} = j\frac{\partial \widetilde{\mathbf{f}}(\underline{\boldsymbol{\theta}})}{\partial (\boldsymbol{\theta}_c)^T} - j\frac{\partial \widetilde{\mathbf{f}}(\underline{\boldsymbol{\theta}})}{\partial (\boldsymbol{\theta}_c^*)^T} \end{cases} \Leftrightarrow \begin{cases} \frac{\partial \widetilde{\mathbf{f}}(\underline{\boldsymbol{\theta}})}{\partial (\boldsymbol{\theta}_c^*)^T} = \frac{1}{2}\left(\frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \operatorname{Re}\{\boldsymbol{\theta}_c\}^T} + j\frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \operatorname{Im}\{\boldsymbol{\theta}_c\}^T}\right) \\ \frac{\partial \widetilde{\mathbf{f}}(\underline{\boldsymbol{\theta}})}{\partial (\boldsymbol{\theta}_c)^T} = \frac{1}{2}\left(\frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \operatorname{Re}\{\boldsymbol{\theta}_c\}^T} - j\frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \operatorname{Im}\{\boldsymbol{\theta}_c\}^T}\right) \end{cases}$$

- and, as a consequence, whatever $\mathbf{f}(\ )$ is real-valued or complex-valued:

$$\left(\frac{\partial \widetilde{\mathbf{f}}(\underline{\boldsymbol{\theta}})}{\partial (\boldsymbol{\theta}_c)^T}\right)^* = \frac{\partial \widetilde{\mathbf{f}}^*(\underline{\boldsymbol{\theta}})}{\partial (\boldsymbol{\theta}_c^*)^T}, \quad \left(\frac{\partial \widetilde{\mathbf{f}}^*(\underline{\boldsymbol{\theta}})}{\partial (\boldsymbol{\theta}_c)^T}\right)^* = \frac{\partial \widetilde{\mathbf{f}}(\underline{\boldsymbol{\theta}})}{\partial (\boldsymbol{\theta}_c^*)^T}$$

For sake of legibility, in the following $\widetilde{\mathbf{f}}(\ )$ will be simply denoted $\mathbf{f}(\ )$: if the input argument of $\mathbf{f}(\ )$ is of the form $\underline{\boldsymbol{\theta}}$ given by (1), then reference to the dual form of the function is implicit.

[1]The derivation of the CRB requires only the differentiability of some functions which do not need to be analytic [24].

B. On MSE lower Bounds as a Gram matrix minimization problem

Throughout the present paper, unless otherwise stated, $\mathbf{x}$ denotes the random observation vector of dimension N, Ω denotes the observations space and $L^2(\Omega)$ denotes the complex Hilbert space of square integrable functions over Ω. The probability density function (p.d.f.) of $\mathbf{x}$ is denoted $p(\mathbf{x};\boldsymbol{\theta})$ and depends on a vector of P real parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_P) \in \Theta$, where Θ denotes the parameter space. $p(\mathbf{x};\boldsymbol{\theta})$ is "regular" in the following sense:
- $\forall \boldsymbol{\theta} \in \Theta, \forall \mathbf{x} \in \Omega, p(\mathbf{x};\boldsymbol{\theta}) > 0$ for almost every $\mathbf{x}$ in the observation space Ω,
- $\forall \boldsymbol{\theta} \in \Theta, \forall \mathbf{x} \in \Omega, p(\mathbf{x};\boldsymbol{\theta})$ is continuous and differentiable with respect to $\boldsymbol{\theta}$,
- $p(\mathbf{x};\boldsymbol{\theta})$ does not incorporate any probability mass function[2].

Additionally, we assume that the observation vector $\mathbf{x}$ corresponds to a parametric observation model involving $P_r \geq 0$ real unknown parameters (delays, directions of arrival, ...) and $P_c \geq 0$ complex unknown parameters (spatial transfer functions components, complex amplitudes, ...) where $2P_c + P_r = P$, leading to a p.d.f. of the form:

$$p(\mathbf{x};\boldsymbol{\theta}), \quad \boldsymbol{\theta} = (\theta_1, \ldots, \theta_P)^T = \left(\mathrm{Re}\left\{\boldsymbol{\theta}_c^T\right\}, \mathrm{Im}\left\{\boldsymbol{\theta}_c^T\right\}, \boldsymbol{\theta}_r^T\right)^T \in \mathbb{R}^P, \quad \boldsymbol{\theta}_c \in \mathbb{C}^{P_c},\ \boldsymbol{\theta}_r \in \mathbb{R}^{P_r},\ 2P_c + P_r = P$$

Then, according to the previous section, the p.d.f. of $\mathbf{x}$ can be parameterized in a dual form:

$$p(\mathbf{x};\boldsymbol{\theta}), \quad \boldsymbol{\theta} = \left(\mathrm{Re}\left\{\boldsymbol{\theta}_c^T\right\}, \mathrm{Im}\left\{\boldsymbol{\theta}_c^T\right\}, \boldsymbol{\theta}_r^T\right)^T \in \mathbb{R}^P \tag{8}$$

$$p(\mathbf{x};\underline{\boldsymbol{\theta}}), \quad \underline{\boldsymbol{\theta}} = \left(\boldsymbol{\theta}_c^T, (\boldsymbol{\theta}_c^*)^T, \boldsymbol{\theta}_r^T\right)^T \in \mathbb{C}^{2P_c} \times \mathbb{R}^{P_r}, \quad 2P_c + P_r = P, \quad P_c \geq 0,\ P_r \geq 0 \tag{9}$$

In the following we will only consider the form (9) since it includes (8) when $P_c = 0$.
Let $\underline{\boldsymbol{\theta}}^0$ be a selected value of the parameter $\underline{\boldsymbol{\theta}}$, and $\widehat{\mathbf{g}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x})$ an estimator of $\mathbf{g}(\underline{\boldsymbol{\theta}}^0)$ where $\mathbf{g}(\underline{\boldsymbol{\theta}}) = (g_1(\underline{\boldsymbol{\theta}}), \ldots, g_{Q_c}(\underline{\boldsymbol{\theta}}), g_{Q_c+1}(\underline{\boldsymbol{\theta}}), \ldots, g_Q(\underline{\boldsymbol{\theta}}))^T$ is a vector of Q functions of $\underline{\boldsymbol{\theta}}$, the first Q_c ones being complex-valued functions, the last $Q_r = Q - Q_c$ being real-valued functions, where $Q_c \in [0, Q]$. For any selected value $\underline{\boldsymbol{\theta}}^0$, $\widehat{\mathbf{g}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x})$ stands for a mapping of the observation space Ω into an estimate of $\mathbf{g}(\underline{\boldsymbol{\theta}}^0)$, designed to have a low MSE at $\underline{\boldsymbol{\theta}}^0$ (possibly the lowest) and some relevant properties for other values of $\underline{\boldsymbol{\theta}}$, as unbiasedness for instance. Let us recall that $\widehat{\mathbf{g}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x})$ is an unbiased estimator of $\mathbf{g}(\underline{\boldsymbol{\theta}}^0)$ at the selected value $\underline{\boldsymbol{\theta}}^0$ if [7][10]:

$$E_{\underline{\boldsymbol{\theta}}^0}\left[\widehat{\mathbf{g}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x})\right] = \mathbf{g}(\underline{\boldsymbol{\theta}}^0) = \int_{\Omega} \widehat{\mathbf{g}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x})\, p(\mathbf{x};\underline{\boldsymbol{\theta}}^0)\, d\mathbf{x}, \tag{10}$$

where $E_{\underline{\boldsymbol{\theta}}}[\mathbf{g}(\mathbf{x})]$ is the statistical expectation of the vector of functions $\mathbf{g}(\ )$ with respect to $\mathbf{x}$ parameterized by $\underline{\boldsymbol{\theta}}$. Then $\left(\widehat{\mathbf{g}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x})\right)^*$ is as well an unbiased estimator of $\mathbf{g}^*(\underline{\boldsymbol{\theta}}^0)$ at $\underline{\boldsymbol{\theta}}^0$, i.e.,:

$$E_{\underline{\boldsymbol{\theta}}^0}\left[\left(\widehat{\mathbf{g}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x})\right)^*\right] = \mathbf{g}(\underline{\boldsymbol{\theta}}^0)^* \quad \Leftrightarrow \quad E_{\underline{\boldsymbol{\theta}}^0}\left[\underline{\widehat{\mathbf{g}(\underline{\boldsymbol{\theta}}^0)}}(\mathbf{x})\right] = \underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0) \tag{11}$$

and therefore (linearity of unbiasedness property [7][10]):

$$\begin{aligned}
\forall \boldsymbol{\delta} &\in \mathbb{C}^{2Q_c+Q_r},\ E_{\underline{\boldsymbol{\theta}}^0}\left[\boldsymbol{\delta}^T \underline{\widehat{\mathbf{g}(\underline{\boldsymbol{\theta}}^0)}}(\mathbf{x})\right] = \boldsymbol{\delta}^T \underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0) \qquad (12)\\
\underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0) &= \left(g_1(\underline{\boldsymbol{\theta}}^0), \ldots, g_{Q_c}(\underline{\boldsymbol{\theta}}^0), g_1^*(\underline{\boldsymbol{\theta}}^0), \ldots, g_{Q_c}^*(\underline{\boldsymbol{\theta}}^0), g_{Q_c+1}(\underline{\boldsymbol{\theta}}^0), \ldots, g_Q(\underline{\boldsymbol{\theta}}^0)\right)^T \\
\underline{\widehat{\mathbf{g}(\underline{\boldsymbol{\theta}}^0)}}(\mathbf{x}) &= \left(\widehat{g_1(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x}), \ldots, \widehat{g_{Q_c}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x}), \left(\widehat{g_1(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x})\right)^*, \ldots, \left(\widehat{g_{Q_c}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x})\right)^*,\right. \\
&\qquad \left.\widehat{g_{Q_c+1}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x}), \ldots, \widehat{g_Q(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x})\right)^T
\end{aligned}$$

Actually, if the exhaustive characterization - in the sense of statistical performance - of an estimator $\widehat{\mathbf{g}(\boldsymbol{\theta}^0)}(\mathbf{x})$ containing complex-valued components, is supposed to include the characterization of all its individual components, that is real and imaginary parts, then the characterization of $\underline{\widehat{\mathbf{g}(\underline{\boldsymbol{\theta}}^0)}}(\mathbf{x})$ is necessary as:

$$\begin{aligned}
\boldsymbol{\lambda}^T \mathrm{Re}\left\{\widehat{\mathbf{g}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x})\right\} &= \boldsymbol{\delta}^T \underline{\widehat{\mathbf{g}(\underline{\boldsymbol{\theta}}^0)}}(\mathbf{x}), \quad \boldsymbol{\delta} = \left(\tfrac{\lambda_1}{2}, \ldots, \tfrac{\lambda_{Q_c}}{2}, \tfrac{\lambda_1}{2}, \ldots, \tfrac{\lambda_{Q_c}}{2}, \lambda_{Q_c+1}, \ldots, \lambda_Q\right)^T \\
\boldsymbol{\mu}^T \mathrm{Im}\left\{\widehat{\mathbf{g}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x})\right\} &= \boldsymbol{\delta}^T \underline{\widehat{\mathbf{g}(\underline{\boldsymbol{\theta}}^0)}}(\mathbf{x}), \quad \boldsymbol{\delta} = \left(\tfrac{\mu_1}{2j}, \ldots, \tfrac{\mu_{Q_c}}{2j}, \tfrac{-\mu_1}{2j}, \ldots, \tfrac{-\mu_{Q_c}}{2j}, \mu_{Q_c+1}, \ldots, \mu_Q\right)^T
\end{aligned}$$

[2] The CRB has not originally been designed to cope with discrete distributions and we do not address this issue in the present paper. For discrete distributions, the lower bound used is usually the Hammersley-Chapman-Robbins bound [8].

Indeed, characterization of $\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x})$ only, as introduced in [21], prevents from having insight on the behaviour of individual components and leads to a partial characterization. Therefore the statistical performance of any estimator of $\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)$, unbiased at $\underline{\boldsymbol{\theta}}^0$, is fully characterized, in the MSE sense, by:

$$MSE_{\underline{\boldsymbol{\theta}}^0}\left[\boldsymbol{\delta}^T\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x})\right] = \int_{\Omega}\left|\boldsymbol{\delta}^T\left(\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x}) - \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)\right)\right|^2 p\left(\mathbf{x};\underline{\boldsymbol{\theta}}^0\right)d\mathbf{x}, \tag{13}$$

which is a norm deriving from an Hermitian product $\langle \mid \rangle_{\underline{\boldsymbol{\theta}}^0}$ depending on $p\left(\mathbf{x};\underline{\boldsymbol{\theta}}^0\right)$:

$$\begin{aligned} MSE_{\underline{\boldsymbol{\theta}}^0}\left[\boldsymbol{\delta}^T\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x})\right] &= \left\|\boldsymbol{\delta}^T\left(\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x}) - \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)\right)\right\|^2_{\underline{\boldsymbol{\theta}}^0} \\ \langle g(\mathbf{x}) \mid h(\mathbf{x})\rangle_{\underline{\boldsymbol{\theta}}^0} = E_{\underline{\boldsymbol{\theta}}^0}\left[g(\mathbf{x})h^*(\mathbf{x})\right] &= \int_{\Omega} g(\mathbf{x})h^*(\mathbf{x})p\left(\mathbf{x};\underline{\boldsymbol{\theta}}^0\right)d\mathbf{x} \end{aligned} \tag{14}$$

Finally, by (4):

$$\forall\boldsymbol{\delta}\in\mathbb{C}^{2Q_c+Q_r},\ MSE_{\underline{\boldsymbol{\theta}}^0}\left[\boldsymbol{\delta}^T\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x})\right] = \boldsymbol{\delta}^H\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x}) - \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)\right\}\right)\boldsymbol{\delta} \tag{15}$$

where:

$$\{\mathbf{h}(\mathbf{x})\} = \left\{\left(h_1(\mathbf{x}),\ldots,h_Q(\mathbf{x})\right)^T\right\} = \left\{h_1(\mathbf{x}),\ldots,h_Q(\mathbf{x})\right\} \tag{16}$$

denotes a family of vectors whose elements are the vector components, and:

$$\left(\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\{\mathbf{u}(\mathbf{x})\}_{[1,Q]},\{\mathbf{c}(\mathbf{x})\}_{[1,P]}\right)\right)_{p,q} = \langle\mathbf{u}_q(\mathbf{x})\mid\mathbf{c}_p(\mathbf{x})\rangle_{\underline{\boldsymbol{\theta}}^0} = E_{\underline{\boldsymbol{\theta}}^0}\left[\mathbf{u}_q(\mathbf{x})\mathbf{c}_p^*(\mathbf{x})\right]. \tag{17}$$

Expression (15) clearly shows that the statistical performance of any subset of components of $\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x})$ is fully characterized by the knowledge of $\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x}) - \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)\right\}\right)$. Hence the interest of finding a matrix $\mathbf{B}_{\underline{\boldsymbol{\theta}}^0}$ independent of $\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x})$ and able to lower bound expression (15) leading to:

$$MSE_{\underline{\boldsymbol{\theta}}^0}\left[\boldsymbol{\delta}^T\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x})\right] \geq \boldsymbol{\delta}^H\mathbf{B}_{\underline{\boldsymbol{\theta}}^0}\boldsymbol{\delta} \quad\Leftrightarrow\quad \mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x}) - \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)\right\}\right) \geq \mathbf{B}_{\underline{\boldsymbol{\theta}}^0}$$

In the search for a lower bound on the MSE, the property (14) allows the use of two equivalent fundamental results: the generalization of the Cauchy-Schwartz inequality to Gram matrices (generally referred to as the "covariance inequality" [22][30]) and the solution of the minimization of a Gram matrix (with respect to the Löwner ordering [25, §7.7]) under linear constraints [7][10][11][31] as defined in the following lemma:

Lemma 1: Let $\mathbb{U}$ be an Hermitian vector space of any dimension (finite or infinite) on $\mathbb{C}$. Let $\{\mathbf{c}\}_{[1,K]}$ be a given family of K vectors of $\mathbb{U}$ such that $\mathbf{G}\left(\{\mathbf{c}\}_{[1,K]}\right) = \mathbf{MDM}^H$, $\mathbf{M}\in\mathcal{M}_{\mathbb{C}}\left(K,\widetilde{K}\right)$, $\mathbf{M}^H\mathbf{M} = \mathbf{I}_{\widetilde{K}}$, $\mathbf{D} = \mathbf{Diag}\left(d_1,\ldots,d_{\widetilde{K}}\right)$, $d_1\geq\ldots\geq d_{\widetilde{K}}>0$, $\widetilde{K}\leq K$. Let $\{\mathbf{u}\}_{[1,Q]}$ be an unknown family of Q vectors of $\mathbb{U}$. Then the solution of the minimization of $\mathbf{G}\left(\{\mathbf{u}\}_{[1,Q]}\right)$ with respect to the Löwner ordering under the set of linear constraints $\mathbf{G}\left(\{\mathbf{u}\}_{[1,Q]},\{\mathbf{c}\}_{[1,K]}\right) = \mathbf{V}$, i.e.:

$$\min_{\{\mathbf{u}\}_{[1,Q]}}\left\{\mathbf{G}\left(\{\mathbf{u}\}_{[1,Q]}\right)\right\} \text{ under } \mathbf{G}\left(\{\mathbf{u}\}_{[1,Q]},\{\mathbf{c}\}_{[1,K]}\right) = \mathbf{V}\in\mathcal{M}_{\mathbb{C}}(K,Q)$$

is:

$$\min_{\{\mathbf{u}\}_{[1,Q]}}\left\{\mathbf{G}\left(\{\mathbf{u}\}_{[1,Q]}\right)\right\} = \left(\mathbf{M}^H\mathbf{V}\right)^H\mathbf{D}^{-1}\left(\mathbf{M}^H\mathbf{V}\right),\ \mathbf{u}_q = \sum_{k=1}^{K}(\mathbf{A})_{k,q}\mathbf{c}_k,\ \mathbf{A} = \mathbf{MD}^{-1}\mathbf{M}^H\mathbf{V} \tag{18}$$

if and only if $\mathbf{V}$ verifies $\left(\mathbf{MM}^H\right)\mathbf{V} = \mathbf{V}$, i.e. if and only if $\mathbf{V}\in Span\left\{\mathbf{G}\left(\{\mathbf{c}\}_{[1,K]}\right)\right\}$.

If $\mathbf{G}\left(\{\mathbf{c}\}_{[1,K]}\right)$ is a full rank matrix then (18) reduces to:

$$\min_{\{\mathbf{u}\}_{[1,Q]}}\left\{\mathbf{G}\left(\{\mathbf{u}\}_{[1,Q]}\right)\right\} = \mathbf{V}^H\mathbf{G}\left(\{\mathbf{c}\}_{[1,K]}\right)^{-1}\mathbf{V}$$

We recommend the use of lemma 1 where $\{\mathbf{u}\}_{[1,Q]} = \left\{\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x}) - \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)\right\}$, as it provides a better understanding of the hypotheses associated with the Cramér-Rao (and other) lower bound on the MSE [10][11][31].
In the rest of the paper, the minimization of a Gram matrix is always with respect to the Löwner ordering, and for sake of legibility, we will simply write $\min\left\{\mathbf{G}\left(\{\mathbf{u}\}_{[1,Q]}\right)\right\}$ to denote $\min\limits_{\{\mathbf{u}\}_{[1,Q]}}\left\{\mathbf{G}\left(\{\mathbf{u}\}_{[1,Q]}\right)\right\}$.

Under the constraint (10) or (12), the minimization of $\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x}) - \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)\right\}\right)$ is trivial and is given by:

$$\mathbf{B}_{\underline{\boldsymbol{\theta}}^0} = \mathbf{0}, \quad \widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x}) = \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)$$

which provides both a trivial bound ($\mathbf{0}$) and an unacceptable solution, because it not only requires *a priori* knowledge of selected value of the parameter $\left(\underline{\boldsymbol{\theta}}^0\right)$, but is also independent of the observations. To render the lower bound dependent on the observations, all that is required is to define a constraint that is not satisfied by the trivial solution, and compatible with uniform unbiasedness [7]:

$$E_{\underline{\boldsymbol{\theta}}}\left[\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x})\right] = \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}\right), \forall \underline{\boldsymbol{\theta}} \in \Theta \tag{19}$$

which is unbiasedness in a wide sense (for all possible value of $\underline{\boldsymbol{\theta}}^0$). Unfortunately, it is almost always impossible to find an analytical solution of the minimization of $\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x}) - \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)\right\}\right)$ under the Barankin constraint (19)[7][10]. Nevertheless, if $\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}\right)$ is differentiable at $\underline{\boldsymbol{\theta}}^0$, it is possible to find an analytical solution of the minimization of $\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x}) - \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)\right\}\right)$ under a locally restricted version of (19):

$$E_{\underline{\boldsymbol{\theta}}^0 + d\underline{\boldsymbol{\theta}}}\left[\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x})\right] = \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0 + d\underline{\boldsymbol{\theta}}\right) + \mathbf{o}\left(d\underline{\boldsymbol{\theta}}\right) \tag{20}$$

meaning that, up to the first order and in the neighbourhood of $\underline{\boldsymbol{\theta}}^0$, $\widehat{g_q\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x})$ remains an unbiased estimator of $g_q\left(\underline{\boldsymbol{\theta}}^0\right)$ independently of a - small - variation of $\underline{\boldsymbol{\theta}}$. It is quite straightforward to understand that local unbiasedness is the "weakest" constraint deriving from uniform unbiasedness (19) not leading to the trivial solution. Indeed, if $\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}\right)$ is differentiable at $\underline{\boldsymbol{\theta}}^0$, (20) can be rewritten in terms of Taylor expansion of each side:

$$E_{\underline{\boldsymbol{\theta}}^0}\left[\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x})\right] + \left(\int_{\Omega} \widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x}) \frac{\partial p\left(\mathbf{x}; \underline{\boldsymbol{\theta}}^0\right)}{\partial \underline{\boldsymbol{\theta}}^T} d\mathbf{x}\right) d\underline{\boldsymbol{\theta}} + \mathbf{o}\left(d\underline{\boldsymbol{\theta}}\right) = \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right) + \frac{\partial \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}{\partial \underline{\boldsymbol{\theta}}^T} d\underline{\boldsymbol{\theta}} + \mathbf{o}\left(d\underline{\boldsymbol{\theta}}\right)$$

leading to the constraint (uniqueness of Taylor expansion):

$$\int_{\Omega} \widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x}) \frac{\partial p\left(\mathbf{x}; \underline{\boldsymbol{\theta}}^0\right)}{\partial \underline{\boldsymbol{\theta}}^T} d\mathbf{x} = \frac{\partial \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}{\partial \underline{\boldsymbol{\theta}}^T} \quad \Leftrightarrow \quad E_{\underline{\boldsymbol{\theta}}^0}\left[\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x}) \frac{\partial \ln p\left(\mathbf{x}; \underline{\boldsymbol{\theta}}^0\right)}{\partial \underline{\boldsymbol{\theta}}^T}\right] = \frac{\partial \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}{\partial \underline{\boldsymbol{\theta}}^T}$$

equivalent to:

$$E_{\underline{\boldsymbol{\theta}}^0}\left[\left(\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x}) - \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)\right) \frac{\partial \ln p\left(\mathbf{x}; \underline{\boldsymbol{\theta}}^0\right)}{\partial \underline{\boldsymbol{\theta}}^T}\right] = \frac{\partial \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}{\partial \underline{\boldsymbol{\theta}}^T} \tag{21}$$

since:

$$E_{\underline{\boldsymbol{\theta}}}\left[\frac{\partial \ln p\left(\mathbf{x}; \underline{\boldsymbol{\theta}}\right)}{\partial \underline{\boldsymbol{\theta}}^T}\right] = \int_{\Omega} \frac{\partial p\left(\mathbf{x}; \underline{\boldsymbol{\theta}}\right)}{\partial \underline{\boldsymbol{\theta}}^T} d\mathbf{x} = \mathbf{0}. \tag{22}$$

Therefore, the problem of finding a lower bound of $\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x})-\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)\right\}\right)$ for locally unbiased estimators is actually the problem of minimizing a Gram matrix under a set of linear constraints (lemma 1):

$$\min\left\{\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\{\mathbf{u}\}_{[1,Q]}\right)\right\} \text{ under (11) and (21)}$$
$$\Updownarrow$$
$$\min\left\{\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\{\mathbf{u}\}_{[1,Q]}\right)\right\} \text{ under } \mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\{\mathbf{u}\}_{[1,Q]},\{\mathbf{c}\}_{[1,P+1]}\right)=\begin{pmatrix}\left(\frac{\partial\underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)}{\partial\underline{\boldsymbol{\theta}}^T}\right)^T\\ \mathbf{0}^T\end{pmatrix} \tag{23}$$
$$\{\mathbf{u}\}_{[1,Q]}=\left\{\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x})-\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)\right\},\quad \{\mathbf{c}\}_{[1,P+1]}=\left\{\left\{\left(\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\theta}}\right)}{\partial\underline{\boldsymbol{\theta}}}\right)^*\right\},1\left(\mathbf{x}\right)\right\}$$

The solution of (23) is easily obtained in two steps by resorting to lemma 1 and to the following lemma [32]:

Lemma 2: Let $\mathbb{U}$ be an Hermitian vector space of any dimension (finite or infinite) on $\mathbb{C}$. Let $\{\mathbf{c}\}_{[1,K+1]}$ be a given family of $K+1$ independent vectors of $\mathbb{U}$ and let $\widetilde{\mathbf{V}}$ be a given matrix of $\mathcal{M}_{\mathbb{C}}(K,Q)$. Then the problem of the minimization of $\mathbf{G}\left(\{\mathbf{u}\}_{[1,N]}\right)$, where $\{\mathbf{u}\}_{[1,N]}$ is an unknown family of N vectors of $\mathbb{U}$, verifies the following equivalence:

$$\min\left\{\mathbf{G}\left(\{\mathbf{u}\}_{[1,N]}\right)\right\} \text{ under } \begin{cases}\mathbf{G}\left(\{\mathbf{u}\}_{[1,N]},\{\mathbf{c}\}_{[1,K+1]}\right)=\begin{bmatrix}\widetilde{\mathbf{V}}\\ \mathbf{0}^T\end{bmatrix}\\ \mathbf{G}\left(\{\mathbf{c}\}_{[1,K]},\mathbf{c}_{K+1}\right)=\mathbf{0}^T\end{cases}$$
$$\Updownarrow \tag{24}$$
$$\min\left\{\mathbf{G}\left(\{\mathbf{u}\}_{[1,N]}\right)\right\} \text{ under } \mathbf{G}\left(\{\mathbf{u}\}_{[1,N]},\{\mathbf{c}\}_{[1,K]}\right)=\widetilde{\mathbf{V}}$$

The 2 previous fundamental lemmas can be completed with a third useful lemma [32]:

Lemma 3: Let $\mathbb{U}$ be an Hermitian vector space of any dimension (finite or infinite) on $\mathbb{C}$. Let $\{\mathbf{c}\}_{[1,K]}$ be a given family of K independent vectors of $\mathbb{U}$ and let $\mathbf{V}$ be a given matrix of $\mathcal{M}_{\mathbb{C}}(K,Q)$. Then the problem of the minimization of $\mathbf{G}\left(\{\mathbf{u}\}_{[1,Q]}\right)$, where $\{\mathbf{u}\}_{[1,Q]}$ is an unknown family of Q vectors of $\mathbb{U}$, verifies the following equivalence:

$$\min\left\{\mathbf{G}\left(\{\mathbf{u}\}_{[1,Q]}\right)\right\} \text{ under } \mathbf{G}\left(\{\mathbf{u}\}_{[1,Q]},\{\mathbf{c}\}_{[1,K]}\right)=\mathbf{V}\in\mathcal{M}_{\mathbb{C}}(K,Q)$$
$$\Updownarrow$$
$$\min\left\{\mathbf{G}\left(\{\mathbf{u}\}_{[1,Q]}\right)\right\} \text{ under } \begin{cases}\mathbf{T}^H\mathbf{G}\left(\{\mathbf{u}\}_{[1,Q]},\{\mathbf{c}\}_{[1,K]}\right)=\mathbf{T}^H\mathbf{V}\in\mathcal{M}_{\mathbb{C}}(K,Q)\\ \mathbf{T}\in\mathcal{M}_{\mathbb{C}}(K,K),\quad|\mathbf{T}|\neq 0\end{cases}\quad, \tag{25}$$
$$\Updownarrow$$
$$\min\left\{\mathbf{G}\left(\{\mathbf{u}\}_{[1,Q]}\right)\right\} \text{ under } \begin{cases}\mathbf{G}\left(\{\mathbf{u}\}_{[1,Q]},\{\mathbf{c}'\}_{[1,K]}\right)=\mathbf{T}^H\mathbf{V}\in\mathcal{M}_{\mathbb{C}}(K,Q)\\ \mathbf{c}'_k=\sum_{l=1}^{K}T_{kl}\mathbf{c}_l,\quad \mathbf{T}\in\mathcal{M}_{\mathbb{C}}(K,K),\quad|\mathbf{T}|\neq 0\end{cases}$$

i.e., the minimization problem is invariant by bijective transformation of the K linear constraints.

From (22), $\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\{\mathbf{c}\}_{[1,P]},\mathbf{c}_{P+1}\right)=\mathbf{0}^T$. Then, according to lemma 2, (23) is equivalent to:

$$\min\left\{\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\{\mathbf{u}\}_{[1,Q]}\right)\right\} \text{ under } \mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\{\mathbf{u}\}_{[1,Q]},\{\mathbf{c}\}_{[1,P]}\right)=\left(\frac{\partial\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}{\partial\underline{\boldsymbol{\theta}}^T}\right)^T \tag{26}$$
$$\{\mathbf{u}\}_{[1,Q]}=\left\{\widehat{\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}(\mathbf{x})-\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)\right\},\quad \{\mathbf{c}\}_{[1,P]}=\left\{\left(\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\theta}}\right)}{\partial\underline{\boldsymbol{\theta}}}\right)^*\right\}$$

whose solution is provided by lemma 1:

$$\min\left\{\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\underline{\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}}\left(\mathbf{x}\right)-\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)\right\}\right)\right\}=\left(\frac{\partial\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)^T}{\partial\underline{\boldsymbol{\theta}}}\right)^H\mathbf{F}_{\underline{\boldsymbol{\theta}}^0}^{-1}\left(\frac{\partial\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)^T}{\partial\underline{\boldsymbol{\theta}}}\right) \tag{27}$$

$$\mathbf{F}_{\underline{\boldsymbol{\theta}}^0}=\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\left(\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\theta}}\right)}{\partial\underline{\boldsymbol{\theta}}}\right)^*\right\}\right)=E_{\underline{\boldsymbol{\theta}}^0}\left[\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\theta}}\right)}{\partial\underline{\boldsymbol{\theta}}}\left(\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\theta}}\right)}{\partial\underline{\boldsymbol{\theta}}}\right)^H\right]$$

$$\left(\underline{\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}}\left(\mathbf{x}\right)-\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)\right)_{\text{eff}}^T=\left(\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\theta}}\right)}{\partial\underline{\boldsymbol{\theta}}}\right)^H\mathbf{F}_{\underline{\boldsymbol{\theta}}^0}^{-1}\left(\frac{\partial\underline{\mathbf{g}}^T\left(\underline{\boldsymbol{\theta}}^0\right)}{\partial\underline{\boldsymbol{\theta}}}\right) \tag{28}$$

provided that $\mathbf{F}_{\underline{\boldsymbol{\theta}}^0}=\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\left(\frac{\partial p(\mathbf{x};\underline{\boldsymbol{\theta}})}{\partial\underline{\boldsymbol{\theta}}}\right)^*\right\}\right)$ is invertible, that is provided that $\left\{\frac{\partial p(\mathbf{x};\underline{\boldsymbol{\theta}})}{\partial\underline{\boldsymbol{\theta}}}\right\}$ is an independent family of vectors. Maybe unexpectedly, it appears that the definition of the Hermitian product used in books of mathematics (2) leads to a Fisher Information Matrix (FIM) $\mathbf{F}_{\underline{\boldsymbol{\theta}}^0}$ that is not only a Gram matrix (an expected result) but also the correlation matrix of vector $\frac{\partial p(\mathbf{x};\underline{\boldsymbol{\theta}})}{\partial\underline{\boldsymbol{\theta}}}$, in "signal processing" notation. Let us note that (27) is a generalization of the CRB derived in [21, (11)(13)]. Indeed, the CRB derived in [21] does not allow for an exhaustive MSE characterization of any component of $\widehat{\mathbf{g}\left(\underline{\boldsymbol{\theta}}^0\right)}\left(\mathbf{x}\right)$. Last, an important limitation of the CRB (and of any lower bound deriving from the Barankin rationale) is its inability to take into account the support of the parameters. This major drawback derives from the expression of the efficient estimator (the one reaching the CRB) given in (28) which clearly does not incorporate any information relative to the parameter support. It is the reason why the CRB must be used only for interior points of the parameter domain and never for boundary points, where it is uninformative. To fill this lack, when the parameters are assumed to be random, other bounds have been derived: the so-called Bayesian bounds. These bounds take into account the support of the parameters throughout an a priori probability density function [1]. However they are pointless when the parameters are deterministic.

C. Corollaries

There are some corollaries on the derivation of the CRB that are completely straightforward with the used approach, condensing previous derivations (mostly restricted to real parameters) to a few lines with a wider scope.

1) On FIM singularity (preliminary results):

One of the key feature of the used approach is that all the information required to solve the problem is contained in the constraints (see lemma 1):

$$\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\{\mathbf{u}\}_{[1,Q]},\{\mathbf{c}\}_{[1,P]}\right)=\mathbf{V}\in\mathcal{M}_{\mathbb{C}}\left(P,Q\right) \tag{29}$$

that is, in the case of CRB (26):

$$\mathbf{V}=\left(\frac{\partial\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)}{\partial\underline{\boldsymbol{\theta}}^T}\right)^T,\ \ \{\mathbf{u}\}_{[1,Q]}=\left\{\underline{\widehat{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}}\left(\mathbf{x}\right)-\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)\right\},\ \ \{\mathbf{c}\}_{[1,P]}=\left\{\left(\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\theta}}\right)}{\partial\underline{\boldsymbol{\theta}}}\right)^*\right\}$$

Therefore, as the solution of the minimization of $\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\{\mathbf{u}\}_{[1,Q]}\right)$ must satisfy the linear constraints (29), the first question that naturally arises is to derive the condition of existence of such a solution. A classical result in algebra (used in lemma 1) is that (29) admits a solution:

- $\forall\mathbf{V}$ if $\mathbf{F}_{\underline{\boldsymbol{\theta}}^0}=\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\{\mathbf{c}\}_{[1,P]}\right)$ is invertible, i.e., if $\{\mathbf{c}\}_{[1,P]}$ is a family of independent vectors,
- for a restricted subset of $\mathbf{V}$ if $\mathbf{F}_{\underline{\boldsymbol{\theta}}^0}$ is singular, i.e., if $\{\mathbf{c}\}_{[1,P]}$ is a family of dependent vectors.

Therefore a singular FIM matrix corresponds simply to the case where the family of vectors $\left\{\frac{\partial p(\mathbf{x};\underline{\boldsymbol{\theta}})}{\partial\underline{\boldsymbol{\theta}}}\right\}$ is dependent. In that case, only a subset of constraint matrices $\mathbf{V}$ are compatible with the "hidden" linear relation between the vectors $\left\{\frac{\partial p(\mathbf{x};\underline{\boldsymbol{\theta}})}{\partial\underline{\boldsymbol{\theta}}}\right\}$.

The nature of the restriction is given by lemma 1: if $\mathbf{F}_{\underline{\boldsymbol{\theta}}^0}=\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\{\mathbf{c}\}_{[1,P]}\right)=\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}\mathbf{D}_{\underline{\boldsymbol{\theta}}^0}\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}^H$, $\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}\in\mathcal{M}_{\mathbb{C}}\left(P,\widetilde{P}\right)$, $\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}^H\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}=\mathbf{I}_{\widetilde{P}}$, $\mathbf{D}_{\underline{\boldsymbol{\theta}}^0}=\mathbf{Diag}\left(d_1\left(\underline{\boldsymbol{\theta}}^0\right),\ldots,d_{\widetilde{P}}\left(\underline{\boldsymbol{\theta}}^0\right)\right)$, $d_1\left(\underline{\boldsymbol{\theta}}^0\right)\geq\ldots\geq d_{\widetilde{P}}\left(\underline{\boldsymbol{\theta}}^0\right)>0$, $\widetilde{P}<P$, then $\mathbf{V}$ must satisfy

$\left(\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}{}^H\right)\mathbf{V}=\mathbf{V}$, that is $\mathbf{V}\in Span\left\{\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}\right\}$. It becomes obvious that if $\underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}\right)=\underline{\boldsymbol{\theta}}$, then $\mathbf{V}=\left(\frac{\partial\underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)}{\partial\underline{\boldsymbol{\theta}}^T}\right)^T=\mathbf{I}_P$ and the locally unbiased constraint defined by:

$$\left(\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}{}^H\right)\mathbf{V}=\left(\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}{}^H\right)\mathbf{I}_P=\mathbf{I}_P\Leftrightarrow\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}{}^H=\mathbf{I}_P$$

can not be satisfied. Therefore no locally unbiased estimator of $\underline{\boldsymbol{\theta}}$ at $\underline{\boldsymbol{\theta}}^0$can exist if the FIM is singular.
An additional insight is obtained if we consider the function:

$$h_{\underline{\boldsymbol{\theta}}^0}\left(\underline{\boldsymbol{\theta}}\right)=E_{\underline{\boldsymbol{\theta}}^0}\left[\left(\frac{p\left(\mathbf{x};\underline{\boldsymbol{\theta}}\right)}{p\left(\mathbf{x};\underline{\boldsymbol{\theta}}^0\right)}-1\right)^2\right]\geq 0$$

Then, at the vicinity of $\underline{\boldsymbol{\theta}}^0$:

$$h_{\underline{\boldsymbol{\theta}}^0}\left(\underline{\boldsymbol{\theta}}^0+d\underline{\boldsymbol{\theta}}\right)=d\underline{\boldsymbol{\theta}}^H\mathbf{F}_{\underline{\boldsymbol{\theta}}^0}d\underline{\boldsymbol{\theta}}+o\left(\left\|d\underline{\boldsymbol{\theta}}\right\|^2\right),$$

and $h_{\underline{\boldsymbol{\theta}}^0}\left(\underline{\boldsymbol{\theta}}\right)$ reaches the minimum value 0 at $\underline{\boldsymbol{\theta}}=\underline{\boldsymbol{\theta}}^0$. Therefore if $\mathbf{F}_{\underline{\boldsymbol{\theta}}^0}$ is non singular, $h_{\underline{\boldsymbol{\theta}}^0}\left(\underline{\boldsymbol{\theta}}\right)$ has a strict extremum at the vicinity of $\underline{\boldsymbol{\theta}}^0$. On the opposite, if $\mathbf{F}_{\underline{\boldsymbol{\theta}}^0}$ is singular, then the extremum is not strict at the vicinity of $\underline{\boldsymbol{\theta}}^0$ and therefore :

$$\exists d\underline{\boldsymbol{\theta}}\ /\ h_{\underline{\boldsymbol{\theta}}^0}\left(\underline{\boldsymbol{\theta}}^0+d\underline{\boldsymbol{\theta}}\right)=0\ \Leftrightarrow\ \exists d\underline{\boldsymbol{\theta}}\ /\ p\left(\mathbf{x};\underline{\boldsymbol{\theta}}^0+d\underline{\boldsymbol{\theta}}\right)=p\left(\mathbf{x};\underline{\boldsymbol{\theta}}^0\right)\ \text{(almost everywhere)}$$

what implies that:

$$\forall\widehat{\underline{\boldsymbol{\theta}}^0}\left(\mathbf{x}\right),\quad\exists d\underline{\boldsymbol{\theta}}\ /\ E_{\underline{\boldsymbol{\theta}}^0}\left[\widehat{\underline{\boldsymbol{\theta}}^0}\left(\mathbf{x}\right)\right]=E_{\underline{\boldsymbol{\theta}}^0+d\underline{\boldsymbol{\theta}}}\left[\widehat{\underline{\boldsymbol{\theta}}^0}\left(\mathbf{x}\right)\right]$$

which is impossible for a locally unbiased estimator of $\underline{\boldsymbol{\theta}}^0$. Nevertheless according to lemma 1, there may exist $\mathbf{g}\left(\underline{\boldsymbol{\theta}}\right)$ that have locally unbiased estimates at $\underline{\boldsymbol{\theta}}^0$ even if the FIM is singular, provided that $\left(\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}^H\right)\frac{\partial\mathbf{g}(\underline{\boldsymbol{\theta}}^0)^T}{\partial\underline{\boldsymbol{\theta}}}=\frac{\partial\mathbf{g}(\underline{\boldsymbol{\theta}}^0)^T}{\partial\underline{\boldsymbol{\theta}}}$, which requires that $\mathbf{g}\left(\underline{\boldsymbol{\theta}}\right)$ have at most $\widetilde{P}$ components. As a consequence, a singular FIM at $\underline{\boldsymbol{\theta}}^0$ generates a CRB singularity at $\underline{\boldsymbol{\theta}}^0$ in the sense that the CRB is unbounded at the vicinity of $\underline{\boldsymbol{\theta}}^0$. Indeed, if the FIM is non singular, then the CRB is bounded by:

$$\forall g\left(\underline{\boldsymbol{\theta}}\right),\ \frac{\partial g\left(\underline{\boldsymbol{\theta}}^0\right)}{\partial\underline{\boldsymbol{\theta}}}^H\mathbf{F}_{\underline{\boldsymbol{\theta}}^0}^{-1}\frac{\partial g\left(\underline{\boldsymbol{\theta}}^0\right)}{\partial\underline{\boldsymbol{\theta}}}\leq\frac{1}{d_P\left(\underline{\boldsymbol{\theta}}^0\right)}\left\|\frac{\partial g\left(\underline{\boldsymbol{\theta}}^0\right)}{\partial\underline{\boldsymbol{\theta}}}\right\|^2$$

where $d_P\left(\underline{\boldsymbol{\theta}}^0\right)>0$ is the smallest eigenvalue of the FIM $\mathbf{F}_{\underline{\boldsymbol{\theta}}^0}$. Therefore it is not thoroughly accurate to say that a singular FIM at $\underline{\boldsymbol{\theta}}^0$ leads to locally unbiased estimates with infinite variance as mentioned in [19][33]. A preliminary correct assertion is that a singular FIM at $\underline{\boldsymbol{\theta}}^0$ leads to a CRB singularity at $\underline{\boldsymbol{\theta}}^0$ which results in an unbounded CRB at the vicinity of $\underline{\boldsymbol{\theta}}^0$ and the non existence of locally unbiased estimates for a subspace of functions $\mathbf{g}\left(\underline{\boldsymbol{\theta}}\right)$, including $\underline{\boldsymbol{\theta}}$. A more precise result will be given in §VI-D.

2) On constrained FIM:

The addition of given constraints on the unknown deterministic parameters leads to the constrained FIM and constrained CRB [15][17][18][20][33], whose derivation is straightforward by resorting to the present approach. Let us consider the subset $\mathcal{C}$ of Θ described by a set of K equality constraints:

$$\mathbf{f}\left(\boldsymbol{\theta}\right)=\mathbf{f}\left(\underline{\boldsymbol{\theta}}\right)=\mathbf{0}\in\mathbb{C}^{K_c}\times\mathbb{R}^{K_r},\ 2K_c+K_r=K,\ 1\leq K\leq P-1 \tag{30}$$

where the matrix $\frac{\partial}{\partial\underline{\boldsymbol{\theta}}^T}\begin{pmatrix}\text{Re}\left\{\mathbf{f}\left(\underline{\boldsymbol{\theta}}\right)\right\}\\ \text{Im}\left\{\mathbf{f}\left(\underline{\boldsymbol{\theta}}\right)\right\}\end{pmatrix}\in\mathcal{M}_{\mathbb{C}}\left(K,P\right)$ has full row rank (K), which is equivalent to requiring that the constraints are not redundant [15][33]. Thus, with some manipulation:

$$\left\{\begin{array}{c}\underline{\boldsymbol{\theta}}^0\in\mathcal{C}\\ \underline{\boldsymbol{\theta}}^0+d\underline{\boldsymbol{\theta}}\in\mathcal{C}\end{array}\right.\Leftrightarrow\left\{\begin{array}{c}\underline{\mathbf{f}}\left(\underline{\boldsymbol{\theta}}^0\right)=\mathbf{0}\\ \frac{\partial\underline{\mathbf{f}}(\underline{\boldsymbol{\theta}}^0)}{\partial\underline{\boldsymbol{\theta}}^T}d\underline{\boldsymbol{\theta}}=\mathbf{0}\end{array}\right.\Leftrightarrow\left\{\begin{array}{l}\underline{\mathbf{f}}\left(\underline{\boldsymbol{\theta}}^0\right)=\mathbf{0}\\ d\underline{\boldsymbol{\theta}}\in\ker\left\{\frac{\partial\underline{\mathbf{f}}(\underline{\boldsymbol{\theta}}^0)}{\partial\underline{\boldsymbol{\theta}}^T}\right\}\end{array}\right.\Leftrightarrow\left\{\begin{array}{l}\underline{\mathbf{f}}\left(\underline{\boldsymbol{\theta}}^0\right)=\mathbf{0}\\ d\underline{\boldsymbol{\theta}}=\mathbf{U}_{\underline{\boldsymbol{\theta}}^0}d\boldsymbol{\lambda},\ d\boldsymbol{\lambda}\in\mathbb{R}^{P-K}\end{array}\right.$$

where $\mathbf{U}_{\underline{\boldsymbol{\theta}}^0} \in \mathcal{M}_{\mathbb{C}}(P, P-K)$ is a basis of $\ker\left\{\frac{\partial \mathbf{f}(\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\theta}}^T}\right\}$. Therefore a locally unbiased estimate of $\mathbf{g}(\underline{\boldsymbol{\theta}}^0)$ is now required to be locally unbiased only on $\mathcal{C}$, what means at any selected value $\underline{\boldsymbol{\theta}}^0 \in \mathcal{C}$ and locally at the vicinity of $\underline{\boldsymbol{\theta}}^0 \in \mathcal{C}$, that is for any $d\underline{\boldsymbol{\theta}} = \mathbf{U}_{\underline{\boldsymbol{\theta}}^0} d\boldsymbol{\lambda}$:

$$\left\{\begin{array}{l} E_{\underline{\boldsymbol{\theta}}^0}\left[\left(\widehat{\underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x}) - \underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)\right)\right] = \mathbf{0} \\ E_{\underline{\boldsymbol{\theta}}^0}\left[\left(\widehat{\underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x}) - \underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)\right)\frac{\partial \ln p(\mathbf{x};\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\theta}}^T}\right] d\underline{\boldsymbol{\theta}} = \frac{\partial \underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\theta}}^T} d\underline{\boldsymbol{\theta}} \end{array}\right., \ \forall \underline{\boldsymbol{\theta}}^0 \in \mathcal{C}, \ \forall d\underline{\boldsymbol{\theta}} = \mathbf{U}_{\underline{\boldsymbol{\theta}}^0} d\boldsymbol{\lambda}, \ d\boldsymbol{\lambda} \in \mathbb{R}^{P-K},$$

which is equivalent to:

$$\left\{\begin{array}{l} E_{\underline{\boldsymbol{\theta}}^0}\left[\left(\widehat{\underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x}) - \underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)\right)\right] = \mathbf{0} \\ E_{\underline{\boldsymbol{\theta}}^0}\left[\left(\widehat{\underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x}) - \underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)\right)\frac{\partial \ln p(\mathbf{x};\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\theta}}^T}\mathbf{U}_{\underline{\boldsymbol{\theta}}^0}\right] = \left(\frac{\partial \underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\theta}}^T}\mathbf{U}_{\underline{\boldsymbol{\theta}}^0}\right) \end{array}\right., \ \forall \underline{\boldsymbol{\theta}}^0 \in \mathcal{C}. \tag{31}$$

The solution of the minimization of $\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\widehat{\underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x}) - \underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)\right\}\right)$ under linear constraints (31) is of the same form as the one previously derived for (23) except that:

$$\left.\begin{array}{l} \mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\{\mathbf{u}\}_{[1,Q]}, \{\mathbf{c}\}_{[1,P]}\right) = \left(\frac{\partial \underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\theta}}^T}\right)^T \\ \{\mathbf{c}\}_{[1,P]} = \left\{\left(\frac{\partial \ln p(\mathbf{x};\underline{\boldsymbol{\theta}})}{\partial \underline{\boldsymbol{\theta}}}\right)^*\right\} \end{array}\right\} \text{becomes} \left\{\begin{array}{l} \mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\{\mathbf{u}\}_{[1,Q]}, \{\mathbf{c}\}_{[1,P-K]}\right) = \mathbf{U}_{\underline{\boldsymbol{\theta}}^0}^T\left(\frac{\partial \underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\theta}}^T}\right)^T \\ \{\mathbf{c}\}_{[1,P-K]} = \left\{\mathbf{U}_{\underline{\boldsymbol{\theta}}^0}^H\left(\frac{\partial \ln p(\mathbf{x};\underline{\boldsymbol{\theta}})}{\partial \underline{\boldsymbol{\theta}}}\right)^*\right\} \end{array}\right. \tag{32}$$

Then, if $\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\{\mathbf{c}\}_{[1,P-K]}\right)$ is non singular:

$$\min\left\{\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\widehat{\underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x}) - \underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)\right\}\right)\right\} = \left(\frac{\partial \underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)^T}{\partial \underline{\boldsymbol{\theta}}}\right)^H \mathbf{U}_{\underline{\boldsymbol{\theta}}^0}^* \left(\mathbf{F}_{\underline{\boldsymbol{\theta}}^0}^c\right)^{-1} \mathbf{U}_{\underline{\boldsymbol{\theta}}^0}^T \left(\frac{\partial \underline{\mathbf{g}}^T(\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\theta}}}\right) \tag{33}$$

$$\mathbf{F}_{\underline{\boldsymbol{\theta}}^0}^c = \mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\mathbf{U}_{\underline{\boldsymbol{\theta}}^0}^H\left(\frac{\partial \ln p(\mathbf{x};\underline{\boldsymbol{\theta}})}{\partial \underline{\boldsymbol{\theta}}}\right)^*\right\}\right) = \mathbf{U}_{\underline{\boldsymbol{\theta}}^0}^T \mathbf{F}_{\underline{\boldsymbol{\theta}}^0} \mathbf{U}_{\underline{\boldsymbol{\theta}}^0}^*$$

$$\left(\widehat{\underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x}) - \underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)\right)_{\text{eff}}^T = \left(\frac{\partial \ln p(\mathbf{x};\underline{\boldsymbol{\theta}})}{\partial \underline{\boldsymbol{\theta}}}\right)^H \mathbf{U}_{\underline{\boldsymbol{\theta}}^0}^* \left(\mathbf{F}_{\underline{\boldsymbol{\theta}}^0}^c\right)^{-1} \mathbf{U}_{\underline{\boldsymbol{\theta}}^0}^T \left(\frac{\partial \underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)^T}{\partial \underline{\boldsymbol{\theta}}}\right)$$

where $\mathbf{F}_{\underline{\boldsymbol{\theta}}^0}^c$ denotes the constrained FIM. When $\mathbf{F}_{\underline{\boldsymbol{\theta}}^0}^c$ is singular, the general form of (33) is given by lemma 1 (18) and yields to [33, (8)(9)] which encompass all results previously released on this topic [15][17][18][20]. It is worth noticing that Lemma 3 - and the form of (33) - allows to assert that the constrained CRB does not depend on the choice of $\mathbf{U}_{\underline{\boldsymbol{\theta}}^0}$, provided that $\mathbf{U}_{\underline{\boldsymbol{\theta}}^0}$ is a basis of $\ker\left\{\frac{\partial \mathbf{f}(\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\theta}}^T}\right\}$. Moreover, any matrix basis, $\mathbf{U}_{\underline{\boldsymbol{\theta}}^0}$, of $\ker\left\{\frac{\partial \mathbf{f}(\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\theta}}^T}\right\}$ can be complemented by a family of K independent vectors of $\mathbb{C}^P$ in order to obtain a basis $\mathbf{B}_{\underline{\boldsymbol{\theta}}^0}$ of the parameter space $\Theta \equiv \mathbb{C}^P$ [25]. Therefore, according to lemma 3, the following transformations of constraints (26):

$$E_{\underline{\boldsymbol{\theta}}^0}\left[\left(\widehat{\underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x}) - \underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)\right)\frac{\partial \ln p(\mathbf{x};\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\theta}}^T}\mathbf{B}_{\underline{\boldsymbol{\theta}}^0}\right] = \left(\frac{\partial \underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\theta}}^T}\mathbf{B}_{\underline{\boldsymbol{\theta}}^0}\right) \tag{34}$$

will leave the minimum of $\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\widehat{\underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x}) - \underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)\right\}\right)$ unchanged and equal to (27). As:

$$E_{\underline{\boldsymbol{\theta}}^0}\left[\left(\widehat{\underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x}) - \underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)\right)\frac{\partial \ln p(\mathbf{x};\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\theta}}^T}\mathbf{U}_{\underline{\boldsymbol{\theta}}^0}\right] = \left(\frac{\partial \underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\theta}}^T}\mathbf{U}_{\underline{\boldsymbol{\theta}}^0}\right) \tag{35}$$

is a subset of constraints (34), we have:

$$\min\left\{\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\widehat{\underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x}) - \underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)\right\}\right)\right\} \text{ under (34)} \ \geq \ \min\left\{\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\widehat{\underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)}(\mathbf{x}) - \underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}^0)\right\}\right)\right\} \text{ under (35)}$$

that is:

$$\mathbf{F}_{\underline{\boldsymbol{\theta}}^0}^{-1} \geq \mathbf{U}_{\underline{\boldsymbol{\theta}}^0}^* \left(\mathbf{F}_{\underline{\boldsymbol{\theta}}^0}^c\right)^{-1} \mathbf{U}_{\underline{\boldsymbol{\theta}}^0}^T = \mathbf{U}_{\underline{\boldsymbol{\theta}}^0}^* \left(\mathbf{U}_{\underline{\boldsymbol{\theta}}^0}^T \mathbf{F}_{\underline{\boldsymbol{\theta}}^0} \mathbf{U}_{\underline{\boldsymbol{\theta}}^0}^*\right)^{-1} \mathbf{U}_{\underline{\boldsymbol{\theta}}^0}^T \tag{36}$$

This simple derivation shows that the addition of constraints on parameters decreases the value of the CRB, a normal consequence since $\underline{\boldsymbol{\theta}}$ spans a subset of Θ [33].

3) On a necessary regularity condition:

The present approach allows to recall that there is no need to interchange integration with respect to $\mathbf{x}$ and differentiation with respect to $\underline{\boldsymbol{\theta}}$ [7][8][34], contrarily to what is claimed by most of the authors in the field resorting to the covariance inequality [1][19][20][30][33]. Indeed the only requirement is:

$$\forall p \in [1, P], \; E_{\underline{\boldsymbol{\theta}}^0}\left[\left|\frac{\partial \ln p\left(\mathbf{x}; \underline{\boldsymbol{\theta}}\right)}{\partial \theta_p}\right|^2\right] = \int_\Omega \left|\frac{\partial p\left(\mathbf{x}; \underline{\boldsymbol{\theta}}\right)}{\partial \theta_p}\right|^2 \frac{1}{p\left(\mathbf{x}; \underline{\boldsymbol{\theta}}\right)} d\mathbf{x} < \infty \;\Leftrightarrow\; \frac{1}{\sqrt{p\left(\mathbf{x}; \underline{\boldsymbol{\theta}}\right)}} \frac{\partial p\left(\mathbf{x}; \underline{\boldsymbol{\theta}}\right)}{\partial \theta_p} \in L^2\left(\Omega\right)$$

4) On a sufficient regularity condition:

For sake of legibility we will consider in this section only vectors of real parameters: $\underline{\boldsymbol{\theta}} = \boldsymbol{\theta} \in \mathbb{R}^P$.

It is quite straightforward to realize with the present approach, that the usual regularity condition on the differentiability of both $p\left(\mathbf{x}; \boldsymbol{\theta}\right)$ and $\mathbf{g}\left(\boldsymbol{\theta}\right)$ at $\boldsymbol{\theta}^0$ can be relaxed to semi-differentiability (left and right differentiability) only, under certain conditions. Let us first denote $\frac{\partial f(\boldsymbol{\theta}^+)}{\partial \boldsymbol{\theta}}$ and $\frac{\partial f(\boldsymbol{\theta}^-)}{\partial \boldsymbol{\theta}}$ the vectors of right derivatives and left derivatives, respectively. Then, the first order constraint (21) derived under differentiability assumption can be written as well for left and right differentiability, that is:

$$\begin{array}{c} E_{\underline{\boldsymbol{\theta}}^0}\left[\left(\widehat{\underline{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}}\left(\mathbf{x}\right) - \underline{\mathbf{g}}\left(\boldsymbol{\theta}^0\right)\right) \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta}^{0-})}{\partial \boldsymbol{\theta}^T}\right] = \frac{\partial \underline{\mathbf{g}}(\boldsymbol{\theta}^{0-})}{\partial \boldsymbol{\theta}^T} \\ E_{\underline{\boldsymbol{\theta}}^0}\left[\left(\widehat{\underline{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}}\left(\mathbf{x}\right) - \underline{\mathbf{g}}\left(\boldsymbol{\theta}^0\right)\right) \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta}^{0+})}{\partial \boldsymbol{\theta}^T}\right] = \frac{\partial \underline{\mathbf{g}}(\boldsymbol{\theta}^{0+})}{\partial \boldsymbol{\theta}^T} \end{array}$$

which is equivalent to the set of linear constraints:

$$\begin{array}{c} \mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\mathbf{u}\right\}_{[1,Q]}, \left\{\mathbf{c}\right\}_{[1,2P]}\right) = \left[\frac{\partial \underline{\mathbf{g}}(\boldsymbol{\theta}^{0-})}{\partial \boldsymbol{\theta}^T}, \frac{\partial \underline{\mathbf{g}}(\boldsymbol{\theta}^{0+})}{\partial \boldsymbol{\theta}^T}\right]^T \\ \left\{\mathbf{u}\right\}_{[1,Q]} = \left\{\widehat{\underline{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}}\left(\mathbf{x}\right) - \underline{\mathbf{g}}\left(\boldsymbol{\theta}^0\right)\right\}, \quad \left\{\mathbf{c}\right\}_{[1,2P]} = \left\{\left(\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta}^{0-})}{\partial \boldsymbol{\theta}^T}\right)^*, \left(\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta}^{0+})}{\partial \boldsymbol{\theta}^T}\right)^*\right\} \end{array} \quad (37)$$

Then the search for the minimization of $\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\mathbf{u}\right\}_{[1,Q]}\right)$ under (37) still follows the same scheme, where $\left\{\mathbf{c}\right\}_{[1,2P]}$ must be reduced to a subset of independent vectors. This can be done in three steps:

• 1) for every parameter θ_p where both $\mathbf{g}\left(\boldsymbol{\theta}\right)$ and $p\left(\mathbf{x}; \boldsymbol{\theta}\right)$ are differentiable at $\boldsymbol{\theta}^0$, the subset of 2 associated constraints (local unbiasedness of θ_p):

$$\begin{array}{c} E_{\underline{\boldsymbol{\theta}}^0}\left[\left(\widehat{\underline{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}}\left(\mathbf{x}\right) - \underline{\mathbf{g}}\left(\boldsymbol{\theta}^0\right)\right) \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta}^{0-})}{\partial \theta_p}\right] = \frac{\partial \underline{\mathbf{g}}(\boldsymbol{\theta}^{0-})}{\partial \theta_p} \\ E_{\underline{\boldsymbol{\theta}}^0}\left[\left(\widehat{\underline{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}}\left(\mathbf{x}\right) - \underline{\mathbf{g}}\left(\boldsymbol{\theta}^0\right)\right) \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta}^{0+})}{\partial \theta_p}\right] = \frac{\partial \underline{\mathbf{g}}(\boldsymbol{\theta}^{0+})}{\partial \theta_p} \end{array} \quad (38)$$

must be replaced by the equivalent reduced constraint:

$$E_{\underline{\boldsymbol{\theta}}^0}\left[\left(\widehat{\underline{\mathbf{g}\left(\boldsymbol{\theta}^0\right)}}\left(\mathbf{x}\right) - \underline{\mathbf{g}}\left(\boldsymbol{\theta}^0\right)\right) \frac{\partial \ln p\left(\mathbf{x}; \boldsymbol{\theta}^0\right)}{\partial \theta_p}\right] = \frac{\partial \underline{\mathbf{g}}\left(\boldsymbol{\theta}^0\right)}{\partial \theta_p}$$

• 2) for every parameter θ_p where $p\left(\mathbf{x}; \boldsymbol{\theta}\right)$ or $\mathbf{g}\left(\boldsymbol{\theta}\right)$ is not differentiable at $\boldsymbol{\theta}^0$, the subset of 2 associated constraints (38) must be examined since they are 3 possible cases:

first case: $\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta}^{0-})}{\partial \theta_p}$ and $\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta}^{0+})}{\partial \theta_p}$ are independent vectors, then the 2 associated constraints (38) are independent and must be kept both.

second case: $\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta}^{0-})}{\partial \theta_p} = \alpha \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta}^{0+})}{\partial \theta_p}$ (dependent vectors, possibly equal) and $\frac{\partial \underline{\mathbf{g}}(\boldsymbol{\theta}^{0-})}{\partial \theta_p} = \alpha \frac{\partial \underline{\mathbf{g}}(\boldsymbol{\theta}^{0+})}{\partial \theta_p}$, then the 2 associated constraints (38) reduce to a single one, and we can decide arbitrarily to keep the left locally unbiased constraint.

third case: $\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta}^{0-})}{\partial \theta_p} = \alpha \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta}^{0+})}{\partial \theta_p}$ (dependent vectors, possibly equal) and $\frac{\partial \underline{\mathbf{g}}(\boldsymbol{\theta}^{0-})}{\partial \theta_p} \neq \alpha \frac{\partial \underline{\mathbf{g}}(\boldsymbol{\theta}^{0+})}{\partial \theta_p}$, then the 2 associated constraints are incompatible and a locally unbiased estimated at $\boldsymbol{\theta}^0$ can not exist for $\mathbf{g}\left(\boldsymbol{\theta}\right)$.

• 3) Last, one must check that the remaining set of $\widetilde{P}$ vectors $\left\{\mathbf{c}\right\}_{[1,\widetilde{P}]}$ resulting from the analysis performed in step 1) and 2) leads to a non singular FIM $\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\mathbf{c}\right\}_{[1,\widetilde{P}]}\right)$.

As an example let us consider the simple case where:

$$p\left(x;\theta\right)=\frac{e^{-\frac{1}{2}(x-|\theta|)^{2}}}{\sqrt{2\pi}}\Leftrightarrow\left\{\begin{array}{l}\ln p\left(x;\theta\right)=-\frac{1}{2}\left(x-\theta\right)^{2}-\frac{1}{2}\ln\left(2\pi\right),\theta\geq0\\ \ln p\left(x;\theta\right)=-\frac{1}{2}\left(x+\theta\right)^{2}-\frac{1}{2}\ln\left(2\pi\right),\theta\leq0\end{array}\right.\quad,\quad g\left(\theta\right)=|\theta|$$

Then neither $\ln p\left(x;\theta\right)$ nor $g\left(\theta\right)$ are differentiable in θ at $\theta^{0}=0$, and according to all textbooks the CRB of $|\theta|$ does not exist at $\theta^{0}=0$. However, as:

$$\left\{\begin{array}{l}\frac{\partial\ln p(\mathbf{x};0^{+})}{\partial\theta}=2x\\ \frac{\partial\ln p(\mathbf{x};0^{-})}{\partial\theta}=-2x\end{array}\right.\text{ and }\left\{\begin{array}{l}\frac{\partial g(0^{+})}{\partial\theta}=1\\ \frac{\partial g(0^{-})}{\partial\theta}=-1\end{array}\right.\quad\Rightarrow\quad\left\{\begin{array}{l}\frac{\partial\ln p(\mathbf{x};0^{+})}{\partial\theta}=-\frac{\partial\ln p(\mathbf{x};0^{-})}{\partial\theta}\\ \frac{\partial g(0^{+})}{\partial\theta}=-\frac{\partial g(0^{-})}{\partial\theta}\end{array}\right.,$$

therefore the CRB does exist at $\theta^{0}=0$ and its value can be computed as:

$$CRB_{0}=\left(\frac{\partial g\left(0^{-}\right)}{\partial\theta}\right)E_{0}\left[\left(\frac{\partial\ln p\left(\mathbf{x};0-\right)}{\partial\theta}\right)^{2}\right]^{-1}\left(\frac{\partial g\left(0^{-}\right)}{\partial\theta}\right)=\frac{1}{4E_{0}\left[x^{2}\right]}=\frac{1}{4}$$

On the opposite, the CRB at $\theta^{0}=0$ no longer exists if $g\left(\theta\right)=\theta$.
This relaxation on differentiability property can be extended to complex parameters, where for each complex parameter θ_{p}, semi-differentiability can be defined in terms of a partitioning of the complex plane at the vicinity of θ_{p}^{0}, where for each subset of the partition there is a given differential function. Then, one must follow the 3 steps rationale described for real parameters except that, for each complex parameter θ_{p}, the analysis of the system of 2 equations (38) must be replaced by the analysis of a system of L equations, where L is the number of different "semi-derivatives" at the vicinity of θ_{p}^{0}.

IV. Reparameterization and constraints for CRB: a unified approach

Let us define the following notation convention:

$$\mathbf{CRB}_{\mathbf{g}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}^{0}\right)=\left(\frac{\partial\mathbf{g}\left(\underline{\boldsymbol{\theta}}^{0}\right)^{T}}{\partial\underline{\boldsymbol{\theta}}}\right)^{H}\mathbf{CRB}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}^{0}\right)\left(\frac{\partial\mathbf{g}\left(\underline{\boldsymbol{\theta}}^{0}\right)^{T}}{\partial\underline{\boldsymbol{\theta}}}\right),\quad\mathbf{CRB}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}^{0}\right)=\mathbf{F}_{\underline{\boldsymbol{\theta}}^{0}}^{-1}\qquad(39)$$

A. A reparameterization inequality

Let us consider a reparameterization of the p.d.f. $p\left(\mathbf{x};\underline{\boldsymbol{\theta}}\right)$ for the unknown parameters $\underline{\boldsymbol{\theta}}$ where:

$$\underline{\boldsymbol{\theta}}=\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}\right),\quad\dim\left\{\underline{\boldsymbol{\omega}}\right\}=P',\quad\dim\left\{\underline{\boldsymbol{\theta}}\right\}=P.$$

Then $p\left(\mathbf{x};\underline{\boldsymbol{\omega}}\right)=p\left(\mathbf{x};\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}\right)\right)$ and by use of the derivation chain rule identity:

$$\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\omega}}^{0}\right)}{\partial\underline{\boldsymbol{\omega}}^{T}}=\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)\right)}{\partial\underline{\boldsymbol{\theta}}^{T}}\frac{\partial\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)}{\partial\underline{\boldsymbol{\omega}}^{T}}\Leftrightarrow\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\omega}}^{0}\right)}{\partial\underline{\boldsymbol{\omega}}}=\frac{\partial\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)^{T}}{\partial\underline{\boldsymbol{\omega}}}\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)\right)}{\partial\underline{\boldsymbol{\theta}}},$$

we easily obtain that (27):

$$\mathbf{F}_{\underline{\boldsymbol{\omega}}^{0}}=E_{\underline{\boldsymbol{\omega}}^{0}}\left[\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\omega}}^{0}\right)}{\partial\underline{\boldsymbol{\omega}}}\left(\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\omega}}^{0}\right)}{\partial\underline{\boldsymbol{\omega}}}\right)^{H}\right]=\frac{\partial\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)^{T}}{\partial\underline{\boldsymbol{\omega}}}\mathbf{F}_{\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)}\left(\frac{\partial\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)^{T}}{\partial\underline{\boldsymbol{\omega}}}\right)^{H}\qquad(40)$$

which is a trivial generalization of a well known identity for real parameters [35, p125]. In the rest of the present paper, we only consider injective reparameterization, what implies that $P'\leq P$ and $rank\left(\frac{\partial\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)}{\partial\underline{\boldsymbol{\omega}}^{T}}\right)=P'$ (full rank matrix). Under this assumption $\mathbf{F}_{\underline{\boldsymbol{\omega}}^{0}}$ is singular if and only if $\mathbf{F}_{\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)}$ is singular. Now, let us consider the problem of estimating $\mathbf{h}\left(\underline{\boldsymbol{\omega}}\right)=\mathbf{g}\left(\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}\right)\right)$. According to (27), the CRB associated to any locally unbiased estimator of $\mathbf{h}\left(\underline{\boldsymbol{\omega}}^{0}\right)=\mathbf{g}\left(\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)\right)$ is given by:

$$\mathbf{G}_{\underline{\boldsymbol{\omega}}^{0}}\left(\left\{\underline{\widehat{\mathbf{h}\left(\boldsymbol{\omega}^{0}\right)}}\left(\mathbf{x}\right)-\underline{\mathbf{h}}\left(\underline{\boldsymbol{\omega}}^{0}\right)\right\}\right)\geq\mathbf{CRB}_{\underline{\mathbf{h}}|\underline{\boldsymbol{\omega}}}\left(\underline{\boldsymbol{\omega}}^{0}\right)=\left(\frac{\partial\underline{\mathbf{h}}\left(\underline{\boldsymbol{\omega}}^{0}\right)^{T}}{\partial\underline{\boldsymbol{\omega}}}\right)^{H}\mathbf{CRB}_{\underline{\boldsymbol{\omega}}|\underline{\boldsymbol{\omega}}}\left(\underline{\boldsymbol{\omega}}^{0}\right)\left(\frac{\partial\underline{\mathbf{h}}\left(\underline{\boldsymbol{\omega}}^{0}\right)^{T}}{\partial\underline{\boldsymbol{\omega}}}\right)\qquad(41)$$

Additionally, as:

$$\frac{\partial \underline{\mathbf{h}}\left(\underline{\boldsymbol{\omega}}^{0}\right)^{T}}{\partial \underline{\boldsymbol{\omega}}}=\frac{\partial \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)\right)^{T}}{\partial \underline{\boldsymbol{\omega}}}=\frac{\partial \underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)^{T}}{\partial \underline{\boldsymbol{\omega}}}\frac{\partial \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)\right)^{T}}{\partial \underline{\boldsymbol{\theta}}}$$

and taking into account (40), the above CRB can also be expressed as:

$$\mathbf{CRB}_{\underline{\mathbf{h}}|\underline{\boldsymbol{\omega}}}\left(\underline{\boldsymbol{\omega}}^{0}\right)=\left(\frac{\partial \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)\right)^{T}}{\partial \underline{\boldsymbol{\theta}}}\right)^{H}\mathbf{CRB}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\omega}}}\left(\underline{\boldsymbol{\omega}}^{0}\right)\left(\frac{\partial \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)\right)^{T}}{\partial \underline{\boldsymbol{\theta}}}\right)=\mathbf{CRB}_{\underline{\mathbf{g}}(\underline{\boldsymbol{\theta}})|\underline{\boldsymbol{\omega}}}\left(\underline{\boldsymbol{\omega}}^{0}\right),$$

$$\mathbf{CRB}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\omega}}}\left(\underline{\boldsymbol{\omega}}^{0}\right)=\left(\frac{\partial \underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)^{T}}{\partial \underline{\boldsymbol{\omega}}}\right)^{H}\left(\frac{\partial \underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)^{T}}{\partial \underline{\boldsymbol{\omega}}}\mathbf{F}_{\underline{\boldsymbol{\theta}}(\underline{\boldsymbol{\omega}}^{0})}\left(\frac{\partial \underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)^{T}}{\partial \underline{\boldsymbol{\omega}}}\right)^{H}\right)^{-1}\left(\frac{\partial \underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)^{T}}{\partial \underline{\boldsymbol{\omega}}}\right) \tag{42}$$

Now, by application of the following lemma:

Lemma 4: Let $\mathbf{A}\in\mathcal{M}_{\mathbb{C}}(P,P)$ be a positive definite Hermitan matrix ($\mathbf{A}=\mathbf{A}^{H}$, $\mathbf{A}>\mathbf{0}$, $\mathbf{A}=\mathbf{L}_{\mathbf{A}}^{H}\mathbf{L}_{\mathbf{A}}$), let $\mathbf{B}\in\mathcal{M}_{\mathbb{C}}(P,K)$ where $K\leq P$ be a full rank matrix ($rank(\mathbf{B})=K$) then:

$$\mathbf{\Pi}_{\mathbf{L}_{\mathbf{A}}\mathbf{B}}^{\perp}=\mathbf{I}-\mathbf{L}_{\mathbf{A}}\mathbf{B}\left(\mathbf{B}^{H}\mathbf{A}\mathbf{B}\right)^{-1}\mathbf{B}^{H}\mathbf{L}_{\mathbf{A}}^{H}\geq\mathbf{0}\quad\Rightarrow\quad\mathbf{A}^{-1}\geq\mathbf{B}\left(\mathbf{B}^{H}\mathbf{A}\mathbf{B}\right)^{-1}\mathbf{B}^{H} \tag{43}$$

we can derive a specific form - for $\underline{\mathbf{g}}(\underline{\boldsymbol{\theta}})=\underline{\boldsymbol{\theta}}$ - of the reparameterization inequality:

$$\mathbf{CRB}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)\right)\geq\mathbf{CRB}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\omega}}}\left(\underline{\boldsymbol{\omega}}^{0}\right) \tag{44}$$

leading to its most general form, $\forall\underline{\mathbf{g}}(\underline{\boldsymbol{\theta}})$:

$$\mathbf{CRB}_{\underline{\mathbf{g}}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)\right)\geq\mathbf{CRB}_{\underline{\mathbf{g}}(\underline{\boldsymbol{\theta}})|\underline{\boldsymbol{\omega}}}\left(\underline{\boldsymbol{\omega}}^{0}\right) \tag{45}$$

We believe that this inequality is new to the best of our knowledge, since it has neither been released in most reference books [1][27][28][35] nor in recent papers in the field [15][33]. The reparamaterization inequality (44)(45) expresses analytically a quite intuitive estimation principle: when the total number of unknown parameters decreases in an observation model ($P'<P$), the asymptotic quality of estimation increases (or remain unchanged), in the sense that the CRB decreases (or remain equal), whatever the function $\underline{\mathbf{g}}(\ )$ of the unknown parameters considered. If $P'=P$, then the reparameterization has no effect on the asymptotic quality of estimation since then:

$$\mathbf{CRB}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)\right)=\mathbf{CRB}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\omega}}}\left(\underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)\right)$$

B. Reparameterization as constraints

A particular case in reparameterization occurs when $\underline{\boldsymbol{\omega}}$ is a subset of $\underline{\boldsymbol{\theta}}$. Then, arbitrarily we can rearrange the components of $\underline{\boldsymbol{\theta}}$ in order to obtain:

$$\underline{\boldsymbol{\theta}}(\underline{\boldsymbol{\omega}})=\left(\underline{\boldsymbol{\omega}}^{T},\underline{\boldsymbol{\varepsilon}}^{T}=\underline{\mathbf{h}}(\underline{\boldsymbol{\omega}})^{T}\right)^{T},\ \dim\{\underline{\boldsymbol{\theta}}\}=P,\ \dim\{\underline{\boldsymbol{\varepsilon}}\}=K,\ \dim\{\underline{\boldsymbol{\omega}}\}=P-K,\ 1\leq K\leq P-1 \tag{46}$$

where the function $\underline{\boldsymbol{\theta}}(\underline{\boldsymbol{\omega}})$ is assumed injective. This kind of reparameterization is equivalent to the introduction of the following set of K constraints:

$$\underline{\boldsymbol{\varepsilon}}-\underline{\mathbf{h}}(\underline{\boldsymbol{\omega}})=\mathbf{0}=\underline{\mathbf{f}}(\underline{\boldsymbol{\theta}})\in\mathbb{C}^{K} \tag{47}$$

Indeed, one can check that:

$$\frac{\partial \underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)}{\partial \underline{\boldsymbol{\omega}}^{T}}=\begin{bmatrix}\mathbf{I}_{P-K}\\ \frac{\partial \underline{\mathbf{h}}(\underline{\boldsymbol{\omega}}^{0})}{\partial \underline{\boldsymbol{\omega}}^{T}}\end{bmatrix}\in\mathcal{M}_{\mathbb{C}}(P,P-K) \tag{48}$$

and:

$$\frac{\partial \underline{\mathbf{f}}\left(\underline{\boldsymbol{\theta}}^{0}\right)}{\partial \underline{\boldsymbol{\theta}}^{T}}=\left[\frac{\partial \underline{\mathbf{f}}\left(\underline{\boldsymbol{\theta}}^{0}\right)}{\partial \underline{\boldsymbol{\omega}}^{T}},\frac{\partial \underline{\mathbf{f}}\left(\underline{\boldsymbol{\theta}}^{0}\right)}{\partial \underline{\boldsymbol{\varepsilon}}^{T}}\right]=\left[-\frac{\partial \underline{\mathbf{h}}\left(\underline{\boldsymbol{\omega}}^{0}\right)}{\partial \underline{\boldsymbol{\omega}}^{T}},\mathbf{I}_{K}\right]\in\mathcal{M}_{\mathbb{C}}(K,P)$$

Moreover, as:

$$\mathbf{U}_{\underline{\boldsymbol{\theta}}^{0}}=\begin{bmatrix}\mathbf{I}_{P-K}\\ \frac{\partial \underline{\mathbf{h}}(\underline{\boldsymbol{\omega}}^{0})}{\partial \underline{\boldsymbol{\omega}}^{T}}\end{bmatrix}\ \Rightarrow\ \left[-\frac{\partial \underline{\mathbf{h}}\left(\underline{\boldsymbol{\omega}}^{0}\right)}{\partial \underline{\boldsymbol{\omega}}^{T}},\mathbf{I}_{K}\right]\mathbf{U}_{\underline{\boldsymbol{\theta}}^{0}}=-\frac{\partial \underline{\mathbf{h}}\left(\underline{\boldsymbol{\omega}}^{0}\right)}{\partial \underline{\boldsymbol{\omega}}^{T}}+\frac{\partial \underline{\mathbf{h}}\left(\underline{\boldsymbol{\omega}}^{0}\right)}{\partial \underline{\boldsymbol{\omega}}^{T}}=\mathbf{0}, \tag{49}$$

therefore $\mathbf{U}_{\underline{\boldsymbol{\theta}}^{0}}=\frac{\partial \underline{\boldsymbol{\theta}}(\underline{\boldsymbol{\omega}}^{0})}{\partial \underline{\boldsymbol{\omega}}^{T}}$ is a basis of $\ker\left\{\frac{\partial \underline{\mathbf{f}}(\underline{\boldsymbol{\theta}}^{0})}{\partial \underline{\boldsymbol{\theta}}^{T}}\right\}$, what shows that the use of the constrained CRB (33) with $\mathbf{U}_{\underline{\boldsymbol{\theta}}^{0}}$ as in (49) or the use of the reparameterized CRB (40) with reparameterization derivatives as in (48) leads to the same CRB.

C. Constraints as reparameterization

Similarly to subsection III-C2, let us consider the subset $\mathcal{C}$ of Θ defined by:

$$\underline{\mathbf{f}}\left(\underline{\boldsymbol{\theta}}\right) = \mathbf{0} \in \mathbb{C}^{K}, \quad 1 \leq K \leq P-1,$$

where the matrix $\frac{\partial \underline{\mathbf{f}}(\underline{\boldsymbol{\theta}})}{\partial \underline{\boldsymbol{\theta}}^{T}} \in \mathcal{M}_{\mathbb{C}}\left(K,P\right)$ has full row rank K. Then, in the simpler cases, for example if $\underline{\mathbf{f}}\left(\underline{\boldsymbol{\theta}}\right)$ is a linear function of $\underline{\boldsymbol{\theta}}$, $\underline{\boldsymbol{\theta}}$ can be readily expressed as an explicit function of $P-K$ components $\underline{\boldsymbol{\omega}}$ in the following form:

$$\underline{\boldsymbol{\theta}} = \left(\underline{\boldsymbol{\omega}}^{T}, \underline{\boldsymbol{\varepsilon}}^{T}\right)^{T}, \quad \underline{\boldsymbol{\varepsilon}} = \underline{\mathbf{h}}\left(\underline{\boldsymbol{\omega}}\right), \quad \dim\left\{\underline{\boldsymbol{\theta}}\right\} = P, \ \dim\left\{\underline{\boldsymbol{\varepsilon}}\right\} = K, \ \dim\left\{\underline{\boldsymbol{\omega}}\right\} = P-K,$$

which is a particular reparameterization case (46). However in most cases, it is not possible to obtain an explicit analytical expression for $\underline{\mathbf{h}}\left(\ \right)$, although $\underline{\mathbf{h}}\left(\ \right)$ exists under an implicit formulation given by the implicit function theorem for real [36] and complex variables [24].

Theorem 1: implicit function theorem.

Let $\mathbf{f}\left(\boldsymbol{\varepsilon},\boldsymbol{\omega}\right) = \left[f_{1}\left(\boldsymbol{\varepsilon},\boldsymbol{\omega}\right), \cdots, f_{K}\left(\boldsymbol{\varepsilon},\boldsymbol{\omega}\right)\right]^{T}$ be a function of $\mathbb{C}^{K} \times \mathbb{C}^{P'} \rightarrow \mathbb{C}^{K}$. Let us assume that:
Assumption A1: $f_{k}\left(\boldsymbol{\varepsilon},\boldsymbol{\omega}\right)$ for $k = 1, \ldots, K$ are differentiable functions on a neighbourhood of the point $\left(\boldsymbol{\varepsilon}^{0},\boldsymbol{\omega}^{0}\right)$ in $\mathbb{C}^{K} \times \mathbb{C}^{P'}$.
Assumption A2: $\mathbf{f}\left(\boldsymbol{\varepsilon}^{0},\boldsymbol{\omega}^{0}\right) = \mathbf{0}$.
Assumption A3: the $K \times K$ Jacobian matrix $\frac{\partial \mathbf{f}(\boldsymbol{\varepsilon},\boldsymbol{\omega})}{\partial \boldsymbol{\varepsilon}^{T}}$ is non singular at $\left(\boldsymbol{\varepsilon}^{0},\boldsymbol{\omega}^{0}\right)$.
Then there is a neighbourhood $\mathcal{V}$ of the point $\boldsymbol{\omega}^{0}$ in $\mathbb{C}^{P'}$, there is a neighbourhood $\mathcal{U}$ of the point $\boldsymbol{\varepsilon}^{0}$ in $\mathbb{C}^{K}$, and there is a unique mapping h: $\mathcal{V} \rightarrow \mathcal{U}$ such that $h\left(\boldsymbol{\omega}^{0}\right) = \boldsymbol{\varepsilon}^{0}$ and $\mathbf{f}\left(\mathbf{h}\left(\boldsymbol{\omega}\right),\boldsymbol{\omega}\right) = \mathbf{0}$ for all $\boldsymbol{\omega}$ in $\mathcal{V}$. Furthermore, h is differentiable and we have:

$$\mathbf{h}\left(\boldsymbol{\omega}\right) - \boldsymbol{\varepsilon}^{0} = -\left(\frac{\partial \mathbf{f}\left(\boldsymbol{\varepsilon}^{0},\boldsymbol{\omega}^{0}\right)}{\partial \boldsymbol{\varepsilon}^{T}}\right)^{-1} \frac{\partial \mathbf{f}\left(\boldsymbol{\varepsilon}^{0},\boldsymbol{\omega}^{0}\right)}{\partial \boldsymbol{\omega}^{T}} \left(\boldsymbol{\omega}-\boldsymbol{\omega}^{0}\right) + \mathbf{r}\left(\boldsymbol{\omega}-\boldsymbol{\omega}^{0}\right), \tag{50}$$

where the remainder $\mathbf{r}\left(\boldsymbol{\omega}-\boldsymbol{\omega}^{0}\right) = \mathbf{o}\left(\left\|\boldsymbol{\omega}-\boldsymbol{\omega}^{0}\right\|\right)$.

Indeed, as $\frac{\partial \underline{\mathbf{f}}(\underline{\boldsymbol{\theta}})}{\partial \underline{\boldsymbol{\theta}}^{T}}$ has full row rank K, then there exists a subset of K independent columns of partial derivatives with respect to K components of $\underline{\boldsymbol{\theta}}$, which can be gathered in a subvector $\underline{\boldsymbol{\varepsilon}}$. Then, the components of $\underline{\boldsymbol{\theta}}$ can be rearranged accordingly in order to obtain:

$$\underline{\boldsymbol{\theta}} = \left(\underline{\boldsymbol{\omega}}^{T}, \underline{\boldsymbol{\varepsilon}}^{T}\right)^{T}, \quad \underline{\mathbf{f}}\left(\underline{\boldsymbol{\theta}}\right) = \mathbf{0}, \quad \dim\left\{\underline{\boldsymbol{\theta}}\right\} = P, \quad \dim\left\{\underline{\boldsymbol{\varepsilon}}\right\} = K, \quad \dim\left\{\underline{\boldsymbol{\omega}}\right\} = P-K,$$

and the implicit function theorem not only states the existence of an implicit function $\underline{\boldsymbol{\varepsilon}} = \underline{\mathbf{h}}\left(\underline{\boldsymbol{\omega}}\right)$ but also provides an expression to compute its derivatives:

$$\frac{\partial \underline{\mathbf{h}}\left(\underline{\boldsymbol{\omega}}^{0}\right)}{\partial \underline{\boldsymbol{\omega}}^{T}} = -\left(\frac{\partial \underline{\mathbf{f}}\left(\underline{\boldsymbol{\theta}}^{0}\right)}{\partial \underline{\boldsymbol{\varepsilon}}^{T}}\right)^{-1} \frac{\partial \underline{\mathbf{f}}\left(\underline{\boldsymbol{\theta}}^{0}\right)}{\partial \underline{\boldsymbol{\omega}}^{T}} \quad \Rightarrow \quad \frac{\partial \underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)}{\partial \underline{\boldsymbol{\omega}}^{T}} = \begin{bmatrix} \mathbf{I}_{P-K} \\ \frac{\partial \underline{\mathbf{h}}(\underline{\boldsymbol{\omega}}^{0})}{\partial \underline{\boldsymbol{\omega}}^{T}} \end{bmatrix} \tag{51}$$

and a basis $\mathbf{U}_{\underline{\boldsymbol{\theta}}^{0}}$ of $\ker\left\{\frac{\partial \underline{\mathbf{f}}(\underline{\boldsymbol{\theta}}^{0})}{\partial \underline{\boldsymbol{\theta}}^{T}}\right\}$ (49):

$$\mathbf{U}_{\underline{\boldsymbol{\theta}}^{0}} = \begin{bmatrix} \mathbf{I}_{P-K} \\ -\left(\frac{\partial \underline{\mathbf{f}}(\underline{\boldsymbol{\theta}}^{0})}{\partial \underline{\boldsymbol{\varepsilon}}^{T}}\right)^{-1} \frac{\partial \underline{\mathbf{f}}(\underline{\boldsymbol{\theta}}^{0})}{\partial \underline{\boldsymbol{\omega}}^{T}} \end{bmatrix} = \frac{\partial \underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)}{\partial \underline{\boldsymbol{\omega}}^{T}} \tag{52}$$

Finally, the use of the constrained CRB (33) with $\mathbf{U}_{\underline{\boldsymbol{\theta}}^{0}}$ as in (52) or the use of the reparameterized CRB (40) with reparameterization derivatives as in (51) leads to the same CRB.

D. Reparameterization and constraints: an unified approach

In the two previous sections we have shown that, regarding the computation of the CRB, equality constraints on parameters:

$$\underline{\mathbf{f}}\left(\underline{\boldsymbol{\theta}}\right) = \mathbf{0} \in \mathbb{C}^{K}, \quad \underline{\boldsymbol{\theta}} \in \mathbb{C}^{P}, \quad 1 \leq K \leq P$$

is a particular case of the reparameterization of the unknown parameters $\underline{\boldsymbol{\theta}}$, provided that the set of constraints is not redundant. Then, the reparameterization inequality (45) holds provided that:

$$\frac{\partial \underline{\boldsymbol{\theta}}\left(\underline{\boldsymbol{\omega}}^{0}\right)}{\partial \underline{\boldsymbol{\omega}}^{T}} \equiv \mathbf{U}_{\underline{\boldsymbol{\theta}}^{0}} \tag{53}$$

where $\mathbf{U}_{\underline{\boldsymbol{\theta}}^{0}} \in \mathcal{M}_{\mathbb{C}}\left(P, P-K\right)$ is a basis of $\ker\left\{\frac{\partial \underline{\mathbf{f}}(\underline{\boldsymbol{\theta}}^{0})}{\partial \underline{\boldsymbol{\theta}}^{T}}\right\}$. $\mathbf{U}_{\underline{\boldsymbol{\theta}}^{0}}$ can always be computed - after rearrangement of $\underline{\boldsymbol{\theta}}$ - as:

$$\mathbf{U}_{\underline{\boldsymbol{\theta}}^{0}} = \left[\begin{array}{c} \mathbf{I}_{P-K} \\ -\left(\frac{\partial \underline{\mathbf{f}}(\underline{\boldsymbol{\theta}}^{0})}{\partial \underline{\boldsymbol{\varepsilon}}^{T}}\right)^{-1} \frac{\partial \underline{\mathbf{f}}(\underline{\boldsymbol{\theta}}^{0})}{\partial \underline{\boldsymbol{\omega}}^{T}} \end{array}\right], \quad \underline{\boldsymbol{\theta}} = \left(\begin{array}{c} \underline{\boldsymbol{\omega}} \\ \underline{\boldsymbol{\varepsilon}} \end{array}\right) \tag{54}$$

where $\underline{\boldsymbol{\varepsilon}}$ is a subvector (subset) of K components of $\underline{\boldsymbol{\theta}}$ which K columns of partial derivatives - columns of matrix $\frac{\partial \underline{\mathbf{f}}(\underline{\boldsymbol{\theta}}^{0})}{\partial \underline{\boldsymbol{\theta}}^{T}}$ - are independent. Nevertheless, according to the form of $\underline{\mathbf{f}}\left(\underline{\boldsymbol{\theta}}\right)$**,** $\mathbf{U}_{\underline{\boldsymbol{\theta}}^{0}}$ may be derived by another rationale possibly simpler to implement than the use of (54) (see [15] for an example).

V. Reparameterization and constraints: application examples

A. Influence of parameters state - known or unknown - on the CRB

In many parametric observation models some parameters can be either known or unknown according to the experimental conditions. When the value of a parameter is known during an observation, then its value can be incorporated into any expression involving the parametric model, such as MLEs, lower bounds, etc... Otherwise it must be estimated. To take into account the two possible states (known or unknown) of each parameter, the p.d.f. of the observation vector $\mathbf{x}$ is denoted $p\left(\mathbf{x}; \underline{\boldsymbol{\theta}}; \underline{\boldsymbol{\kappa}}\right) \equiv p\left(\mathbf{x}; \underline{\boldsymbol{\theta}}_{unknown}; \underline{\boldsymbol{\kappa}}_{known}\right)$ in the following, where $\underline{\boldsymbol{\theta}}$ is the set (vector) of unknown parameters and $\underline{\boldsymbol{\kappa}}$ is the set (vector) of known parameters. In absence of known parameters, the notation can be reduced to $p\left(\mathbf{x}; \underline{\boldsymbol{\theta}}\right) \equiv p\left(\mathbf{x}; \underline{\boldsymbol{\theta}}_{unknown}\right) \equiv p\left(\mathbf{x}; \underline{\boldsymbol{\theta}}_{unknown}; \varnothing\right)$, and the proposed generalized p.d.f. notation has a downward compatibility with the standard notation $p\left(\mathbf{x}; \underline{\boldsymbol{\theta}}\right) \equiv p\left(\mathbf{x}; \underline{\boldsymbol{\theta}}_{unknown}\right)$. Then notations for CRB (39) becomes:

$$\mathbf{CRB}_{\underline{\mathbf{g}}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}^{0}; \underline{\boldsymbol{\kappa}}^{0}\right) = \left(\frac{\partial \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^{0}\right)^{T}}{\partial \underline{\boldsymbol{\theta}}}\right)^{H} \mathbf{CRB}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}^{0}; \underline{\boldsymbol{\kappa}}^{0}\right) \left(\frac{\partial \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^{0}\right)^{T}}{\partial \underline{\boldsymbol{\theta}}}\right) \tag{55}$$

where:

$$\mathbf{CRB}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}^{0}; \underline{\boldsymbol{\kappa}}^{0}\right) = \mathbf{F}_{\underline{\boldsymbol{\theta}}^{0}; \underline{\boldsymbol{\kappa}}^{0}}^{-1}, \quad \mathbf{F}_{\underline{\boldsymbol{\theta}}^{0}; \underline{\boldsymbol{\kappa}}^{0}} = E_{\underline{\boldsymbol{\theta}}^{0}; \underline{\boldsymbol{\kappa}}^{0}}\left[\frac{\partial \ln p\left(\mathbf{x}; \underline{\boldsymbol{\theta}}^{0}; \underline{\boldsymbol{\kappa}}^{0}\right)}{\partial \underline{\boldsymbol{\theta}}} \left(\frac{\partial \ln p\left(\mathbf{x}; \underline{\boldsymbol{\theta}}^{0}; \underline{\boldsymbol{\kappa}}^{0}\right)}{\partial \underline{\boldsymbol{\theta}}}\right)^{H}\right] \tag{56}$$

and $\underline{\boldsymbol{\theta}}^{0}$, $\underline{\boldsymbol{\kappa}}^{0}$ are selected values of $\underline{\boldsymbol{\theta}}$ and $\underline{\boldsymbol{\kappa}}$. In fact, a known parameter of a deterministic parametric model is just an unknown parameter whose value is known at the time of the experimentation, which can be rewritten as a constraint (beware of commas and semicolons in formulas):

$$p\left(\mathbf{x}; \underline{\boldsymbol{\theta}}; \underline{\boldsymbol{\kappa}}^{0}\right) \Leftrightarrow \left\{\begin{array}{l} p\left(\mathbf{x}; \underline{\boldsymbol{\eta}}\right), \; \underline{\boldsymbol{\eta}}^{T} = \left(\underline{\boldsymbol{\theta}}^{T}, \underline{\boldsymbol{\kappa}}^{T}\right) \\ \underline{\boldsymbol{\kappa}} = \underline{\boldsymbol{\kappa}}^{0} \end{array}\right. \Leftrightarrow \left\{\begin{array}{l} p\left(\mathbf{x}; \underline{\boldsymbol{\eta}}\right), \; \underline{\boldsymbol{\eta}}^{T} = \left(\underline{\boldsymbol{\theta}}^{T}, \underline{\boldsymbol{\kappa}}^{T}\right) \\ \underline{\mathbf{f}}\left(\underline{\boldsymbol{\eta}}\right) = \underline{\boldsymbol{\kappa}} - \underline{\boldsymbol{\kappa}}^{0} = \mathbf{0} \end{array}\right., \; \dim\left\{\underline{\boldsymbol{\theta}}\right\} = P_{u}, \; \dim\left\{\underline{\boldsymbol{\kappa}}\right\} = P_{k}$$

Then:

$$\frac{\partial \underline{\mathbf{f}}\left(\underline{\boldsymbol{\eta}}^{0}\right)}{\partial \underline{\boldsymbol{\eta}}^{T}} = \left[\mathbf{0}, \mathbf{I}_{P_{k}}\right] \in \mathcal{M}_{\mathbb{C}}\left(P_{k}, P_{u}+P_{k}\right) \quad \Rightarrow \quad \mathbf{U}_{\underline{\boldsymbol{\theta}}^{0}} = \left[\begin{array}{c} \mathbf{I}_{P_{u}} \\ \mathbf{0} \end{array}\right] \in \mathcal{M}_{\mathbb{C}}\left(P_{u}+P_{k}, P_{u}\right)$$

and according to (33):

$$\begin{aligned}\mathbf{CRB}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}^0;\underline{\boldsymbol{\kappa}}^0\right) &= \mathbf{U}^*_{\underline{\boldsymbol{\theta}}^0}\left(\mathbf{U}^T_{\underline{\boldsymbol{\theta}}^0}\mathbf{F}_{\underline{\boldsymbol{\eta}}^0}\mathbf{U}^*_{\underline{\boldsymbol{\theta}}^0}\right)^{-1}\mathbf{U}^T_{\underline{\boldsymbol{\theta}}^0}\\ &= E_{\underline{\boldsymbol{\eta}}^0}\left[\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\eta}}^0\right)}{\partial\underline{\boldsymbol{\theta}}}\left(\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\eta}}^0\right)}{\partial\underline{\boldsymbol{\theta}}}\right)^H\right]^{-1}\\ &= E_{\underline{\boldsymbol{\theta}}^0;\underline{\boldsymbol{\kappa}}^0}\left[\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\theta}}^0;\underline{\boldsymbol{\kappa}}^0\right)}{\partial\underline{\boldsymbol{\theta}}}\left(\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\theta}}^0;\underline{\boldsymbol{\kappa}}^0\right)}{\partial\underline{\boldsymbol{\theta}}}\right)^H\right]^{-1}\end{aligned}$$

Therefore, by application of (45) and (53), one obtains:

$$\mathbf{CRB}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}^0;\underline{\boldsymbol{\kappa}}^0\right)\leq\mathbf{CRB}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\eta}}}\left(\underline{\boldsymbol{\eta}}^0\right)$$

A close look at $\mathbf{F}_{\underline{\boldsymbol{\eta}}^0}$ expression allows to state more precisely:

$$\left\{\begin{array}{l}\mathbf{CRB}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}^0;\underline{\boldsymbol{\kappa}}^0\right)=\mathbf{CRB}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\eta}}}\left(\underline{\boldsymbol{\eta}}^0\right)\text{ if and only if }E_{\underline{\boldsymbol{\eta}}^0}\left[\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\eta}}^0\right)}{\partial\underline{\boldsymbol{\theta}}}\left(\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\eta}}^0\right)}{\partial\underline{\boldsymbol{\kappa}}}\right)^H\right]=\mathbf{0}\\ \mathbf{CRB}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}^0;\underline{\boldsymbol{\kappa}}^0\right)<\mathbf{CRB}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\eta}}}\left(\underline{\boldsymbol{\eta}}^0\right)\text{ if and only if }E_{\underline{\boldsymbol{\eta}}^0}\left[\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\eta}}^0\right)}{\partial\underline{\boldsymbol{\theta}}}\left(\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\eta}}^0\right)}{\partial\underline{\boldsymbol{\kappa}}}\right)^H\right]\neq\mathbf{0}\end{array}\right.,\underline{\boldsymbol{\eta}}^T=\left(\underline{\boldsymbol{\theta}}^T,\underline{\boldsymbol{\kappa}}^T\right)\tag{57}$$

This rationale offers a simple and sound derivation on the influence on the CRB of the change of state - known or unknown - for parameters. A first corollary of this derivation is actually implicitly stated in textbooks as follows: whatever the parametric observation model, only the derivative with respect to the unknown parameters must be taken into account in the computation of the FIM. A second corollary, also well known, follows now straightforwardly: if $E_{\underline{\boldsymbol{\eta}}^0}\left[\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\eta}}^0\right)}{\partial\underline{\boldsymbol{\theta}}}\left(\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\eta}}^0\right)}{\partial\underline{\boldsymbol{\kappa}}}\right)^H\right]=\mathbf{0}$ then the asymptotic estimation performances - in the CRB sense - of $\underline{\boldsymbol{\theta}}$ and $\underline{\boldsymbol{\kappa}}$ from observation $\mathbf{x}$ with p.d.f. $p\left(\mathbf{x};\underline{\boldsymbol{\eta}}\right)$, $\underline{\boldsymbol{\eta}}^T=\left(\underline{\boldsymbol{\theta}}^T,\underline{\boldsymbol{\kappa}}^T\right)$ are decorrelated : asymptotic estimation performance of $\underline{\boldsymbol{\theta}}$ is independent of the state - known or unknown - of $\underline{\boldsymbol{\kappa}}$, and vice versa. Last, an equivalent formulation of the second corollary is obtained when among the parameters of a parametric observation model some are considered of "interest" - let us gather them in a vector $\underline{\boldsymbol{\theta}}$ - since they are always unknown and they are the ones we are interested in estimating. The other remaining parameters - let us gather them in a vector $\underline{\boldsymbol{\kappa}}$ - are not of interest although they are a part of the parametric model. Therefore, due to (57), they are expected to degrade the CRB of parameters of interest $\underline{\boldsymbol{\theta}}$ when their state change from known to unknown, since the condition $E_{\underline{\boldsymbol{\eta}}^0}\left[\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\eta}}^0\right)}{\partial\underline{\boldsymbol{\theta}}}\left(\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\eta}}^0\right)}{\partial\underline{\boldsymbol{\kappa}}}\right)^H\right]=\mathbf{0}$ has no reason to hold in general. Hence the name of "nuisance" parameters and its meaning [1][27][28][35].

B. The Gaussian observation model: worst and best cases for CRB

In many practical problems of interest (radar, sonar, communication, ...), the complex observation vector $\mathbf{x}$ of dimension N consists of a bandpass signal which is the output of an Hilbert filtering leading to an "in-phase" real part associated to a "quadrature" imaginary part [1], i.e., a complex circular vector $\mathbf{x}\sim\mathcal{CN}_M\left(\mathbf{m_x},\mathbf{C_x}\right)$ satisfying $\mathbf{C}_{\mathbf{x},\mathbf{x}^*}=E\left[\left(\mathbf{x}-\mathbf{m_x}\right)\left(\mathbf{x}-\mathbf{m_x}\right)^T\right]=\mathbf{0}$ and leading to a compact p.d.f. expression:

$$p\left(\mathbf{x};\mathbf{m_x},\mathbf{C_x}\right)=\frac{e^{-\left(\mathbf{x}-\mathbf{m_x}\right)^H\mathbf{C}_{\mathbf{x}}^{-1}\left(\mathbf{x}-\mathbf{m_x}\right)}}{\pi^N\left|\mathbf{C_x}\right|}\tag{58}$$

Among any other estimation problem based on the Gaussian complex circular observation model, mostly two different signal models are considered: the deterministic (conditional) signal model and the stochastic (unconditional) signal model [37]. The discussed signal models are Gaussian and the parameter of interest dependency is given by parameters which are connected with the expectation value in the deterministic case and with the covariance matrix in the stochastic one. Within this framework, the most studied observation is the following linear one, generalization of [1][38, (2.2)]:

$$\mathbf{x}^l = \mathbf{A}^l \left(\Xi^l\right) \ \mathbf{s}^l + \mathbf{n}^l, \quad l \in [1, L] \tag{59}$$

where L is the number of observation models, M is the number of signal sources, and:
• $\left(\Xi^l\right)^T = \left(\left(\varepsilon_1^l\right)^T, \ldots, \left(\varepsilon_M^l\right)^T\right)$ where ε_m^l is the vector of parameters of dimension P^l for the l^{th} observation model and the m^{th} source,
• $\mathbf{s}^l = \left(s_1^l, \ldots, s_M^l\right)^T$ is the vector of complex amplitudes of the M sources for the l^{th} observation model,
• $\mathbf{A}^l\left(\Xi^l\right) = \left[\mathbf{a}^l\left(\varepsilon_1^l\right), \ldots, \mathbf{a}^l\left(\varepsilon_M^l\right)\right]$ where $\mathbf{a}^l\left(\ \right)$ is a vector of N parametric functions depending on a vector of P^l parameters ε_m^l,
• $\mathbf{n}^l$ are Gaussian complex circular independent noises with spatially white covariance matrix: $\mathbf{C}_{\mathbf{n}^l} = \sigma_l^2 \mathbf{I}_N$, independent from the M signal sources.
Additionally:
• if $\left\{\mathbf{s}^1, \ldots, \mathbf{s}^L\right\}$ are deterministic vectors, then (59) is a deterministic observation model with p.d.f.:

$$p\left(\mathbf{x}^1, \ldots, \mathbf{x}^L\right) = \prod_{l=1}^{L} p\left(\mathbf{x}^l\right), \quad p\left(\mathbf{x}^l\right) = \frac{e^{-\frac{1}{\sigma_l^2}\|\mathbf{x}^l - \mathbf{A}^l(\Xi^l)\mathbf{s}^l\|^2}}{\pi^N\left(\sigma_l^2\right)^N} \tag{60}$$

• if $\left\{\mathbf{s}^1, \ldots, \mathbf{s}^L\right\}$ are mutually Gaussian complex circular i.i.d. with covariance matrix $\mathbf{C}_{\mathbf{s}^l}$, then (59) is a stochastic observation model with p.d.f.:

$$p\left(\mathbf{x}^1, \ldots, \mathbf{x}^L\right) = \prod_{l=1}^{L} p\left(\mathbf{x}^l\right), \ p\left(\mathbf{x}^l\right) = \frac{e^{-(\mathbf{x}^l)^H \mathbf{C}_{\mathbf{x}^l}^{-1} \mathbf{x}^l}}{\pi^N \left|\mathbf{C}_{\mathbf{x}^l}\right|}, \ \mathbf{C}_{\mathbf{x}^l} = \mathbf{A}^l\left(\Xi^l\right)\mathbf{C}_{\mathbf{s}^l}\mathbf{A}^l\left(\Xi^l\right)^H + \sigma_l^2\mathbf{I}_N \tag{61}$$

In most applications and reference papers [16][37][39] or textbooks [1][27][28][40], a more restrictive model is generally considered where the vector of N parametric functions $\mathbf{a}^l\left(\varepsilon_m^l\right)$, the noise power σ_l^2 and the source covariance matrix $\mathbf{C}_{\mathbf{s}^l}$ are invariant during the L observations:

$$\mathbf{a}^l\left(\varepsilon_m^l\right) = \mathbf{a}\left(\varepsilon_m\right), \ P^l = P, \ \sigma_l^2 = \sigma^2, \ \mathbf{C}_{\mathbf{s}^l} = \mathbf{C}_{\mathbf{s}}, \ \mathbf{C}_{\mathbf{x}^l} = \mathbf{C}_{\mathbf{x}} = \mathbf{A}\left(\Xi\right)\mathbf{C}_{\mathbf{s}}\mathbf{A}\left(\Xi\right)^H + \sigma^2\mathbf{I}_N. \tag{62}$$

Then the CRB for parameters of interest Ξ is of the following form [1][16][37]:

$$\mathbf{CRB}_{\Xi|\underline{\boldsymbol{\theta}}}^{-1}\left(\underline{\boldsymbol{\theta}}\right) = \frac{2L}{\sigma^2}\,\mathrm{Re}\left\{\mathbf{H}\left(\Xi\right) \odot \left(\mathbf{R}_{\mathbf{s}}^T \otimes \mathbf{1}_{P\times P}\right)\right\} \tag{63}$$

where $\mathbf{1}_{P\times P} \in \mathcal{M}_{\mathbb{C}}\left(P, P\right)$ is a matrix full of ones,

$$\begin{aligned}
\mathbf{H}\left(\Xi\right) &= \mathbf{D}\left(\Xi\right)^H \mathbf{\Pi}_{\mathbf{A}(\Xi)}^{\perp} \mathbf{D}\left(\Xi\right), \quad \mathbf{\Pi}_{\mathbf{A}(\Xi)}^{\perp} = \mathbf{I} - \mathbf{A}\left(\Xi\right)\left(\mathbf{A}\left(\Xi\right)^H \mathbf{A}\left(\Xi\right)\right)^{-1}\mathbf{A}\left(\Xi\right)^H \\
\mathbf{D}\left(\Xi\right) &= \left[\mathbf{D}_1\left(\varepsilon_1\right), \ldots, \mathbf{D}_M\left(\varepsilon_M\right)\right], \quad \mathbf{D}_m\left(\varepsilon_m\right) = \frac{\partial \mathbf{a}\left(\varepsilon_m\right)}{\partial \varepsilon^T},
\end{aligned}$$

$\underline{\boldsymbol{\theta}}$ and $\mathbf{R_S}$ depend on the type of observation model:
• if the observation model is deterministic (60), then [1][16][37]:

$$\underline{\boldsymbol{\theta}}^T = \left(\left(\mathbf{s}^1\right)^T, \ldots, \left(\mathbf{s}^L\right)^T, \left(\mathbf{s}^{*1}\right)^T, \ldots, \left(\mathbf{s}^{*L}\right)^T, \Xi, \sigma^2\right), \quad \mathbf{R}_{\mathbf{s}} = \frac{1}{L}\sum_{l=1}^{L} \mathbf{s}^l\left(\mathbf{s}^l\right)^H$$

• if the observation model is stochastic (61), then [1][37][39]:

$$\underline{\boldsymbol{\theta}}^T = \left(vec\left(\mathbf{C}_{\mathbf{s}}\right), \Xi, \sigma^2\right), \quad \mathbf{R}_{\mathbf{s}} = \mathbf{C}_{\mathbf{s}}\left(\mathbf{A}\left(\Xi\right)^H \mathbf{C}_{\mathbf{x}}^{-1} \mathbf{A}\left(\Xi\right)\right)\mathbf{C}_{\mathbf{s}}$$

Given expression (63) of the CRB, it is natural to explore the dependence of the CRB on the signal parameters. Typically, in parametric estimation of superimposed signals, the amplitudes of individual signals are considered to be nuisance parameters, the focus being on the signal parameter estimates Ξ. Therefore a relevant question is the following: what is the effect of the correlation matrix $\mathbf{R}_{\mathbf{s}}$ on the magnitude of CRB? Question that can be reformulated from a quantitative viewpoint as: which correlation matrix $\mathbf{R_S}$ leads to the best or worst CRB (63)?

1) Deterministic observation model:
For the deterministic observation model, the answer has been provided by authors in [38] but at the cost of a complicated derivation, whereas it can be derived within a few lines of rationale by application of the reparameterization inequality (45). If all the ε_l are known except the one of the m^{th} signal source ε_m, then the related CRB is given by (63) computed for the single m^{th} signal source:

$$\mathbf{CRB}^{-1}_{\varepsilon_m|\underline{\boldsymbol{\theta}_m}}\left(\underline{\boldsymbol{\theta}_m};\underline{\boldsymbol{\kappa}_m}\right) = \frac{2L}{\sigma^2}\operatorname{Re}\left\{\mathbf{H}\left(\varepsilon_m\right)\right\}\left(\mathbf{R_S}\right)_{m,m}, \quad \mathbf{H}\left(\varepsilon_m\right) = \mathbf{D}_m\left(\varepsilon_m\right)^H \boldsymbol{\Pi}^{\perp}_{\mathbf{A}(\boldsymbol{\Xi})}\mathbf{D}_m\left(\varepsilon_m\right)$$
$$\underline{\boldsymbol{\theta}_m}^T = \left(\left(\mathbf{s}^1\right)^T,\ldots,\left(\mathbf{s}^L\right)^T,\left(\mathbf{s}^{*1}\right)^T,\ldots,\left(\mathbf{s}^{*L}\right)^T,\varepsilon_m,\sigma^2\right), \quad \underline{\boldsymbol{\kappa}_m} = \boldsymbol{\Xi}\setminus\{\varepsilon_m\}$$

and according to (57):

$$\mathbf{CRB}_{\varepsilon_m|\underline{\boldsymbol{\theta}_m}}\left(\underline{\boldsymbol{\theta}_m};\underline{\boldsymbol{\kappa}_m}\right) \leq \mathbf{CRB}_{\varepsilon_m|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}\right) \quad \Leftrightarrow \quad \mathbf{CRB}^{-1}_{\varepsilon_m|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}\right) \leq \mathbf{CRB}^{-1}_{\varepsilon_m|\underline{\boldsymbol{\theta}_m}}\left(\underline{\boldsymbol{\theta}_m};\underline{\boldsymbol{\kappa}_m}\right)$$

If $\mathbf{R_S}$ is diagonal, then $\mathbf{CRB}^{-1}_{\boldsymbol{\Xi}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}\right)$ is block-diagonal and the above inequality becomes an equality: for each source the lowest (best) CRB is obtained when the sources amplitudes are uncorrelated.
Let us now come back to the general observation model (59) and assume that only the N parametric functions $\mathbf{a}^l\left(\;\right)$ are invariant during the L observations: $\mathbf{a}^l\left(\varepsilon^l_m\right) = \mathbf{a}\left(\varepsilon^l_m\right)$, $P^l = P$. Then if $\overline{\boldsymbol{\Xi}} = \frac{1}{L}\sum_{l=1}^{L}\boldsymbol{\Xi}^l$, due to the independence of the L observations and the nature of the FIM (covariance matrix), a few lines of calculus show that:

$$\mathbf{CRB}_{\overline{\boldsymbol{\Xi}}|\underline{\boldsymbol{\eta}}}\left(\underline{\boldsymbol{\eta}}\right) = \frac{1}{L^2}\left(\sum_{l=1}^{L}\mathbf{CRB}_{\boldsymbol{\Xi}^l|\underline{\boldsymbol{\theta}^l}}\left(\underline{\boldsymbol{\theta}^l}\right)\right), \quad \mathbf{CRB}^{-1}_{\boldsymbol{\Xi}^l|\underline{\boldsymbol{\theta}^l}}\left(\underline{\boldsymbol{\theta}^l}\right) = \frac{2}{\sigma^2_l}\operatorname{Re}\left\{\mathbf{H}\left(\boldsymbol{\Xi}^l\right)\odot\left(\left(\mathbf{s}^l\right)^*\left(\mathbf{s}^l\right)^T\otimes\mathbf{1}_{P\times P}\right)\right\}$$
$$\underline{\boldsymbol{\eta}}^T = \left(\left(\underline{\boldsymbol{\theta}}^1\right)^T,\ldots,\left(\underline{\boldsymbol{\theta}}^L\right)^T\right), \quad \left(\underline{\boldsymbol{\theta}^l}\right)^T = \left(\left(\mathbf{s}^l\right)^T,\left(\mathbf{s}^l\right)^H,\boldsymbol{\Xi}^l,\sigma^2_l\right).$$

Actually, the usual invariance hypotheses (62) are a set of parameters constraints:

$$\varepsilon^l_m = \varepsilon^1_m = \varepsilon_m,\; P^l = P^1 = P,\; \sigma^2_l = \left(\sigma^2\right)^1 = \sigma^2,\; l\in[1,L].$$

Under this set of parameters constraints $\overline{\boldsymbol{\Xi}} = \boldsymbol{\Xi}$ and $\mathbf{CRB}_{\overline{\boldsymbol{\Xi}}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}\right)$ expression is given by (63). Then, according to (45):

$$\mathbf{CRB}_{\overline{\boldsymbol{\Xi}}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}\right) \leq \mathbf{CRB}_{\overline{\boldsymbol{\Xi}}|\underline{\boldsymbol{\eta}}}\left(\underline{\boldsymbol{\eta}}\right),\; \forall\underline{\boldsymbol{\theta}},\forall\underline{\boldsymbol{\eta}}.$$

For the particular value of $\underline{\boldsymbol{\eta}}$ satisfying $\varepsilon^l_m = \varepsilon^1_m = \varepsilon_m$, $P^l = P^1 = P$, $\sigma^2_l = \left(\sigma^2\right)^1 = \sigma^2$, $l\in[1,L]$, we therefore obtain:

$$\frac{\sigma^2}{2L}\operatorname{Re}\left\{\mathbf{H}\left(\boldsymbol{\Xi}\right)\odot\left(\mathbf{R}^T_{\mathbf{s}}\otimes\mathbf{1}_{P\times P}\right)\right\}^{-1} \leq \frac{1}{L^2}\left(\sum_{l=1}^{L}\frac{\sigma^2}{2}\operatorname{Re}\left\{\mathbf{H}\left(\boldsymbol{\Xi}\right)\odot\left(\left(\mathbf{s}^l\right)^*\left(\mathbf{s}^l\right)^T\otimes\mathbf{1}_{P\times P}\right)\right\}^{-1}\right)$$

or equivalently:

$$\operatorname{Re}\left\{\mathbf{H}\left(\boldsymbol{\Xi}\right)\odot\left(\mathbf{R}^T_{\mathbf{s}}\otimes\mathbf{1}_{P\times P}\right)\right\}^{-1} \leq \frac{1}{L}\left(\sum_{l=1}^{L}\operatorname{Re}\left\{\mathbf{H}\left(\boldsymbol{\Xi}\right)\odot\left(\left(\mathbf{s}^l\right)^*\left(\mathbf{s}^l\right)^T\otimes\mathbf{1}_{P\times P}\right)\right\}^{-1}\right)$$

Since $\operatorname{Re}\left\{\mathbf{H}\left(\boldsymbol{\Xi}\right)\odot\left(\left(\mathbf{s}^l\right)^*\left(\mathbf{s}^l\right)^T\otimes\mathbf{1}_{P\times P}\right)\right\}^{-1} \geq 0$, $\forall l\in[1,L]$, then:

$$\operatorname{Re}\left\{\mathbf{H}\left(\boldsymbol{\Xi}\right)\odot\left(\mathbf{R}^T_{\mathbf{s}}\otimes\mathbf{1}_{P\times P}\right)\right\}^{-1} \leq \operatorname{Re}\left\{\mathbf{H}\left(\boldsymbol{\Xi}\right)\odot\left(\left(\mathbf{s}^{l_{\max}}\right)^*\left(\mathbf{s}^{l_{\max}}\right)^T\otimes\mathbf{1}_{P\times P}\right)\right\}^{-1}$$
$$l_{\max} = \arg\left(\max_{l\in[1,L]}\left\{\operatorname{Re}\left\{\mathbf{H}\left(\boldsymbol{\Xi}\right)\odot\left(\left(\mathbf{s}^l\right)^*\left(\mathbf{s}^l\right)^T\otimes\mathbf{1}_{P\times P}\right)\right\}^{-1}\right\}\right),$$

and there is equality if and only if $\mathbf{R_s} = \left(\mathbf{s}^{l_{\max}}\right)^*\left(\mathbf{s}^{l_{\max}}\right)^T$, i.e. if and only if $\mathbf{R_s}$ is a rank one matrix. Therefore, for each source:
• the highest (worst) CRB is obtained when the sources amplitudes are fully correlated,
• the lowest (best) CRB is obtained when the sources amplitudes are uncorrelated.

2) Stochastic observation model:
Due to the identity of CRB expressions for both deterministic and stochastic observation models (63), results obtained above in the deterministic case are applicable to $\mathbf{R_s}$ where:

$$\mathbf{R_s} = \mathbf{C_s}\left(\mathbf{A}^H\mathbf{C_x}^{-1}\mathbf{A}\right)\mathbf{C_s},\ \mathbf{C_x} = \mathbf{AC_sA}^H + \sigma^2\mathbf{I}_N,\ \mathbf{A} \equiv \mathbf{A}\left(\mathbf{\Xi}\right)$$

First, since $rank\left(\mathbf{R_s}\right) = rank\left(\mathbf{C_s}\right)$ under the standard implicit assumption of independent steering vectors [1][40], the highest (worst) CRB is obtained when $rank\left(\mathbf{R_s}\right) = 1 = rank\left(\mathbf{C_s}\right)$, i.e. when the sources amplitudes are fully correlated. Second, for each source the lowest (best) CRB is obtained when $\mathbf{R_s}$ is diagonal: $\mathbf{R_s} = \boldsymbol{\lambda}$, $\left(\boldsymbol{\lambda}\right)_{m,l} = \left(\boldsymbol{\lambda}\right)_{m,m}\delta_m^l$. Then, necessarily, $\mathbf{C_s}$ is an invertible (full rank) Hermitian matrix satisfying $\boldsymbol{\lambda} = \mathbf{C_s}\left(\mathbf{A}^H\mathbf{C_x}^{-1}\mathbf{A}\right)\mathbf{C_s}^H$, what means that $\mathbf{C_s}$ diagonalizes $\left(\mathbf{A}^H\mathbf{C_x}^{-1}\mathbf{A}\right)$, which is a positive definite Hermitian matrix. As a consequence $\frac{\mathbf{C_s}}{\alpha}$, $\alpha > 0$, must be positive definite, unitary and Hermitian:

$$\frac{\mathbf{C_s}}{\alpha} = \mathbf{U} > \mathbf{0}\ /\ \mathbf{U}^H\mathbf{U} = \mathbf{I} \text{ and } \mathbf{U} = \mathbf{U}^H \quad \Rightarrow \quad \mathbf{U} > 0 \text{ and } \mathbf{U}^2 = \mathbf{I} \quad \Rightarrow \quad \mathbf{U} = \mathbf{I} \quad \Rightarrow \quad \mathbf{C_s} = \alpha\mathbf{I}$$

Reciprocally, if $\mathbf{C_s} = \alpha\mathbf{I}$ then

$$\boldsymbol{\lambda} = \mathbf{C_s}\left(\mathbf{A}^H\mathbf{C_x}^{-1}\mathbf{A}\right)\mathbf{C_s}^H \Leftrightarrow \frac{\boldsymbol{\lambda}}{\alpha^2} = \mathbf{A}^H\mathbf{C_x}^{-1}\mathbf{A},$$

what means that $\mathbf{A}$ partially diagonalizes $\mathbf{C_x}^{-1}$. Therefore the column vectors of $\mathbf{A}$ must be orthogonal. We believe that this result is new to the best of our knowledge, as it does not appear in [1, p 950], whereas the condition $\left(rank\left(\mathbf{C_s}\right) = 1\right)$ to reach the worst CRB case is mentioned [1, p 954] with reference to [38]. To summarize, for each source:
- the lowest (best) CRB is obtained when $\mathbf{C_s} = \alpha\mathbf{I}$ and the column vectors of $\mathbf{A}\left(\mathbf{\Xi}\right)$ are orthogonal,
- the highest (worst) CRB is obtained when the sources amplitudes are fully correlated $\left(rank\left(\mathbf{C_s}\right) = 1\right)$.

VI. Singular FIM and unidentifiability

In this section for sake of legibility, $\left\{\mathbf{f}\left(\mathbf{x}\right) = \mathbf{g}\left(\mathbf{x}\right)\right\} \equiv \left\{\mathbf{f}\left(\mathbf{x}\right) = \mathbf{g}\left(\mathbf{x}\right),\ \text{almost everywhere}\right\}$, that is, the set of $\mathbf{x}$ values where the equality does not hold has measure zero (Lebesgue measure).

A. On unidentifiability and singular FIM

A definition of unidentifiability in a wide sense is given in [35, p24]: if $\mathbf{x}$ is distributed according to $p\left(\mathbf{x};\underline{\boldsymbol{\theta}}\right)$, then $\underline{\boldsymbol{\theta}}$ is said to be unidentifiable on the basis of $\mathbf{x}$ if there exist $\underline{\boldsymbol{\theta}}^1 \neq \underline{\boldsymbol{\theta}}^2$ for which $p\left(\mathbf{x};\underline{\boldsymbol{\theta}}^1\right) = p\left(\mathbf{x};\underline{\boldsymbol{\theta}}^2\right)$. Nevertheless, both from a practical and theoretical point of view, as we will show it in the following, it is necessary to restrict the scope of the previous definition to a subset $\mathcal{C}$ of the parameter space Θ, leading to the following definition: if $\mathbf{x}$ is distributed according to $p\left(\mathbf{x};\underline{\boldsymbol{\theta}}\right)$, then $\underline{\boldsymbol{\theta}}$ is said to be unidentifiable (on the basis of $\mathbf{x}$) over the subset $\mathcal{C}$ of the parameter space Θ if $\forall\underline{\boldsymbol{\theta}}^1, \underline{\boldsymbol{\theta}}^2 \in \mathcal{C},\ p\left(\mathbf{x};\underline{\boldsymbol{\theta}}^1\right) = p\left(\mathbf{x};\underline{\boldsymbol{\theta}}^2\right)$.
Let us now consider a subset $\mathcal{C}$ of the parameter space Θ defined by K $(1 \leq K \leq P-1)$ non redundant constraints $\mathcal{C} = \left\{\underline{\boldsymbol{\theta}} \in \Theta\ /\ \underline{\mathbf{f}}\left(\underline{\boldsymbol{\theta}}\right) = \mathbf{0} \in \mathbb{C}^K\right\}$, and let us assume that $p\left(\mathbf{x};\underline{\boldsymbol{\theta}}\right)$ is unidentifiable over $\mathcal{C}$. Then:

$$\forall\underline{\boldsymbol{\theta}}^0, \underline{\boldsymbol{\theta}}^0 + d\underline{\boldsymbol{\theta}} \in \mathcal{C} : \frac{\partial p\left(\mathbf{x};\underline{\boldsymbol{\theta}}^0\right)}{\partial\underline{\boldsymbol{\theta}}^T}d\underline{\boldsymbol{\theta}} = 0 \text{ and } d\underline{\boldsymbol{\theta}} = \mathbf{U}_{\underline{\boldsymbol{\theta}}^0}d\boldsymbol{\lambda},\quad \mathbf{U}_{\underline{\boldsymbol{\theta}}^0} \in \mathcal{M}_{\mathbb{C}}\left(P, P-K\right),\ d\boldsymbol{\lambda} \in \mathbb{R}^{P-K},$$

where $\mathbf{U}_{\underline{\boldsymbol{\theta}}^0}$ is a basis of $\ker\left\{\frac{\partial\underline{\mathbf{f}}\left(\underline{\boldsymbol{\theta}}^0\right)}{\partial\underline{\boldsymbol{\theta}}^T}\right\}$. Therefore, if $p\left(\mathbf{x};\underline{\boldsymbol{\theta}}\right)$ is unidentifiable over $\mathcal{C}$:

$$\forall\underline{\boldsymbol{\theta}}^0 \in \mathcal{C}, \forall\boldsymbol{\lambda} \in \mathbb{R}^{P-K},\ \frac{\partial p\left(\mathbf{x};\underline{\boldsymbol{\theta}}^0\right)}{\partial\underline{\boldsymbol{\theta}}^T}\mathbf{U}_{\underline{\boldsymbol{\theta}}^0}\boldsymbol{\lambda} = \mathbf{0} \ \Leftrightarrow\ \forall\underline{\boldsymbol{\theta}}^0 \in \mathcal{C},\ \frac{\mathbf{U}_{\underline{\boldsymbol{\theta}}^0}^T}{p\left(\mathbf{x};\underline{\boldsymbol{\theta}}^0\right)}\frac{\partial p\left(\mathbf{x};\underline{\boldsymbol{\theta}}^0\right)}{\partial\underline{\boldsymbol{\theta}}} = \mathbf{0}.$$

As a consequence, if $p\left(\mathbf{x};\underline{\boldsymbol{\theta}}\right)$ is unidentifiable over $\mathcal{C}$:

$$\forall\underline{\boldsymbol{\theta}}^0 \in \mathcal{C},\ E_{\underline{\boldsymbol{\theta}}^0}\left[\left(\mathbf{U}_{\underline{\boldsymbol{\theta}}^0}^T\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\theta}}^0\right)}{\partial\underline{\boldsymbol{\theta}}}\right)\left(\mathbf{U}_{\underline{\boldsymbol{\theta}}^0}^T\frac{\partial\ln p\left(\mathbf{x};\underline{\boldsymbol{\theta}}^0\right)}{\partial\underline{\boldsymbol{\theta}}}\right)^H\right] = \mathbf{U}_{\underline{\boldsymbol{\theta}}^0}^T\mathbf{F}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}^0\right)\mathbf{U}_{\underline{\boldsymbol{\theta}}^0}^* = \mathbf{0}$$

what means that, $\forall\underline{\boldsymbol{\theta}}^0 \in \mathcal{C}$, the FIM $\mathbf{F}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}^0\right)$ is singular with $rank\left(\mathbf{F}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}^0\right)\right) \leq K$.

B. On singular FIM and unidentifiability

Let us assume that the FIM $\mathbf{F}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}^0\right)$ is singular with rank $1 \leq K \leq P-1$ for $\underline{\boldsymbol{\theta}}^0 \in \mathcal{D} \subset \Theta$. Then $\mathbf{F}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}^0\right) = \mathbf{M}_{\underline{\boldsymbol{\theta}}^0}\mathbf{D}_{\underline{\boldsymbol{\theta}}^0}\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}^H$, $\mathbf{M}_{\underline{\boldsymbol{\theta}}^0} \in \mathcal{M}_{\mathbb{C}}(P,K)$, $\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}^H\mathbf{M}_{\underline{\boldsymbol{\theta}}^0} = \mathbf{I}_K$, $\mathbf{D}_{\underline{\boldsymbol{\theta}}^0} = \mathbf{Diag}\left(d_1\left(\underline{\boldsymbol{\theta}}^0\right),\ldots,d_K\left(\underline{\boldsymbol{\theta}}^0\right)\right)$, $d_1\left(\underline{\boldsymbol{\theta}}^0\right) \geq \ldots \geq d_K\left(\underline{\boldsymbol{\theta}}^0\right) > 0$. Let $\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}^{\perp} \in \mathcal{M}_{\mathbb{C}}(P,P-K)$ be the orthonormal matrix such that its column vectors form a basis of $span\left\{\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}\right\}^{\perp}$. Then:

$$\left(\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}^{\perp}\right)^H \mathbf{F}_{\underline{\boldsymbol{\theta}}|\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{\theta}}^0\right)\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}^{\perp} = E_{\underline{\boldsymbol{\theta}}^0}\left[\left(\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}^{\perp}\right)^H \frac{\partial \ln p(\mathbf{x};\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\theta}}}\left(\left(\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}^{\perp}\right)^H \frac{\partial \ln p(\mathbf{x};\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\theta}}}\right)^H\right] = \mathbf{0}$$
$$\Downarrow$$
$$E_{\underline{\boldsymbol{\theta}}^0}\left[\left\|\left(\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}^{\perp}\right)^H \frac{\partial \ln p(\mathbf{x};\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\theta}}}\right\|^2\right] = 0 \quad \Leftrightarrow \quad \left(\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}^{\perp}\right)^H \frac{\partial p(\mathbf{x};\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\theta}}} = \mathbf{0}$$

Therefore:

$$\forall d\boldsymbol{\lambda} \in \mathbb{C}^{P-K},\ \frac{\partial p\left(\mathbf{x};\underline{\boldsymbol{\theta}}^0\right)}{\partial \underline{\boldsymbol{\theta}}^T}\left(\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}^{\perp}\right)^* d\boldsymbol{\lambda} = 0 \tag{64}$$

Let $\mathcal{C} = \left\{\underline{\boldsymbol{\theta}} \in \Theta \ / \ \underline{\mathbf{f}}\left(\underline{\boldsymbol{\theta}}\right) = \mathbf{0} \in \mathbb{C}^K\right\}$ such that $\ker\left\{\frac{\partial \underline{\mathbf{f}}(\underline{\boldsymbol{\theta}}^0)}{\partial \underline{\boldsymbol{\theta}}^T}\right\} = span\left\{\left(\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}^{\perp}\right)^*\right\}, \forall \underline{\boldsymbol{\theta}}^0 \in \mathcal{D}$. Then (64) can be rewritten as:

$$\frac{\partial p\left(\mathbf{x};\underline{\boldsymbol{\theta}}^0\right)}{\partial \underline{\boldsymbol{\theta}}^T} d\underline{\boldsymbol{\theta}} = 0 \text{ and } d\underline{\boldsymbol{\theta}} = \left(\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}^{\perp}\right)^* d\boldsymbol{\lambda} \quad \Leftrightarrow \quad p\left(\mathbf{x};\underline{\boldsymbol{\theta}}^0 + d\underline{\boldsymbol{\theta}}\right) = p\left(\mathbf{x};\underline{\boldsymbol{\theta}}^0\right) \text{ and } \underline{\boldsymbol{\theta}}^0, \underline{\boldsymbol{\theta}}^0 + d\underline{\boldsymbol{\theta}} \in \mathcal{C}$$

that is: $\underline{\boldsymbol{\theta}}$ is unidentifiable over $\mathcal{C}$ and $\mathcal{D} \subset \mathcal{C}$. Additionally:

$$\frac{\partial \underline{\mathbf{f}}\left(\underline{\boldsymbol{\theta}}^0\right)}{\partial \underline{\boldsymbol{\theta}}^T}\left(\mathbf{M}_{\underline{\boldsymbol{\theta}}^0}^{\perp}\right)^* = \mathbf{0} \ \Rightarrow \ \mathcal{D} \subset \mathcal{C} = \left\{\underline{\boldsymbol{\theta}} \in \Theta \ / \ \underline{\mathbf{f}}\left(\underline{\boldsymbol{\theta}}\right) = \mathbf{0} \in \mathbb{C}^K,\ \frac{\partial \underline{\mathbf{f}}\left(\underline{\boldsymbol{\theta}}^0\right)}{\partial \underline{\boldsymbol{\theta}}^T} = \mathbf{M}_{\underline{\boldsymbol{\theta}}^0}^T,\ \forall \underline{\boldsymbol{\theta}}^0 \in \mathcal{D}\right\}$$

C. Biased estimates

Actually, biased estimates are the general case of estimates since they simply satisfy:

$$E_{\underline{\boldsymbol{\theta}}}\left[\underline{\widehat{\mathbf{g}\left(\underline{\boldsymbol{\theta}}^0\right)}}(\mathbf{x})\right] = \underline{\mathbf{h}}\left(\underline{\boldsymbol{\theta}}\right) = \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}\right) + \underline{\mathbf{b}}\left(\underline{\boldsymbol{\theta}}\right),\ \forall \underline{\boldsymbol{\theta}} \in \Theta \tag{65}$$

where $\underline{\mathbf{b}}\left(\underline{\boldsymbol{\theta}}\right)$ is called the bias. When $\underline{\mathbf{b}}\left(\underline{\boldsymbol{\theta}}\right) = \mathbf{0}$, then the estimates are unbiased and (65) reduces to (19). If $\underline{\widehat{\mathbf{g}\left(\underline{\boldsymbol{\theta}}^0\right)}}(\mathbf{x})$ satisfies (65) then:

$$\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\underline{\widehat{\mathbf{g}\left(\underline{\boldsymbol{\theta}}^0\right)}}(\mathbf{x}) - \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)\right\}\right) = \mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\underline{\widehat{\mathbf{g}\left(\underline{\boldsymbol{\theta}}^0\right)}}(\mathbf{x}) - \underline{\mathbf{h}}\left(\underline{\boldsymbol{\theta}}^0\right)\right\}\right) + \underline{\mathbf{b}}\left(\underline{\boldsymbol{\theta}}^0\right)^* \underline{\mathbf{b}}\left(\underline{\boldsymbol{\theta}}^0\right)^T$$

Therefore, it is well known that, since $\underline{\mathbf{b}}\left(\underline{\boldsymbol{\theta}}^0\right)^* \underline{\mathbf{b}}\left(\underline{\boldsymbol{\theta}}^0\right)^T \geq \mathbf{0}$ [1]:

$$\min_{\underline{\widehat{\mathbf{g}(\underline{\boldsymbol{\theta}}^0)}}(\mathbf{x})}\left\{\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\underline{\widehat{\mathbf{g}\left(\underline{\boldsymbol{\theta}}^0\right)}}(\mathbf{x}) - \underline{\mathbf{g}}\left(\underline{\boldsymbol{\theta}}^0\right)\right\}\right)\right\} \text{ under } E_{\underline{\boldsymbol{\theta}}}\left[\underline{\widehat{\mathbf{g}\left(\underline{\boldsymbol{\theta}}^0\right)}}(\mathbf{x})\right] = \underline{\mathbf{h}}\left(\underline{\boldsymbol{\theta}}\right),\ \forall \underline{\boldsymbol{\theta}} \in \Theta$$
$$\Updownarrow$$
$$\begin{cases} \min_{\underline{\widehat{\mathbf{h}(\underline{\boldsymbol{\theta}}^0)}}(\mathbf{x})}\left\{\mathbf{G}_{\underline{\boldsymbol{\theta}}^0}\left(\left\{\underline{\widehat{\mathbf{h}\left(\underline{\boldsymbol{\theta}}^0\right)}}(\mathbf{x}) - \underline{\mathbf{h}}\left(\underline{\boldsymbol{\theta}}^0\right)\right\}\right)\right\} \text{ under } E_{\underline{\boldsymbol{\theta}}}\left[\underline{\widehat{\mathbf{h}\left(\underline{\boldsymbol{\theta}}^0\right)}}(\mathbf{x})\right] = \underline{\mathbf{h}}\left(\underline{\boldsymbol{\theta}}\right),\ \forall \underline{\boldsymbol{\theta}} \in \Theta \\ \underline{\widehat{\mathbf{g}\left(\underline{\boldsymbol{\theta}}^0\right)}}(\mathbf{x})_{\text{eff}} = \underline{\widehat{\mathbf{h}\left(\underline{\boldsymbol{\theta}}^0\right)}}(\mathbf{x})_{\text{eff}} \end{cases}, \tag{66}$$

i.e., the minimization of the MSE of biased estimates is obtained by the minimization of the MSE of unbiased estimates, provided the vector of functions to be estimated is updated in order to incorporate the bias.

D. Summary

Finally we have shown the following equivalence - new to the best of our knowledge - between a singular FIM and unidentifiability over a subset $\mathcal{C}$ of the parameter space Θ:

$$\left\{\begin{array}{l}\mathbf{F}_{\underline{\theta}|\underline{\theta}}\left(\underline{\theta}^0\right)=\mathbf{M}_{\underline{\theta}^0}\mathbf{D}_{\underline{\theta}^0}\mathbf{M}_{\underline{\theta}^0}^H,\ \mathbf{M}_{\underline{\theta}^0}\in\mathcal{M}_{\mathbb{C}}\left(P,K\right),\ \mathbf{D}_{\underline{\theta}^0}=\mathbf{Diag}\left(d_1\left(\underline{\theta}^0\right),\ldots,d_K\left(\underline{\theta}^0\right)\right)\\ \mathbf{M}_{\underline{\theta}^0}^H\mathbf{M}_{\underline{\theta}^0}=\mathbf{I}_K,\ d_1\left(\underline{\theta}^0\right)\geq\ldots\geq d_K\left(\underline{\theta}^0\right)>0,\ \underline{\theta}^0\in\mathcal{D}\subset\Theta\end{array}\right.$$
$$\Updownarrow \tag{67}$$
$$\left\{\begin{array}{l}\mathcal{D}\subset\mathcal{C}=\left\{\underline{\theta}\in\Theta\ /\ \underline{\mathbf{f}}\left(\underline{\theta}\right)=\mathbf{0}\in\mathbb{C}^K,\ 1\leq K\leq P-1,\ \frac{\partial\underline{\mathbf{f}}(\underline{\theta}^0)}{\partial\underline{\theta}^T}=\mathbf{M}_{\underline{\theta}^0}^T,\ \forall\underline{\theta}^0\in\mathcal{D}\right\}\\ p\left(\mathbf{x};\underline{\theta}\right)\ \text{is unidentifiable over}\ \mathcal{C}\end{array}\right.$$

Additionally, $\forall\underline{\mathbf{g}}\left(\underline{\theta}\right)\ /\ \frac{\partial\underline{\mathbf{g}}(\underline{\theta}^0)^T}{\partial\underline{\theta}}\in span\left\{\mathbf{M}_{\underline{\theta}^0}\right\}$, the CRB exists and is given by lemma 1. Then $\frac{\partial\underline{\mathbf{g}}(\underline{\theta}^0)^T}{\partial\underline{\theta}}=\mathbf{M}_{\underline{\theta}^0}\mathbf{B},$ $\mathbf{B}\in\mathcal{M}_{\mathbb{C}}\left(K,K\right)$, and:

$$\forall\underline{\theta}^0,\underline{\theta}^0+d\underline{\theta}\in\mathcal{C},\ \frac{\partial\underline{\mathbf{g}}\left(\underline{\theta}^0\right)}{\partial\underline{\theta}^T}d\underline{\theta}=\mathbf{B}^T\mathbf{M}_{\underline{\theta}^0}^T\left(\mathbf{M}_{\underline{\theta}^0}^{\perp}\right)^*d\boldsymbol{\lambda}=\mathbf{0},$$

that is: $\underline{\mathbf{g}}\left(\underline{\theta}\right)$ is constant over $\mathcal{C}$. Therefore, the general result is the following:

Proposition 1: The FIM is always singular on a subset $\mathcal{C}$ of the parameter space deriving from a set of equality constraints. The singularity of the FIM over $\mathcal{C}$ is equivalent to the unidentifiability of $\underline{\theta}$ over $\mathcal{C}$, what leads to the inexistence of the CRB over $\mathcal{C}$ resulting in:
- the non existence of locally unbiased estimates of $\underline{\mathbf{g}}\left(\underline{\theta}\right)$ when $\underline{\theta}\in\mathcal{C}$, except if $\underline{\mathbf{g}}\left(\underline{\theta}\right)$ is constant over $\mathcal{C}$,
- an unbounded CRB at the vicinity of any $\underline{\theta}\in\mathcal{C}$.

This result encompasses - and clarifies - previous works in the field [19][33] that have only partly investigated the relationship between unidentifiability and FIM singularity.
Last, let us consider the subset $\mathcal{R}\subset\Theta$ defined by the set of $P-K$ following constraints:

$$\underline{\mathbf{r}}\left(\underline{\theta}\right)=\mathbf{0}\ /\ \mathbf{M}_{\underline{\theta}^0}^*\ \text{is a basis of}\ \ker\left\{\frac{\partial\underline{\mathbf{r}}\left(\underline{\theta}^0\right)}{\partial\underline{\theta}^T}\right\} \tag{68}$$

Equations (68) are a regularization equations set, regularization in the sense that this set allows to define the greatest subset - with respect to inclusion - of the parameter space Θ where an unbiased estimate exist for any function of the unknown parameters [33]. Indeed the constrained CRB is then given by (see §III-C2):

$$\forall\underline{\theta}^0\in\mathcal{R},\ \mathbf{CRB}_{\underline{\theta}|\underline{\theta}}^c\left(\underline{\theta}^0\right)=\mathbf{M}_{\underline{\theta}^0}\mathbf{D}_{\underline{\theta}^0}^{-1}\mathbf{M}_{\underline{\theta}^0}^H>0$$

This regularization equations (68) are rather theoretical results than practical ones, as it is unlikely that an analytical expression for eigenvectors of the FIM exists in most cases. Anyway, its existence suggests that, when a FIM is singular with a rank $K<P$, it is worth looking for a subset of the initial observation model for which a "meaningful" intra-parameters relation of the form:

$$\underline{\theta}\left(\underline{\omega}\right)=\left(\underline{\omega}^T,\underline{\varepsilon}^T=\underline{\mathbf{r}}'\left(\underline{\omega}\right)^T\right),\ \dim\left\{\underline{\theta}\right\}=P,\ \dim\left\{\underline{\omega}\right\}=K,\ \dim\left\{\underline{\varepsilon}\right\}=P-K, \tag{69}$$

exists and satisfies:

$$\left[\begin{array}{c}\mathbf{I}_{P-K}\\ \frac{\partial\underline{\mathbf{h}}(\underline{\omega}^0)}{\partial\underline{\omega}^T}\end{array}\right]=\mathbf{M}_{\underline{\theta}^0}^*\mathbf{\Gamma}_{\underline{\theta}^0},\ \forall\underline{\theta}^0\in\mathcal{D},\ \mathbf{\Gamma}_{\underline{\theta}^0}\in\mathcal{M}_{\mathbb{C}}\left(P-K,P-K\right),\ \left|\mathbf{\Gamma}_{\underline{\theta}^0}\right|\neq 0,$$

and to work with this subset of observation models, whenever it is possible and relevant (see [33, §IV] for an example).

VII. Conclusion

We have provided in the present paper a unique simple derivation of the CRB whatever the nature (complex or real) of the unknown parameters based on the norm minimization approach. The norm minimization approach makes it easy to understand more advanced concepts like FIM singularity, constrained CRB, regularity conditions, which become straightforward corollaries of the derivation.

We have introduced a general reparameterization/constraints inequality that can be useful both for practical applications (system design and optimization) and theoretical derivations (CRB inequalities). Moreover, we have explicitly formulated the relationship between unidentifiability and a singular FIM, which is a key feature for system design. Indeed, it is of first importance to know if the parameters of interest are identifiable, that is, if they can be estimated (with or without bias) whatever their values.

Last, this paper should allow the reader to understand easily the derivation of any kind - conditional/unconditional, biased/unbiased - of deterministic MSE lower bounds for real/complex parameters [10][14], since all known bounds on the MSE are different solutions of the same norm minimization problem under sets of appropriate linear constraints.

References

[1] H.L. Van Trees, *Optimum Array Processing*, New-York, Wiley-Interscience, 2002

[2] H. Cramér, *Mathematical Methods of Statistics. Princeton*. NJ: Princeton Univ. Press, 1946.

[3] M. Fréchet, "Sur l'extension de certaines évaluations statistiques au cas de petits échantillons", Rev. Int. Stat, vol. 11, pp. 182-205, 1943

[4] C.R. Rao, "Information and accuracy attainable in the estimation of statistical parameters," Bull. Calcutta Math. Soc., vol. 37, pp. 81-91, 1945.

[5] A. Renaux, P. Forster, E. Chaumette, P. Larzabal, "On the High RSB CML Estimator Full Statistical Characterization", IEEE Trans. on SP, vol 54, no. 12, Dec. 2006

[6] R.A. Fisher, "On an absolute criterion for fitting frequency curves", Mess. of Math., vol 41, p. 155-160, 1912

[7] E.W. Barankin, "Locally best unbiased estimates", Ann. Math. Stat., vol. 20, no. 4, pp. 477-501, 1949.

[8] D.G. Chapman and H. Robbins, "Minimum variance estimation without regularity assumptions", Ann. Math. Stat., vol. 22, no. 4, pp. 581-586, 1951.

[9] R. McAulay and L.P. Seidman, "A useful form of the Barankin lower bound and its application to PPM threshold analysis", IEEE Trans. on IT, vol. 15, no. 2, pp. 273-279, Mar. 1969.

[10] E. Chaumette, J. Galy, A. Quinlan, P. Larzabal, "A New Barankin Bound Approximation for the Prediction of the Threshold Region Performance of Maximum-Likelihood Estimators", IEEE Trans. on SP, vol. 56, no. 11, pp. 5319-5333, Nov. 2008

[11] P. Forster, P. Larzabal, "On lower bounds for deterministic parameter estimation", in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., pp.1137-1140, 2002.

[12] K. Todros and J. Tabrikian, "General Classes of Performance Lower Bounds for Parameter Estimation-Part I: Non-Bayesian Bounds for Unbiased Estimators", IEEE Trans. on IT, vol. 56, no. 10, Oct. 2010

[13] D.C. Rife and R.R. Boorstyn, "Single tone parameter estimation from discrete-time observations", IEEE Trans. on IT, vol. 20, no.5, pp. 591-598, Sep. 1974.

[14] E. Chaumette, P. Larzabal, P. Forster, "On the Influence of a Detection Step on Lower Bounds for Deterministic Parameters Estimation", IEEE Trans. on SP, vol 53, pp. 4080-4090, Nov. 2005

[15] A.K. Jagannatham, B.D. Rao, "Cramér–Rao Lower Bound for Constrained Complex Parameters", IEEE SP letters, vol. 11, no. 11, Nov. 2004

[16] S. F. Yau and Y. Bresler, "A Compact Cram&-Rao Bound Expression for Parametric Estimation of Superimposed Signals", IEEE Trans. on SP, vol. 40, no. 5, pp 1226-1230, May 1992.

[17] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," IEEE Trans. on IT, vol. 26, no. 6, pp.1285–1301, Nov. 1990

[18] T.L. Marzetta, "A simple derivation of the constrained multiple parameter Cramér–Rao bound", IEEE Trans. on SP, vol. 41, no. 6, pp. 2247–2249, Jun. 1993.

[19] P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular information matrices", IEEE Trans. on SP, vol. 49, no1, pp. 87-90, Jan. 2001

[20] P. Stoica and B. C. Ng, "On the Cramér–Rao bound under parametric constraints", IEEE SP Letters, vol. 5, no. 7, pp. 177-179, Jul. 1998

[21] A. Van Den Bos, "A Cramér-Rao Lower Bound for Complex Parameters", IEEE Trans. on SP, vol. 42, no. 10, Oct. 1994

[22] C. R. Blyth, "Necessary and sufficient conditions for inequalities of Cramér-Rao type," Ann. Math. Stat., vol. 2, no. 3, pp. 464-473, 1974.

[23] N. Bourbaki, *Elements of mathematics. Algebra: Algebraic structures. Linear algebra, 1*, Addison-Wesley (1974)

[24] M. Hazewinkel, *Encyclopaedia of Mathematics*, Springer-Verlag Berlin Heidelberg New York, 2002

[25] R.A. Horn, C.R. Johnson, *Matrix Analysis*. Cambridge University Press, 1999

[26] Eric W. Weisstein, *MathWorld–A Wolfram Web Resource*, http://mathworld.wolfram.com/HermitianInnerProduct.html

[27] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[28] L. Scharf, *Statistical Signal Processing*, Addison-Wiley, 1991

[29] D. H. Brandwood, “A complex gradient operator and its application in adaptive array theory” ,Proc. IEE, vol. 130, no. 1, pp. 11–16, Feb. 1983.
[30] J.S. Abel, ”A bound on mean-square-estimate error”, IEEE Trans. on IT, vol. 39, no. 5, pp. 1675-1680, Sep. 1993.
[31] F.E. Glave, ”A new look at the Barankin lower bound”, IEEE Trans. on IT, vol. 18, no. 3, pp. 349-356, May 1972
[32] E. Chaumette and P. Larzabal, ”Three useful algebraic lemmas”, Technical Note, ENS-CACHAN, France.
[33] Z. Ben-Haim, Y. C. Eldar, ”On the Constrained Cramér–Rao Bound With a Singular Fisher Information Matrix”, IEEE SP Letters, vol. 16, no. 6, Jun. 2009
[34] D.A.S. Fraser and I. Guttman, ”Bhattacharyya bounds without regularity assumptions,” Ann. Math. Stat., vol. 23, no. 4., pp. 629-632, 1952.
[35] E. L. Lehmann and G. Casella, *Theory of Point Estimation (2nd ed.)*. Springer 1998
[36] W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, New York, 1976
[37] P. Stoica and A. Nehorai, ”Performances study of conditional and unconditional direction of arrival estimation”, IEEE Trans. on SP, vol. 38, no. 10, pp. 1783-1795, Oct. 1990.
[38] S. F. Yau and Y. Bresler, ”Worst Case Cramér-Rao Bounds for Parametric Estimation of Superimposed Signals”, IEEE Trans. on SP, vol. 40. no. 12, pp 2973-2986, Dec. 1992
[39] P. Stoica, E. G. Larson, and A. B. Gershman, ”The stochastic CRB for array processing: A textbook derivation”, IEEE SP Letters, vol.8, pp. 148-150, May 2001
[40] Ottersten and M. Viberg and P. Stoica and A. Nehorai, ”Exact and Large Sample Maximum Likelihood Techniques for Parameter Estimation and Detection in Array Processing”, Radar Array Processing”, §4, 1993, Springer-Verlag

I. Detection and Parameter Estimation of Extended Targets Using the Generalized Monopulse Estimator (IEEE TAES)

IEEE Trans. on AES, 48(4): 3389-3417, 2012.

Detection and Parameter Estimation of Extended Targets Using the Generalized Monopulse Estimator

Eric Chaumette, Ulrich Nickel and Pascal Larzabal

Abstract

In many radar applications extended targets appear together with point targets. An important problem in particular for tracking algorithms is to rapidly detect these extended targets and to determine their centroid and shape. We present a detector for extended targets and Maximum-Likelihood estimators for the centroid and the extension which are based only on the outputs of the generalized monopulse ratio. The generalized monopulse ratio is based on the sum and difference beams generated from the digital outputs of arbitrary subarrays of a planar array. It is therefore applicable to various kinds of modern array antennas. The extended target is characterized by an ellipse in the azimuth-elevation plane. The statistical performance of the centroid and the extension estimators is characterized by analytically their mean and variance. Numerical simulations indicate that these statistical characterizations are accurate and that small extended targets can be detected early. The statistical performance measures can therefore also be used for parametric system studies.

Key words: Generalized monopulse, extended target tracking, statistical performance description, maximum likelihood estimator

I. List of Symbols

$\mathbf{a}$ array response vector of plane wave. For a planar array and omni-directional element pattern $a_i = e^{j2\pi f \frac{x_i u + y_i v}{c}}$
u, v components of the unit direction vector of the source in the x, y antenna coordinate system, (direction cosines)
x_i, y_i positions of array antenna elements
c velocity of light
f frequency
$\mathbf{z}$ array output data snapshot
$\mathbf{Z}$ array output data matrix, $\mathbf{Z} = (\mathbf{z}_1, .., \mathbf{z}_K)$
$vec\{\mathbf{Z}\}$ arranging a matrix in vector form by stacking columns on top
b target complex amplitude
$\mathbf{d}$ general vector for difference beamforming
$\mathbf{w}$ general vector for sum beamforming
D difference beam output, e.g. $D = \mathbf{d}^H \mathbf{z}$
$\mathbf{D}$ vector of multiple difference beams for planar or volume arrays, $\mathbf{D} = (D_1, ..D_M)^T$
$\mathbf{G}$ general symbol for covariance matrix of any kind of beam outputs
$\mathbf{\Gamma}$ general symbol for the sample covariance matrix of any kind of beam outputs
K number of time samples
N number of array elements
$\mathbf{n}$ noise or interference vector
R monopulse ratio, $R = Re\{\frac{D}{S}\}$ for difference, sum beam outputs D, S

Eric Chaumette is with the Office National d'Etudes et de Recherches Aérospatiales, the French Aerospace Lab, DEMR/TSI, Chemin de la Hunière 91120 Palaiseau, France. E-mail: eric.chaumette@onera.fr
U. Nickel is with Fraunhofer-FKIE, Neuenahrer Str. 20, 53343 Wachtberg, Germany. E-mail: ulrich.nickel@fkie.fraunhofer.de
P. Larzabal is with SATIE/CNRS, Ecole Normale Supérieure de Cachan, 61 avenue du Président Wilson - 94235 Cachan Cedex, France. E-mail: pascal.larzabal@satie.ens-cachan.fr
This work has been partly funded by the European Network of excellence NEWCOM#

$\mathbf{R}$ vector of monopulse ratios for multiple difference beams
S sum beam output
$\boldsymbol{\theta}$ general parameter vector
$\mathbf{z}^H$ conjugate-transposed vector
$\overline{z}$ complex conjugate number
$\otimes$ Kronecker matrix product

II. Introduction

Position and length estimation of extended targets is of interest in many radar applications, e.g. tracking of groups of targets, convoys on a road or for determination of the intercept point in missile guidance. This is a particular problem for tracking applications: The extended target has to be detected, measurements have to be associated correctly, and the trajectory of the whole extended object must be estimated. The straight forward approach at the signal processing level is to apply high resolution methods (conventional or superresolution methods) to resolve a complex extended target into individual scatterers. This is the approach in radar imaging. Typically, multiple beams are formed or sophisticated superresolution methods are applied. This requires an array antenna with digital beamforming and the techniques are often based on Maximum-Likelihood estimation of multiple targets, e.g. [1], or non-parametric methods like the MUSIC method [2]. An overview of these methods is given in [3] Chapter 3. In general, these are quite time consuming methods.

In many cases resolution into individual scatterers is not necessary and the object of interest has the shape of a line, e.g. in ballistic missile defense and in GMTI radar convoy tracking. In these cases one simply wants to estimate the centroid and the extent of such a line target. In [4] the problem of convoy tracking was solved by estimating the target length at signal processing level by a number of beams with varying separation. This was done for a linear array by varying the beams in azimuth angle only. The extension to azimuth and elevation measurement would require a more time consuming search. In [5] a solution to the extended target problem was sought at the tracking level by modeling the distribution of the tracking input data in Cartesian coordinates by a Gaussian. The uncertainty ellipse is then taken as a measure of the extension. These two papers document the recent interest in extended target estimation for tracking applications.

Monopulse processing for angle estimation is implemented in most modern radar systems. The monopulse ratio is a quantity which is very sensitive against deviations from the point target assumption. Therefore it is a very sensitive criterion to detect extended targets. There have been attempts in the 70s and 80s years to estimate the parameters of an extended target directly from the monopulse outputs [6], [7]. These estimators were found for a single sum-difference beam pair and their statistical description was either approximative [6] or too complex to be used as an analysis tool [7]. More recently, in [8], [9] monopulse estimates in different range cells have been combined to solve the problem of unresolved point targets representing possibly an extended target.

The distribution of the monopulse estimates for multiple targets have been investigated in a variety of papers [10], [11], [12], [13], [14]. What is missing is a neat formula for an estimator of the centroid and the extent. In particular, for a 2D-monopulse antenna with azimuth and elevation estimates one would like to estimate directly the orientation of an extended target in two dimensions. Such an estimator does not exist yet.

The monopulse principle [15]] has been extended to antennas of arbitrary dimension (linear, planar, volume ...) with digital beamforming based on arbitrary subarrays [16]. This general monopulse form includes also all kinds of adaptive monopulse using adaptive sum and difference beams for interference suppression. For this general monopulse procedure, a statistical description of the distribution of the estimates has been given for various signal models in [17]. In this paper we use this statistical description to derive closed form estimators for the length and centroid of extended targets. We also derive a test statistic for detecting the presence of an extended target. Such a criterion has not been presented yet. From an applications viewpoint early detection of an extended target (an approaching extended target at far range) is of interest to initiate some actions.

By using the generalized monopulse procedure these results can be applied for monopulse ratios formed from an arbitrary number of difference beams. It is thus applicable for planar or volume arrays or adaptive arrays. In particular, it delivers in this way directly the orientation of a line target in the azimuth-elevation angle space which is in some application of importance. Superresolution methods would require in this case a multi-dimensional search.

This paper is organized as follows. Section III describes the considered problem and the generalized monopulse procedure based on a set of difference beams. Both cases, the complex and real monopulse estimator are considered. In Section IV and V, we derive and characterize our new estimator for the centroid and the extent of an extended target. The underlying signal model is that of a number of independent point scatterers with Rayleigh amplitude fluctuation (Swerling I-II model). For complex targets this may be a reasonable approximation. It is known that this model does not very well describe deterministic effects like glint. However, a serious glint effect over a longer time window is not very likely, except if it is artificially produced (deception jammer). In addition we propose a test statistic to detect the presence of an extended target in Section VI. Section VII presents numerical studies showing the performance of the proposed detector and estimators and conclusions are presented finally.

III. Problem Formulation and Generalized Monopulse Estimation (GME)

A. Observation model of an extended Swerling I-II target

The reader is advised to read sections II and III of paper [17] where GME is introduced in more details. In this section, to avoid overlapping with [17], we will briefly introduce the synthesis of the generalized beam outputs $\mathbf{B}_k = \left((\mathbf{D}_k)^T, S_k\right)^T$ used in GME.
The observations consists of K independent data snapshots $\mathbf{z}(t_k)$ at time $\{t_1, \ldots, t_K\}$ of an array of N antennas. If we have a target with parameters $\boldsymbol{\theta}$, then the (space/time/subarrayed) [17] snapshot at time t_k has the general structure:

$$\mathbf{z}(t_k) = b(t_k)\,\mathbf{a}(\boldsymbol{\theta}) + \mathbf{n}(t_k),\ \ \mathbf{z}(t_k) = (z_1(t_k), \ldots, z_N(t_k))^T,\ \ \mathbf{n}(t_k) = (n_1(t), \ldots, n_N(t_k))^T, \tag{1}$$

where $\mathbf{a}(\boldsymbol{\theta})$ describes the plane wave model, $b(t_k)$ the complex target amplitude at time t_k, and $\mathbf{n}(t_k)$ the noise component which may be composed of receiver noise plus external interference. For the time series we write shortly $\mathbf{z}(t_k) = \mathbf{z}_k$, and $\mathbf{z}_k = b_k\mathbf{a}(\boldsymbol{\theta}) + \mathbf{n}_k$. For each snapshot $\mathbf{z}(t_k)$ we form M difference beams $\mathbf{D}(t_k) = (D_1(t_k), \ldots, D_M(t_k))^T$ and a sum beam $S(t_k)$ resulting in a vector of $M+1$ beams $\mathbf{B}(t_k) = \left(\mathbf{D}(t_k)^T, S(t_k)\right)^T$. Indicating the elements of the time series by indices as before we obtain a beam output vector $\mathbf{B}_k = \left((\mathbf{D}_k)^T, S_k\right)^T$. This beam output vector has the structure:

$$\mathbf{B}_k = b_k\boldsymbol{\alpha}(\boldsymbol{\theta}) + \mathbf{v}_k, \quad \boldsymbol{\alpha}(\boldsymbol{\theta}) = \left(\boldsymbol{\alpha}_{\mathbf{D}}^T(\boldsymbol{\theta}), \alpha_S(\boldsymbol{\theta})\right)^T = \left(\mathbf{d}_1^H\mathbf{a}(\boldsymbol{\theta}), \ldots, \mathbf{d}_M^H\mathbf{a}(\boldsymbol{\theta}), \mathbf{w}^H\mathbf{a}(\boldsymbol{\theta})\right)^T \tag{2}$$

where variables $\mathbf{d}_1$,...,$\mathbf{d}_M$ are the weight vectors for difference beam forming and $\mathbf{w}$ for the sum beam forming, and the noise contribution is $\mathbf{v}_k = (\mathbf{d}_1, \ldots, \mathbf{d}_M, \mathbf{w})^H\,\mathbf{n}_k$. The mean of this beam output vector is denoted by

$$E\{\mathbf{B}_k\} = \mathbf{t}_{\mathbf{B}_k} = \left(\mathbf{t}_{\mathbf{D}_k}^T, t_{S_k}\right)^T = \left(E\{\mathbf{D}_k\}^T, E\{S_k\}\right)^T,$$

the covariance is assumed to be time invariant and is denoted by

$$cov\{\mathbf{B}_k\} = cov\{\mathbf{B}\} = \mathbf{G} = \begin{bmatrix} \mathbf{G}_{\mathbf{D}} & \mathbf{G}_{\mathbf{D}S} \\ \mathbf{G}_{\mathbf{D}S}^H & G_S \end{bmatrix}$$

Covariance $\mathbf{G}$ decomposes into $\mathbf{G} = \mathbf{G}^{\text{signal}} + \mathbf{G}^{\mathbf{v}}$. We assume that the noise and interference contributions given by the vector $\mathbf{v}$ (or $\mathbf{n}$) are independent identical Gaussian complex circular distributed.
We now model for K observations the vector of sum and difference beams as complex Gaussian distributed. This corresponds to a Swerling-I target model if the snapshots are taken at a long time interval (scan-to-scan) or to a Swerling-II model if snapshots are taken at short interval (pulse-to-pulse). For a modern multifunction radar with different tasks performed in time multiplex the distinction between scans is no more appropriate. In fact, the validity of this assumption depends very much on the variability of the target RCS and the radar revisit time and has to be checked for specific application. However, in [17] it was shown that the error ellipses for different fluctuation models have the same orientation in space and differ only little in size. This indicates a certain robustness against model deviations. For a set of N independent point scatterers with power $\sigma_n^2, n \in [1, N]$, the observed signals at $(M+1)$ beams output are thus modelled as (2):

$$\mathbf{B}_k = \sum_{n=1}^{N} b_k(n)\,\boldsymbol{\alpha}(\boldsymbol{\theta}_n) + \mathbf{v}_k, \qquad E\left\{|b_k(n)|^2\right\} = \sigma_n^2 \tag{3}$$

with p.d.f.:

$$p\left(\mathbf{B}_1,\ldots,\mathbf{B}_K\right)=\prod_{k=1}^{K}p_{\mathcal{CN}_{M+1}}\left(\mathbf{B}_k;\mathbf{0},\mathbf{G}\right)=\frac{e^{-Ktr\left(\mathbf{G}^{-1}\boldsymbol{\Gamma}\right)}}{\left(\pi^{M+1}\left|\mathbf{G}\right|\right)^K},\tag{4}$$

$$\boldsymbol{\Gamma}=\frac{1}{K}\sum_{k=1}^{K}\mathbf{B}_k\mathbf{B}_k^H=\left[\begin{array}{cc}\boldsymbol{\Gamma}_{\mathbf{D}} & \boldsymbol{\Gamma}_{\mathbf{D}S}\\ \boldsymbol{\Gamma}_{\mathbf{D}S}^H & \Gamma_S\end{array}\right],\quad \mathbf{G}=\left[\begin{array}{cc}\mathbf{G}_{\mathbf{D}} & \mathbf{G}_{\mathbf{D}S}\\ \mathbf{G}_{\mathbf{D}S}^H & G_S\end{array}\right]=\sum_{n=1}^{N}\sigma_n^2\boldsymbol{\alpha}\left(\theta_n\right)\boldsymbol{\alpha}\left(\theta_n\right)^H+\mathbf{G}^{\mathbf{v}}$$

By introducing the vector of true monopulse ratios [17]:

$$\mathbf{R}\left(\boldsymbol{\theta}\right)=\left(\frac{\alpha_{d_1}\left(\boldsymbol{\theta}\right)}{\alpha_s\left(\boldsymbol{\theta}\right)},\ldots,\frac{\alpha_{d_M}\left(\boldsymbol{\theta}\right)}{\alpha_s\left(\boldsymbol{\theta}\right)}\right)^T,\tag{5}$$

we obtain:

$$\begin{aligned}\mathbf{G}&=P_s\left[\begin{array}{cc}E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\mathbf{R}\left(\boldsymbol{\theta}\right)^H\right\} & E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}\\ E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}^H & 1\end{array}\right]+\mathbf{G}^{\mathbf{v}}\\ \mathbf{G}&=P_s\left[\begin{array}{cc}E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}^H+cov_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\} & E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}\\ E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}^H & 1\end{array}\right]+\mathbf{G}^{\mathbf{v}}\end{aligned}\tag{6}$$

where:

$$P_s=\sum_{n=1}^{N}\sigma_n^2\left|\alpha_s\left(\boldsymbol{\theta}_n\right)\right|^2\tag{7}$$

$$E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}=\sum_{n=1}^{N}\frac{\sigma_n^2\left|\alpha_s\left(\boldsymbol{\theta}_n\right)\right|^2}{P_s}\mathbf{R}\left(\boldsymbol{\theta}_n\right)\tag{8}$$

$$cov_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}=E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\mathbf{R}\left(\boldsymbol{\theta}\right)^H\right\}-E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}^H\tag{9}$$

Let us notice that P_s and $E_{\boldsymbol{\theta}}\left\{\right\}$ can be straightforwardly extended to any continuous distribution of point scatterers having a power density $\rho\left(\boldsymbol{\theta}\right)$, i.e. such as $\sigma^2\left(\boldsymbol{\theta}\right)=\rho\left(\boldsymbol{\theta}\right)d\boldsymbol{\theta}$:

$$\begin{aligned}P_s&=\lim_{N\to\infty}\sum_{n=1}^{N}\rho\left(\boldsymbol{\theta}_n\right)d\boldsymbol{\theta}\left|\alpha_s\left(\boldsymbol{\theta}_n\right)\right|^2=\int_{\Theta}\rho\left(\boldsymbol{\theta}\right)\left|\alpha_s\left(\boldsymbol{\theta}\right)\right|^2d\boldsymbol{\theta}\\ E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}&=\lim_{N\to\infty}\sum_{n=1}^{N}\frac{\sigma_n^2\left|\alpha_s\left(\boldsymbol{\theta}\right)\right|^2}{P_s}\mathbf{R}\left(\boldsymbol{\theta}_n\right)=\int_{\Theta}\mathbf{R}\left(\boldsymbol{\theta}\right)p\left(\boldsymbol{\theta}\right)d\boldsymbol{\theta}\\ p\left(\boldsymbol{\theta}\right)&=\frac{\rho\left(\boldsymbol{\theta}\right)\left|\alpha_s\left(\boldsymbol{\theta}\right)\right|^2}{\int_{\Theta}\rho\left(\boldsymbol{\theta}\right)\left|\alpha_s\left(\boldsymbol{\theta}\right)\right|^2d\boldsymbol{\theta}},\quad\rho\left(\boldsymbol{\theta}\right)\geq0,\quad\boldsymbol{\theta}\in\boldsymbol{\Theta}\end{aligned}\tag{10}$$

where Θ is a subset of $\mathbb{R}^M$ representing the M-dimensional angular domain of the extended target. This result generalizes the continuous single angle uniform distribution introduced in [7]. Additionally, definitions (7)-(9) allow as well to take into account a distribution of N distinct targets consisting of a single point scatterer.

B. Angular centroid and extent as a function of the true monopulse ratios

In standard monopulse estimation [7][15], it is assumed that:

• the M true monopulse ratios $R_m\left(\boldsymbol{\theta}\right)$ are real,

• the M monopulse curves (deviation angle function) are decoupled and linear in the vicinity of the sum beam boresight,

leading to simple relations between the M (off-boresight) angular centroids $\left(E_{\boldsymbol{\theta}}\left\{\Delta\boldsymbol{\theta}\right\}\right)_m$ and (off-boresight) extents $\sqrt{\left(cov_{\boldsymbol{\theta}}\left\{\Delta\boldsymbol{\theta}\right\}\right)_{m,m}}$, and the knowledge of $E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}$ and $cov_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}$:

$$\Delta\boldsymbol{\theta}=\boldsymbol{\theta}-\boldsymbol{\theta}_0\simeq\boldsymbol{\lambda}\mathbf{R}\left(\boldsymbol{\theta}\right),\ \lambda_{m,l}=\lambda_{l,l}\delta_l^m\ \Rightarrow\ E_{\boldsymbol{\theta}}\left\{\Delta\boldsymbol{\theta}\right\}=\boldsymbol{\lambda}E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\},\ cov_{\boldsymbol{\theta}}\left\{\Delta\boldsymbol{\theta}\right\}=\boldsymbol{\lambda}cov_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}\boldsymbol{\lambda},$$

where $\Delta\boldsymbol{\theta}$ is the off-boresight angle and $\boldsymbol{\theta}_0$ is an initial estimate of the desired angles (steering direction). These relations can be generalized when the M monopulse curves are no longer decoupled [16]:

$$\text{(GME-1)} \quad : \quad \Delta\boldsymbol{\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}_0 = -\mathbf{C}\left(\operatorname{Re}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\} - \boldsymbol{\mu}\right) \tag{11}$$

$$\text{(GME-2)} \quad : \quad \Delta\boldsymbol{\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}_0 = -\operatorname{Re}\left\{\mathbf{C}\left(\mathbf{R}\left(\boldsymbol{\theta}\right) - \boldsymbol{\mu}\right)\right\} \tag{12}$$

where (GME-1) is designed for real-valued $\mathbf{R}\left(\boldsymbol{\theta}\right)$, (GME-2) is designed for complex-valued $\mathbf{R}\left(\boldsymbol{\theta}\right)$; and $\mathbf{C}, \boldsymbol{\mu}$ are (real or complex) correction quantities. . The correction quantities are calculated from the response of the antenna to a plane wave $\mathbf{a}_0 = \mathbf{a}\left(\boldsymbol{\theta}_0\right)$ and the beamforming vectors $\mathbf{d}_1$,...,$\mathbf{d}_M, \mathbf{w}$. This allows using this procedure for varying adaptive weights. Following [16] correction values are calculated for (GME-1) as:

$$\mu_i = \operatorname{Re}\left\{\frac{\mathbf{d}_i^H \mathbf{a}_0}{\mathbf{w}^H \mathbf{a}_0}\right\}, \quad i = 1 \ldots M$$

and the elements of the inverse correction matrix by:

$$\left(\mathbf{C}^{-1}\right)_{i,k} = \frac{\operatorname{Re}\left\{\mathbf{d}_i^H \mathbf{a}_{k,0} \mathbf{a}_0^H \mathbf{w} + \mathbf{d}_i^H \mathbf{a}_0 \mathbf{a}_{k,0}^H \mathbf{w}\right\}}{\left|\mathbf{w}^H \mathbf{a}_0\right|^2} - \mu_i \operatorname{Re}\left\{\frac{\mathbf{w}^H \mathbf{a}_{k,0}}{\mathbf{w}^H \mathbf{a}_0}\right\}, \quad i, k = 1 \ldots M$$

with the abbreviation $\mathbf{a}_{k,0} = \frac{\partial \mathbf{a}}{\partial u_k}\left(\boldsymbol{\theta}_0\right)$ for the directions characterized by a vector $\boldsymbol{\theta} = \left(u_1, \ldots, u_M\right)$. For (GME-2) it can be shown that these correction values can be used by omitting real part operation. Then we have:

$$\text{(GME-1)} \quad : \quad E_{\boldsymbol{\theta}}\left\{\Delta\boldsymbol{\theta}\right\} = -\mathbf{C}\left(E_{\boldsymbol{\theta}}\left\{\operatorname{Re}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}\right\} - \boldsymbol{\mu}\right)$$

$$\text{(GME-2)} \quad : \quad E_{\boldsymbol{\theta}}\left\{\Delta\boldsymbol{\theta}\right\} = -\operatorname{Re}\left\{\mathbf{C}\left(E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\} - \boldsymbol{\mu}\right)\right\}$$

and (see Appendix IX-A for details):

$$\text{(GME-1)} \quad : \quad cov_{\boldsymbol{\theta}}\left\{\Delta\boldsymbol{\theta}\right\} = \mathbf{C} cov_{\boldsymbol{\theta}}\left\{\operatorname{Re}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}\right\} \mathbf{C}^T \tag{13}$$

$$\begin{aligned}\text{(GME-2)} \quad : \quad cov_{\boldsymbol{\theta}}\left\{\Delta\boldsymbol{\theta}\right\} = &\frac{1}{2}\operatorname{Re}\left\{\mathbf{C} cov_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\} \mathbf{C}^H\right\} + \frac{1}{2}\operatorname{Re}\left\{\mathbf{C} E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\mathbf{R}\left(\boldsymbol{\theta}\right)^T\right\} \mathbf{C}^T\right\} \\ &-\frac{1}{2}\operatorname{Re}\left\{\mathbf{C}\left(E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\} E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}^T\right)\mathbf{C}^T\right\}\end{aligned} \tag{14}$$

IV. GME OF THE ANGULAR CENTROID AND EXTENT OF A SWERLING I-II TARGET

A. *MLE of angular centroid and extent*

The application of the ML invariant principle [18] to (11)(12)(13)(14) leads to:

$$\text{(GME-1)} : \mathbf{R} \in \mathbb{R}^M, \begin{cases} \widehat{E_{\boldsymbol{\theta}}\left\{\Delta\boldsymbol{\theta}\right\}} = -\mathbf{C}\left(\widehat{E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}} - \boldsymbol{\mu}\right) \\ \widehat{cov_{\boldsymbol{\theta}}\left\{\Delta\boldsymbol{\theta}\right\}} = \mathbf{C}\widehat{cov_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}}\mathbf{C}^T \end{cases} \tag{15}$$

and

$$\text{(GME-2)} : \mathbf{R} \in \mathbb{C}^M, \begin{cases} \widehat{E_{\boldsymbol{\theta}}\left\{\Delta\boldsymbol{\theta}\right\}} = -\operatorname{Re}\left\{\mathbf{C}\left(\widehat{E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}} - \boldsymbol{\mu}\right)\right\} \\ \widehat{cov_{\boldsymbol{\theta}}\left\{\Delta\boldsymbol{\theta}\right\}} = \frac{1}{2}\operatorname{Re}\left\{\mathbf{C}\widehat{cov_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}}\mathbf{C}^H\right\} + \frac{1}{2}\operatorname{Re}\left\{\mathbf{C}\widehat{E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\mathbf{R}\left(\boldsymbol{\theta}\right)^T\right\}}\mathbf{C}^T\right\} \\ \qquad -\frac{1}{2}\operatorname{Re}\left\{\mathbf{C}\left(\widehat{E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}}\widehat{E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}}^T\right)\mathbf{C}^T\right\} \end{cases} \tag{16}$$

where $\widehat{x}$ stands for the MLE of unknown value x (scalar, vector or matrix).
Therefore MLEs of angular centroid and extent and their statistical prediction are thoroughly characterized by the knowledge and the statistical prediction of $\widehat{E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}}$, $\widehat{cov_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}}$ and $\widehat{E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\mathbf{R}\left(\boldsymbol{\theta}\right)^T\right\}}$.

B. MLE of $E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}$, $cov_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}$ and P_s

1) The standard single angle case in presence of spatially white noise of known power:
In Appendix IX-B1, we present a simple derivation scheme that can be used not only to obtain the results previously released by Milstein [7] without demonstration (" .. solutions are straightforward but exceedingly lengthy ...") but also its generalization to the estimation of a complex monopulse ratio centroid $E_\theta\left\{R\left(\theta\right)\right\}$. The main results are the following:

• if $R\left(\theta\right), E_\theta\left\{R\left(\theta\right)\right\} \in \mathbb{R}$, then:

$$\mathbf{G} = \begin{bmatrix} P_s E_\theta\left\{R\left(\theta\right)\right\}^2 + P_s cov_\theta\left\{R\left(\theta\right)\right\} & P_s E_\theta\left\{R\left(\theta\right)\right\} \\ P_s E_\theta\left\{R\left(\theta\right)\right\} & P_s \end{bmatrix} + G_S^{\mathbf{v}}\mathbf{I}$$

$$\widehat{P_s} = \Gamma_{22} - G_S^{\mathbf{v}} \tag{17}$$

$$\widehat{E_\theta\left\{R\left(\theta\right)\right\}} = \mathrm{Re}\left\{\frac{\Gamma_{12}}{\Gamma_{22} - G_S^{\mathbf{v}}}\right\} \tag{18}$$

$$\widehat{cov_\theta\left\{R\left(\theta\right)\right\}} = \frac{\Gamma_{11} - G_S^{\mathbf{v}}}{\Gamma_{22} - G_S^{\mathbf{v}}} - \left(\widehat{E_\theta\left\{R\left(\theta\right)\right\}}\right)^2 \tag{19}$$

• if $R\left(\theta\right), E_\theta\left\{R\left(\theta\right)\right\} \in \mathbb{C}$, then:

$$\mathbf{G} = \begin{bmatrix} P_s \left|E_\theta\left\{R\left(\theta\right)\right\}\right|^2 + P_s cov_\theta\left\{R\left(\theta\right)\right\} & P_s E_\theta\left\{R\left(\theta\right)\right\} \\ P_s E_\theta\left\{R\left(\theta\right)\right\}^* & P_s \end{bmatrix} + G_S^{\mathbf{v}}\mathbf{I}$$

$$\widehat{P_s} = \Gamma_{22} - G_S^{\mathbf{v}} \tag{20}$$

$$\widehat{E_\theta\left\{R\left(\theta\right)\right\}} = \frac{\Gamma_{12}}{\Gamma_{22} - G_S^{\mathbf{v}}} \tag{21}$$

$$\widehat{cov_\theta\left\{R\left(\theta\right)\right\}} = \frac{\Gamma_{11} - G_S^{\mathbf{v}}}{\Gamma_{22} - G_S^{\mathbf{v}}} - \left|\widehat{E_\theta\left\{R\left(\theta\right)\right\}}\right|^2 \tag{22}$$

Additionally it can be shown (see Appendix IX-B3) that the noise power $G_S^{\mathbf{v}}$ and the variance term $cov_\theta\left\{R\left(\theta\right)\right\}$ can not be simultaneously identified. Therefore, the hypothesis of a known noise covariance matrix is compulsory. Nevertheless, as mentioned by Milstein [7], it is a realistic hypothesis in radar, since the thermal noise power can be measured separately from the target power, and in principle can be measured as accurately as desired.

2) The general M angles case in presence of noise of known covariance matrix:
Let us denote by $\boldsymbol{\omega}$ the vector of unknown parameters.

• If $\mathbf{R}\left(\boldsymbol{\theta}\right), E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\} \in \mathbb{C}^M$, then one can notice that in this case, the MLE's derived in the previous section lead to:

$$\widehat{\mathbf{G}} = \mathbf{G}\left(\widehat{\boldsymbol{\omega}}\right) = \mathbf{\Gamma}. \tag{23}$$

Therefore a generalization of the MLE's derived in the single angle case compliant with (23) is:

$$\widehat{P_s} = \Gamma_S - G_S^{\mathbf{v}} \tag{24}$$

$$\widehat{E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}} = \frac{\mathbf{\Gamma}_{\mathbf{D}S} - \mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}}{\Gamma_S - G_S^{\mathbf{v}}} \tag{25}$$

$$\widehat{cov_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}} = \frac{\mathbf{\Gamma}_{\mathbf{D}} - \mathbf{G}_{\mathbf{D}}^{\mathbf{v}}}{\Gamma_S - G_S^{\mathbf{v}}} - \left(\widehat{E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}}\right)\left(\widehat{E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}}\right)^H \tag{26}$$

Indeed, as these estimators verify (23), they are also the solutions of the following constraints:

$$tr\left(\frac{\partial \mathbf{G}\left(\boldsymbol{\omega}\right)^{-1}}{\partial \omega_l}\left(\mathbf{\Gamma} - \mathbf{G}\left(\boldsymbol{\omega}\right)\right)\right) = 0, \tag{27}$$

provided that $\widehat{\mathbf{G}}$ is invertible, that is, provided that $\mathbf{\Gamma}$ is invertible. And the solutions of (27) are the MLEs since (4):

$$\frac{\partial \ln p\left(\mathbf{B}_1, \ldots, \mathbf{B}_K\right)}{\partial \omega_l} = K\left[tr\left(\frac{\partial \mathbf{G}\left(\boldsymbol{\omega}\right)^{-1}}{\partial \omega_l}\left(\mathbf{\Gamma} - \mathbf{G}\left(\boldsymbol{\omega}\right)\right)\right)\right].$$

Therefore estimators (24)-(26) are exactly the MLEs.

• If $\mathbf{R}(\boldsymbol{\theta}), E_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\} \in \mathbb{R}^M$, then one can notice that, in this case, the MLE's derived in the previous section lead to:

$$\widehat{\mathbf{G}} = \mathbf{G}(\widehat{\boldsymbol{\omega}}) = \mathrm{Re}\{\boldsymbol{\Gamma}\} \tag{28}$$

Therefore a generalization of the MLE's derived in the single angle case compliant with (28) is:

$$\begin{aligned} \widehat{P_s} &= \Gamma_S - G_S^{\mathbf{v}} && (29)\\ \widehat{E_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\}} &= \mathrm{Re}\left\{\frac{\boldsymbol{\Gamma}_{\mathbf{D}S} - \mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}}{\Gamma_S - G_S^{\mathbf{v}}}\right\} && (30)\\ \widehat{cov_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\}} &= \mathrm{Re}\left\{\frac{\boldsymbol{\Gamma}_{\mathbf{D}} - \mathbf{G}_{\mathbf{D}}^{\mathbf{v}}}{\Gamma_S - G_S^{\mathbf{v}}}\right\} - \left(\widehat{E_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\}}\right)\left(\widehat{E_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\}}\right)^T && (31) \end{aligned}$$

Indeed, as these estimators verify (28), they are also the solutions of the following constraints:

$$tr\left(\frac{\partial \mathbf{G}(\boldsymbol{\omega})^{-1}}{\partial \omega_l}\mathrm{Re}\{\boldsymbol{\Gamma} - \mathbf{G}(\boldsymbol{\omega})\}\right) = 0, \tag{32}$$

provided that $\widehat{\mathbf{G}} = \mathrm{Re}\{\boldsymbol{\Gamma}\} + j\,\mathrm{Im}\{\mathbf{G}^{\mathbf{v}}\}$ is invertible (see Appendix IX-B4 for details). And the solutions of (32) are the MLEs since (4):

$$\frac{\partial \ln p(\mathbf{B}_1, \ldots, \mathbf{B}_K)}{\partial \omega_l} = Ktr\left(\frac{\partial \mathbf{G}(\boldsymbol{\omega})^{-1}}{\partial \omega_l}\mathrm{Re}\{\boldsymbol{\Gamma} - \mathbf{G}(\boldsymbol{\omega})\}\right).$$

Therefore estimators (29)-(31) are exactly the MLEs.

C. MLE of $E_{\boldsymbol{\theta}}\left\{\mathbf{R}(\boldsymbol{\theta})\mathbf{R}(\boldsymbol{\theta})^T\right\}, \mathbf{R}(\boldsymbol{\theta}) \in \mathbb{C}^M$

If $\mathbf{R}(\boldsymbol{\theta}), E_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\} \in \mathbb{C}^M$, then the MLE of $E_{\boldsymbol{\theta}}\left\{\mathbf{R}(\boldsymbol{\theta})\mathbf{R}(\boldsymbol{\theta})^T\right\}$ can be derived by analogy with $E_{\boldsymbol{\theta}}\left\{\widehat{\mathbf{R}(\boldsymbol{\theta})\mathbf{R}(\boldsymbol{\theta})}\right.$ Indeed as (24):

$$E_{\boldsymbol{\theta}}\left\{\widehat{\mathbf{R}(\boldsymbol{\theta})\mathbf{R}(\boldsymbol{\theta})^H}\right\} = \frac{\boldsymbol{\Gamma}_{\mathbf{D}} - \mathbf{G}_{\mathbf{D}}^{\mathbf{v}}}{\Gamma_S - G_S^{\mathbf{v}}} = \frac{\frac{1}{K}\sum_{k=1}^{K}\mathbf{D}_k\mathbf{D}_k^H - E\left\{\mathbf{D}_k^{\mathbf{v}}(\mathbf{D}_k^{\mathbf{v}})^H\right\}}{\Gamma_S - G_S^{\mathbf{v}}}$$

then:

$$E_{\boldsymbol{\theta}}\left\{\widehat{\mathbf{R}(\boldsymbol{\theta})\mathbf{R}(\boldsymbol{\theta})^T}\right\} = \frac{\frac{1}{K}\sum_{k=1}^{K}\mathbf{D}_k\mathbf{D}_k^T - E\left\{\mathbf{D}_k^{\mathbf{v}}(\mathbf{D}_k^{\mathbf{v}})^T\right\}}{\Gamma_S - G_S^{\mathbf{v}}} = \frac{\frac{1}{K}\sum_{k=1}^{K}\mathbf{D}_k\mathbf{D}_k^T}{\Gamma_S - G_S^{\mathbf{v}}} \tag{33}$$

since the noise samples $(\mathbf{v}_k)_{k\in[1,K]}$ are complex Gaussian circular.

V. Statistical Prediction

As mentioned by Milstein [7], the MLEs obtained have a meaning only if:

$$\widehat{P_s} = \Gamma_S - G_S^{\mathbf{v}} > 0 \Leftrightarrow \|\mathbf{S}\|^2 > KG_S^{\mathbf{v}}$$

what is another intuitive way to introduce the requirement of a detection test on the sum beam. Indeed, one should not try to estimate parameters of a target if the target is not present. Therefore, let us denote by $\Sigma = \left\{\mathbf{S} = (S_1, \ldots, S_k) \mid \|\mathbf{S}\|^2 > \eta > KG_S^{\mathbf{v}}\right\}$ the event of a detection on the sum beam, and by P_Σ its probability (Probability of Detection). Additionally, let us denote:

$$\widehat{\mathbf{m}_{\mathbf{R}}} = \frac{\boldsymbol{\Gamma}_{\mathbf{D}S} - \mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}}{\Gamma_S - G_S^{\mathbf{v}}} \in \mathbb{C}^M \tag{34}$$

As:

$$\begin{aligned}
cov\left\{\widehat{\mathbf{m_R}} \mid \Sigma\right\} &= E\left\{\widehat{\mathbf{m_R}}\widehat{\mathbf{m_R}}^H \mid \Sigma\right\} - E\left\{\widehat{\mathbf{m_R}} \mid \Sigma\right\} E\left\{\widehat{\mathbf{m_R}} \mid \Sigma\right\}^H \\
cov\left\{\mathrm{Re}\left\{\widehat{\mathbf{m_R}}\right\} \mid \Sigma\right\} &= E\left\{\mathrm{Re}\left\{\widehat{\mathbf{m_R}}\right\}\mathrm{Re}\left\{\widehat{\mathbf{m_R}}\right\}^T \mid \Sigma\right\} - \mathrm{Re}\left\{E\left\{\widehat{\mathbf{m_R}} \mid \Sigma\right\}\right\}\mathrm{Re}\left\{E\left\{\widehat{\mathbf{m_R}} \mid \Sigma\right\}\right\}^T \\
cov\left\{\mathrm{Im}\left\{\widehat{\mathbf{m_R}}\right\} \mid \Sigma\right\} &= E\left\{\mathrm{Im}\left\{\widehat{\mathbf{m_R}}\right\}\mathrm{Im}\left\{\widehat{\mathbf{m_R}}\right\}^T \mid \Sigma\right\} - \mathrm{Im}\left\{E\left\{\widehat{\mathbf{m_R}} \mid \Sigma\right\}\right\}\mathrm{Im}\left\{E\left\{\widehat{\mathbf{m_R}} \mid \Sigma\right\}\right\}^T
\end{aligned}$$

and:

$$\begin{aligned}
\mathrm{Re}\left\{\widehat{\mathbf{m_R}}\right\}\mathrm{Re}\left\{\widehat{\mathbf{m_R}}\right\}^T &= \frac{1}{2}\left(\mathrm{Re}\left\{\widehat{\mathbf{m_R}}\widehat{\mathbf{m_R}}^H\right\} + \mathrm{Re}\left\{\widehat{\mathbf{m_R}}\widehat{\mathbf{m_R}}^T\right\}\right) \\
\mathrm{Im}\left\{\widehat{\mathbf{m_R}}\right\}\mathrm{Im}\left\{\widehat{\mathbf{m_R}}\right\}^T &= \frac{1}{2}\left(\mathrm{Re}\left\{\widehat{\mathbf{m_R}}\widehat{\mathbf{m_R}}^H\right\} - \mathrm{Re}\left\{\widehat{\mathbf{m_R}}\widehat{\mathbf{m_R}}^T\right\}\right)
\end{aligned}$$

the computation of $E\left\{\widehat{\mathbf{m_R}} \mid \Sigma\right\}$, $E\left\{\widehat{\mathbf{m_R}}\widehat{\mathbf{m_R}}^T \mid \Sigma\right\}$ and $E\left\{\widehat{\mathbf{m_R}}\widehat{\mathbf{m_R}}^H \mid \Sigma\right\}$ allows a complete statistical description of $\widehat{\mathbf{m_R}}$, $\mathrm{Re}\left\{\widehat{\mathbf{m_R}}\right\}$ and $\mathrm{Im}\left\{\widehat{\mathbf{m_R}}\right\}$. As shown in Appendix IX-C:

$$E\left\{\widehat{\mathbf{m_R}} \mid \Sigma\right\} = \boldsymbol{\rho} + \left(G_S^{\mathbf{v}}\boldsymbol{\rho} - \mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right) K P\left(K,1\right) \tag{35}$$

$$\begin{aligned}
E\left\{\widehat{\mathbf{m_R}}\widehat{\mathbf{m_R}}^T \mid \Sigma\right\} = \; & \boldsymbol{\rho}\boldsymbol{\rho}^T\left(1 + 2G_S^{\mathbf{v}}KP\left(K,1\right) + \left(G_S^{\mathbf{v}}\right)^2 K^2 P\left(K,2\right)\right) + \left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)^T\right) K^2 P\left(K,2\right) \\
& - \left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\boldsymbol{\rho}^T + \boldsymbol{\rho}\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)^T\right)\left(KP\left(K,1\right) + K^2 G_S^{\mathbf{v}} P\left(K,2\right)\right)
\end{aligned} \tag{36}$$

$$\begin{aligned}
E\left\{\widehat{\mathbf{m_R}}\widehat{\mathbf{m_R}}^H \mid \Sigma\right\} = \; & \left(\mathbf{G_D} - G_S\boldsymbol{\rho}\boldsymbol{\rho}^H - K\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\boldsymbol{\rho}^H - K\boldsymbol{\rho}\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)^H\right)\left(P\left(K,1\right) + KG_S^{\mathbf{v}}P\left(K,2\right)\right) \\
& + \boldsymbol{\rho}\boldsymbol{\rho}^H\left(1 + 2KG_S^{\mathbf{v}}P\left(K,1\right) + \left(KG_S^{\mathbf{v}}\right)^2 P\left(K,2\right)\right) + K^2\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)^H P\left(K,2\right)
\end{aligned} \tag{37}$$

where:

$$P\left(K,L\right) = \int\limits_{t\geq\eta} \frac{p_{\chi^2_{2K}}\left(t;0,\frac{G_S}{2}\right)}{\left(t - KG_S^{\mathbf{v}}\right)^L}\frac{dt}{P_\Sigma}, \quad P_\Sigma = e^{-\frac{\eta}{G_S}} e_{K-1}\left(\frac{\eta}{G_S}\right), \quad e_N\left(\eta\right) = \sum_{n=0}^{N}\frac{\eta^n}{n!}, \quad \boldsymbol{\rho} = \frac{\mathbf{G}_{\mathbf{D}S}}{G_S} \tag{38}$$

Additionally, see Appendix IX-C:

$$E\left\{\widehat{P_s} \mid \Sigma\right\} = G_S\frac{e_K\left(\frac{\eta}{G_S}\right)}{e_{K-1}\left(\frac{\eta}{G_S}\right)} - G_S^{\mathbf{v}} \tag{39a}$$

$$cov\left(\widehat{P_s} \mid \Sigma\right) = \frac{G_S^2}{K}\left(\left(K+1\right)\frac{e_{K+1}\left(\frac{\eta}{G_S}\right)}{e_{K-1}\left(\frac{\eta}{G_S}\right)} - \left(\frac{e_K\left(\frac{\eta}{G_S}\right)}{e_{K-1}\left(\frac{\eta}{G_S}\right)}\right)^2\right) \tag{39b}$$

$$E\left\{\frac{\mathbf{\Gamma_D} - \mathbf{G}_{\mathbf{D}}^{\mathbf{v}}}{\Gamma_S - G_S^{\mathbf{v}}} \mid \Sigma\right\} = \left(\mathbf{G_D} - \mathbf{G}_{\mathbf{D}}^{\mathbf{v}} - \left(G_S - G_S^{\mathbf{v}}\right)\boldsymbol{\rho}\boldsymbol{\rho}^H\right) KP\left(K,1\right) + \boldsymbol{\rho}\boldsymbol{\rho}^H \tag{40}$$

$$E\left\{\frac{\frac{1}{K}\sum\limits_{k=1}^{K}\mathbf{D}_k\mathbf{D}_k^T}{\Gamma_S - G_S^{\mathbf{v}}} \mid \Sigma\right\} = \mathbf{0} \tag{41}$$

This gives the following results:

- for the case $\mathbf{R}(\boldsymbol{\theta}), \widehat{E_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\}} \in \mathbb{C}^M$, we have:

$$\widehat{E_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\}} = \widehat{\mathbf{m}_{\mathbf{R}}} \tag{42a}$$

$$\mathbf{E}_1 = E\left\{\widehat{E_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\}} \mid \Sigma\right\} = E\{\widehat{\mathbf{m}_{\mathbf{R}}} \mid \Sigma\} \tag{42b}$$

$$\mathbf{E}_2 = cov\left\{\widehat{E_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\}} \mid \Sigma\right\} = E\left\{\widehat{\mathbf{m}_{\mathbf{R}}}\widehat{\mathbf{m}_{\mathbf{R}}}^H \mid \Sigma\right\} - E\{\widehat{\mathbf{m}_{\mathbf{R}}} \mid \Sigma\} E\{\widehat{\mathbf{m}_{\mathbf{R}}} \mid \Sigma\}^H \tag{42c}$$

$$\mathbf{E}_3 = E\left\{\widehat{cov_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\}} \mid \Sigma\right\} = E\left\{\frac{\boldsymbol{\Gamma}_{\mathbf{D}} - \mathbf{G}_{\mathbf{D}}^{\mathbf{v}}}{\Gamma_S - G_S^{\mathbf{v}}} \mid \Sigma\right\} - E\left\{\widehat{\mathbf{m}_{\mathbf{R}}}\widehat{\mathbf{m}_{\mathbf{R}}}^H \mid \Sigma\right\} \tag{42d}$$

and (GME-2) (16) is finally characterized by:

$$\widehat{E_{\boldsymbol{\theta}}\{\Delta\boldsymbol{\theta}\}} = -\operatorname{Re}\{\mathbf{C}(\widehat{\mathbf{m}_{\mathbf{R}}} - \boldsymbol{\mu})\}$$

$$\widehat{cov_{\boldsymbol{\theta}}\{\Delta\boldsymbol{\theta}\}} = \frac{1}{2}\operatorname{Re}\left\{\mathbf{C}\left(\frac{\boldsymbol{\Gamma}_{\mathbf{D}} - \mathbf{G}_{\mathbf{D}}^{\mathbf{v}}}{\Gamma_S - G_S^{\mathbf{v}}} - \widehat{\mathbf{m}_{\mathbf{R}}}\widehat{\mathbf{m}_{\mathbf{R}}}^H\right)\mathbf{C}^H + \mathbf{C}\left(\frac{\frac{\mathbf{1}}{K}\sum_{k=1}^{K}\mathbf{D}_k\mathbf{D}_k^T}{\Gamma_S - G_S^{\mathbf{v}}} - \widehat{\mathbf{m}_{\mathbf{R}}}\widehat{\mathbf{m}_{\mathbf{R}}}^T\right)\mathbf{C}^T\right\}$$

$$E\left\{\widehat{E_{\boldsymbol{\theta}}\{\Delta\boldsymbol{\theta}\}} \mid \Sigma\right\} = -\operatorname{Re}\{\mathbf{C}(\mathbf{E}_1 - \boldsymbol{\mu})\} \tag{43a}$$

$$cov\left\{\widehat{E_{\boldsymbol{\theta}}\{\Delta\boldsymbol{\theta}\}} \mid \Sigma\right\} = \frac{1}{2}\operatorname{Re}\left\{\mathbf{C}\mathbf{E}_2\mathbf{C}^H + \mathbf{C}E\left\{\widehat{\mathbf{m}_{\mathbf{R}}}\widehat{\mathbf{m}_{\mathbf{R}}}^T \mid \Sigma\right\}\mathbf{C}^T\right\} - \frac{1}{2}\operatorname{Re}\left\{\mathbf{C}\mathbf{E}_1(\mathbf{C}\mathbf{E}_1)^T\right\} \tag{43b}$$

$$E\left\{\widehat{cov_{\boldsymbol{\theta}}\{\Delta\boldsymbol{\theta}\}} \mid \Sigma\right\} = \frac{1}{2}\operatorname{Re}\{\mathbf{C}\mathbf{E}_3\mathbf{C}^H\} - \frac{1}{2}\operatorname{Re}\left\{\mathbf{C}E\left\{\widehat{\mathbf{m}_{\mathbf{R}}}\widehat{\mathbf{m}_{\mathbf{R}}}^T \mid \Sigma\right\}\mathbf{C}^T\right\} \tag{43c}$$

- for the case $\mathbf{R}(\boldsymbol{\theta}), \widehat{E_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\}} \in \mathbb{R}^M$, we have:

$$\widehat{E_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\}} = \operatorname{Re}\{\widehat{\mathbf{m}_{\mathbf{R}}}\} \tag{44a}$$

$$\mathbf{E}_1 = E\left\{\widehat{E_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\}} \mid \Sigma\right\} = \operatorname{Re}\{E\{\widehat{\mathbf{m}_{\mathbf{R}}} \mid \Sigma\}\} \tag{44b}$$

$$\mathbf{E}_2 = cov\left\{\widehat{E_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\}} \mid \Sigma\right\} = \begin{array}{l}\frac{1}{2}\operatorname{Re}\left\{E\left\{\widehat{\mathbf{m}_{\mathbf{R}}}\widehat{\mathbf{m}_{\mathbf{R}}}^H \mid \Sigma\right\} + E\left\{\widehat{\mathbf{m}_{\mathbf{R}}}\widehat{\mathbf{m}_{\mathbf{R}}}^T \mid \Sigma\right\}\right\} \\ -\operatorname{Re}\{E\{\widehat{\mathbf{m}_{\mathbf{R}}} \mid \Sigma\}\}\operatorname{Re}\{E\{\widehat{\mathbf{m}_{\mathbf{R}}} \mid \Sigma\}\}^T\end{array} \tag{44c}$$

$$\mathbf{E}_3 = E\left\{\widehat{cov_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\}} \mid \Sigma\right\} = \begin{array}{l}\operatorname{Re}\left\{E\left\{\frac{\boldsymbol{\Gamma}_{\mathbf{D}} - \mathbf{G}_{\mathbf{D}}^{\mathbf{v}}}{\Gamma_S - G_S^{\mathbf{v}}} \mid \Sigma\right\}\right\} \\ -\frac{1}{2}\operatorname{Re}\left\{E\left\{\widehat{\mathbf{m}_{\mathbf{R}}}\widehat{\mathbf{m}_{\mathbf{R}}}^H \mid \Sigma\right\} + E\left\{\widehat{\mathbf{m}_{\mathbf{R}}}\widehat{\mathbf{m}_{\mathbf{R}}}^T \mid \Sigma\right\}\right\}\end{array} \tag{44d}$$

and (GME-1) (15) is finally characterized by:

$$\widehat{E_{\boldsymbol{\theta}}\{\Delta\boldsymbol{\theta}\}} = -\mathbf{C}(\operatorname{Re}\{\widehat{\mathbf{m}_{\mathbf{R}}}\} - \boldsymbol{\mu})$$

$$\widehat{cov_{\boldsymbol{\theta}}\{\Delta\boldsymbol{\theta}\}} = \mathbf{C}\left(\operatorname{Re}\left\{\frac{\boldsymbol{\Gamma}_{\mathbf{D}} - \mathbf{G}_{\mathbf{D}}^{\mathbf{v}}}{\Gamma_S - G_S^{\mathbf{v}}}\right\} - \operatorname{Re}\{\widehat{\mathbf{m}_{\mathbf{R}}}\}\operatorname{Re}\{\widehat{\mathbf{m}_{\mathbf{R}}}\}^T\right)\mathbf{C}^T$$

$$E\left\{\widehat{E_{\boldsymbol{\theta}}\{\Delta\boldsymbol{\theta}\}} \mid \Sigma\right\} = -\mathbf{C}(\mathbf{E}_1 - \boldsymbol{\mu}) \tag{45a}$$

$$cov\left\{\widehat{E_{\boldsymbol{\theta}}\{\Delta\boldsymbol{\theta}\}} \mid \Sigma\right\} = \mathbf{C}\mathbf{E}_2\mathbf{C}^T \tag{45b}$$

$$E\left\{\widehat{cov_{\boldsymbol{\theta}}\{\Delta\boldsymbol{\theta}\}} \mid \Sigma\right\} = \mathbf{C}\mathbf{E}_3\mathbf{C}^T \tag{45c}$$

Note that expressions (35)-(40) are far more compact that the ones given by Milstein [7] in the single angle case.

Finally we have:

$$P(K,L) = \frac{e^{\frac{\eta - KG_S^{\mathbf{v}}}{G_S}}}{e_{K-1}\left(\frac{\eta}{G_S}\right)(G_S)^L}\left(\sum_{k=0}^{K-1}\frac{1}{k!(K-1-k)!}\left(\frac{KG_S^{\mathbf{v}}}{G_S}\right)^{K-1-k}\int_{t\geq\frac{\eta - KG_S^{\mathbf{v}}}{G_S}} t^{k-L}e^{-t}dt\right) \tag{46}$$

where:

$$\int\limits_{t\geq\eta} t^N e^{-t} dt = \begin{cases} e^{-\eta} e_N(\eta) N!, & N \geq 0 \\ E_1(\eta), & N = -1 \\ \frac{e^{-\eta}}{\eta} - E_1(\eta), & N = -2 \\ \frac{e^{-\eta}}{\eta}\left(\frac{1}{2\eta} - \frac{1}{2}\right) + \frac{1}{2}E_1(\eta), & N = -3 \\ \frac{e^{-\eta}}{\eta}\left(\frac{1}{3\eta^2} - \frac{1}{6\eta} + \frac{1}{6}\right) - \frac{1}{6}E_1(\eta), & N = -4 \end{cases}, \quad e_N(\eta) = \sum_{n=0}^{N} \frac{\eta^n}{n!}, \quad E_1(\eta) = \int\limits_{t\geq\eta} \frac{e^{-t}}{t} dt$$

VI. Multiple (Signal) Sources Indicator

One of the key result of the parameterization introduced is the ability to estimate $cov_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\}$. In fact, if there is a single signal source then $cov_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\} = \mathbf{0}$. Therefore we propose as a Multiple (signal) Sources Indicator (MSI) the test:

$$\widehat{MSI} = \frac{1}{M} tr\left(\widehat{cov_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\}}\right) \lessgtr T_{MS}, \quad \mathbf{R}(\boldsymbol{\theta}) \in \mathbb{C}^M \tag{47}$$

that is:

$$\widehat{MSI} = \frac{1}{M} tr\left(\frac{\mathbf{\Gamma_D} - \mathbf{G_D^v}}{\Gamma_S - G_S^{\mathbf{v}}} - \widehat{\mathbf{m_R}}\widehat{\mathbf{m_R}}^H\right) \lessgtr T_{MS}, \quad \widehat{\mathbf{m_R}} = \frac{\mathbf{\Gamma_{D}}_S - \mathbf{G_{D}^v}_S}{\Gamma_S - G_S^{\mathbf{v}}}, \tag{48}$$

The motivation is as follows:

- there is a difference between the estimation of angular extents provided by $\left(E\left\{\widehat{cov_{\boldsymbol{\theta}}\{\Delta\boldsymbol{\theta}\}} \mid \Sigma\right\}\right)_{m,m}$ and the detection of the existence of angular extent, that is the detection of multiple signal sources. The angular extent estimates should be taken into account only if multiple signal sources are detected.
- the use of a complex $\mathbf{R}(\boldsymbol{\theta})$ allows to have an indicator immune to beam phase calibration problem (once again, the difference between detection of angle extents and estimation of angle extents).
- a careful examination of $cov_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\}$ (42a) shows that if $K < M$, the estimator is no longer a MLE as it is no longer able to estimate all the components of $cov_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\}$ (rank deficiency). Nevertheless, $tr\left(E\left\{\widehat{cov_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\}} \mid \Sigma\right\}\right)$ can be estimated whatever the value of K. Therefore, from a practical point of view, that is, in operational conditions where $K \leq M$ is likely, there is no point in expressing the MSI in the angle domain.

In order to predict the correct value for the threshold T_{MS} to achieve a given probability of false alarm, i.e. the probability to detect multiple sources when a single source is actually present, it is necessary to know the p.d.f. $p\left(\widehat{MSI} \mid \Sigma\right)$. When an expression of $p\left(\widehat{MSI} \mid \Sigma\right)$ is not available, a sensible value for T_{MS} can be obtained if the first two moments $E\left\{\widehat{MSI} \mid \Sigma\right\}$ and $cov\left\{\widehat{MSI} \mid \Sigma\right\}$ are available, by using either the Bienaymé - Tchebychev inequality or p.d.f. model fitting, as we propose in Section VI (Gaussian p.d.f. fitting). Actually, from (37)(40), one can already compute $E\left\{\widehat{MSI} \mid \Sigma\right\}$ since:

$$E\left\{\widehat{MSI} \mid \Sigma\right\} = \frac{1}{M} tr\left(E\left\{\frac{\mathbf{\Gamma_D} - \mathbf{G_D^v}}{\Gamma_S - G_S^{\mathbf{v}}} \mid \Sigma\right\} - E\left\{\widehat{\mathbf{m_R}}\widehat{\mathbf{m_R}}^H \mid \Sigma\right\}\right) \tag{49}$$

The derivation of $cov\left\{\widehat{MSI} \mid \Sigma\right\}$ requires additional computations detailed in Appendix IX-D. To summarize, since:

$$\widehat{MSI} = \frac{1}{M} tr\left(\widehat{cov_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\}}\right) = \frac{1}{M}\sum_{m=1}^{M}\left(\widehat{cov_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\}}\right)_{m,m}$$

then:

$$\begin{aligned} cov\left\{\widehat{MSI} \mid \Sigma\right\} &= \frac{1}{M^2}\sum_{m=1}^{M}\sum_{m'=1}^{M} E\left\{\widehat{M}_{m,m}^{m',m'} \mid \Sigma\right\} - E\left\{\widehat{MSI} \mid \Sigma\right\}^2 \\ \widehat{M}_{m,l}^{m',l'} &= \left(\widehat{cov_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\}}\right)_{m,l}\left(\widehat{cov_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\}}\right)_{m',l'} \end{aligned} \tag{50}$$

where:

$$E\left[\widehat{M}_{m,l}^{m',l'} \mid \Sigma\right] = \begin{array}{l} \left(\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'} + \frac{1}{K}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l'}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l}\right) K^2 P\left(K,2\right) + \\ \left(\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{B}\right)_{m',l'} + \left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'}\left(\mathbf{B}\right)_{m,l}\right) K^3 P\left(K,3\right) + \\ \left(\mathbf{B}\right)_{m,l}\left(\mathbf{B}\right)_{m',l'} K^4 P\left(K,4\right) + \\ \left(\begin{array}{l} K^2\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{A}\right)_{m',l'} + K^2\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'}\left(\mathbf{A}\right)_{m,l} \\ -2K\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'} - 2\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l'}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l} \end{array}\right)\left(P\left(K,2\right) + KG_S^{\mathbf{v}}P\left(K,3\right)\right) + \\ \left(\begin{array}{l} \left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l'}\left(\boldsymbol{\nu}\right)_l^*\left(\boldsymbol{\nu}\right)_{m'} + \left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l}\left(\boldsymbol{\nu}\right)_m\left(\boldsymbol{\nu}\right)_{l'}^* \\ -\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{B}\right)_{m',l'} - \left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'}\left(\mathbf{B}\right)_{m,l} \\ +K\left(\left(\mathbf{A}\right)_{m,l}\left(\mathbf{B}\right)_{m',l'} + \left(\mathbf{A}\right)_{m',l'}\left(\mathbf{B}\right)_{m,l}\right) \end{array}\right) K^2\left(P\left(K,3\right) + KG_S^{\mathbf{v}}P\left(K,4\right)\right) + \\ \left(\begin{array}{l} \left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'} + \left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l'}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l} \\ -K\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{A}\right)_{m',l'} - K\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'}\left(\mathbf{A}\right)_{m,l} \\ +K^2\left(\mathbf{A}\right)_{m,l}\left(\mathbf{A}\right)_{m',l'} \end{array}\right)\left(\begin{array}{c} P\left(K,2\right) + 2KG_S^{\mathbf{v}}P\left(K,3\right) \\ +\left(KG_S^{\mathbf{v}}\right)^2 P\left(K,4\right) \end{array}\right) \end{array} \tag{51}$$

$$\begin{aligned} \boldsymbol{\nu} &= \mathbf{G}_{\mathbf{D}S}^{\mathbf{v}} - \boldsymbol{\rho}G_S^{\mathbf{v}}, \quad \mathbf{A} = \mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\boldsymbol{\rho}^H + \boldsymbol{\rho}\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)^H - G_S^{\mathbf{v}}\boldsymbol{\rho}\boldsymbol{\rho}^H - \mathbf{G}_{\mathbf{D}}^{\mathbf{v}}, \quad \mathbf{B} = G_S^{\mathbf{v}}\mathbf{G}_{\mathbf{D}}^{\mathbf{v}} - \mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)^H \\ \mathbf{C}_{\mathbf{D}|S} &= \mathbf{G}_{\mathbf{D}} - G_S\boldsymbol{\rho}\boldsymbol{\rho}^H \end{aligned}$$

VII. Numerical Performance Study

To demonstrate the applicability of the method for arbitrary arrays, we show performance results for an array composed of fairly irregular subarrays. We consider a fully filled planar array with 902 elements on a regular triangular grid, but beamforming is done digitally from the outputs of 32 irregular subarrays. This can be viewed as an array of 32 “super-elements” with very unequal patterns. This type of array has been used in previous papers and it is a generic antenna for an airborne radar with low sidelobes and digital beamforming at subarray level. For sum beam forming we apply a -35dB Taylor weighting at the elements. The difference beams are formed by approximating a -35dB Bayliss weighting at subarray level for the given Taylor weighting at the elements. Details of difference beamforming at subarray level and approximating low sidelobe patterns can be found in [19]. Figure 1 shows the array with subarrays, Figure 2 the resulting patterns of the sum and difference beam (azimuth cut only). We parameterize all spatial angles by the components of the unit direction vector (u,v-coordinates for a planar array or direction sines/cosines ranging from -1 to 1). By approximating the Bayliss weighting at subarray level it is of course not possible to attain the -35dB sidelobe level; a suboptimum level of -30dB is obtained. This array is used for all subsequent plots. We have to consider the averaged monopulse ratio to detect or measure extended targets. For the following simulation we took always the number of snapshots $K = 10$. The given SNR values are taken at the sum beam output. The SNR at the array element is then for the given antenna with Taylor weighting 28dB lower.

As we are interested in early detection of extended targets (small targets), we use a simple two point target model with fairly small separation. Figure 3 shows the performance of the proposed estimator for the challenging case of two point scatterers separated only by 0.3 of a beamwidth (BW) representing the extended target. The dashed circle indicates the 3dB beamwidth contour. The SNR is at 28dB. The detection threshold entering the monopulse ratio was set at 13dB. The blue ellipses indicate the estimated target size based on 10 snapshots . The expectation of the target size ellipse according to (45c) is shown by the red ellipse. The centroid (45a) is marked by the grey circle. From the performance characterization (45b) one can derive an uncertainty ellipse for the centroid marked by the dashed grey ellipse. This shows that the given performance prediction is quite accurate. A remarkable property is however, that the orientation in the 2D space is quite well determined despite the small extent of the target.

If the two point scatterers are wider separated, the attenuation of the sum pattern comes into play more and more and this biases the estimate. This is shown in Figure 4 for a target consisting of two point scatterers at 0.5BW separation at SNR of 28dB. In fact, the second point scatterer is already on the 3dB contour of the sum beam and is therefore attenuated by 3dB. Clearly, the diameter of the ellipse is not necessarily exactly equal to the scatterer separation.

The underlying Swerling I-II model is a simple approximation. Real extended targets can have any kind of fluctuation. Figure 5 shows an example that at least the centroid and the orientation is quite well estimated for correlated scatterers. We selected the case of two anti-phase targets fluctuating around 170° with a standard deviation of 20°.

We now compare the formulas for the mean centroid (45a) and extension (45c) with simulations for some varying parameters. We study in a joint scenario the dependency on SNR, length of extended target and position relative to the antenna look direction where the monopulse is evaluated. We consider 20 point scatterers forming a diagonal line in the u,v-plane at various SNR with line length varying from zero to 0.6BW. The north-east scatterer is always fixed at an offset position $(u_0, v_0) = 0.1\left(\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$. Again the monopulse estimate is based on 10 snapshots $(K = 10)$. Figure 6 shows the configuration when the extended target has the maximum length of 0.6BW. For target length zero this will reduce to a single point scatterer at $(u_0, v_0) = 0.1\left(\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$. The dashed circle indicates the 3dB contour of the sum beam. The extended target reaches far outside the 3dB beamwidth. All axes are normalized to the antenna beamwidth.

Figure 7 shows the developed theoretical descriptors of the estimators according to (45a), (45b), (45c). For the extent we have taken twice the length of the maximum half axis (or diameter) of the ellipse. Clearly, bias and variance of the centroid and extent estimator increase with the asymmetry with respect to the look direction of the antenna. However, for the line target touching the 3dB contour (at $\delta u = 0.4$) the centroid bias is below 0.05BW and the extension bias is about 0.13BW. The extension estimate bias is in most cases negative meaning that the length is estimated too short (due to the higher mass in the centre of the line). A measure of $1.2\times$(ellipse diameter) would in this case be a better description of the extent.

Figure 8 shows the simulated values. The simulations support pretty well the theoretical results (except for the low SNR case 13dB). The bias in the length estimate can be mitigated by repeating the monopulse measurement with an antenna look direction equal to the centroid. The portion of the line within the main beam will then be more symmetrical.

Figures 9 and 10 show the dependence on the number of snapshots K used in the averaged monopulse ratio to estimate the extended target. This simulation was done for the scenario as before for a length of the extended target (20 scatterers) of $\delta u = 0.4$BW, and antenna look direction offset -0.1BW as in Figure 6. Again the simulation supports the theoretical results (except for the low SNR case 13dB). Note that a fairly small number of about four snapshots is already sufficient to estimate the extension. This is important for rapid estimation.

Next we study the performance of the MSI detector given in Section V. We use the theoretical values for the mean (49) and variance (50) of the MSI to calculate a threshold, such that under Gaussian assumptions we have a false alarm rate of 10^{-3} and 10^{-5}. The null hypothesis is taken as the weakest detectable point target in the antenna look direction, corresponding to 13dB sum beam SNR, because this is the monopulse detection threshold. Figure 11 shows the detection probability for this threshold setting for various values of SNR over the length of the line target considered before ranging now from 0 to 2BW and the offset (u_0, v_0) set to zero. The detector is calculated from 10 snapshots $(K = 10)$. The detector decides for an extended target fairly late, at lengths where a noticable ellipse is already estimated.

Figure 12 shows the case of an extended target consisting of two point scatterers with increasing separation. One can see that such a type of extended target is detected earlier than a real line target as in Figure 10, e.g. at about 0.4BW separation. At about 1.5BW the detector separates the two targets into individual point targets. This large separation stems from the fact that the maximum of the difference beam is approximately at the first null of the sum beam, which means that, although the second target is already attenuated by the sum beam, it is still fully present in the difference beam. It is interesting to check what the centroid and extent measurements are in this case. Figures 13 and 14 show the ellipses for the separations of 1.5BW and 2BW (at SNR = 20dB). One can see that at 2BW separation, a point target is practically estimated.

If we have detected and measured a long extension outside the main beam, a new measurement for the centroid and extension should be performed with a new pointing direction of the antenna. This way, we can compose larger extended target of any shape by a sequence of ellipses (similar to a Gaussian sum approximation).

We have assumed here an extended target in white noise only. If there is non-white noise, e.g. a sidelobe target, it is a matter of sidelobe level and threshold design to ensure acceptable performance. The procedure can even be applied to severe non-white noise by using adaptive beams as we use the generalized monopulse ratio. Performance studies for this case are a topic of future research, because this requires adjustments and simulation of relevant

target-jamming scenarios.
The goodness of the Gaussian approximation for the test statistic is shown by the simulated distribution of the MSI. Figure 15 shows the histogram of the MSI values for a single point scatterer at look direction at 13dB SNR, which is just the smallest detectable target with the monopulse detection threshold at 13dB (the null hypothesis used before). We used 10^5 Monte-Carlo runs for these plots. The figure also shows a Gaussian fit to this histogram. The theoretical mean and variance and the corresponding values for the fitted Gaussian density are shown for comparison. Figure 16 shows the goodness of this fit against a Gaussian by the Matlab normplot display (data obeying a Gaussian should lie on the dashed line). Except for the tails of the distribution the approximation is quite good (in the probability interval $[0.01, 0.9]$). Hence this approximation is useful to determine the threshold for a desired false alarm level.
Figure 17 shows the corresponding histogram and the resulting Gaussians for two targets separated by 0.5BW at 13dB SNR. Obviously, the density is not a Gaussian, but the first two moments fit quite well. As we always require for an extended target a positive MSI value we have a very low probability of false alarm.
The Gaussian distribution is a good approximation for this density only for the null hypothesis. Its use to predict theoretically the probability of detection is not recommended. For this purpose a Rayleigh- or Gamma-distribution would be more appropriate. To verify this, extensive simulations with different scenarios would be necessary and this should therefore be the topic of further studies.

VIII. Conclusion

We have developed a detector for extended targets and estimators for the centroid and the extension of such targets based only on the outputs of the generalized monopulse ratio. The statistical performance of these quantities is characterized by their mean and variance. Numerical simulations indicated that these statistical characterizations are accurate such that these quantities can be used for system studies.
For the simulations we used a realistic antenna array with low sidelobes and digital subarray outputs as a prototype for antennas with various subarray shapes. For this antenna the detector showed a good sensitivity in detecting small extended targets of length at 0.3 of a beamwidth with a fairly small number of snapshots for the averaged monopulse ratio while well separated point targets are also detected as point targets by the detector (i.e. do not give false alarms).
The estimators for centroid and extent were particularly reliable in displaying the orientation of the extended target in the two-dimensional angle space. The direct calculation is much faster than multi-beam search techniques proposed before. This is of importance in practical applications of early of extended targets. The estimate of the centroid and length of the extended target can be slightly biased depending on the position of the extended target relative to the sum beam shape and the related attenuation of scatterers on the skirt of the beam. However this bias is below $1/10$ of a beamwidth. This monopulse based length and centroid estimation is limited to an area of about the antenna beamwidth. Targets with larger extensions have to be estimated by repeating the procedure for a new antenna look direction and by concatenating the resulting ellipses.

IX. Appendix

A. Angular centroid and extent as a function of the true monopulse ratios

- (GME-1):

$$\Delta\boldsymbol{\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}_0 = -\mathbf{C}\left(\mathrm{Re}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\} - \boldsymbol{\mu}\right)$$

Then:

$$E_{\boldsymbol{\theta}}\left\{\Delta\boldsymbol{\theta}\right\} = -\mathbf{C}\left(E_{\boldsymbol{\theta}}\left\{\mathrm{Re}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}\right\} - \boldsymbol{\mu}\right), \quad \Delta\boldsymbol{\theta} - E_{\boldsymbol{\theta}}\left\{\Delta\boldsymbol{\theta}\right\} = -\mathbf{C}\left(\mathrm{Re}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\} - E_{\boldsymbol{\theta}}\left\{\mathrm{Re}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}\right\}\right).$$

Therefore:

$$cov_{\boldsymbol{\theta}}\left\{\Delta\boldsymbol{\theta}\right\} = \mathbf{C}cov_{\boldsymbol{\theta}}\left\{\mathrm{Re}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}\right\}\mathbf{C}^T$$

- (GME-2):

$$\Delta\boldsymbol{\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}_0 = -\,\mathrm{Re}\left\{\mathbf{C}\left(\mathbf{R}\left(\boldsymbol{\theta}\right) - \boldsymbol{\mu}\right)\right\}$$

Then:

$$E_{\boldsymbol{\theta}}\left\{\Delta\boldsymbol{\theta}\right\} = -\,\mathrm{Re}\left\{\mathbf{C}\left(E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\} - \boldsymbol{\mu}\right)\right\}, \quad \Delta\boldsymbol{\theta} - E_{\boldsymbol{\theta}}\left\{\Delta\boldsymbol{\theta}\right\} = -\,\mathrm{Re}\left\{\mathbf{C}\left(\mathbf{R}\left(\boldsymbol{\theta}\right) - E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}\right)\right\}$$

Moreover, as:

$$\mathrm{Re}\left\{\mathbf{x}\right\}\mathrm{Re}\left\{\mathbf{x}\right\}^T = \frac{1}{2}\left[\mathrm{Re}\left\{\mathbf{x}\mathbf{x}^H\right\} + \mathrm{Re}\left\{\mathbf{x}\mathbf{x}^T\right\}\right]$$

therefore:

$$\begin{aligned}
cov_{\boldsymbol{\theta}}\left\{\Delta\boldsymbol{\theta}\right\} &= E_{\boldsymbol{\theta}}\left\{\mathrm{Re}\left\{\mathbf{C}\left(\mathbf{R}\left(\boldsymbol{\theta}\right) - E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}\right)\right\}\mathrm{Re}\left\{\mathbf{C}\left(\mathbf{R}\left(\boldsymbol{\theta}\right) - E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}\right)\right\}^T\right\} \\
cov_{\boldsymbol{\theta}}\left\{\Delta\boldsymbol{\theta}\right\} &= \frac{1}{2}E_{\boldsymbol{\theta}}\left\{\begin{array}{l}\mathrm{Re}\left\{\mathbf{C}\left(\mathbf{R}\left(\boldsymbol{\theta}\right) - E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}\right)\left(\mathbf{R}\left(\boldsymbol{\theta}\right) - E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}\right)^H\mathbf{C}^H\right\} + \\ \mathrm{Re}\left\{\mathbf{C}\left(\mathbf{R}\left(\boldsymbol{\theta}\right) - E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}\right)\left(\mathbf{R}\left(\boldsymbol{\theta}\right) - E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}\right)^T\mathbf{C}^T\right\}\end{array}\right\} \\
cov_{\boldsymbol{\theta}}\left\{\Delta\boldsymbol{\theta}\right\} &= \frac{1}{2}\left(\begin{array}{l}\mathrm{Re}\left\{\mathbf{C}cov_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}\mathbf{C}^H\right\} + \\ \mathrm{Re}\left\{\mathbf{C}\left(E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\mathbf{R}\left(\boldsymbol{\theta}\right)^T\right\} - E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}^T\right)\mathbf{C}^T\right\}\end{array}\right)
\end{aligned}$$

B. MLE of $\widehat{P_s}$, $\widehat{E_\theta\left\{R\right\}}$, $\widehat{cov_\theta\left\{R\right\}}$

1) 2 beams - complex $\widehat{E_\theta\left\{R\right\}}$ - spatially white noise of known power:

For sake of legibility, let $\sigma^2 = G_S^{\mathbf{v}}$, $p = P_s \in \mathbb{R} > 0$, $e = E\left\{R\left(\theta\right)\right\} \in \mathbb{C}$, $v = \widehat{cov_\theta\left\{R\right\}} \in \mathbb{R} \geq 0$. Then:

$$\mathbf{G} = \left[\begin{array}{cc} pee^* + pv + \sigma^2 & pe \\ pe^* & p + \sigma^2 \end{array}\right]$$

We use the ML invariance principle to reparameterize $\mathbf{G}$ as:

$$\mathbf{G}\left(\boldsymbol{\omega}\right) = \left[\begin{array}{cc} pee^* + w + \sigma^2 & pe \\ pe^* & p + \sigma^2 \end{array}\right], \quad \boldsymbol{\omega} = \left(p, e, w, \sigma^2\right)^T, \quad w = pv$$

Then:

$$\left(p, e, v, \sigma^2\right) = \left(p, e, \frac{w}{p}, \sigma^2\right) = f\left(\boldsymbol{\omega}\right) \quad \Rightarrow \quad \left(\widehat{p, e, v, \sigma^2}\right) = f\left(\widehat{\boldsymbol{\omega}}\right)$$

From (4):

$$\ln p\left(\mathbf{B}_1, \ldots, \mathbf{B}_K\right) = -K\left[tr\left(\mathbf{G}\left(\boldsymbol{\omega}\right)^{-1}\left(\boldsymbol{\Gamma}\right)\right) + \ln\left|\mathbf{G}\left(\boldsymbol{\omega}\right)\right| + 2\ln\pi\right]$$

where:

$$
\begin{aligned}
|\mathbf{G}(\boldsymbol{\omega})| &= \sigma^2 pee^* + \left(w+\sigma^2\right)\left(p+\sigma^2\right), \quad \mathbf{G}(\boldsymbol{\omega})^{-1} = \frac{1}{|\mathbf{G}(\boldsymbol{\omega})|}\left[\begin{array}{cc} p+\sigma^2 & -pe^* \\ -pe & pee^*+w+\sigma^2 \end{array}\right] \\
tr\left(\mathbf{G}(\boldsymbol{\omega})^{-1}\boldsymbol{\Gamma}\right) &= Ktr\left(\frac{1}{|\mathbf{G}(\boldsymbol{\omega})|}\left[\begin{array}{cc} p+\sigma^2 & -pe \\ -pe^* & pee^*+w+\sigma^2 \end{array}\right]\left[\begin{array}{cc} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{12}^* & \Gamma_{22} \end{array}\right]\right) \\
&= \frac{1}{|\mathbf{G}(\boldsymbol{\omega})|} tr\left(\left[\begin{array}{cc} \left(p+\sigma^2\right)\Gamma_{11} - pe\Gamma_{12}^* & \\ & -pe^*\Gamma_{12} + \left(pee^*+w+\sigma^2\right)\Gamma_{22} \end{array}\right]\right) \\
&= \frac{p\Gamma_{11} - pe\Gamma_{12}^* - pe^*\Gamma_{12} + (pee^*+w)\Gamma_{22} + \sigma^2(\Gamma_{22}+\Gamma_{11})}{\sigma^2 pee^* + (w+\sigma^2)(p+\sigma^2)}
\end{aligned}
$$

Therefore the MLE's are obtained by solving the optimization problem:

$$
\widehat{(p,e,w)} = \min_{(p,e,w)} \{L(\boldsymbol{\omega})\}, \quad L(\boldsymbol{\omega}) = \frac{F(\boldsymbol{\omega})}{|\mathbf{G}(\boldsymbol{\omega})|} + \ln(|\mathbf{G}(\boldsymbol{\omega})|)
$$

where:

$$
\begin{aligned}
F(\boldsymbol{\omega}) &= p\Gamma_{11} - pe\Gamma_{12}^* - pe^*\Gamma_{12} + (pee^*+w)\Gamma_{22} + \sigma^2(\Gamma_{22}+\Gamma_{11}) \\
|\mathbf{G}(\boldsymbol{\omega})| &= \sigma^2 pee^* + \left(w+\sigma^2\right)\left(p+\sigma^2\right)
\end{aligned}
$$

Then:

$$
\begin{aligned}
\frac{\partial L(\boldsymbol{\omega})}{\partial \omega_k} &= \frac{\partial F(\boldsymbol{\omega})}{\partial \omega_k}\frac{1}{|\mathbf{G}(\boldsymbol{\omega})|} - \frac{F(\boldsymbol{\omega})}{|\mathbf{G}(\boldsymbol{\omega})|^2}\frac{\partial |\mathbf{G}(\boldsymbol{\omega})|}{\partial \omega_k} + \frac{1}{|\mathbf{G}(\boldsymbol{\omega})|}\frac{\partial |\mathbf{G}(\boldsymbol{\omega})|}{\partial \omega_k} \\
&= \frac{1}{|\mathbf{G}(\boldsymbol{\omega})|}\left(\frac{\partial F(\boldsymbol{\omega})}{\partial \omega_k} + \frac{\partial |\mathbf{G}(\boldsymbol{\omega})|}{\partial \omega_k}\left(1 - \frac{F(\boldsymbol{\omega})}{|\mathbf{G}(\boldsymbol{\omega})|}\right)\right)
\end{aligned}
$$

and:

$$
\frac{\partial L(\boldsymbol{\omega})}{\partial \omega_k} = 0 \quad \Leftrightarrow \quad \frac{\frac{\partial F(\boldsymbol{\omega})}{\partial \omega_k}}{\frac{\partial |\mathbf{G}(\boldsymbol{\omega})|}{\partial \omega_k}} = \frac{F(\boldsymbol{\omega})}{|\mathbf{G}(\boldsymbol{\omega})|} - 1
$$

Therefore, extremality of $L(\boldsymbol{\omega})$ can be obtained by solving the following (simplified) equations:

$$
(E1): \frac{\frac{\partial F(\boldsymbol{\omega})}{\partial w}}{\frac{\partial |\mathbf{G}(\boldsymbol{\omega})|}{\partial w}} = \frac{\frac{\partial F(\boldsymbol{\omega})}{\partial p}}{\frac{\partial |\mathbf{G}(\boldsymbol{\omega})|}{\partial p}}, \quad (E2): \frac{\frac{\partial F(\boldsymbol{\omega})}{\partial w}}{\frac{\partial |\mathbf{G}(\boldsymbol{\omega})|}{\partial w}} = \frac{\frac{\partial F(\boldsymbol{\omega})}{\partial e^*}}{\frac{\partial |\mathbf{G}(\boldsymbol{\omega})|}{\partial e^*}}, \quad (E3): \frac{\frac{\partial F(\boldsymbol{\omega})}{\partial v}}{\frac{\partial |\mathbf{G}(\boldsymbol{\omega})|}{\partial v}} = \frac{F(\boldsymbol{\omega})}{|\mathbf{G}(\boldsymbol{\omega})|} - 1 \tag{52}
$$

• Exploitation of $(E1)$:

$$
\frac{\frac{\partial F(\boldsymbol{\omega})}{\partial w}}{\frac{\partial |\mathbf{G}(\boldsymbol{\omega})|}{\partial w}} = \frac{\frac{\partial F(\boldsymbol{\omega})}{\partial e^*}}{\frac{\partial |\mathbf{G}(\boldsymbol{\omega})|}{\partial e^*}} \Leftrightarrow \frac{\Gamma_{22}}{p+\sigma^2} = \frac{e\Gamma_{22} - \Gamma_{12}}{\sigma^2 e} \quad \Leftrightarrow \quad \sigma^2 e\Gamma_{22} = pe\Gamma_{22} + \sigma^2 e\Gamma_{22} - \left(p+\sigma^2\right)\Gamma_{12}
$$

$$
\Updownarrow
$$

$$
e = \left(\frac{p+\sigma^2}{p}\right)\frac{\Gamma_{12}}{\Gamma_{22}} \tag{53}
$$

• Exploitation of $(E2)$:

$$
\frac{\frac{\partial F(\boldsymbol{\omega})}{\partial w}}{\frac{\partial |\mathbf{G}(\boldsymbol{\omega})|}{\partial w}} = \frac{\frac{\partial F(\boldsymbol{\omega})}{\partial p}}{\frac{\partial |\mathbf{G}(\boldsymbol{\omega})|}{\partial p}} \quad \Leftrightarrow \quad \frac{\Gamma_{22}}{p+\sigma^2} = \frac{\Gamma_{11} - e\Gamma_{12}^* - e^*\Gamma_{12} + ee^*\Gamma_{22}}{\sigma^2 ee^* + w + \sigma^2}
$$

$$
\Updownarrow
$$

$$
\Gamma_{22}\sigma^2 ee^* + \Gamma_{22}\left(w+\sigma^2\right) = \left(\Gamma_{11} - e^*\Gamma_{12}^* - e^*\Gamma_{12}\right)\left(p+\sigma^2\right) + pee^*\Gamma_{22} + \sigma^2 ee^*\Gamma_{22}
$$

$$
\Updownarrow
$$

$$
w+\sigma^2 = \frac{\Gamma_{11} - e\Gamma_{12}^* - e^*\Gamma_{12}}{\Gamma_{22}}\left(p+\sigma^2\right) + pee^* \Leftrightarrow w+\sigma^2 = \frac{\Gamma_{11}}{\Gamma_{22}}\left(p+\sigma^2\right) - p\frac{\left(p+\sigma^2\right)^2}{p^2}\frac{|\Gamma_{12}|^2}{(\Gamma_{22})^2}
$$

$$
\Updownarrow
$$

$$
w+\sigma^2 = \frac{\Gamma_{11}}{\Gamma_{22}}\left(p+\sigma^2\right) - pee^* \tag{54}
$$

• Exploitation of $(E3)$:

$$1=\frac{F(\boldsymbol{\omega})}{|\mathbf{G}(\boldsymbol{\omega})|}-\frac{\frac{\partial F(\boldsymbol{\omega})}{\partial v}}{\frac{\partial|\mathbf{G}(\boldsymbol{\omega})|}{\partial v}}=\frac{p\left(\Gamma_{11}-e\Gamma_{12}^{*}-e^{*}\Gamma_{12}+ee^{*}\Gamma_{22}\right)+\left(w+\sigma^{2}\right)\Gamma_{22}+\sigma^{2}\Gamma_{11}}{p\left(\sigma^{2}ee^{*}+\left(w+\sigma^{2}\right)\right)+\left(w+\sigma^{2}\right)\sigma^{2}}-\frac{\Gamma_{22}}{p+\sigma^{2}}$$

$$\Updownarrow$$

$$1=\frac{\frac{\Gamma_{22}}{p+\sigma^{2}}\left[p\left(\sigma^{2}ee^{*}+\left(w+\sigma^{2}\right)\right)+\sigma^{2}\left(w+\sigma^{2}\right)\right]+\frac{\Gamma_{22}}{p+\sigma^{2}}p\left(w+\sigma^{2}\right)+\sigma^{2}\Gamma_{11}}{p\left(\sigma^{2}ee^{*}+\left(w+\sigma^{2}\right)\right)+\left(w+\sigma^{2}\right)\sigma^{2}}-\frac{\Gamma_{22}}{p+\sigma^{2}}$$

$$\Updownarrow$$

$$1=\frac{\frac{\Gamma_{22}}{p+\sigma^{2}}p\left(w+\sigma^{2}\right)+\sigma^{2}\Gamma_{11}}{p\left(\sigma^{2}ee^{*}+\left(w+\sigma^{2}\right)\right)+\left(w+\sigma^{2}\right)\sigma^{2}}\quad\Leftrightarrow\quad 1=\frac{\frac{\Gamma_{22}}{p+\sigma^{2}}\left[\left(w+\sigma^{2}\right)p+\sigma^{2}\frac{\Gamma_{11}}{\Gamma_{22}}\left(p+\sigma^{2}\right)\right]}{p\left(\sigma^{2}ee^{*}+\left(w+\sigma^{2}\right)\right)+\left(w+\sigma^{2}\right)\sigma^{2}}$$

$$\Updownarrow$$

$$1=\frac{\frac{\Gamma_{22}}{p+\sigma^{2}}\left[\left(w+\sigma^{2}\right)p+\sigma^{2}\left(w+\sigma^{2}\right)+\sigma^{2}pee^{*}\right]}{\left(w+\sigma^{2}\right)p+\sigma^{2}\left(w+\sigma^{2}\right)+\sigma^{2}pee^{*}}\quad\Leftrightarrow\quad 1=\frac{\Gamma_{22}}{p+\sigma^{2}}\tag{55}$$

Finally:

$$\widehat{p}=\Gamma_{22}-\sigma^{2},\quad \widehat{e}=\frac{\Gamma_{12}}{\Gamma_{22}-\sigma^{2}},\quad \widehat{v}=\frac{\Gamma_{11}-\sigma^{2}}{\Gamma_{22}-\sigma^{2}}-|\widehat{e}|^{2}\tag{56}$$

2) 2 beams - real $\widehat{E_\theta\{R\}}$ - spatially white noise of known power:
Then $e=E\{R(\theta)\}\in\mathbb{R}$, and:

$$\mathbf{G}=\begin{bmatrix} pe^{2}+pv+\sigma^{2} & pe \\ pe & p+\sigma^{2}\end{bmatrix}$$

Applying the same derivation scheme used in Appendix IX-B1 but for e real, one obtains:

$$\widehat{p}=\Gamma_{22}-\sigma^{2},\quad \widehat{e}=\mathrm{Re}\left\{\frac{\Gamma_{12}}{\Gamma_{22}-\sigma^{2}}\right\},\quad \widehat{v}=\frac{\Gamma_{11}-\sigma^{2}}{\Gamma_{22}-\sigma^{2}}-\left(\widehat{e}\right)^{2}\tag{57}$$

3) 2 beams - complex $\widehat{E_\theta\{R\}}$ - spatially white noise of unknown power:
Then, one have to add to (52) the following additional extremality constraint:

$$\frac{\frac{\partial F(\boldsymbol{\omega})}{\partial w}}{\frac{\partial|\mathbf{G}(\boldsymbol{\omega})|}{\partial w}}=\frac{\frac{\partial F(\boldsymbol{\omega})}{\partial \sigma^{2}}}{\frac{\partial|\mathbf{G}(\boldsymbol{\omega})|}{\partial \sigma^{2}}}\Leftrightarrow\frac{\Gamma_{22}}{p+\sigma^{2}}=\frac{\Gamma_{22}+\Gamma_{11}}{pee^{*}+p+w+2\sigma^{2}}$$

$$\Updownarrow$$

$$\Gamma_{22}pee^{*}+\Gamma_{22}p+\Gamma_{22}w+\Gamma_{22}2\sigma^{2}=\sigma^{2}\Gamma_{22}+\sigma^{2}\Gamma_{11}+p\Gamma_{22}+p\Gamma_{11}$$

$$\Updownarrow$$

$$\Gamma_{22}pee^{*}+\Gamma_{22}\left(w+\sigma^{2}\right)=\Gamma_{11}\left(p+\sigma^{2}\right)\quad\Leftrightarrow\quad w+\sigma^{2}=\frac{\Gamma_{11}}{\Gamma_{22}}\left(p+\sigma^{2}\right)-pee^{*}\tag{58}$$

which appears to be equivalent to the already existing constraint:

$$\frac{\frac{\partial F(\boldsymbol{\omega})}{\partial w}}{\frac{\partial|\mathbf{G}(\boldsymbol{\omega})|}{\partial w}}=\frac{\frac{\partial F(\boldsymbol{\omega})}{\partial p}}{\frac{\partial|\mathbf{G}(\boldsymbol{\omega})|}{\partial p}},$$

what shows that w and σ^2 can not be identified simultaneously.

4) $(M+1)$ beams - real $\widehat{E_{\boldsymbol{\theta}}\{\mathbf{R}\}}$ - noise of known covariance matrix:
As $\ln p(\mathbf{B}_1,\ldots,\mathbf{B}_K)$ is a real-valued function so is $\frac{\partial \ln p(\mathbf{B}_1,\ldots,\mathbf{B}_K)}{\partial\omega_l}$, where ω_l is any of the unknown real parameters, hence:

$$tr\left(\frac{\partial\mathbf{G}(\boldsymbol{\omega})^{-1}}{\partial\omega_l}\left(\boldsymbol{\Gamma}-\mathbf{G}(\boldsymbol{\omega})\right)\right)=\mathrm{Re}\left\{tr\left(\frac{\partial\mathbf{G}(\boldsymbol{\omega})^{-1}}{\partial\omega_l}\left(\boldsymbol{\Gamma}-\mathbf{G}(\boldsymbol{\omega})\right)\right)\right\}=tr\left(\mathrm{Re}\left\{\frac{\partial\mathbf{G}(\boldsymbol{\omega})^{-1}}{\partial\omega_l}\left(\boldsymbol{\Gamma}-\mathbf{G}(\boldsymbol{\omega})\right)\right\}\right)$$

Let us consider the particular case where:
• $(P1)$: $\mathrm{Im}\left\{\mathbf{G}\left(\boldsymbol{\omega}\right)\right\}$ does not depend on the unknown vector of parameters $\boldsymbol{\omega}$,
• $(P2)$: $\mathrm{Re}\left\{\mathbf{G}\left(\boldsymbol{\omega}\right)\right\}$ depends symmetrically on the unknown vector of parameters $\boldsymbol{\omega}$,
then:

$$\frac{\partial \mathbf{G}\left(\boldsymbol{\omega}\right)^{-1}}{\partial \omega_l} = -\mathbf{G}\left(\boldsymbol{\omega}\right)^{-1} \frac{\partial \mathbf{G}\left(\boldsymbol{\omega}\right)}{\partial \omega_l} \mathbf{G}\left(\boldsymbol{\omega}\right)^{-1} = -\mathbf{G}\left(\boldsymbol{\omega}\right)^{-1} \frac{\partial \, \mathrm{Re}\left\{\mathbf{G}\left(\boldsymbol{\omega}\right)\right\}}{\partial \omega_l} \mathbf{G}\left(\boldsymbol{\omega}\right)^{-1}$$

and:

$$\left(\frac{\partial \mathbf{G}\left(\boldsymbol{\omega}\right)^{-1}}{\partial \omega_l}\right)^H = -\mathbf{G}\left(\boldsymbol{\omega}\right)^{-1} \left(\frac{\partial \, \mathrm{Re}\left\{\mathbf{G}\left(\boldsymbol{\omega}\right)\right\}}{\partial \omega_l}\right)^T \mathbf{G}\left(\boldsymbol{\omega}\right)^{-1} = \frac{\partial \mathbf{G}\left(\boldsymbol{\omega}\right)^{-1}}{\partial \omega_l}$$

Therefore, in this particular case:

$$\begin{aligned} tr\left(\frac{\partial \mathbf{G}\left(\boldsymbol{\omega}\right)^{-1}}{\partial \omega_l}\left(\boldsymbol{\Gamma} - \mathbf{G}\left(\boldsymbol{\omega}\right)\right)\right) &= tr\left(\mathrm{Re}\left\{\frac{\partial \mathbf{G}\left(\boldsymbol{\omega}\right)^{-1}}{\partial \omega_l}\left(\boldsymbol{\Gamma} - \mathbf{G}\left(\boldsymbol{\omega}\right)\right)\right\}\right) \\ &= tr\left(\frac{\partial \mathbf{G}\left(\boldsymbol{\omega}\right)^{-1}}{\partial \omega_l} \mathrm{Re}\left\{\boldsymbol{\Gamma} - \mathbf{G}\left(\boldsymbol{\omega}\right)\right\}\right) \end{aligned}$$

In the problem at hand $\mathbf{G}\left(\boldsymbol{\omega}\right)$ (6) is compliant with $(P1)$ and $(P2)$ since:

$$\mathbf{G}\left(\boldsymbol{\omega}\right) = P_s \begin{bmatrix} E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\mathbf{R}\left(\boldsymbol{\theta}\right)^T\right\} & E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\} \\ E_{\boldsymbol{\theta}}\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}^T & 1 \end{bmatrix} + \mathbf{G}^{\mathbf{v}}.$$

Let us denote $\widehat{\mathbf{G}} = \mathbf{G}\left(\widehat{\boldsymbol{\omega}}\right)$. If $\widehat{P_s}$, $\widehat{E_{\boldsymbol{\theta}}\left\{\mathbf{R}\right\}}$, $\widehat{cov_{\boldsymbol{\theta}}\left\{\mathbf{R}\right\}}$ are given by (29)-(31) then:

$$\widehat{\mathbf{G}} = \mathrm{Re}\left\{\boldsymbol{\Gamma}\right\} - \mathrm{Re}\left\{\mathbf{G}^{\mathbf{v}}\right\} + \mathbf{G}^{\mathbf{v}} = \mathrm{Re}\left\{\boldsymbol{\Gamma}\right\} + j\,\mathrm{Im}\left\{\mathbf{G}^{\mathbf{v}}\right\}$$

and:

$$\mathrm{Re}\left\{\boldsymbol{\Gamma} - \widehat{\mathbf{G}}\right\} = \mathrm{Re}\left\{\boldsymbol{\Gamma}\right\} - \mathrm{Re}\left\{\widehat{\mathbf{G}}\right\} = 0 \quad \Rightarrow \quad tr\left(\frac{\partial \mathbf{G}\left(\boldsymbol{\omega}\right)^{-1}}{\partial \omega_l}\left(\boldsymbol{\Gamma} - \widehat{\mathbf{G}}\right)\right) = 0$$

C. Statistical prediction

The derivation of $E\left\{\widehat{P_s} \mid \Sigma\right\}$, $E\left\{\widehat{\mathbf{m_R}} \mid \Sigma\right\}$, $E\left\{\widehat{\mathbf{m_R}}\widehat{\mathbf{m_R}}^T \mid \Sigma\right\}$ and $E\left\{\widehat{\mathbf{m_R}}\widehat{\mathbf{m_R}}^H \mid \Sigma\right\}$ is an extension of the results introduced in [11][17]. The first step of computation of the three conditional expectations still relies on conditional Gaussian distribution applied to $\mathbf{B} = \left(\mathbf{B}_1^T \ \ldots \ \mathbf{B}_K^T\right)^T$ where $\mathbf{B}_k = \left(\mathbf{D}_k^T, S_k\right)^T$, $\mathbf{D} = \left(\mathbf{D}_1^T \ \ldots \ \mathbf{D}_K^T\right)^T$, $\mathbf{S} = \left(S_1, \ldots, S_K\right)^T$. Then $\mathbf{B} \sim \mathcal{CN}_{(M+1)K}\left(\mathbf{0}, \mathbf{C_B}\right)$ and:

$$p_{\mathcal{CN}_{(M+1)K}}\left(\mathbf{B}; \mathbf{0}, \mathbf{C_B}\right) = \frac{e^{-\mathbf{B}^H \mathbf{C}_\mathbf{B}^{-1} \mathbf{B}}}{\pi^{(M+1)K}\left|\mathbf{C_B}\right|} = p_{\mathcal{CN}_{MK}}\left(\mathbf{D}; \mathbf{t_{D|S}}, \mathbf{C_{D|S}}\right) p_{\mathcal{CN}_K}\left(\mathbf{S}; \mathbf{0}, \mathbf{G_S}\right) \tag{59}$$

$$\mathbf{C_B} = \mathbf{I}_K \otimes \mathbf{G}, \quad \mathbf{C_{D|S}} = \mathbf{I}_K \otimes \mathbf{G}_{\mathbf{D}|S}, \quad \mathbf{G}_{\mathbf{D}|S} = \mathbf{G_D} - G_S \boldsymbol{\rho}\boldsymbol{\rho}^H, \quad \mathbf{t_{D|S}} = \mathbf{S} \otimes \boldsymbol{\rho}, \quad \boldsymbol{\rho} = \frac{\mathbf{G}_{\mathbf{D}S}}{G_S}.$$

Identity (59) allows the computation of the required conditional expectations in two steps:

$$\begin{aligned} E\left\{q\left(\mathbf{D}, \mathbf{S}\right) \mid \Sigma\right\} &= \iint_{\Sigma} q\left(\mathbf{D}, \mathbf{S}\right) \frac{p_{\mathcal{CN}_{(M+1)K}}\left(\mathbf{B}; \mathbf{0}, \mathbf{C_B}\right)}{P_\Sigma} d\mathbf{B} \\ &= \int_{\Sigma} \left\{\int q\left(\mathbf{D}, \mathbf{S}\right) p_{\mathcal{CN}_{MK}}\left(\mathbf{D}; \mathbf{t_{D|S}}, \mathbf{C_{D|S}}\right) d\mathbf{D}\right\} \frac{p_{\mathcal{CN}_K}\left(\mathbf{S}; \mathbf{0}, \mathbf{G_S}\right)}{P_\Sigma} d\mathbf{S} \\ E\left\{q\left(\mathbf{D}, \mathbf{S}\right) \mid \Sigma\right\} &= \int_{\Sigma} E\left\{q\left(\mathbf{D}, \mathbf{S}\right) \mid \mathbf{S}\right\} \frac{p_{\mathcal{CN}_K}\left(\mathbf{S}; \mathbf{0}, \mathbf{G_S}\right)}{P_\Sigma} d\mathbf{S} \end{aligned}$$

Using the abbreviation for the set of m^{th} difference beam samples $\mathbf{D}^m = \left(D_m\left(\mathbf{z}_1\right), \ldots, D_m\left(\mathbf{z}_K\right)\right)^T$, we denote $q\left(\mathbf{D}, \mathbf{S}\right) = \frac{\mathbf{S}^H \mathbf{D}^m - K\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)_m}{\left(\mathbf{S}^H \mathbf{S} - K G_S^{\mathbf{v}}\right)^L}$, $q\left(\mathbf{D}, \mathbf{S}\right) = \frac{\mathbf{S}^H \mathbf{D}^m - K\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)_m}{\left(\mathbf{S}^H \mathbf{S} - K G_S^{\mathbf{v}}\right)^L} \frac{\mathbf{S}^H \mathbf{D}^l - K\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)_l}{\left(\mathbf{S}^H \mathbf{S} - K G_S^{\mathbf{v}}\right)^L}$ or $q\left(\mathbf{D}, \mathbf{S}\right) = \frac{\mathbf{S}^H \mathbf{D}^m - K\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)_m}{\left(\mathbf{S}^H \mathbf{S} - K G_S^{\mathbf{v}}\right)^L} \left(\frac{\mathbf{S}^H \mathbf{D}^l - K\left(\mathbf{G}_{\mathbf{D}}^{\mathbf{v}}\right.}{\left(\mathbf{S}^H \mathbf{S} - K G_S^{\mathbf{v}}\right)}\right.$

$L = 1, 2$.
Let us define:

$$P\left(K, L\right) = \int\limits_{t \geq \eta} \frac{p_{\chi^2_{2K}}\left(t; 0, \frac{G_S}{2}\right)}{P_\Sigma} \frac{dt}{\left(t - KG^{\mathbf{v}}_S\right)^L}$$

1) 1rst step:

- **Calculation of** $E\left\{\widehat{P_s} \mid \Sigma\right\}$:

$$E\left\{\widehat{P_s} \mid \Sigma\right\} = \frac{1}{K}E\left\{\mathbf{S}^H\mathbf{S} - KG^{\mathbf{v}}_S \mid \Sigma\right\} = \frac{1}{K}E\left\{\mathbf{S}^H\mathbf{S} \mid \Sigma\right\} - G^{\mathbf{v}}_S \quad (60)$$

$$E\left\{\mathbf{S}^H\mathbf{S} \mid \Sigma\right\} = \int\limits_{t \geq \eta} t p_{\chi^2_{2K}}\left(t; 0, \frac{G_S}{2}\right) \frac{dt}{P_\Sigma} \quad (61)$$

- **Calculation of** $cov\left\{\widehat{P_s} \mid \Sigma\right\}$:

$$cov\left\{\widehat{P_s} \mid \Sigma\right\} = \frac{1}{K^2}E\left\{\left(\mathbf{S}^H\mathbf{S}\right)^2 \mid \Sigma\right\} - \left(\frac{1}{K}E\left\{\mathbf{S}^H\mathbf{S} \mid \Sigma\right\}\right)^2 \quad (62)$$

$$E\left\{\left(\mathbf{S}^H\mathbf{S}\right)^2 \mid \Sigma\right\} = \int\limits_{t \geq \eta} t^2 p_{\chi^2_{2K}}\left(t; 0, \frac{G_S}{2}\right) \frac{dt}{P_\Sigma} \quad (63)$$

- **Calculation of** $E\left\{\widehat{\mathbf{m_R}} \mid \Sigma\right\}$:

$$E\left\{\widehat{m_{R_m}} \mid \Sigma\right\} = E\left\{\frac{\mathbf{S}^H\mathbf{D}^m - K\left(\mathbf{G}^{\mathbf{v}}_{\mathbf{D}S}\right)_m}{\mathbf{S}^H\mathbf{S} - KG^{\mathbf{v}}_S} \mid \Sigma\right\}$$

$$E\left\{\widehat{m_{R_m}} \mid \Sigma\right\} = E\left\{\frac{\mathbf{S}^H\mathbf{D}^m}{\mathbf{S}^H\mathbf{S} - KG^{\mathbf{v}}_S} \mid \Sigma\right\} - K\left(\mathbf{G}^{\mathbf{v}}_{\mathbf{D}S}\right)_m P\left(K, 1\right) \quad (64)$$

- **Calculation of** $E\left\{\widehat{\mathbf{m_R}}\widehat{\mathbf{m_R}}^T \mid \Sigma\right\}$:

$$E\left\{\widehat{m_{R_m}}\widehat{m_{R_l}} \mid \Sigma\right\} = E\left\{\left(\frac{\mathbf{S}^H\mathbf{D}^m - K\left(\mathbf{G}^{\mathbf{v}}_{\mathbf{D}S}\right)_m}{\mathbf{S}^H\mathbf{S} - KG^{\mathbf{v}}_S}\right)\left(\frac{\mathbf{S}^H\mathbf{D}^l - K\left(\mathbf{G}^{\mathbf{v}}_{\mathbf{D}S}\right)_l}{\mathbf{S}^H\mathbf{S} - KG^{\mathbf{v}}_S}\right) \mid \Sigma\right\}$$

$$E\left\{\widehat{m_{R_m}}\widehat{m_{R_l}} \mid \Sigma\right\} = E\left\{\frac{\mathbf{S}^H\mathbf{D}^m\mathbf{S}^H\mathbf{D}^l}{\left(\mathbf{S}^H\mathbf{S} - KG^{\mathbf{v}}_S\right)^2} \mid \Sigma\right\} + K^2\left(\mathbf{G}^{\mathbf{v}}_{\mathbf{D}S}\right)_m\left(\mathbf{G}^{\mathbf{v}}_{\mathbf{D}S}\right)_l P\left(K, 2\right) \quad (65)$$

$$-K\left(\mathbf{G}^{\mathbf{v}}_{\mathbf{D}S}\right)_m E\left\{\frac{\mathbf{S}^H\mathbf{D}^l}{\left(\mathbf{S}^H\mathbf{S} - KG^{\mathbf{v}}_S\right)^2} \mid \Sigma\right\} - K\left(\mathbf{G}^{\mathbf{v}}_{\mathbf{D}S}\right)_l E\left\{\frac{\mathbf{S}^H\mathbf{D}^m}{\left(\mathbf{S}^H\mathbf{S} - KG^{\mathbf{v}}_S\right)^2} \mid \Sigma\right\}$$

- **Calculation of** $E\left\{\widehat{\mathbf{m_R}}\widehat{\mathbf{m_R}}^H \mid \Sigma\right\}$:

$$E\left\{\widehat{m_{R_m}}\widehat{m_{R_l}}^* \mid \Sigma\right\} = E\left\{\left(\frac{\mathbf{S}^H\mathbf{D}^m - K\left(\mathbf{G}^{\mathbf{v}}_{\mathbf{D}S}\right)_m}{\mathbf{S}^H\mathbf{S} - KG^{\mathbf{v}}_S}\right)\left(\frac{\mathbf{S}^H\mathbf{D}^l - K\left(\mathbf{G}^{\mathbf{v}}_{\mathbf{D}S}\right)_l}{\mathbf{S}^H\mathbf{S} - KG^{\mathbf{v}}_S}\right)^* \mid \Sigma\right\}$$

$$E\left\{\widehat{m_{R_m}}\widehat{m_{R_l}}^* \mid \Sigma\right\} = E\left\{\frac{\mathbf{S}^H\mathbf{D}^m\left(\mathbf{D}^l\right)^H\mathbf{S}}{\left(\mathbf{S}^H\mathbf{S} - KG^{\mathbf{v}}_S\right)^2} \mid \Sigma\right\} + K^2\left(\mathbf{G}^{\mathbf{v}}_{\mathbf{D}S}\right)_m\left(\mathbf{G}^{\mathbf{v}}_{\mathbf{D}S}\right)_l^* P\left(K, 2\right) \quad (66)$$

$$-K\left(\mathbf{G}^{\mathbf{v}}_{\mathbf{D}S}\right)_m E\left\{\frac{\mathbf{S}^H\mathbf{D}^l}{\left(\mathbf{S}^H\mathbf{S} - KG^{\mathbf{v}}_S\right)^2} \mid \Sigma\right\}^* - K\left(\mathbf{G}^{\mathbf{v}}_{\mathbf{D}S}\right)_l^* E\left\{\frac{\mathbf{S}^H\mathbf{D}^m}{\left(\mathbf{S}^H\mathbf{S} - KG^{\mathbf{v}}_S\right)^2} \mid \Sigma\right\}$$

- **Calculation of** $E\left\{\frac{K\mathbf{\Gamma_D}-K\mathbf{G}_{\mathbf{D}}^{\mathbf{v}}}{\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}} \mid \Sigma\right\}$:

$$E\left\{\frac{(K\mathbf{\Gamma_D}-K\mathbf{G}_{\mathbf{D}}^{\mathbf{v}})_{ml}}{\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}} \mid \Sigma\right\} = E\left\{\frac{K(\mathbf{\Gamma_D})_{ml}}{\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}} \mid \Sigma\right\} - K(\mathbf{G}_{\mathbf{D}}^{\mathbf{v}})_{ml}\,E\left\{\frac{1}{\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}} \mid \Sigma\right\} \tag{67}$$
$$= E\left\{\frac{\left(\mathbf{D}^l\right)^H\mathbf{D}^m}{\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}} \mid \Sigma\right\} - K(\mathbf{G}_{\mathbf{D}}^{\mathbf{v}})_{ml}\,P(K,1)$$

- **Calculation of** $E\left\{\sum_{k=1}^{K}\frac{\mathbf{D}_k\mathbf{D}_k^T}{\Gamma_S-G_S^{\mathbf{v}}} \mid \Sigma\right\}$:

$$E\left\{\sum_{k=1}^{K}\frac{\left(\mathbf{D}_k\mathbf{D}_k^T\right)_{ml}}{\Gamma_S-G_S^{\mathbf{v}}} \mid \Sigma\right\} = E\left\{\frac{\left(\mathbf{D}^l\right)^T\mathbf{D}^m}{\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}} \mid \Sigma\right\} \tag{68}$$

2) 2nd step:

- **Calculation of** $E\{\widehat{\mathbf{m_R}} \mid \Sigma\}$:

$$E\left\{\frac{\mathbf{S}^H\mathbf{D}^m}{\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}} \mid \Sigma\right\} = \int_{\Sigma}\frac{\mathbf{S}^H E\{\mathbf{D}^m \mid \mathbf{S}\}}{\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}}\frac{p_{\mathcal{CN}_K}(\mathbf{S};\mathbf{0},\mathbf{G_S})}{P_\Sigma}d\mathbf{S} = \int_{\Sigma}\frac{\mathbf{S}^H(\rho_m\mathbf{S})}{\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}}\frac{p_{\mathcal{CN}_K}(\mathbf{S};\mathbf{0},\mathbf{G_S})}{P_\Sigma}d\mathbf{S}$$
$$= \rho_m E\left\{\frac{\mathbf{S}^H\mathbf{S}}{\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}} \mid \Sigma\right\} = \rho_m\left(1+KG_S^{\mathbf{v}}P(K,1)\right)$$

Therefore:

$$E\{\widehat{\mathbf{m_R}} \mid \Sigma\} = \boldsymbol{\rho} + \left(G_S^{\mathbf{v}}\boldsymbol{\rho} - \mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)KP(K,1) \tag{69}$$

- **Calculation of** $E\left\{\widehat{\mathbf{m_R}}\widehat{\mathbf{m_R}}^T \mid \Sigma\right\}$:

$$E\left\{\frac{\mathbf{S}^H\mathbf{D}^m\mathbf{S}^H\mathbf{D}^l}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^2} \mid \Sigma\right\} = \int_{\Sigma}\frac{\mathbf{S}^H E\left\{\mathbf{D}^m\left(\mathbf{D}^l\right)^T \mid \mathbf{S}\right\}\mathbf{S}^*}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^2}\frac{p_{\mathcal{CN}_K}(\mathbf{S};\mathbf{0},\mathbf{G_S})}{P_\Sigma}d\mathbf{S}$$
$$= \int_{\Sigma}\frac{\mathbf{S}^H\left(\mathbf{t}_{\mathbf{D}^m|\mathbf{S}}\mathbf{t}_{\mathbf{D}^l|\mathbf{S}}^T\right)\mathbf{S}^*}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^2}\frac{p_{\mathcal{CN}_K}(\mathbf{S};\mathbf{0},\mathbf{G_S})}{P_\Sigma}d\mathbf{S}$$
$$= \int_{\Sigma}\frac{\mathbf{S}^H\left(\rho_m\rho_l\mathbf{S}\mathbf{S}^T\right)\mathbf{S}^*}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^2}\frac{p_{\mathcal{CN}_K}(\mathbf{S};\mathbf{0},\mathbf{G_S})}{P_\Sigma}d\mathbf{S}$$
$$= \rho_m\rho_l E\left\{\frac{\left(\mathbf{S}^H\mathbf{S}\right)^2}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^2} \mid \Sigma\right\}$$

and:

$$E\left\{\frac{\left(\mathbf{S}^H\mathbf{S}\right)^2}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^2} \mid \Sigma\right\} = 1 + 2KG_S^{\mathbf{v}}P(K,1) + \left(KG_S^{\mathbf{v}}\right)^2P(K,2)$$

Finally :

$$E\left\{\frac{\mathbf{S}^H\mathbf{D}^m\mathbf{S}^H\mathbf{D}^l}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^2} \mid \Sigma\right\} = \rho_m\rho_l\left(1 + 2KG_S^{\mathbf{v}}P(K,1) + \left(KG_S^{\mathbf{v}}\right)^2P(K,2)\right) \tag{70}$$

$$E\left\{\frac{\mathbf{S}^H\mathbf{D}^m}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^2} \mid \Sigma\right\} = \int_{\Sigma}\frac{\mathbf{S}^H E\{\mathbf{D}^m \mid \mathbf{S}\}}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^2}\frac{p_{\mathcal{CN}_K}(\mathbf{S};\mathbf{0},\mathbf{G_S})}{P_\Sigma}d\mathbf{S}$$
$$= \int_{\Sigma}\frac{\mathbf{S}^H(\rho_m\mathbf{S})}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^2}\frac{p_{\mathcal{CN}_K}(\mathbf{S};\mathbf{0},\mathbf{G_S})}{P_\Sigma}d\mathbf{S}$$
$$= \rho_m E\left\{\frac{\mathbf{S}^H\mathbf{S}}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^2} \mid \Sigma\right\}$$

and:

$$E\left\{\frac{\mathbf{S}^H\mathbf{S}}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^2}\mid\Sigma\right\}=P\left(K,1\right)+KG_S^{\mathbf{v}}P\left(K,2\right)$$

Finally:

$$E\left\{\frac{\mathbf{S}^H\mathbf{D}^m}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^2}\mid\Sigma\right\}=\rho_m\left(P\left(K,1\right)+KG_S^{\mathbf{v}}P\left(K,2\right)\right) \tag{71}$$

Finally from (65)(70)(71):

$$\begin{aligned} E\left\{\widehat{m_{R_m}}\widehat{m_{R_l}}\mid\Sigma\right\} &= \rho_m\rho_l\left(1+2KG_S^{\mathbf{v}}P\left(K,1\right)+\left(KG_S^{\mathbf{v}}\right)^2P\left(K,2\right)\right)+\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)_m\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)_lK^2P\left(K,2\right)\\ &\quad-\left(\rho_m\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)_l+\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)_m\rho_l\right)\left(KP\left(K,1\right)+G_S^{\mathbf{v}}K^2P\left(K,2\right)\right)\end{aligned}$$

$$\begin{aligned} E\left\{\widehat{\mathbf{m_R}}\widehat{\mathbf{m_R}}^T\mid\Sigma\right\} &= \boldsymbol{\rho}\boldsymbol{\rho}^T\left(1+2G_S^{\mathbf{v}}KP\left(K,1\right)+\left(G_S^{\mathbf{v}}\right)^2K^2P\left(K,2\right)\right)+\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)^T\right)K^2P\left(K,2\right)\\ &\quad-\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\boldsymbol{\rho}^T+\boldsymbol{\rho}\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)^T\right)\left(KP\left(K,1\right)+K^2G_S^{\mathbf{v}}P\left(K,2\right)\right)\end{aligned} \tag{72}$$

- **Calculation of** $E\left\{\widehat{\mathbf{m_R}}\widehat{\mathbf{m_R}}^H\mid\Sigma\right\}$:

$$\begin{aligned} E\left\{\frac{\mathbf{S}^H\mathbf{D}^m\left(\mathbf{D}^l\right)^H\mathbf{S}}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^2}\mid\Sigma\right\} &= \int_\Sigma\frac{\mathbf{S}^HE\left\{\mathbf{D}^m\left(\mathbf{D}^l\right)^H\mid\mathbf{S}\right\}\mathbf{S}}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^2}\frac{p_{\mathcal{CN}_K}\left(\mathbf{S};\mathbf{0},\mathbf{G_S}\right)}{P_\Sigma}d\mathbf{S}\\ &= \int_\Sigma\frac{\mathbf{S}^H\left\{\mathbf{C}_{\mathbf{D}^m\mathbf{D}^l|\mathbf{S}}+\mathbf{t}_{\mathbf{D}^m|\mathbf{S}}\mathbf{t}_{\mathbf{D}^l|\mathbf{S}}^H\right\}\mathbf{S}}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^2}\frac{p_{\mathcal{CN}_K}\left(\mathbf{S};\mathbf{0},\mathbf{G_S}\right)}{P_\Sigma}d\mathbf{S}\\ &= \int_\Sigma\frac{\mathbf{S}^H\left(G_{D^mD^l|S}\mathbf{I}_K+\rho_m\rho_l^*\mathbf{S}\mathbf{S}^H\right)\mathbf{S}}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^2}\frac{p_{\mathcal{CN}_K}\left(\mathbf{S};\mathbf{0},\mathbf{G_S}\right)}{P_\Sigma}d\mathbf{S}\\ &= G_{D^mD^l|S}E\left\{\frac{\mathbf{S}^H\mathbf{S}}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^2}\mid\Sigma\right\}+\rho_m\rho_l^*E\left\{\frac{\left(\mathbf{S}^H\mathbf{S}\right)^2}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^2}\mid\Sigma\right\}\end{aligned}$$

$$\begin{aligned} E\left\{\frac{\mathbf{S}^H\mathbf{D}^m\left(\mathbf{D}^l\right)^H\mathbf{S}}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^2}\mid\Sigma\right\} &= G_{D^mD^l|S}\left(P\left(K,1\right)+KG_S^{\mathbf{v}}P\left(K,2\right)\right)\\ &\quad+\rho_m\rho_l^*\left(1+2KG_S^{\mathbf{v}}P\left(K,1\right)+\left(KG_S^{\mathbf{v}}\right)^2P\left(K,2\right)\right)\end{aligned} \tag{73}$$

Finally from (66)(71)(73):

$$\begin{aligned} E\left\{\widehat{m_{R_m}}\widehat{m_{R_l}}^*\mid\Sigma\right\} &= \left(G_{\mathbf{D}|S}\right)_{m,l}\left(P\left(K,1\right)+KG_S^{\mathbf{v}}P\left(K,2\right)\right)\\ &\quad+\rho_m\rho_l^*\left(1+2KG_S^{\mathbf{v}}P\left(K,1\right)+\left(KG_S^{\mathbf{v}}\right)^2P\left(K,2\right)\right)\\ &\quad-\left(\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)_l^*\rho_m+\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)_m\rho_l^*\right)\left(KP\left(K,1\right)+K^2G_S^{\mathbf{v}}P\left(K,2\right)\right)\\ &\quad+\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)_m\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)_l^*K^2P\left(K,2\right)\end{aligned}$$

$$\begin{aligned} E\left\{\widehat{\mathbf{m_R}}\widehat{\mathbf{m_R}}^H\mid\Sigma\right\} &= \left(\mathbf{G_D}-G_S\boldsymbol{\rho}\boldsymbol{\rho}^H-K\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\boldsymbol{\rho}^H-K\boldsymbol{\rho}\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)^H\right)\left(P\left(K,1\right)+KG_S^{\mathbf{v}}P\left(K,2\right)\right)\\ &\quad+\boldsymbol{\rho}\boldsymbol{\rho}^H\left(1+2KG_S^{\mathbf{v}}P\left(K,1\right)+\left(KG_S^{\mathbf{v}}\right)^2P\left(K,2\right)\right)+\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)^HK^2P\left(K,2\right)\end{aligned} \tag{74}$$

• **Calculation of** $E\left\{\frac{K\mathbf{\Gamma_D}-K\mathbf{G}^{\mathbf{v}}_{\mathbf{D}}}{\mathbf{S}^H\mathbf{S}-KG^{\mathbf{v}}_S}\mid\Sigma\right\}$

$$\begin{aligned}
E\left\{\frac{\left(\mathbf{D}^l\right)^H\mathbf{D}^m}{\mathbf{S}^H\mathbf{S}-KG^{\mathbf{v}}_S}\mid\Sigma\right\} &= tr\left(\int_\Sigma \frac{E\left\{\mathbf{D}^m\left(\mathbf{D}^l\right)^H\mid\mathbf{S}\right\}}{\mathbf{S}^H\mathbf{S}-KG^{\mathbf{v}}_S}\frac{p_{\mathcal{CN}_K}\left(\mathbf{S};\mathbf{0},\mathbf{G_S}\right)}{P_\Sigma}d\mathbf{S}\right)\\
&= tr\left(\int_\Sigma \frac{\mathbf{C}_{\mathbf{D}^m\mathbf{D}^l|\mathbf{S}}+\mathbf{t}_{\mathbf{D}^m|\mathbf{S}}\mathbf{t}^H_{\mathbf{D}^l|\mathbf{S}}}{\mathbf{S}^H\mathbf{S}-KG^{\mathbf{v}}_S}\frac{p_{\mathcal{CN}_K}\left(\mathbf{S};\mathbf{0},\mathbf{G_S}\right)}{P_\Sigma}d\mathbf{S}\right)\\
&= tr\left(\int_\Sigma \frac{G_{D^mD^l|S}\mathbf{I}_K+\rho_m\rho_l^*\mathbf{S}\mathbf{S}^H}{\mathbf{S}^H\mathbf{S}-KG^{\mathbf{v}}_S}\frac{p_{\mathcal{CN}_K}\left(\mathbf{S};\mathbf{0},\mathbf{G_S}\right)}{P_\Sigma}d\mathbf{S}\right)\\
&= KG_{D^mD^l|S}E\left\{\frac{1}{\mathbf{S}^H\mathbf{S}-KG^{\mathbf{v}}_S}\mid\Sigma\right\}+\rho_m\rho_l^*E\left\{\frac{\mathbf{S}^H\mathbf{S}}{\mathbf{S}^H\mathbf{S}-KG^{\mathbf{v}}_S}\mid\Sigma\right\}\\
&= G_{D^mD^l|S}KP\left(K,1\right)+\rho_m\rho_l^*\left(1+KG^{\mathbf{v}}_SP\left(K,1\right)\right)
\end{aligned}$$

$$E\left\{\frac{\left(K\mathbf{\Gamma_D}-K\mathbf{G}^{\mathbf{v}}_{\mathbf{D}}\right)_{ml}}{\mathbf{S}^H\mathbf{S}-KG^{\mathbf{v}}_S}\mid\Sigma\right\}=E\left\{\frac{\left(\mathbf{D}^l\right)^H\mathbf{D}^m}{\mathbf{S}^H\mathbf{S}-KG^{\mathbf{v}}_S}\mid\Sigma\right\}-\left(\mathbf{G}^{\mathbf{v}}_{\mathbf{D}}\right)_{ml}KP\left(K,1\right)$$

$$\begin{aligned}
E\left\{\frac{K\mathbf{\Gamma_D}-K\mathbf{G}^{\mathbf{v}}_{\mathbf{D}}}{\mathbf{S}^H\mathbf{S}-KG^{\mathbf{v}}_S}\mid\Sigma\right\} &= \left(\mathbf{G_{D|S}}-\mathbf{G}^{\mathbf{v}}_{\mathbf{D}}+G^{\mathbf{v}}_S\boldsymbol{\rho}\boldsymbol{\rho}^H\right)KP\left(K,1\right)+\boldsymbol{\rho}\boldsymbol{\rho}^H\\
&= \left(\mathbf{G_D}-\mathbf{G}^{\mathbf{v}}_{\mathbf{D}}-\left(G_S-G^{\mathbf{v}}_S\right)\boldsymbol{\rho}\boldsymbol{\rho}^H\right)KP\left(K,1\right)+\boldsymbol{\rho}\boldsymbol{\rho}^H
\end{aligned}$$

• **Calculation of** $E\left\{\sum_{k=1}^{K}\frac{\mathbf{D}_k\mathbf{D}_k^T}{\Gamma_S-G^{\mathbf{v}}_S}\mid\Sigma\right\}$

$$\begin{aligned}
E\left\{\sum_{k=1}^{K}\frac{\left(\mathbf{D}_k\mathbf{D}_k^T\right)_{ml}}{\Gamma_S-G^{\mathbf{v}}_S}\mid\Sigma\right\} &= tr\left(\int_\Sigma \frac{E\left\{\mathbf{D}^m\left(\mathbf{D}^l\right)^T\mid\mathbf{S}\right\}}{\mathbf{S}^H\mathbf{S}-KG^{\mathbf{v}}_S}\frac{p_{\mathcal{CN}_K}\left(\mathbf{S};\mathbf{0},\mathbf{G_S}\right)}{P_\Sigma}d\mathbf{S}\right)\\
&= tr\left(\int_\Sigma \frac{\mathbf{t}_{\mathbf{D}^m|\mathbf{S}}\mathbf{t}^T_{\mathbf{D}^l|\mathbf{S}}}{\mathbf{S}^H\mathbf{S}-KG^{\mathbf{v}}_S}\frac{p_{\mathcal{CN}_K}\left(\mathbf{S};\mathbf{0},\mathbf{G_S}\right)}{P_\Sigma}d\mathbf{S}\right)\\
&= \rho_m\rho_l tr\left(\int_\Sigma \frac{\mathbf{S}\mathbf{S}^T}{\mathbf{S}^H\mathbf{S}-KG^{\mathbf{v}}_S}\frac{p_{\mathcal{CN}_K}\left(\mathbf{S};\mathbf{0},\mathbf{G_S}\right)}{P_\Sigma}d\mathbf{S}\right)\\
&= 0
\end{aligned}$$

D. Statistical prediction (continued)

Another form of (59) is:

$$p_{\mathcal{CN}_{(M+1)K}}\left(\mathbf{B};\mathbf{0},\mathbf{C_B}\right)=\left(\prod_{k=1}^{K}p_{\mathcal{CN}_M}\left(\mathbf{D}_k;\mathbf{t}_{\mathbf{D}_k|S_k},\mathbf{C}_{\mathbf{D}_k|S_k}\right)\right)\left(\prod_{k=1}^{K}p_{\mathcal{CN}_1}\left(S_k;0,G_S\right)\right)$$

$$\mathbf{C}_{\mathbf{D}_k|S_k}=\mathbf{C}_{\mathbf{D}|S}=\mathbf{G_D}-G_S\boldsymbol{\rho}\boldsymbol{\rho}^H,\quad \mathbf{t}_{\mathbf{D}_k|S_k}=S_k\boldsymbol{\rho}$$

Additionally, as $\mathbf{D}-\mathbf{S}\otimes\boldsymbol{\rho}\mid\mathbf{S}$ is complex Gaussian circular centered, then:

• $\{\mathbf{D}_k-\boldsymbol{\rho}S_k\mid S_k\}$ are mutually complex Gaussian circular centered,
• $\{\mathbf{D}^m-\rho_m\mathbf{S}\mid\mathbf{S}\}$ are mutually complex Gaussian circular centered,
• $\left\{\left(\mathbf{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_m=\mathbf{S}^H\left(\mathbf{D}^m-\rho_m\mathbf{S}\right)\mid\mathbf{S}\right\}$ are mutually complex Gaussian circular centered.

Therefore, let us define $\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}$ and $\boldsymbol{\Gamma}_{\mathbf{D}|\mathbf{S}}$ as:

$$
\begin{aligned}
\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}} &= \frac{1}{K}\sum_{k=1}^{K}\left(\mathbf{D}_k-\boldsymbol{\rho}S_k\right)S_k^{*} = \boldsymbol{\Gamma}_{\mathbf{D}S} - \Gamma_S\boldsymbol{\rho} \\
\boldsymbol{\Gamma}_{\mathbf{D}|\mathbf{S}} &= \frac{1}{K}\sum_{k=1}^{K}\left(\mathbf{D}_k-\boldsymbol{\rho}S_k\right)\left(\mathbf{D}_k-\boldsymbol{\rho}S_k\right)^{H} = \boldsymbol{\Gamma}_{\mathbf{D}} - \boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\boldsymbol{\rho}^{H} - \boldsymbol{\rho}\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}^{H} - \Gamma_S\boldsymbol{\rho}\boldsymbol{\rho}^{H}
\end{aligned}
$$

Then:

$$
\begin{aligned}
cov_{\boldsymbol{\theta}}\widehat{\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}} &= \frac{\boldsymbol{\Gamma}_{\mathbf{D}}-\mathbf{G}_{\mathbf{D}}^{\mathbf{v}}}{\Gamma_S-G_S^{\mathbf{v}}} - \left(\frac{\boldsymbol{\Gamma}_{\mathbf{D}S}-\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}}{\Gamma_S-G_S^{\mathbf{v}}}\right)\left(\frac{\boldsymbol{\Gamma}_{\mathbf{D}S}-\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}}{\Gamma_S-G_S^{\mathbf{v}}}\right)^{H} \\
cov_{\boldsymbol{\theta}}\widehat{\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}} &= \frac{\boldsymbol{\Gamma}_{\mathbf{D}|\mathbf{S}}+\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\boldsymbol{\rho}^{H}+\boldsymbol{\rho}\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}^{H}+\left(\Gamma_S\boldsymbol{\rho}\boldsymbol{\rho}^{H}-\mathbf{G}_{\mathbf{D}}^{\mathbf{v}}\right)}{\Gamma_S-G_S^{\mathbf{v}}} \\
&\quad - \left(\frac{\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}+\left(\Gamma_S\boldsymbol{\rho}-\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)}{\Gamma_S-G_S^{\mathbf{v}}}\right)\left(\frac{\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}+\left(\Gamma_S\boldsymbol{\rho}-\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)}{\Gamma_S-G_S^{\mathbf{v}}}\right)^{H} \\
cov_{\boldsymbol{\theta}}\widehat{\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}} &= \frac{\boldsymbol{\Gamma}_{\mathbf{D}|\mathbf{S}}}{\left(\Gamma_S-G_S^{\mathbf{v}}\right)} + \frac{\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\mathbf{v}^{H}+\mathbf{v}\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}^{H}-\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}^{H}+\mathbf{M}_S}{\left(\Gamma_S-G_S^{\mathbf{v}}\right)^{2}}
\end{aligned}
$$

where:

$$
\boldsymbol{\nu} = \mathbf{G}_{\mathbf{D}S}^{\mathbf{v}} - \boldsymbol{\rho}G_S^{\mathbf{v}}, \quad \mathbf{M}_S = \frac{\mathbf{S}^{H}\mathbf{S}}{K}\mathbf{A}+\mathbf{B}
$$

$$
\mathbf{A} = \mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\boldsymbol{\rho}^{H} + \boldsymbol{\rho}\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)^{H} - G_S^{\mathbf{v}}\boldsymbol{\rho}\boldsymbol{\rho}^{H} - \mathbf{G}_{\mathbf{D}}^{\mathbf{v}}, \quad \mathbf{B} = G_S^{\mathbf{v}}\mathbf{G}_{\mathbf{D}}^{\mathbf{v}} - \mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\left(\mathbf{G}_{\mathbf{D}S}^{\mathbf{v}}\right)^{H}
$$

As a consequence:

$$
\begin{aligned}
\widehat{M}_{m,l}^{m',l'} &= \left(cov_{\boldsymbol{\theta}}\widehat{\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}}\right)_{m,l}\left(cov_{\boldsymbol{\theta}}\widehat{\left\{\mathbf{R}\left(\boldsymbol{\theta}\right)\right\}}\right)_{m',l'} \\
\widehat{M}_{m,l}^{m',l'} &= \left(\frac{\left(\boldsymbol{\Gamma}_{\mathbf{D}|\mathbf{S}}\right)_{m,l}}{\left(\Gamma_S-G_S^{\mathbf{v}}\right)} + \frac{\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{m}\left(\boldsymbol{\nu}\right)_{l}^{*}+\left(\boldsymbol{\nu}\right)_{m}\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{l}^{*}-\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{m}\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{l}^{*}+\left(\mathbf{M}_S\right)_{m,l}}{\left(\Gamma_S-G_S^{\mathbf{v}}\right)^{2}}\right)\times \\
&\quad \left(\frac{\left(\boldsymbol{\Gamma}_{\mathbf{D}|\mathbf{S}}\right)_{m',l'}}{\left(\Gamma_S-G_S^{\mathbf{v}}\right)} + \frac{\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{m'}\left(\boldsymbol{\nu}\right)_{l'}^{*}+\left(\boldsymbol{\nu}\right)_{m'}\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{l'}^{*}-\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{m'}\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{l'}^{*}+\left(\mathbf{M}_S\right)_{m',l'}}{\left(\Gamma_S-G_S^{\mathbf{v}}\right)^{2}}\right)
\end{aligned}
$$

that is:

$$
\widehat{M}_{m,l}^{m',l'} = \left|
\begin{array}{l}
\frac{(\boldsymbol{\Gamma}_{\mathbf{D}|\mathbf{S}})_{m,l}(\boldsymbol{\Gamma}_{\mathbf{D}|\mathbf{S}})_{m',l'}}{(\Gamma_S-G_S^{\mathbf{v}})^{2}} \\
+\frac{(\boldsymbol{\Gamma}_{\mathbf{D}|\mathbf{S}})_{m,l}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{m'}(\boldsymbol{\nu})_{l'}^{*}}{(\Gamma_S-G_S^{\mathbf{v}})^{3}} + \frac{(\boldsymbol{\Gamma}_{\mathbf{D}|\mathbf{S}})_{m,l}(\boldsymbol{\nu})_{m'}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{l'}^{*}}{(\Gamma_S-G_S^{\mathbf{v}})^{3}} - \frac{(\boldsymbol{\Gamma}_{\mathbf{D}|\mathbf{S}})_{m,l}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{m'}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{l'}^{*}}{(\Gamma_S-G_S^{\mathbf{v}})^{3}} + \frac{(\boldsymbol{\Gamma}_{\mathbf{D}|\mathbf{S}})_{m,l}(\mathbf{M}_S)_{m',l'}}{(\Gamma_S-G_S^{\mathbf{v}})^{3}} \\
+\frac{(\boldsymbol{\Gamma}_{\mathbf{D}|\mathbf{S}})_{m',l'}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{m}(\boldsymbol{\nu})_{l}^{*}}{(\Gamma_S-G_S^{\mathbf{v}})^{3}} + \frac{(\boldsymbol{\Gamma}_{\mathbf{D}|\mathbf{S}})_{m',l'}(\boldsymbol{\nu})_{m}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{l}^{*}}{(\Gamma_S-G_S^{\mathbf{v}})^{3}} - \frac{(\boldsymbol{\Gamma}_{\mathbf{D}|\mathbf{S}})_{m',l'}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{,}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{l}^{*}}{(\Gamma_S-G_S^{\mathbf{v}})^{3}} + \frac{(\boldsymbol{\Gamma}_{\mathbf{D}|\mathbf{S}})_{m',l'}(\mathbf{M}_S)_{m,l}}{(\Gamma_S-G_S^{\mathbf{v}})^{3}} \\
+\frac{(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{m}(\boldsymbol{\nu})_{l}^{*}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{m'}(\boldsymbol{\nu})_{l'}^{*}}{(\Gamma_S-G_S^{\mathbf{v}})^{4}} + \frac{(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{m}(\boldsymbol{\nu})_{l}^{*}(\boldsymbol{\nu})_{m'}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{l'}^{*}}{(\Gamma_S-G_S^{\mathbf{v}})^{4}} \\
-\frac{(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{m}(\boldsymbol{\nu})_{l}^{*}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{m'}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{l'}^{*}}{(\Gamma_S-G_S^{\mathbf{v}})^{4}} + \frac{(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{m}(\boldsymbol{\nu})_{l}^{*}(\mathbf{M}_S)_{m',l'}}{(\Gamma_S-G_S^{\mathbf{v}})^{4}} \\
+\frac{(\boldsymbol{\nu})_{m}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{l}^{*}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{m'}(\boldsymbol{\nu})_{l'}^{*}}{(\Gamma_S-G_S^{\mathbf{v}})^{4}} + \frac{(\boldsymbol{\nu})_{m}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{l}^{*}(\boldsymbol{\nu})_{m'}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{l'}^{*}}{(\Gamma_S-G_S^{\mathbf{v}})^{4}} \\
-\frac{(\boldsymbol{\nu})_{m}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{l}^{*}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{m'}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{l'}^{*}}{(\Gamma_S-G_S^{\mathbf{v}})^{4}} + \frac{(\boldsymbol{\nu})_{m}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{l}^{*}(\mathbf{M}_S)_{m',l'}}{(\Gamma_S-G_S^{\mathbf{v}})^{4}} \\
-\frac{(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{m}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{l}^{*}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{m'}(\boldsymbol{\nu})_{l'}^{*}}{(\Gamma_S-G_S^{\mathbf{v}})^{4}} - \frac{(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{m}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{l}^{*}(\boldsymbol{\nu})_{m'}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{l'}^{*}}{(\Gamma_S-G_S^{\mathbf{v}})^{4}} \\
+\frac{(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{m}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{l}^{*}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{m'}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{l'}^{*}}{(\Gamma_S-G_S^{\mathbf{v}})^{4}} - \frac{(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{m}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{l}^{*}(\mathbf{M}_S)_{m',l'}}{(\Gamma_S-G_S^{\mathbf{v}})^{4}} \\
+\frac{(\mathbf{M}_S)_{m,l}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{m'}(\boldsymbol{\nu})_{l'}^{*}}{(\Gamma_S-G_S^{\mathbf{v}})^{4}} + \frac{(\mathbf{M}_S)_{m,l}(\boldsymbol{\nu})_{m'}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{l'}^{*}}{(\Gamma_S-G_S^{\mathbf{v}})^{4}} - \frac{(\mathbf{M}_S)_{m,l}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{m'}(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}})_{l'}^{*}}{(\Gamma_S-G_S^{\mathbf{v}})^{4}} + \frac{(\mathbf{M}_S)_{m,l}(\mathbf{M}_S)_{m',l'}}{(\Gamma_S-G_S^{\mathbf{v}})^{4}}
\end{array}
\right.
$$

As:

$$
\begin{array}{rr}
E\left\{\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{m}\mid\mathbf{S}\right\}=0, & E\left\{\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{m}^{*}\mid\mathbf{S}\right\}=0\\
E\left\{\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{m}\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{m'}\mid\mathbf{S}\right\}=0, & E\left\{\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{m}^{*}\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{m'}^{*}\mid\mathbf{S}\right\}=0\\
E\left\{\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{m}\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{l}\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{n}\mid\mathbf{S}\right\}=0, & E\left\{\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{m}^{*}\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{l}\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{n}\mid\mathbf{S}\right\}=0\\
E\left\{\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{m}^{*}\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{l}^{*}\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{n}\mid\mathbf{S}\right\}=0, & E\left\{\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{m}^{*}\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{l}^{*}\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{n}^{*}\mid\mathbf{S}\right\}=0\\
E\left\{\left(\boldsymbol{\Gamma}_{\mathbf{D}|\mathbf{S}}\right)_{m,l}\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{m'}\mid\mathbf{S}\right\}=0, & E\left\{\left(\boldsymbol{\Gamma}_{\mathbf{D}|\mathbf{S}}\right)_{m,l}\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{m'}^{*}\mid\mathbf{S}\right\}=0
\end{array}
$$

and:

$$
\begin{aligned}
E\left\{\left(\boldsymbol{\Gamma}_{\mathbf{D}|\mathbf{S}}\right)_{m,l}\mid\mathbf{S}\right\} &= \left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\\
E\left\{\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{m}\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{l}^{*}\mid\mathbf{S}\right\} &= \frac{\mathbf{S}^{H}\mathbf{S}}{K^{2}}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\\
E\left\{\left(\boldsymbol{\Gamma}_{\mathbf{D}|\mathbf{S}}\right)_{m,l}\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{m'}\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{l'}^{*}\mid\mathbf{S}\right\} &= \frac{\mathbf{S}^{H}\mathbf{S}}{K^{3}}\left(\begin{array}{c}K\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'}\\+\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l'}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l}\end{array}\right)\\
E\left\{\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{m}\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{l}^{*}\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{m'}\left(\boldsymbol{\Gamma}_{\mathbf{D}S|\mathbf{S}}\right)_{l'}^{*}\mid\mathbf{S}\right\} &= \frac{\left(\mathbf{S}^{H}\mathbf{S}\right)^{2}}{K^{4}}\left(\begin{array}{c}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'}\\+\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l'}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l}\end{array}\right)\\
E\left\{\left(\boldsymbol{\Gamma}_{\mathbf{D}|\mathbf{S}}\right)_{m,l}\left(\boldsymbol{\Gamma}_{\mathbf{D}|\mathbf{S}}\right)_{m',l'}\mid\mathbf{S}\right\} &= \left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'}+\frac{1}{K}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l'}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l}
\end{aligned}
$$

one obtains that:

$$
E\left\{\widehat{M}_{m,l}^{m',l'}\mid\mathbf{S}\right\}=\left|\begin{array}{l}
\frac{1}{\left(\mathbf{S}^{H}\mathbf{S}-KG_{S}^{\mathbf{v}}\right)^{2}}K^{2}\left(\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'}+\frac{1}{K}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l'}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l}\right)\\
+\frac{1}{\left(\mathbf{S}^{H}\mathbf{S}-KG_{S}^{\mathbf{v}}\right)^{3}}K^{3}\left(\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{B}\right)_{m',l'}+\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'}\left(\mathbf{B}\right)_{m,l}\right)\\
+\frac{\mathbf{S}^{H}\mathbf{S}}{\left(\mathbf{S}^{H}\mathbf{S}-KG_{S}^{\mathbf{v}}\right)^{3}}\left(\begin{array}{l}K^{2}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{A}\right)_{m',l'}+K^{2}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'}\left(\mathbf{A}\right)_{m,l}\\-2K\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'}-2\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l'}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l}\end{array}\right)\\
+\frac{1}{\left(\mathbf{S}^{H}\mathbf{S}-KG_{S}^{\mathbf{v}}\right)^{4}}K^{4}\left(\mathbf{B}\right)_{m,l}\left(\mathbf{B}\right)_{m',l'}\\
+\frac{\mathbf{S}^{H}\mathbf{S}}{\left(\mathbf{S}^{H}\mathbf{S}-KG_{S}^{\mathbf{v}}\right)^{4}}K^{2}\left(\begin{array}{l}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l'}\left(\boldsymbol{\nu}\right)_{l}^{*}\left(\boldsymbol{\nu}\right)_{m'}+\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l}\left(\boldsymbol{\nu}\right)_{m}\left(\boldsymbol{\nu}\right)_{l'}^{*}\\-\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{B}\right)_{m',l'}-\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'}\left(\mathbf{B}\right)_{m,l}\\+K\left(\left(\mathbf{A}\right)_{m,l}\left(\mathbf{B}\right)_{m',l'}+\left(\mathbf{A}\right)_{m',l'}\left(\mathbf{B}\right)_{m,l}\right)\end{array}\right)\\
+\frac{\left(\mathbf{S}^{H}\mathbf{S}\right)^{2}}{\left(\mathbf{S}^{H}\mathbf{S}-KG_{S}^{\mathbf{v}}\right)^{4}}\left(\begin{array}{l}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'}+\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l'}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l}\\+K^{2}\left(\mathbf{A}\right)_{m,l}\left(\mathbf{A}\right)_{m',l'}-K\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{A}\right)_{m',l'}-K\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'}\left(\mathbf{A}\right)_{m,l}\end{array}\right)
\end{array}\right.
$$

Finally:

$$
E\left\{\widehat{M}_{m,l}^{m',l'}\mid\Sigma\right\}=\left|\begin{array}{l}
E\left\{\frac{1}{\left(\mathbf{S}^{H}\mathbf{S}-KG_{S}^{\mathbf{v}}\right)^{2}}\mid\Sigma\right\}K^{2}\left(\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'}+\frac{1}{K}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l'}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l}\right)\\
+E\left\{\frac{1}{\left(\mathbf{S}^{H}\mathbf{S}-KG_{S}^{\mathbf{v}}\right)^{3}}\mid\Sigma\right\}K^{3}\left(\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{B}\right)_{m',l'}+\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'}\left(\mathbf{B}\right)_{m,l}\right)\\
+E\left\{\frac{1}{\left(\mathbf{S}^{H}\mathbf{S}-KG_{S}^{\mathbf{v}}\right)^{4}}\mid\Sigma\right\}K^{4}\left(\mathbf{B}\right)_{m,l}\left(\mathbf{B}\right)_{m',l'}\\
+E\left\{\frac{\mathbf{S}^{H}\mathbf{S}}{\left(\mathbf{S}^{H}\mathbf{S}-KG_{S}^{\mathbf{v}}\right)^{3}}\mid\Sigma\right\}\left(\begin{array}{l}K^{2}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{A}\right)_{m',l'}+K^{2}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'}\left(\mathbf{A}\right)_{m,l}\\-2K\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'}-2\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l'}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l}\end{array}\right)\\
+E\left\{\frac{\mathbf{S}^{H}\mathbf{S}}{\left(\mathbf{S}^{H}\mathbf{S}-KG_{S}^{\mathbf{v}}\right)^{4}}\mid\Sigma\right\}K^{2}\left(\begin{array}{l}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l'}\left(\boldsymbol{\nu}\right)_{l}^{*}\left(\boldsymbol{\nu}\right)_{m'}+\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l}\left(\boldsymbol{\nu}\right)_{m}\left(\boldsymbol{\nu}\right)_{l'}^{*}\\-\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{B}\right)_{m',l'}-\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'}\left(\mathbf{B}\right)_{m,l}\\+K\left(\left(\mathbf{A}\right)_{m,l}\left(\mathbf{B}\right)_{m',l'}+\left(\mathbf{A}\right)_{m',l'}\left(\mathbf{B}\right)_{m,l}\right)\end{array}\right)\\
+E\left\{\frac{\left(\mathbf{S}^{H}\mathbf{S}\right)^{2}}{\left(\mathbf{S}^{H}\mathbf{S}-KG_{S}^{\mathbf{v}}\right)^{4}}\mid\Sigma\right\}\left(\begin{array}{l}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'}+\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l'}\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l}\\-K\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m,l}\left(\mathbf{A}\right)_{m',l'}-K\left(\mathbf{C}_{\mathbf{D}|S}\right)_{m',l'}\left(\mathbf{A}\right)_{m,l}\\+K^{2}\left(\mathbf{A}\right)_{m,l}\left(\mathbf{A}\right)_{m',l'}\end{array}\right)
\end{array}\right.
$$

where:

$$
\begin{aligned}
E\left\{\frac{1}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^n} \mid \Sigma\right\} &= P\left(K,n\right) \\
E\left\{\frac{\mathbf{S}^H\mathbf{S}}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^3} \mid \Sigma\right\} &= P\left(K,2\right)+KG_S^{\mathbf{v}}P\left(K,3\right) \\
E\left\{\frac{\mathbf{S}^H\mathbf{S}}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^4} \mid \Sigma\right\} &= P\left(K,3\right)+KG_S^{\mathbf{v}}P\left(K,4\right) \\
E\left\{\frac{\left(\mathbf{S}^H\mathbf{S}\right)^2}{\left(\mathbf{S}^H\mathbf{S}-KG_S^{\mathbf{v}}\right)^4} \mid \Sigma\right\} &= P\left(K,2\right)+2KG_S^{\mathbf{v}}P\left(K,3\right)+\left(KG_S^{\mathbf{v}}\right)^2P\left(K,4\right)
\end{aligned}
$$

REFERENCES

[1] A.A. Ksienski, R.B. McGhee, ”A decision theoretic approach to the angular resolution and parameter estimation of multiple targets”, IEEE Trans. AES 4 (3), 1968

[2] R. Schmidt, ”Multiple emitter location and signal parameter estimation”, IEEE Trans. AP 34 (3), pp. 276-280, 1986

[3] S. Haykin, J. Litva, T.J. Shepherd (eds.), *Radar Array Processing*, Springer-Verlag, 1993

[4] R. Klemm, M. Mertens, ”Tracking of convoys by airborne STAP radar”, Proceeding FUSION 2008, Cologne, Germany, ISIF

[5] M. Feldmann, D. Fränken, W. Koch, ”Tracking of Extended Object and Group Targets using Random Matrices”, Accepted for publication in IEEE Trans. SP

[6] G.A. Gordon, ”Monopulse estimation of the centroid of an ensemble of radar scatterers”, IEEE Trans. AES 11 (1), pp. 94-102, 1975

[7] L.B. Milstein, ”Maximum likelihood estimation of the angular position and extend of a target”, IEEE Trans. IT, vol. 27 (2), pp. 187-191, 1981

[8] X. Zhang, P.K. Willett, Y. Bar-Shalom, ”Monopulse Radar Detection and Localization of Multiple Unresolved targets via Joint Bin Processing”, IEEE Trans. AES 53 (4), pp. 1225-1236, 2005

[9] X. Zhang, P. Willett, W.D. Blair, ”The Multitarget Monopulse CRLB for Matched Filter Samples”, IEEE Trans SP 55 (8), pp. 4183-4197, 2007

[10] I. Kanter, ”Multiple Gaussian targets: the track-on-jam problem”, IEEE Trans. AES 13 (6), pp. 620-623, 1977

[11] B.-E. Tullson, ”Monopulse tracking of Rayleigh targets: a simple approach”, IEEE Trans. AES 27 (3), pp. 520-531, 1991

[12] A.D. Seifer, ”Monopulse-radar angle tracking in noise or noise jamming”, IEEE Trans. AES 28 (3), pp. 622-638, 1992

[13] W.D. Blair, M. Brandt-Pearce, ”Statistical Description of Monopulse Parameters for Tracking Rayleigh Targets”, IEEE Trans. AES 34 (2), pp. 597-611, 1998

[14] W.D. Blair, M. Brandt-Pearce, ”Unresolved Rayleigh target detection using monopulse measurements”, IEEE Trans. AES 34 (2), pp. 543-552, 1998

[15] E. Mosca, ”Angle Estimation in Amplitude Comparison monopulse Systems”, IEEE Trans. AES 5 (2), pp. 205-212, 1969

[16] U. Nickel, ”Overview of generalized monopulse estimation.”, IEEE AES Magazine, Vol. 21, Part 2, pp. 27-55, 2006

[17] U. Nickel, E. Chaumette, P. Larzabal, ”Statistical Performance Prediction of Generalized Monopulse Estimation”, IEEE Trans. AES 47 (1), pp. 381-404, 2011

[18] L. Scharf, *Statistical Signal Processing*, Addison-Wiley, 1991

[19] U. Nickel, ”Properties of digital beamforming with subarrays”, CIE Int. Conf. RADAR 2006, Shanghai, China, 6.-19. Oct. 2006, IEEE Cat. Nr. 06TH8857

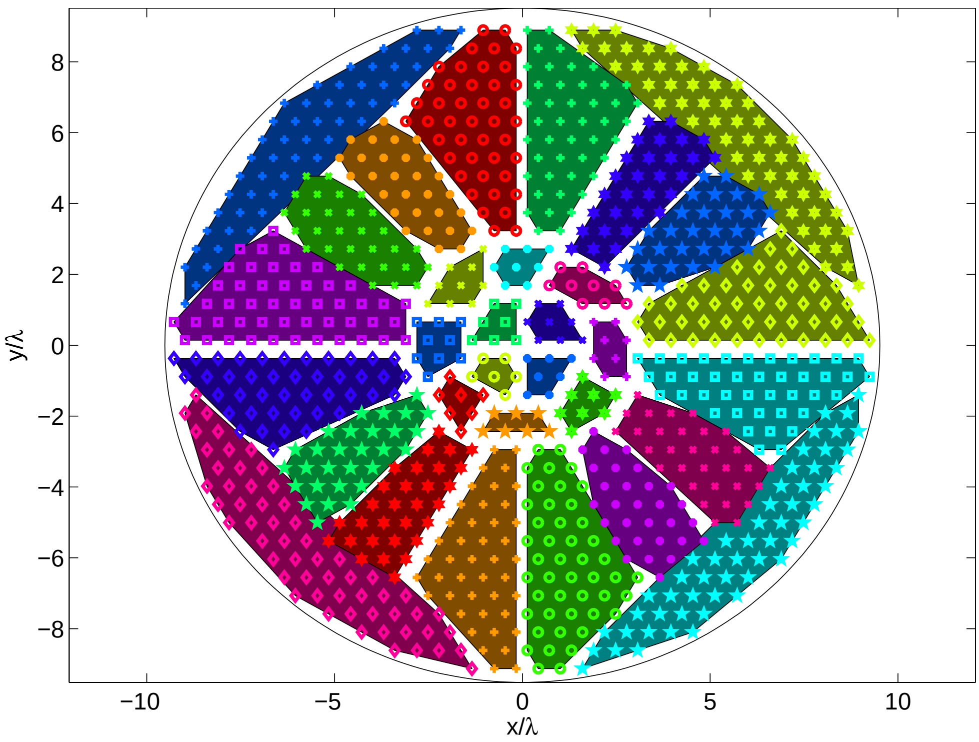


Fig. 1. Generic fully filled array with 902 elements and 32 subarrays.

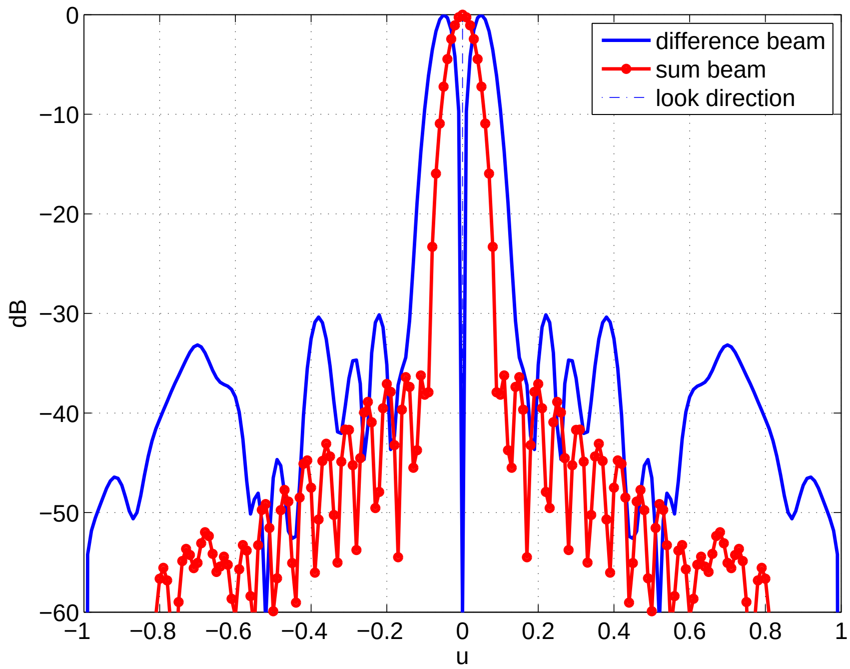


Fig. 2. Azimuth cut of sum and difference pattern with 35dB Taylor and approximated Bayliss weighting

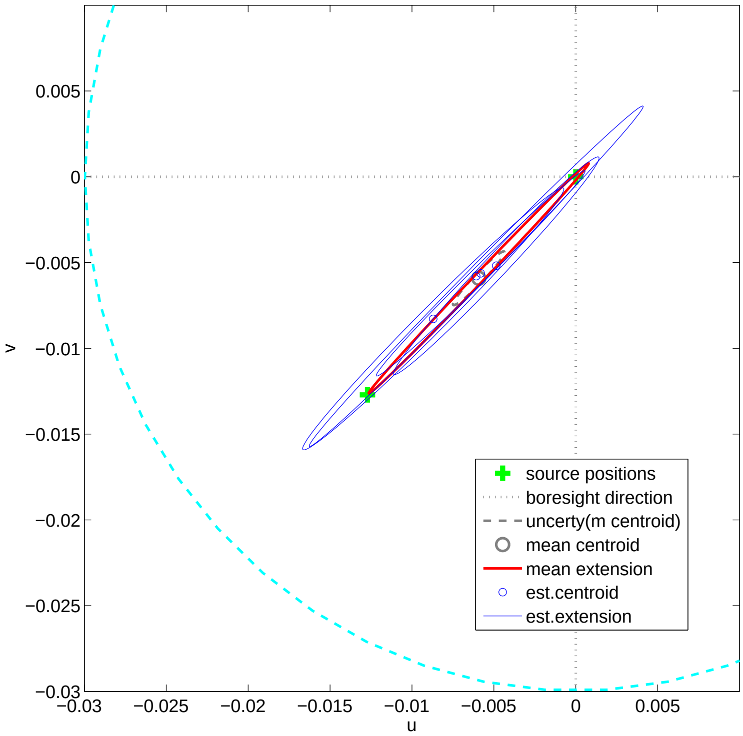


Fig. 3. Estimation of centroid and extension of a target consisting of two point scatterers at 0.3BW separation based on 10 snapshots, SNR 28dB.

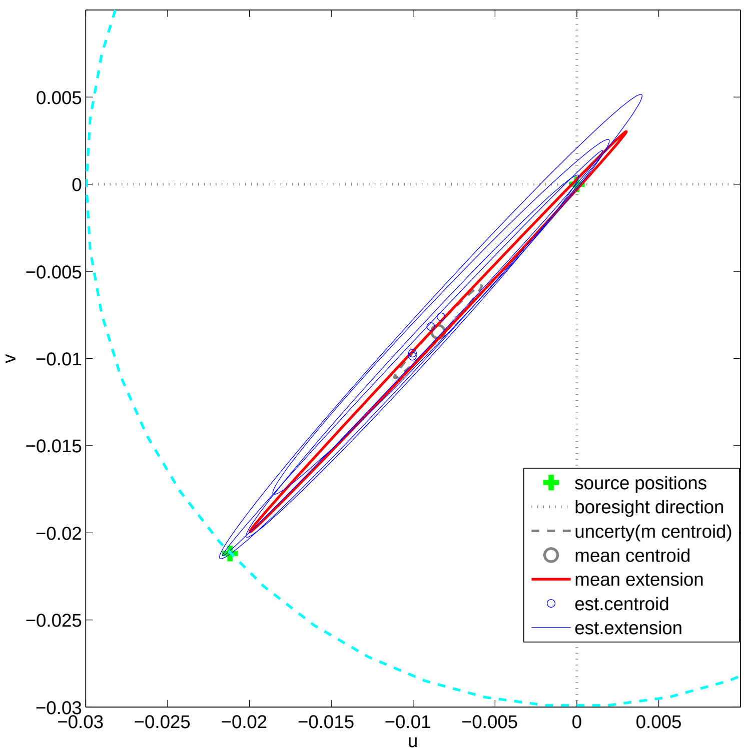
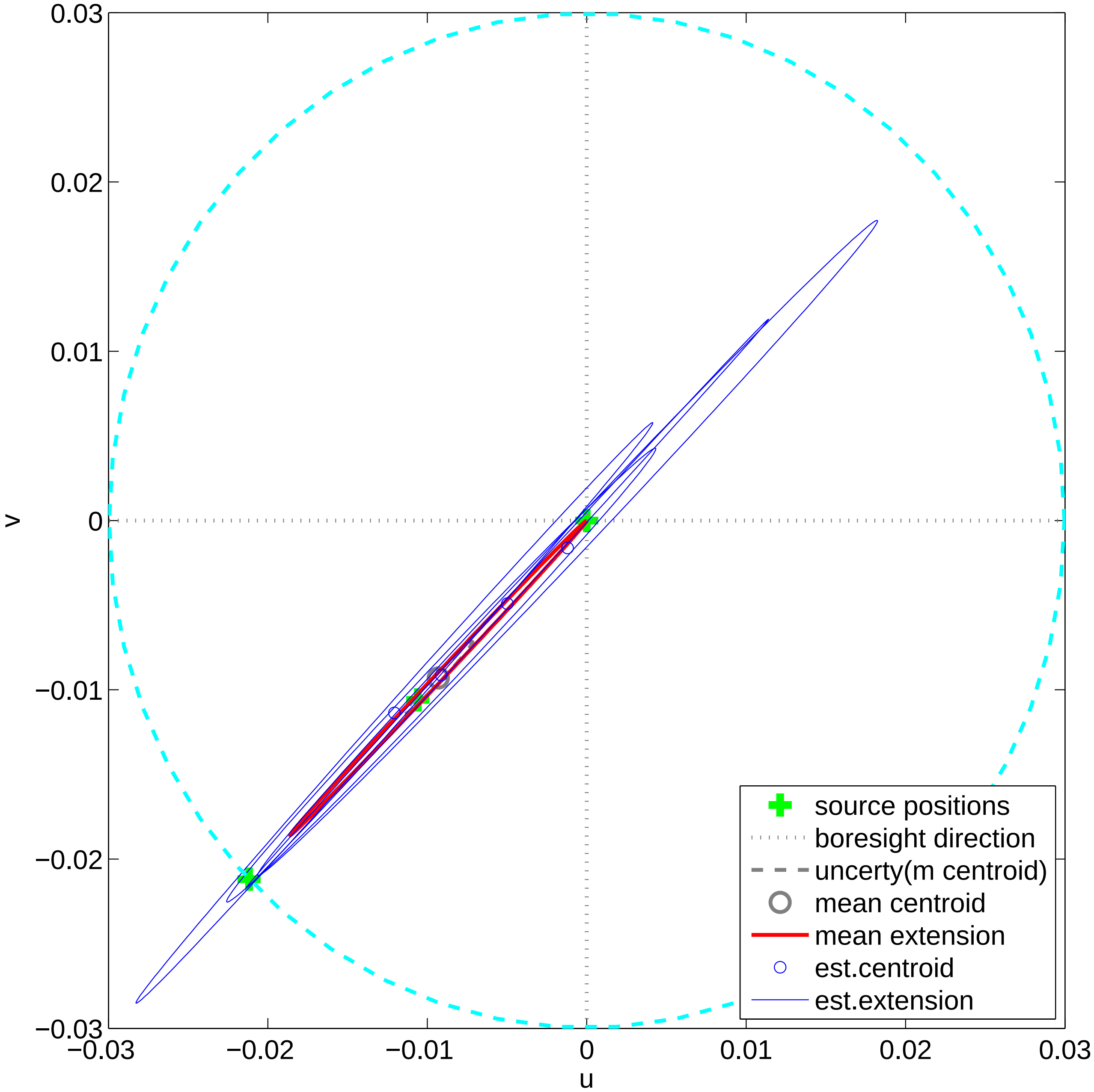


Fig. 4. Estimation of centroid and extension of a target consisting of two point scatterers at 0.5BW separation based on 10 snapshots, SNR 28dB.

Fig. 5. Estimation of centroid and extension of three correlated point scatterers at 0.5BW separation based on 10 snapshots, SNR 28dB. The two outermost scatterers have fixed amplitude and normal distributed phase differences with mean 170° and std 20°, the scatterer in the middle is complex normal distributed.



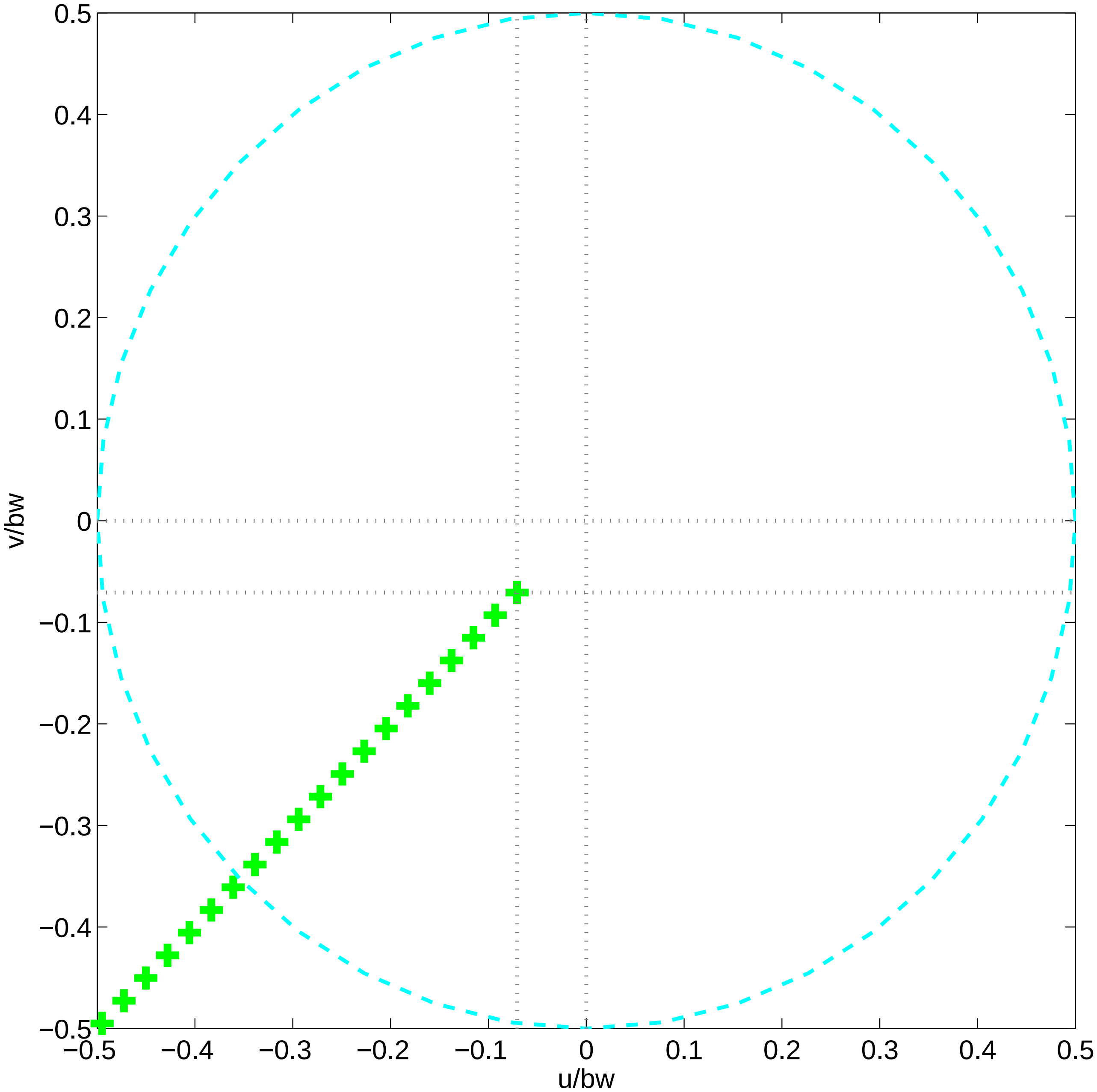


Fig. 6. Extended target scenario: 20 point scatterers at maximum length 0.6BW relative to antenna look direction. Dashed circle indicates 3dB contour of sum beam.

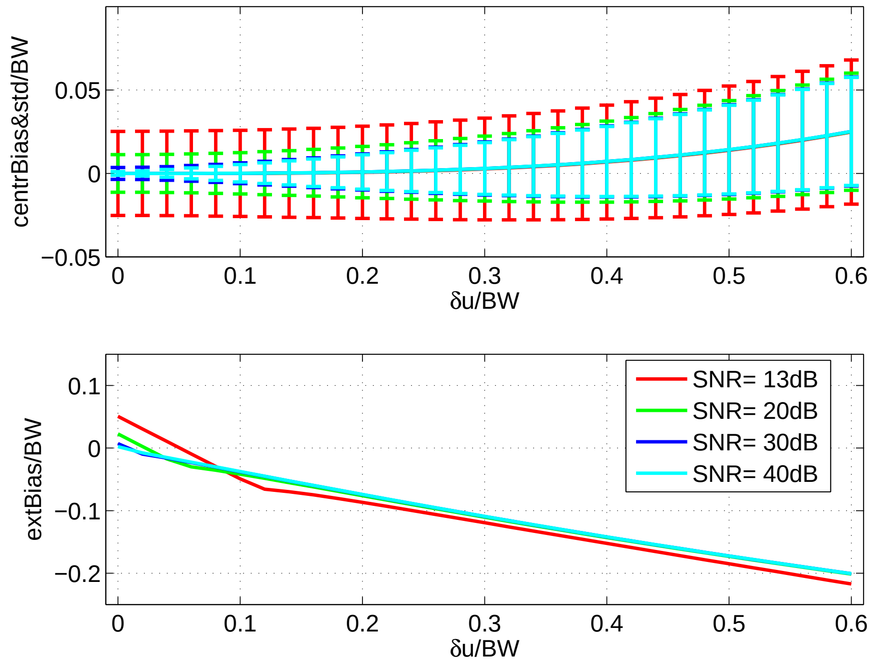


Fig. 7. Error bars for the theoretical centroid and extent of a line target (composed of 20 scatterers arranged as in Figure 6) over length δu.

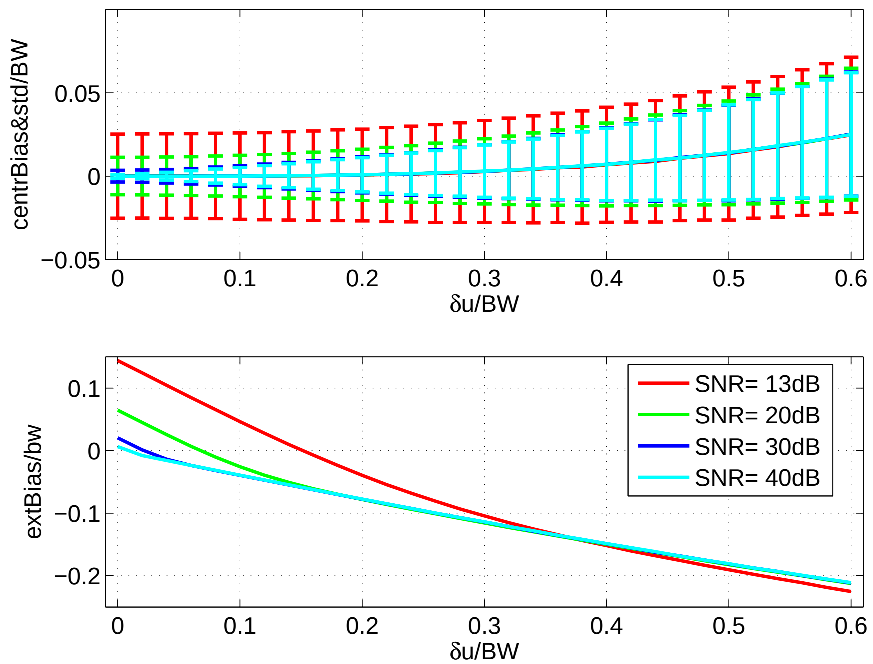


Fig. 8. Simulated bars for centroid and extent of a line target corresponding to theoretical results of Figure 7 based on 10^5 Monte-Carlo runs for each point.

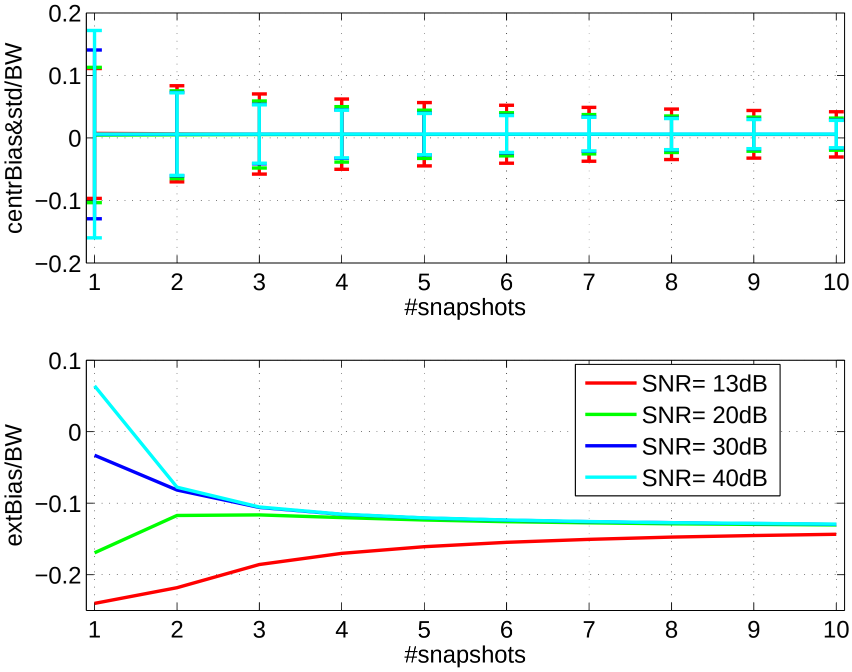


Fig. 9. Error bars for the theoretical centroid and extent of a line target of 0.4BW length (composed of 20 scatterers arranged as in Figure 6) over number of snapshots of averaged monopulse ratio, monopulse detection threshold 13dB.

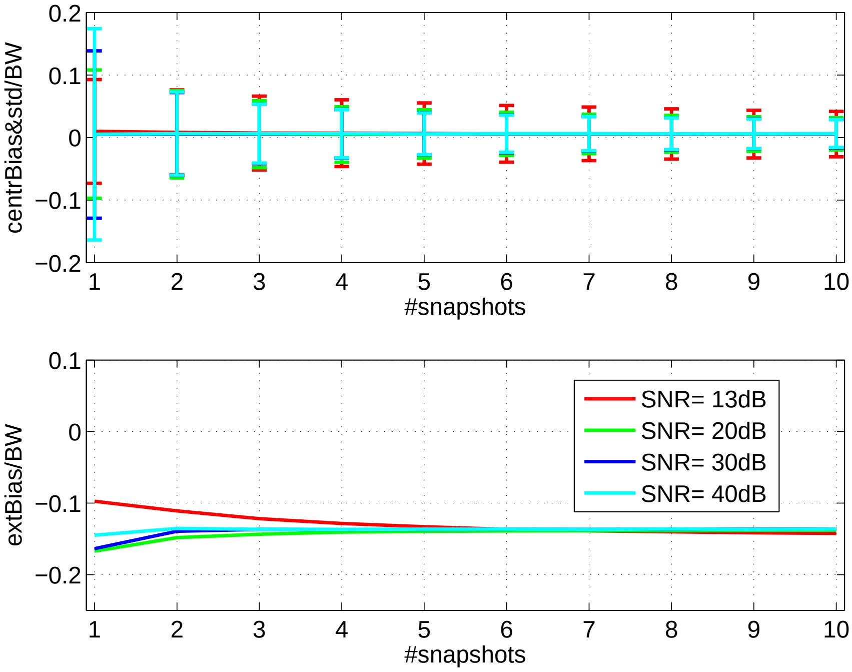


Fig. 10. Simulated error bars for centroid and extension estimators corresponding to theoretical results of Figure 9 based on 10^5 Monte-Carlo runs for each point.

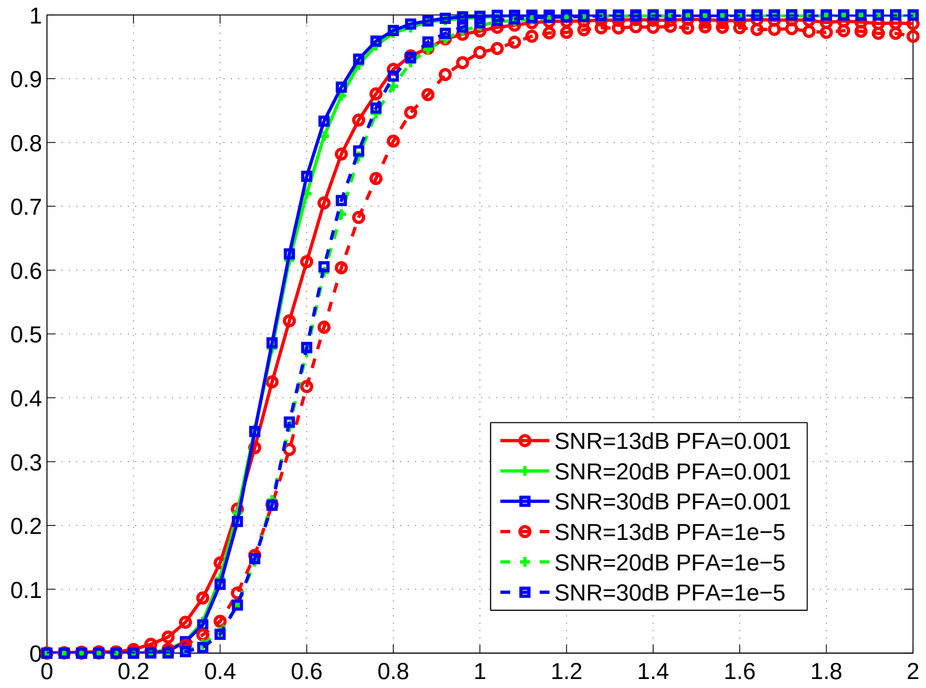


Fig. 11. Detection probability for line target of varying length modelled by 20 point scatterers for various SNR. Threshold selected by Gaussian approximation for $PFA = 10^{-3}$ and 10^{-5}, monopulse detection threshold 13dB, 10^4 Monte-Carlo runs for each point of curve.

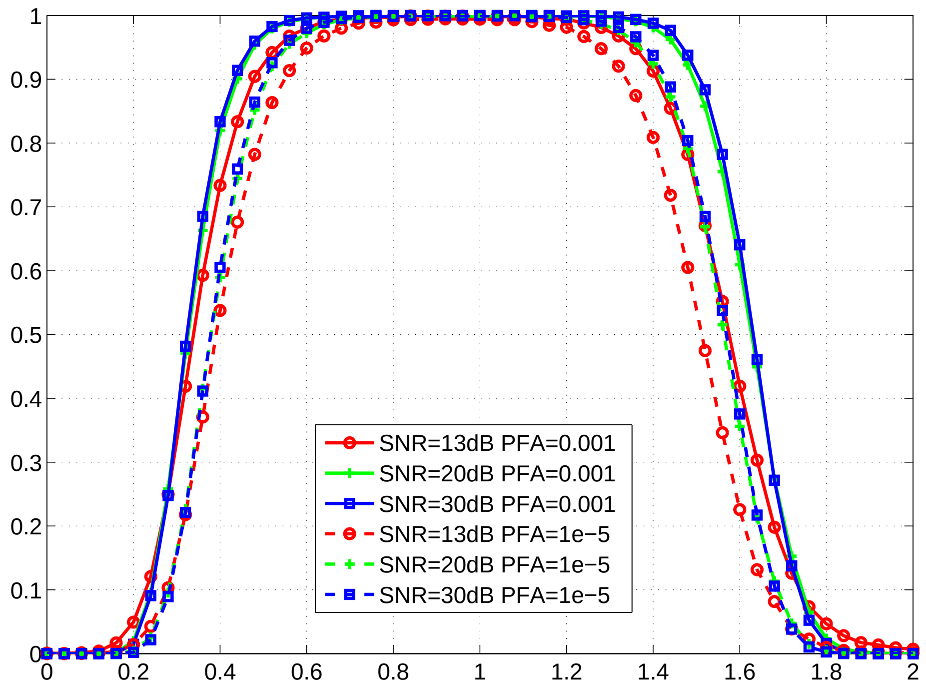


Fig. 12. Detection probability over separation of two point scatterers for various SNR. Threshold selected by Gaussian approximation for $PFA = 10^{-3}$ and 10^{-5}, monopulse detection threshold 13dB,10^4Monte-Carlo runs for each point of curve.

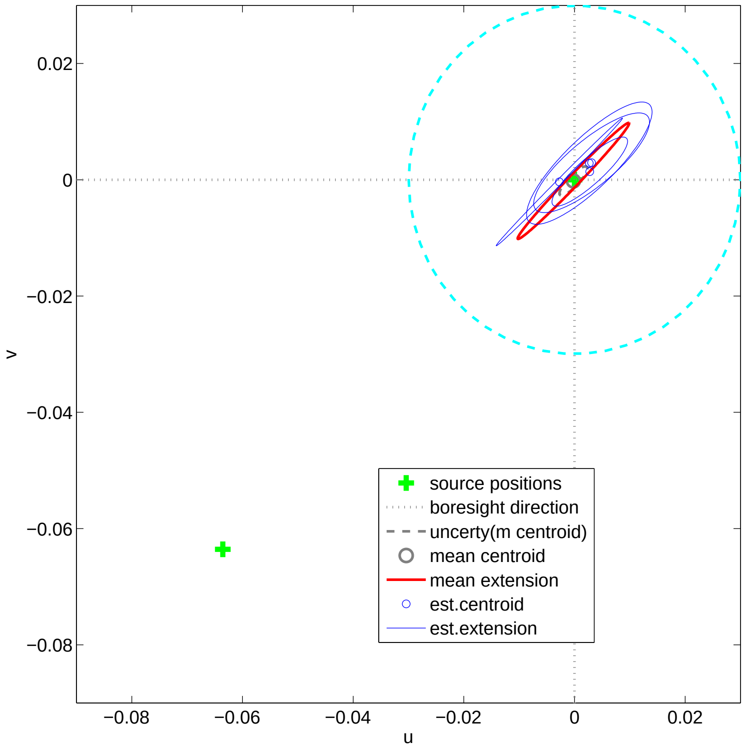


Fig. 13. Estimation of extended target separating into two point targets with 1.5BW separation (SNR = 20dB, monopulse detection threshold 13dB).

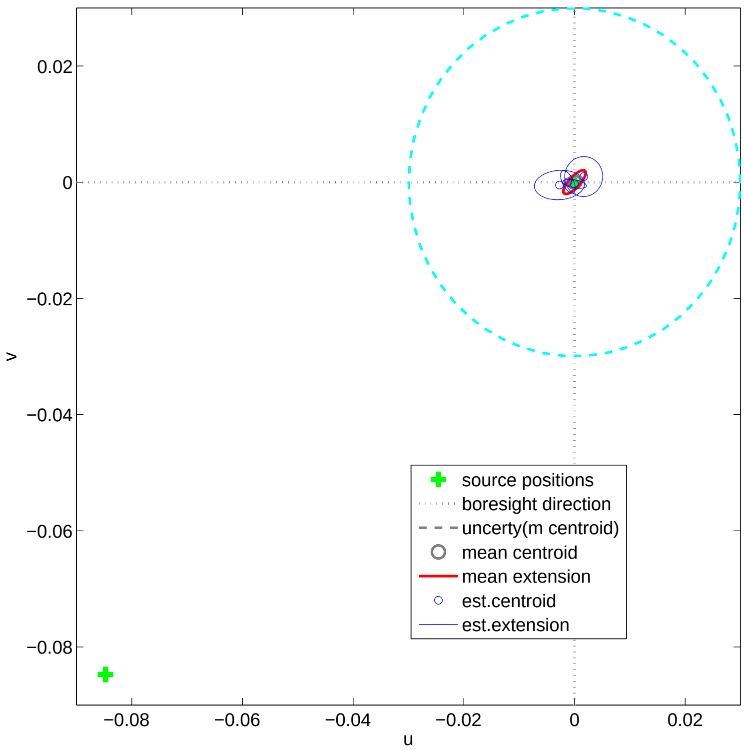


Fig. 14. Estimation of extended target separating into two point targets with 2BW separation (SNR = 20dB, monopulse detection threshold 13dB).

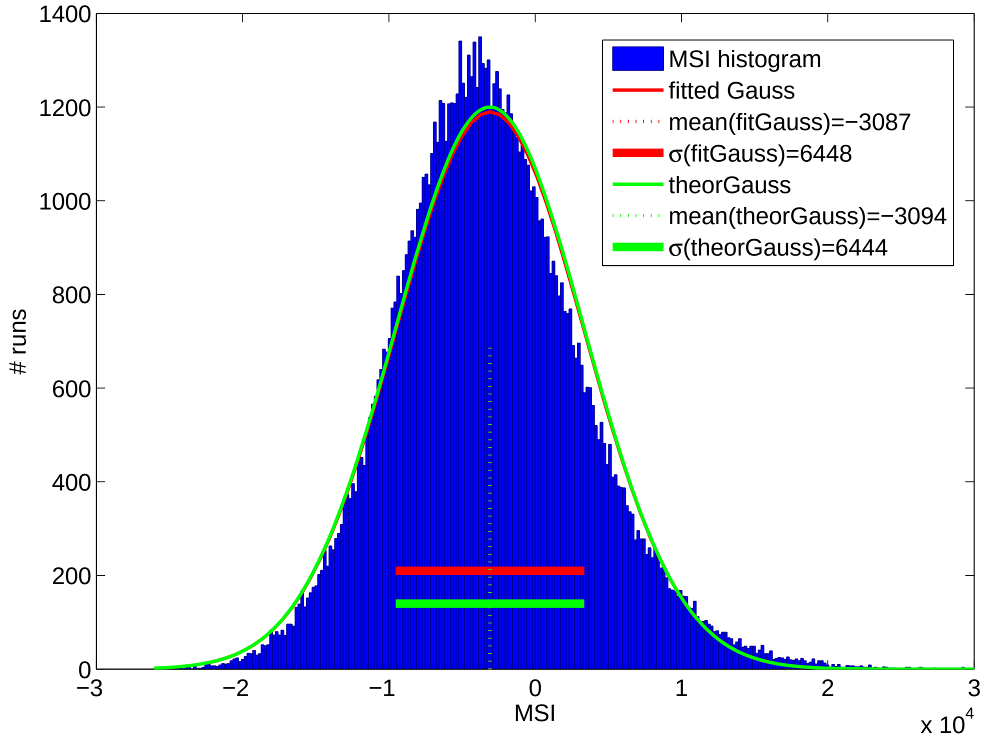


Fig. 15. Histogram of MSI and Gaussian approximation for null hypothesis: a single point target in antenna look direction with 13dB SNR, monopulse detection threshold 13dB, 10^5 Monte-Carlo runs.

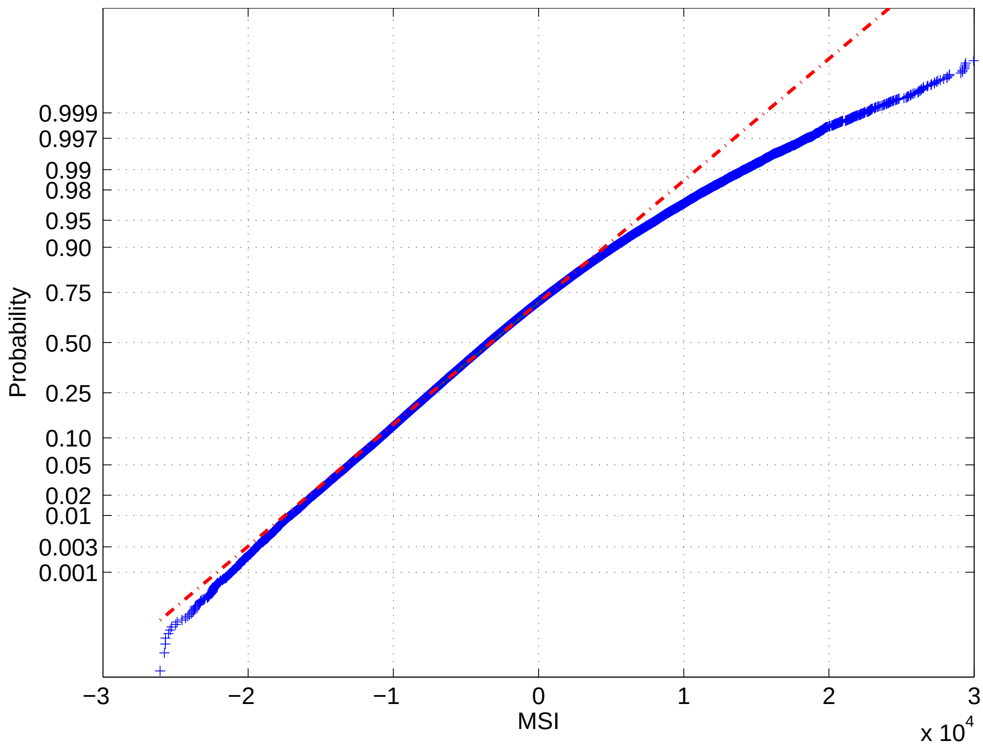


Fig. 16. Matlab Normal probability plot for null hypothesis approximation corresponding to Figure 15. Data obeying a Gaussian should lie on the dashed line.

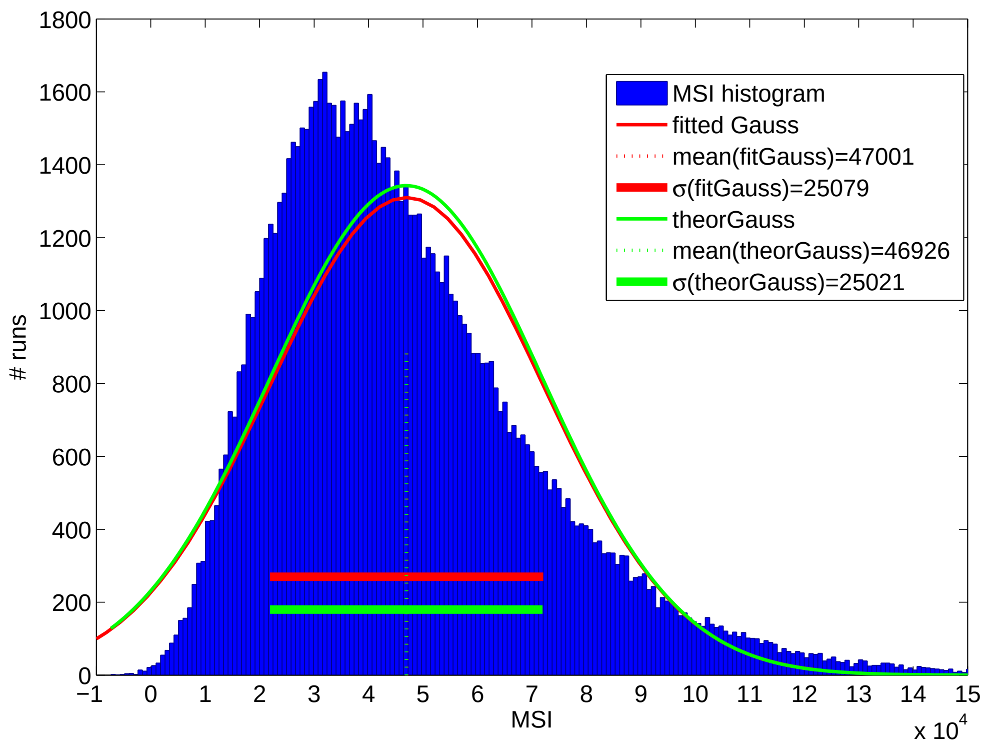


Fig. 17. Histogram of MSI and Gaussian approximation for two targets at 0.5BW separation with 13dB SNR, 10^5 Monte-Carlo runs.

J. Mse lower bounds conditioned by the energy detector (EUSIPCO)

in Proc. Eurasip EUSIPCO, 2007.

MSE LOWER BOUNDS CONDITIONED BY THE ENERGY DETECTOR

(1) *Eric Chaumette,* (2) *Jerome Galy,* (3) *Francois Vincent,* (4) *Alexandre Renaux and* (5) *Pascal Larzabal*

(1) Thales Air Defence, Limours, France - eric.chaumette@fr.thalesgroup.com
(2) LIRMM, Montpellier, France - galy@lirmm.fr
(3) ENSICA, Toulouse, France - vincent@ensica.fr
(4) Washington University, St. Louis, USA - renaux@ese.wustl.edu
(5) Ecole Normale Superieure, Cachan, France - larzabal@satie.ens-cachan.fr

ABSTRACT

A wide variety of processing incorporates a binary detection test that restricts the set of observations for parameter estimation. This statistical conditioning must be taken into account to compute the Cramér-Rao bound [1] (CRB) and more generally, lower bounds on the Mean Square Error (MSE) [2]. Therefore, we propose a derivation of some lower bounds - including the CRB - for the deterministic signal model conditioned by the energy detector [3] widely used in signal processing applications.

1. INTRODUCTION

Lower bounds on the MSE in estimating a set of deterministic parameters [1] from noisy observations provide the best performance of any estimators in terms of the MSE. They allow to investigate fundamental limits of a parameter estimation problem or to assess the relative performance of a specific estimator. Historically the first MSE lower bound for deterministic parameters to be derived was the CRB [1], which has been the most widely used since. Its popularity is largely due to its simplicity of calculation, the fact that in many cases, it can be achieved asymptotically (high SNR [4] and/or large number of snapshots [1]) by maximum likelihood estimators (MLEs), and last but not least, its noticeable property of being the lowest bound on the MSE of locally unbiased estimators. This initial characterization of locally unbiased estimators has been extended first by Bhattacharyya's work, and significantly generalized by Barankin's work which allows the derivation of the highest lower bound on MSE since it takes into account the unbiasedness over the parameter space [1][2][5][6]. Unfortunately the Barankin bound (BB) is generally incomputable [6]. Numerous works (see references in [2][6] and [7]) devoted to the computing and tightness of bounds on MSE have shown that the CRB and BB can be regarded as key representative of two general classes of bounds, respectively the Small-Error bounds and the Large-Error bounds. Indeed, in non-linear estimation problems three distinct regions of operation can be observed. In the asymptotic region, the MSE of estimators is small and, in many cases, close to the Small-Error bounds. In the *a priori* performance region where the number of independent snapshots and/or the SNR are very low, the observations provide little information and the MSE is close to that obtained from the prior knowledge about the problem. Between these two extremes, there is an additional ambiguity region, also called the transition region. In this region, the MSE of MLEs deteriorates rapidly with respect to Small-Error bounds and generally exhibits a threshold behavior corresponding to a "performance breakdown" [8] highlighted by Large-Error bounds. As a result, the search for an easily computable but tight approximation of the BB is still a subject worth investigation. Therefore, Quinlan-Chaumette-Larzabal [6] have suggested a new approximation (QCLB) of the BB that allows a better prediction of the SNR value at the start of the transition region than existing approximations with a comparable computational complexity (CRB, Hammersley-Chapman-Robbins bound (HCRB), McAulay-Seidman bound (MSB), Abel bound of order 1 (AB_1)).
However, in nearly all fields of science and engineering, a wide variety of processing requires a binary detection step designed to decide if a signal is present or not in noise. As a detection step restricts the set of observations available for parameter estimation, any accurate MSE lower bound must take into account this initial statistical conditioning. As a contribution to the theoretical characterization of the joint detection and estimation problem, we propose in the present paper the derivation of above mentioned approximations of the BB (CRB, HCRB, MSB, AB_1, QCLB) for the deterministic signal model conditioned by the energy detector, which is a simple *realizable* test widely used in signal processing applications [3]. We therefore complete the characterization obtained for the CRB in [9].

2. DETERMINISTIC SIGNAL AND ENERGY DETECTOR

In many practical problems of interest, the received data samples is a vector $\mathbf{x}$ consisting of a bandpass signal that can be modelled as a mixture of a complex signal $\mathbf{s}_{\boldsymbol{\theta}}$ and a complex circular zero mean Gaussian noise $\mathbf{n}$: $\mathbf{x} = \mathbf{s}_{\boldsymbol{\theta}} + \mathbf{n}$. We consider the case where the signal of interest $\mathbf{s}_{\boldsymbol{\theta}}$ is dependent upon the vector of unknown deterministic parameters $\boldsymbol{\theta}$. The noise covariance matrix $\mathbf{C}_\mathbf{n}$ does not depend upon $\boldsymbol{\theta}$. Therefore $\mathbf{x} \sim \mathcal{CN}_L(\mathbf{m}_\mathbf{x}, \mathbf{C}_\mathbf{x})$, i.e. is complex circular Gaussian of dimension L with mean $\mathbf{m}_\mathbf{x} = \mathbf{s}_{\boldsymbol{\theta}}$ and covariance matrix $\mathbf{C}_\mathbf{x}$ ($\mathbf{C}_\mathbf{x} = \mathbf{C}_\mathbf{n}$), with p.d.f. [3, §13]:

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = f_{\mathcal{CN}_L}(\mathbf{x}; \mathbf{m}_\mathbf{x}(\boldsymbol{\theta}), \mathbf{C}_\mathbf{x}) = \frac{e^{-(\mathbf{x}-\mathbf{s}_{\boldsymbol{\theta}})^H \mathbf{C}_\mathbf{x}^{-1}(\mathbf{x}-\mathbf{s}_{\boldsymbol{\theta}})}}{\pi^L |\mathbf{C}_\mathbf{x}|} \quad (1)$$

In practical problems, the signal of interest $\mathbf{s}_{\boldsymbol{\theta}}$ is not always present. Such problems require first a binary detection step (decision rule) to decide if the signal of interest $\mathbf{s}_{\boldsymbol{\theta}}$ is present or not in the noise before running an estimation scheme [2]. Let us recall that optimal decision rules are based on the exact statistics of the observations [3, §3]. Their expressions require knowledge of the p.d.f. of observations under each hypothesis and the *a priori* probability of each hypothesis, if known (Bayes criterion). If no *a priori* probability of hypotheses is available, then the likelihood ratio test (LRT) is often used for binary hypothesis testing. Unfortunately these optimal detection tests are generally not *realizable* since they almost always depend at least on one of the unknown parameters $\boldsymbol{\theta}$. The LRTs are intended for providing the best attainable performance of any decision rule for a given problem [3, §3]. Therefore, a common approach to designing *realizable* tests is to replace the unknown parameters by estimates, the detection problem becoming a composite hypothesis testing problem (CHTP) [3, §6]. Although not necessarily optimal for detection performance, the estimates are generally chosen in the maximum likelihood sense, thereby obtaining the generalized likelihood ratio test (GLRT). If $\mathbf{C_x}$ is known and $\mathbf{s}_{\boldsymbol{\theta}}$ supposed to be completely unknown, then the GLRT reduces to the energy detector [3, §7.3]:

$$\left\|\mathbf{W}_{\mathbf{x}}^{-1}\mathbf{x}\right\|^2 = \mathbf{x}^H\mathbf{C}_{\mathbf{x}}^{-1}\mathbf{x} \gtrless T, \quad \mathbf{C_x} = \mathbf{W_x}\mathbf{W_x}^H \qquad (2)$$

where T is the detection threshold. It is a simple practical *realizable* detection test that can be used in any application. Additionally from a theoretical standpoint, one can expect the detection performance of the GLRT derived from the parametric model of $\mathbf{s}_{\boldsymbol{\theta}}$ to be somewhere between that of the Neyman-Pearson detector and the energy detector [3, §7.3].

3. BACKGROUND ON THE QCLB

The general approach lately introduced in [6] allows to revisit existing bounds by exploring the unbiasedness assumptions, from its *weakest* formulation (CRB) to its *strongest* formulation (BB). This approach has suggested a new approximation (QCLB) of the BB that allows a better prediction of the SNR threshold value than existing approximations (CRB, HCRB, MSB, AB_1), with a comparable computational complexity. Indeed, all mentioned lower bounds can be computed via the QCLB. This versatility will be used in §4 to take into account the detection test. For the sake of simplicity, we focus on the estimation of a single real function $g(\theta)$ of a single unknown real deterministic parameter θ. Ω denotes the observation space, Θ the parameter space, $\mathbb{F}_\Omega$ the real vector space of square integrable functions over Ω and $f_\theta(\mathbf{x})$ the p.d.f. of observations. A fundamental property of the MSE of a particular estimator $\widehat{g(\theta_0)}(\mathbf{x}) \in \mathbb{F}_\Omega$ of $g(\theta_0)$, where θ_0 is a selected value of the parameter θ, is that it is a norm associated with a particular scalar product $\langle \mid \rangle_\theta$:

$$MSE_{\theta_0}\left[\widehat{g(\theta_0)}\right] = \left\|\widehat{g(\theta_0)}(\mathbf{x}) - g(\theta_0)\right\|_{\theta_0}^2$$

where:

$$\begin{aligned}\langle g(\mathbf{x}) \mid h(\mathbf{x})\rangle_{\theta_0} &= E_{\theta_0}\left[g(\mathbf{x})\,h(\mathbf{x})\right] \\ &= \int_\Omega g(\mathbf{x})\,h(\mathbf{x})\,f_{\theta_0}(\mathbf{x})\,d\mathbf{x}\end{aligned}$$

In the search for a lower bound on the MSE, this property allows the use of two equivalent fundamental results: the generalization of the Cauchy-Schwartz inequality to Gram matrices (generally referred to as the "covariance inequality") and the minimization of a norm under linear constraints introduced hereinafter. Let $\mathbb{U}$ be an Euclidean vector space of any dimension (finite or infinite) on the body of real numbers $\mathbb{R}$ which has a scalar product $\langle \mid \rangle$. Let $(\mathbf{c}_1, \ldots, \mathbf{c}_K)$ be a free family of K vectors of $\mathbb{U}$ and $\mathbf{v} = (v_1, \ldots, v_K)^T$ a vector of $\mathbb{R}^K$. The problem of the minimization of $\|\mathbf{u}\|^2$ under the K linear constraints $\langle \mathbf{u} \mid \mathbf{c}_k\rangle = v_k$, $k \in [1, K]$ then has the solution:

$$\begin{aligned}\min\left\{\|\mathbf{u}\|^2\right\} &= \mathbf{v}^T\mathbf{G}^{-1}\mathbf{v} \text{ for } \mathbf{u}_{opt} = \sum_{k=1}^K \alpha_k \mathbf{c}_k \qquad (3)\\ (\alpha_1, \ldots, \alpha_K)^T &= \boldsymbol{\alpha} = \mathbf{G}^{-1}\mathbf{v},\ \mathbf{G}_{n,k} = \langle \mathbf{c}_k \mid \mathbf{c}_n\rangle\end{aligned}$$

As formulated by Barankin [5], the ultimate constraint that an unbiased estimator $\widehat{g(\theta_0)}(\mathbf{x})$ of $g(\theta_0)$ should verify is to be unbiased for all possible values of the unknown parameter:

$$E_\theta\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta),\ \forall\theta \in \Theta \qquad (4)$$

In this case the problem of interest becomes:

$$\min\left\{MSE_{\theta_0}\left[\widehat{g(\theta_0)}\right]\right\} \text{ under } E_\theta\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta), \qquad (5)$$

$\forall\theta \in \Theta$ and corresponds to the search for the locally-best unbiased estimator. Unfortunately, it is generally impossible to find an analytical solution of (5) providing the BB. Nevertheless the BB can be approximated by discretization of Barankin unbiasedness constraint (4). A general approach introduced in [6] consists in partitioning the parameter space Θ in N real sub-intervals $I_n = [\theta_n, \theta_{n+1}[$ where (4) is piecewise approximated by the constraints, $\theta_n + d\theta \in I_n$:

$$E_{\theta_n+d\theta}\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta_n + d\theta) + o\left(d\theta^{L_n}\right) \qquad (6)$$

Provided that both $f_\theta(\mathbf{x})$ and $g(\theta)$ can be developed in piecewise series expansions of order L_n, then $\min\left\{MSE_{\theta_0}\left[\widehat{g(\theta_0)}\right]\right\}$ under (6) is easily obtained using (3) [6]. Designating the BB approximations obtained as N-piecewise BB approximation of *homogeneous* order L, if on all sub-intervals I_n the series expansions are of the same order L, and of *heterogeneous* orders $\{L_1, ..., L_N\}$ if otherwise, this approach suggests a straightforward practical BB approximation: the QCLB based on a $N+1$-piecewise BB approximation of *homogeneous* order 1 defined by the constraints:

- $E_{\theta_n+d\theta}\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta_n+d\theta)+o(d\theta), \theta_n+d\theta \in I_n$

The QCLB is therefore a generalization of the CRB based on a 1-piecewise BB approximation of *homogeneous* order 1:

- $E_{\theta_0+d\theta}\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta_0+d\theta)+o(d\theta), \theta_0+d\theta \in \Theta$

is as well a generalization of the usual BB approximation used in the open literature, i.e. the MSB, based on an $N+1$-piecewise BB approximation of *homogeneous* order 0:

- $E_{\theta_n+d\theta}\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta_n+d\theta)+O(d\theta), \theta_n+d\theta \in I_n$

and a generalization of the AB_1 based on a $N+1$-piecewise BB approximation of *heterogeneous* order $\{1,0,...,0\}$:

- $\begin{cases} E_{\theta_0+d\theta}\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta_0+d\theta)+o(d\theta) \\ E_{\theta_n+d\theta}\left[\widehat{g(\theta_0)}(\mathbf{x})\right] = g(\theta_n+d\theta)+O(d\theta) \end{cases}$

where $\theta_0+d\theta \in I_0,\ \ \theta_n+d\theta \in I_{n>1}$.

For any set of $N+1$ test points $\{\theta_n\}_{[1,N+1]} = \{\theta_0\} \cup \{\theta_n\}_{[1,N]}$ (or set of $N+1$ sub-intervals I_n), the QCLB verify $QCLB \geq AB_1 \geq \max\{MSB, CRB\}$ and is given by:

$$QCLB = \mathbf{v}^T \begin{bmatrix} \mathbf{MS} & \mathbf{C} \\ \mathbf{C}^T & \mathbf{EFI} \end{bmatrix}^{-1} \mathbf{v} \tag{7}$$

where:

$$\begin{aligned}
\mathbf{v} &= \left(\Delta\mathbf{g}^T, \left(\ldots, \frac{\partial g(\theta_n)}{\partial\theta}, \ldots\right)\right)^T \\
\Delta\mathbf{g}^T &= (\ldots, g(\theta_n) - g(\theta_0), \ldots) \\
\mathbf{MS}_{n,l} &= E_{\theta_0}\left[\frac{f_{\theta_n}(\mathbf{x}) f_{\theta_l}(\mathbf{x})}{f_{\theta_0}(\mathbf{x})^2}\right] \\
\mathbf{C}_{n,l} &= E_{\theta_0}\left[\frac{\partial \ln f_{\theta_l}(\mathbf{x})}{\partial\theta} \frac{f_{\theta_n}(\mathbf{x}) f_{\theta_l}(\mathbf{x})}{f_{\theta_0}(\mathbf{x})^2}\right] \\
\mathbf{EFI}_{n,l} &= E_{\theta_0}\left[\frac{\partial \ln f_{\theta_n}(\mathbf{x})}{\partial\theta} \frac{\partial \ln f_{\theta_l}(\mathbf{x})}{\partial\theta} \frac{f_{\theta_n}(\mathbf{x}) f_{\theta_l}(\mathbf{x})}{f_{\theta_0}(\mathbf{x})^2}\right]
\end{aligned}$$

MS is the Mac-Aulay Seidman matrix, **EFI** stands for the Extended Fisher Information matrix, as it reduces to the FI (Fisher Information) when the set of test points is reduced to θ_0 only. **C** is a kind of "hybrid" matrix. An immediate generalization consists of taking their supremum over sub-interval definitions (set of test points).

4. CONDITIONAL LOWER BOUNDS

In this section, we provide an extension of QCLB analytical expression - and therefore of the CRB, HCRB, MSB and AB_1- by taking into account the energy detector. Indeed, if $\mathcal{D}$ is a *realizable* conditioning event, conditional bounds are obtained by substituting $\mathcal{D}$ and $f_\theta(\mathbf{x} \mid \mathcal{D})$ for Ω and $f_\theta(\mathbf{x})$ in the various expressions [2]:

$$\mathbf{MS}_{n,l} = E_{\theta_0}\left[\frac{f_{\theta_n}(\mathbf{x} \mid \mathcal{D}) f_{\theta_l}(\mathbf{x} \mid \mathcal{D})}{f_{\theta_0}(\mathbf{x} \mid \mathcal{D})^2} \mid \mathcal{D}\right]$$

$$\mathbf{C}_{n,l} = E_{\theta_0}\left[\frac{\partial \ln f_{\theta_l}(\mathbf{x} \mid \mathcal{D})}{\partial\theta} \frac{f_{\theta_n}(\mathbf{x} \mid D) f_{\theta_l}(\mathbf{x} \mid \mathcal{D})}{f_{\theta_0}(\mathbf{x} \mid D)^2} \mid \mathcal{D}\right]$$

$$\mathbf{EFI}_{n,l} = E_{\theta_0}\left[\frac{\partial \ln f_{\theta_n}(\mathbf{x} \mid \mathcal{D})}{\partial\theta} \frac{\partial \ln f_{\theta_l}(\mathbf{x} \mid \mathcal{D})}{\partial\theta} \frac{f_{\theta_n}(\mathbf{x} \mid \mathcal{D}) f_{\theta_l}(\mathbf{x} \mid \mathcal{D})}{f_{\theta_0}(\mathbf{x} \mid \mathcal{D})^2} \mid \mathcal{D}\right]$$

If $f_{\boldsymbol\theta}(\mathbf{x})$ is given by (1) and $\mathcal{D} = \left\{\mathbf{x} \mid \mathbf{x}^H \mathbf{C}_{\mathbf{x}}^{-1} \mathbf{x} \geq T\right\}$ is the event of the energy detector (2), then [9]:

$$P_{\mathcal{D}}(\mathbf{s}_\theta) = \int_{\mathcal{D}} f_\theta(\mathbf{x})\, d\mathbf{x} = \int_{t \geq T} f_{\mathcal{X}_{2L}^2}\left(t; \mathbf{s}_\theta^H \mathbf{C}_{\mathbf{x}}^{-1} \mathbf{s}_\theta\right) dt \tag{8}$$

where $f_{\mathcal{X}_{2L}^2}(t;\lambda)$ is the p.d.f. of a non central chi-squared random variable with $2L$ degrees of freedom and noncentrality parameter λ:

$$f_{\mathcal{X}_{2L}^2}(t;\lambda) = e^{-(t+\lambda)} I_{L-1}\left(2\sqrt{\lambda t}\right) \left(\sqrt{\frac{t}{\lambda}}\right)^{(L-1)} \tag{9}$$

$I_L(z)$ being the modified Bessel functions of the first kind [3, p 26]. Then a few lines of algebra leads to:

$$\begin{aligned}
\frac{f_{\theta_n}(\mathbf{x} \mid \mathcal{D}) f_{\theta_l}(\mathbf{x} \mid \mathcal{D})}{f_{\theta_0}(\mathbf{x} \mid \mathcal{D})} &= (\mathbf{MS}_{n,l})\, f_{\mathcal{CN}_L}(\mathbf{x} \mid \mathcal{D}; \mathbf{m}_{\mathbf{x}}, \mathbf{C}_{\mathbf{x}}) \\
\mathbf{m}_{\mathbf{x}} &= \mathbf{s}_{\theta_n} + \mathbf{s}_{\theta_l} - \mathbf{s}_{\theta_0}
\end{aligned}$$

$$\mathbf{MS}_{n,l} = e^{2\,\mathrm{Re}\left\{\left(\mathbf{s}_{\theta_n} - \mathbf{s}_{\theta_0}\right)^H \mathbf{C}_{\mathbf{x}}^{-1} \left(\mathbf{s}_{\theta_l} - \mathbf{s}_{\theta_0}\right)\right\}} \frac{P_{\mathcal{D}}(\mathbf{s}_{\theta_0}) P_{\mathcal{D}}(\mathbf{s}_{\theta_n} + \mathbf{s}_{\theta_l} - \mathbf{s}_{\theta_0})}{P_{\mathcal{D}}(\mathbf{s}_{\theta_n}) P_{\mathcal{D}}(\mathbf{s}_{\theta_l})} \tag{10}$$

Let us denote $E[\mathbf{x} \mid \mathcal{D}] = \int_{\mathcal{D}} \mathbf{x} f_{\mathcal{CN}_L}(\mathbf{x} \mid \mathcal{D}; \mathbf{m}_{\mathbf{x}}, \mathbf{C}_{\mathbf{x}})\, d\mathbf{x}$.

Since $\frac{\partial \ln f_\theta(\mathbf{x}|\mathcal{D})}{\partial\theta} = 2\,\mathrm{Re}\left\{\frac{\partial \mathbf{s}_\theta^H}{\partial\theta} \mathbf{C}_{\mathbf{x}}^{-1} (\mathbf{x} - \mathbf{s}_\theta)\right\} - \frac{\partial \ln P_{\mathcal{D}}(\mathbf{s}_\theta)}{\partial\theta}$, then:

$$\begin{aligned}
\mathbf{EFI}_{n,l} &= (\mathbf{MS}_{n,l})\, E\left[\frac{\partial \ln f_{\theta_n}(\mathbf{x} \mid \mathcal{D})}{\partial\theta} \frac{\partial \ln f_{\theta_l}(\mathbf{x} \mid \mathcal{D})}{\partial\theta} \mid \mathcal{D}\right] \\
\mathbf{EFI}_{n,l} &= (\mathbf{MS}_{n,l}) \left[2\,\mathrm{Re}\left\{\frac{\partial \mathbf{s}_{\theta_n}^H}{\partial\theta} \mathbf{C}_{\mathbf{x}}^{-1} \mathbf{A}_{n,l} \mathbf{C}_{\mathbf{x}}^{-1} \frac{\partial \mathbf{s}_{\theta_l}}{\partial\theta}\right\}\right. \\
&\quad + 2\,\mathrm{Re}\left\{\frac{\partial \mathbf{s}_{\theta_n}^H}{\partial\theta} \mathbf{C}_{\mathbf{x}}^{-1} \mathbf{B}_{n,l} \left(\mathbf{C}_{\mathbf{x}}^{-1}\right)^T \frac{\partial \mathbf{s}_{\theta_l}^*}{\partial\theta}\right\} \\
&\quad + \frac{\partial \ln P_{\mathcal{D}}(\mathbf{s}_{\theta_n})}{\partial\theta} \frac{\partial \ln P_{\mathcal{D}}(\mathbf{s}_{\theta_l})}{\partial\theta} \\
&\quad - 2 \frac{\partial \ln P_{\mathcal{D}}(\mathbf{s}_{\theta_l})}{\partial\theta} \mathrm{Re}\left\{\frac{\partial \mathbf{s}_{\theta_n}^H}{\partial\theta} \mathbf{C}_{\mathbf{x}}^{-1} (E[\mathbf{x} \mid \mathcal{D}] - \mathbf{s}_{\theta_n})\right\} \\
&\quad \left. - 2 \frac{\partial \ln P_{\mathcal{D}}(\mathbf{s}_{\theta_n})}{\partial\theta} \mathrm{Re}\left\{\frac{\partial \mathbf{s}_{\theta_l}^H}{\partial\theta} \mathbf{C}_{\mathbf{x}}^{-1} (E[\mathbf{x} \mid \mathcal{D}] - \mathbf{s}_{\theta_l})\right\}\right]
\end{aligned} \tag{11}$$

$$\mathbf{C}_{n,l} = (\mathbf{MS}_{n,l})\, E\left[\frac{\partial \ln f_{\theta_l}(\mathbf{x} \mid \mathcal{D})}{\partial\theta} \mid \mathcal{D}\right]$$

$$\begin{aligned}\mathbf{C}_{n,l} &= (\mathbf{MS}_{n,l})\left[2\,\mathrm{Re}\left\{\frac{\partial \mathbf{s}_{\theta_l}^H}{\partial\theta}\mathbf{C}_{\mathbf{x}}^{-1}\left(E\left[\mathbf{x} \mid \mathcal{D}\right] - \mathbf{s}_{\theta_l}\right)\right\} \right. \quad (12)\\ &\qquad \left. - \frac{\partial \ln P_{\mathcal{D}}\left(\mathbf{s}_{\theta_l}\right)}{\partial\theta}\right]\end{aligned}$$

where:

$$\begin{aligned}\mathbf{A}_{n,l} &= E\left[\left(\mathbf{x} - \mathbf{s}_{\theta_l}\right)\left(\mathbf{x} - \mathbf{s}_{\theta_n}\right)^H \mid \mathcal{D}\right]\\ &= E\left[\mathbf{x}\mathbf{x}^H \mid \mathcal{D}\right] - E\left[\mathbf{x} \mid \mathcal{D}\right]\mathbf{s}_{\theta_n}^H - \mathbf{s}_{\theta_l}E\left[\mathbf{x} \mid \mathcal{D}\right]^H\\ &\quad + \mathbf{s}_{\theta_l}\mathbf{s}_{\theta_n}^H\\ \mathbf{B}_{n,l} &= E\left[\left(\mathbf{x} - \mathbf{s}_{\theta_l}\right)\left(\mathbf{x} - \mathbf{s}_{\theta_n}\right)^T \mid \mathcal{D}\right]\\ &= E\left[\mathbf{x}\mathbf{x}^T \mid \mathcal{D}\right] - E\left[\mathbf{x} \mid \mathcal{D}\right]\mathbf{s}_{\theta_n}^T - \mathbf{s}_{\theta_l}E\left[\mathbf{x} \mid \mathcal{D}\right]^T\\ &\quad + \mathbf{s}_{\theta_l}\mathbf{s}_{\theta_n}^T\end{aligned}$$

and [9]:

$$E\left[\mathbf{x} \mid \mathcal{D}\right] = \frac{1 - P_{L+1}\left(\mathbf{m}_{\mathbf{x}}\right)}{1 - P_L\left(\mathbf{m}_{\mathbf{x}}\right)}\mathbf{m}_{\mathbf{x}}$$

$$E\left[\mathbf{x}\mathbf{x}^H \mid \mathcal{D}\right] = \frac{1 - P_{L+1}\left(\mathbf{m}_{\mathbf{x}}\right)}{1 - P_L\left(\mathbf{m}_{\mathbf{x}}\right)}\mathbf{C}_{\mathbf{x}} + \frac{1 - P_{L+2}\left(\mathbf{m}_{\mathbf{x}}\right)}{1 - P_L\left(\mathbf{m}_{\mathbf{x}}\right)}\mathbf{m}_{\mathbf{x}}\mathbf{m}_{\mathbf{x}}^H$$

$$E\left[\mathbf{x}\mathbf{x}^T \mid \mathcal{D}\right] = \frac{1 - P_{L+2}\left(\mathbf{m}_{\mathbf{x}}\right)}{1 - P_L\left(\mathbf{m}_{\mathbf{x}}\right)}\mathbf{m}_{\mathbf{x}}\mathbf{m}_{\mathbf{x}}^T$$

$$\frac{\partial \ln P_{\mathcal{D}}\left(\mathbf{s}_\theta\right)}{\partial\theta} = \left(\frac{P_L\left(\mathbf{s}_\theta\right) - P_{L+1}\left(\mathbf{s}_\theta\right)}{1 - P_L\left(\mathbf{s}_\theta\right)}\right)\frac{\partial\left(\mathbf{s}_\theta^H\mathbf{C}_{\mathbf{x}}^{-1}\mathbf{s}_\theta\right)}{\partial\theta}$$

$$P_{\mathcal{D}}\left(\mathbf{s}_\theta\right) = 1 - P_L\left(\mathbf{s}_\theta\right)$$

$$P_L\left(\mathbf{s}\right) = \int_0^T f_{\mathcal{X}_{2L}^2}\left(t; \mathbf{s}_\theta^H\mathbf{C}_{\mathbf{x}}^{-1}\mathbf{s}_\theta\right)dt$$

Finally the conditional QCLB is given by (7) computed according to (10)(11)(12) and the conditional MSB, AB_1, CRB are given by:

$$MSB = \Delta\mathbf{g}^T\left[\mathbf{MS}\right]^{-1}\Delta\mathbf{g}$$

$$AB_1 = \mathbf{v}^T\begin{bmatrix}\mathbf{MS} & \mathbf{c}\\ \mathbf{c}^T & \mathbf{EFI}_{0,0}\end{bmatrix}^{-1}\mathbf{v}, \begin{cases}\mathbf{c} = \left(\ldots, \mathbf{C}_{n,0}, \ldots\right)^T\\ \mathbf{v} = \left(\Delta\mathbf{g}^T, \frac{\partial g(\theta_0)}{\partial\theta}\right)^T\end{cases}$$

$$CRB = \frac{\partial g(\theta_0)}{\partial\theta}\left[\mathbf{EFI}_{0,0}\right]^{-1}\frac{\partial g(\theta_0)}{\partial\theta}$$

where [9]:

$$\begin{aligned}\mathbf{EFI}_{0,0} &= 2\,\mathrm{Re}\left\{\frac{\partial \mathbf{s}_{\theta_0}^H}{\partial\theta}\mathbf{C}_{\mathbf{x}}^{-1}\frac{\partial \mathbf{s}_{\theta_0}}{\partial\theta}\right\}\left(\frac{1 - P_{L+1}\left(\theta_0\right)}{1 - P_L\left(\theta_0\right)}\right)\\ &\quad + w_L\left(\theta_0\right)\left(\frac{\partial\left(\mathbf{s}_{\theta_0}^H\mathbf{C}_{\mathbf{x}}^{-1}\mathbf{s}_{\theta_0}\right)}{\partial\theta}\right)^2\end{aligned}$$

$$\begin{aligned}w_L\left(\theta\right) &= \frac{2P_{L+1}\left(\theta\right) - P_L\left(\theta\right) - P_{L+2}\left(\theta\right)}{1 - P_L\left(\theta\right)}\\ &\quad - \left(\frac{P_{L+1}\left(\theta\right) - P_L\left(\theta\right)}{1 - P_L\left(\theta\right)}\right)^2\end{aligned}$$

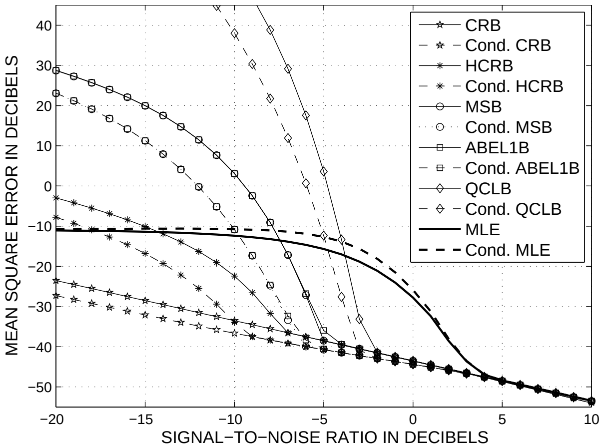


Figure 1: MSE of MLE and MSE Lower Bounds conditioned or not by the Energy Detector versus SNR, $L = 10$, $P_{FA} = 10^{-3}$

5. SINGLE TONE THRESHOLD ANALYSIS

Let us consider the reference estimation problem where the vector $\mathbf{x}$ is modelled by:

$$\begin{aligned}\mathbf{x} &= a\boldsymbol{\psi}(\theta) + \mathbf{n}\\ \boldsymbol{\psi}(\theta) &= \left[1, e^{j2\pi\theta}, ..., e^{j2\pi(L-1)\theta}\right]^T,\ \theta \in \left]-0.5, 0.5\right[\end{aligned}$$

i.e. $\mathbf{s}_\theta = a\boldsymbol{\psi}(\theta)$ and $\mathbf{C}_{\mathbf{x}} = \mathbf{Id}$, a^2 being the SNR ($a > 0$). Then $\frac{\partial \ln P_{\mathcal{D}}(\mathbf{s}_\theta)}{\partial\theta} = 0$ and $\widehat{\theta}_{ML} = \max_\theta\left\{\mathrm{Re}\left[\boldsymbol{\psi}(\theta)^H\mathbf{x}\right]\right\}$

For any set of $N+1$ test points $\{\theta_n\}_{[1,N+1]}$, only the MSB, the AB_1 and the QCLB are of a comparable complexity. Nevertheless, we also include in the comparison the HCRB as it is the simplest representative of Large Errors bounds. For the sake of fair comparison with the HCRB which is the supremum of the MSB where $\{\theta_n\}_{[1,2]} = \{\theta_0, \theta_0 + d\theta\}$, the MSB, AB_1, QCLB are also computed as supremum over the possible values of $\{\theta_n\}_{[1,N+1]}$. For the sake of simplicity $\{\theta_n\}_{[1,3]} = \{\theta_0, \theta_0 + d\theta, \theta_0 - d\theta\}$. We consider the reference estimation case where $\theta_0 = 0$.

Figure (1) compares the various bounds, conditioned or not by the Energy Detector, as a function of SNR in the case of $L = 10$ samples and $P_{FA} = 10^{-3}$. The MSE of the MLE is also shown in order to compare the threshold behaviour of the bounds (10^6 trials). As expected, the QCLB keeps providing a significant improvement in the prediction of the SNR threshold value, whatever the observations are conditioned or not (same results can be observed for $L = 2, 4, ..., 32$ and $P_{FA} = 10^{-1}, 10^{-2}, ..., 10^{-6}$).

A more unexpected and non intuitive result is the increase of the MSE of the MLE in the transition region as the detection threshold increases (as the P_{FA} decreases) highlighted by figure (2). Indeed, intuitively, a detection step is expected to decrease the MSE of the MLE by selecting instances with relatively high signal

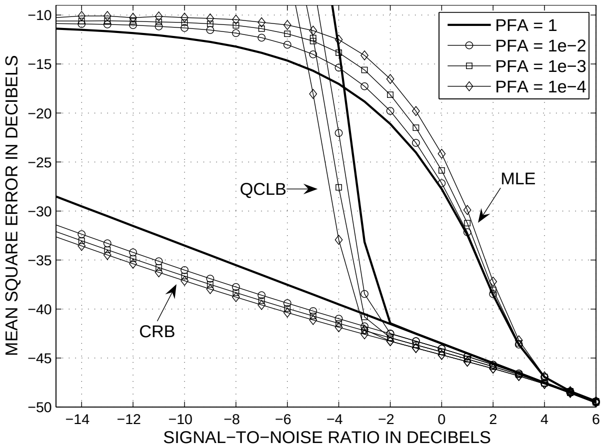


Figure 2: MSE of MLE, CRB and QCLB conditioned or not by the Energy Detector versus SNR, $L = 10$, $P_{FA} = 10^{-2}, 10^{-3}, 10^{-4}$

energy - sufficient to exceed the detection threshold - and disregarding instances belonging to the *a priori* region that deteriorate the MSE. The former analysis is reinforced theoretically by the lower bounds behavior (CRB and QCLB) in figure (2) and has also been reinforced so far practically by results obtained in [2] for the monopulse ratio estimation problem under a stochastic signal model. Again, if we consider the stochastic case, i.e. $a \sim \mathcal{CN}_1(0, snr)$, then $\widehat{\theta}_{ML} = \max_{\theta} \left\{ \left| \boldsymbol{\psi}(\theta)^H \mathbf{x} \right|^2 \right\}$ and one can check that the behavior of its MSE is the opposite and true to the common intuition.

This paradoxical result clearly addresses a challenging theoretical issue that will have to be the subject of further research.

6. CONCLUSION

In the present paper, we have derived lower bounds on MSE (CRB, HCRB, MSB, AB_1, QCLB) for the deterministic signal model conditioned by the Energy Detector. This results will be useful to update the estimation performance analysis for a wide variety of processing including the Energy Detector. Additionally, we have shown that the QCLB keeps providing a significant improvement in the prediction of the SNR threshold value when the observations are conditioned, in comparison with the MSB (the usual BB approximation in the open literature [7]).

REFERENCES

[1] H.L. Van Trees, "Detection, Estimation and Modulation Theory, Part 1", New York, Wiley, 1968

[2] E. Chaumette, P. Larzabal, P. Forster, "On the Influence of a Detection Step on Lower Bounds for Deterministic Parameters Estimation", IEEE Trans. on SP, vol 53, pp 4080-4090, 2005

[3] S.M. Kay, "Fundamentals of Statistical Signal Processing: detection theory", Prentice-Hall, 1998

[4] A. Renaux, P. Forster, E. Chaumette, P. Larzabal, "On the High SNR CML Estimator Full Statistical Characterization", to appear in IEEE Trans. on SP

[5] E.W. Barankin, "Locally best unbiased estimates", Ann. Math. Stat., vol. 20, pp 477-501, 1949

[6] A. Quinlan, E. Chaumette, P. Larzabal, "A direct method to generate approximations of the barankin bound", ICASSP Conf. 2006, Vol III, pp 808-811

[7] J. Tabrikian, J.L. Krolik, "Barankin bounds for source localization in an uncertain ocean environment", IEEE Trans. on SP, vol 47, pp 2917-2927, 1999

[8] D.C. Rife, R.R. Boorstyn,"Single tone parameter estimation from discrete-time observations", IEEE Trans. on IT, vol 20, pp 591-598, 1974

[9] E. Chaumette, P. Larzabal, "Cramér-Rao Bound Conditioned by the Energy Detector", accepted in IEEE SP letters

K. New trends in deterministic lower bounds and SNR threshold estimation: From derivable bounds to conjectural bounds (SAM)

in Proc. IEEE SAM, 2010.

NEW TRENDS IN DETERMINISTIC LOWER BOUNDS AND SNR THRESHOLD ESTIMATION: FROM DERIVABLE BOUNDS TO CONJECTURAL BOUNDS

Eric Chaumette (1), *Alexandre Renaux* (2) *and Pascal Larzabal*(3)

(1) ONERA - DEMR/TSI, The French Aerospace Lab, Chemin de la Hunière, F-91120 Palaiseau, France
(2) Université Paris-Sud 11, L2S, Supelec, 3 rue Joliot Curie, F-91190 Gif-Sur-Yvette, France
(3) SATIE, ENS Cachan, CNRS, UniverSud, 61 av President Wilson, F-94230 Cachan, France

ABSTRACT

It is well known that in non-linear estimation problems the ML estimator exhibits a threshold effect, i.e. a rapid deterioration of estimation accuracy below a certain SNR or number of snapshots. This effect is caused by outliers and is not captured by standard tools such as the Cramér-Rao bound (CRB). The search of the SNR threshold value can be achieved with the help of approximations of the Barankin bound (BB) proposed by many authors. These approximations may result from linear or non-linear transformation (discrete or integral) of the uniform unbiasedness constraint introduced by Barankin. Additionally, the strong analogy between derivations of deterministic bounds and Bayesian bounds of the Weiss-Weinstein family has led us to propose a conjectural bound which outperforms existing ones for SNR threshold prediction.

Index Terms— Parameter estimation, mean-square-error bounds, SNR threshold

1. INTRODUCTION

Minimal performance bounds allow for calculation of the best performance that may be achieved, in the Mean Square Error (MSE) sense, when estimating a set of model parameters from noisy observations. Historically the first MSE lower bound for deterministic parameters to be derived was the Cramér-Rao Bound (CRB) [4], which has been the most widely used since. Its popularity is largely due to its simplicity of calculation leading to closed-form expressions useful for system analysis and design. Additionally, the CRB can be achieved asymptotically (high SNR and/or large number of snapshots) by Maximum Likelihood Estimators (MLE), and last but not least, it is the lowest bound on the MSE of unbiased estimators, since it derives from a local formulation of unbiasedness in the vicinity of the true parameters [2]. This initial characterization of locally unbiased estimators has been improved first by Bhattacharyya's works [4] which refined the characterization of local unbiasedness, and significantly generalized by Barankin works [1], who established the general form of the greatest lower bound on MSE (BB) taking into account a uniform unbiasedness definition (eq. (1)). Unfortunately the BB is the solution of an integral equation with a generally incomputable analytic solution (eq. (8)).

This work has been partly funded by the European Network of excellence NEWCOM++ under the number 216715

This project is partialy funded by both the Région Ile-de-France and the Digiteo Research Park

Therefore, since then, numerous works detailed in [2][3] have been devoted to deriving computable approximations of the BB and have shown that the CRB and the BB can be regarded as key representatives of two general classes of bounds, respectively the Small-Error bounds and the Large-Error bounds. These works have also shown that in non-linear estimation problems three distinct regions of operation can be observed. In the asymptotic region, *i.e.* at a high number of independent snapshots and/or at high SNR, the MSE is small and, in many cases, close to the Small-Error bounds. In the *a priori* performance region where the number of independent snapshots and/or the SNR are very low, the observations provide little information and the MSE is close to that obtained from the prior knowledge about the problem. Between these two extremes, there is a transition region where MSE of MLEs usually deteriorates rapidly with respect to existing MSE lower bounds (Large or Small) and exhibits a threshold behaviour, which corresponds to a "performance breakdown" of the estimators due to the appearance of outliers.
Small-Error bounds are not able to handle the threshold phenomena, whereas it is revealed by Large-Error bounds that can be used to predict the threshold value. On the other hand, Large-Error bounds suffer from their computational cost. Indeed, each BB approximation request the search of an optimum over a set of test points and their tightness depends on the chosen set of test points.
And tightness is the matter, since a more accurate knowledge of the BB allows a better prediction of the SNR threshold value.
Therefore, at least two strategies can be adopted.
The first one is the most consistent with deductive reasoning applied to the unbiasedness paradigm. This strategy, fully mastered from the mathematics and the meaning point of view, provides derivable bounds and relies on the introduction a general class of possible transformations (eq. (11)) of the uniform unbiasedness constraint (eq. (1)), i.e. the mixture of integral linear and non-linear transformations, opening a wide variety of directions in the search of computable tighter BB approximations.
The second one is partially consistent with deductive reasoning since it may be based - see our example Section 3 - on analogies between families of lower bounds without the support of a non-questionnable derivation and interpretation. And yet, but nevertheless this strategy may yield some conjectural bounds tightest than the existing and well-estabished ones.

2. DERIVABLE LOWER BOUNDS

2.1. Linear transformations of the unbiasedness constraint

For the sake of simplicity we will focus on the estimation of a single real function $g(\theta)$ of a single unknown real deterministic parameter θ. In the following, unless otherwise stated, $\mathbf{x}$ denotes the random observation vector of dimension M, Ω the observations space, and $p(\mathbf{x};\theta)$ the probability density function (p.d.f.) of $\mathbf{x}$ depending on $\theta \in \Theta$, where Θ denotes the parameter space. Let $L^2(\Omega)$ be the real Hilbert space of square integrable functions over Ω.
In the search for a lower bound on the MSE of unbiased estimators, two fundamental properties of the problem at hand, introduced by Barankin [1], must be noticed. The first property is that the MSE of a particular estimator $\widehat{g(\theta^0)}(\mathbf{x}) \in L^2(\Omega)$ of $g(\theta^0)$, where θ^0 is a selected value of the parameter θ, is a norm associated with a particular scalar product $\langle \mid \rangle_\theta$:

$$\begin{aligned} MSE_{\theta^0}\left[\widehat{g(\theta^0)}\right] &= \left\|\widehat{g(\theta^0)}(\mathbf{x}) - g(\theta^0)\right\|^2_{\theta^0}, \\ \langle g(\mathbf{x}) \mid h(\mathbf{x})\rangle_{\theta^0} &= E_{\theta^0}\left[g(\mathbf{x})^* h(\mathbf{x})\right]. \end{aligned}$$

The second property is that an unbiased estimator $\widehat{g(\theta^0)}(\mathbf{x})$ of $g(\theta)$ should be uniformly unbiased, i.e. for all possible values of the unknown parameter $\theta \in \Theta$ it must satisfy:

$$E_\theta\left[\widehat{g(\theta^0)}(\mathbf{x})\right] = g(\theta) = E_{\theta^0}\left[\widehat{g(\theta^0)}(\mathbf{x})\,\nu(\mathbf{x};\theta)\right], \quad (1)$$

where $\nu(\mathbf{x};\theta) = \frac{p(\mathbf{x};\theta)}{p(\mathbf{x};\theta^0)}$ denotes the Likelihood Ratio (LR). As a consequence, the locally-best (at θ^0) unbiased estimator is the solution of a norm minimization under linear constraints

$$\min\left\{MSE_{\theta^0}\left[\widehat{g(\theta^0)}\right]\right\} \text{ under } E_{\theta^0}\left[\widehat{g(\theta^0)}(\mathbf{x})\,\nu(\mathbf{x};\theta)\right] = g(\theta),$$

solution that can be obtained by using the norm minimization lemma

$$\begin{aligned} &\min\left\{\mathbf{u}^H\mathbf{u} \text{ under } \mathbf{c}_k^H\mathbf{u} = v_k, 1 \le k \le K\right\} = \mathbf{v}^H\mathbf{G}^{-1}\mathbf{v} \\ &\mathbf{u}_{opt} = \sum_{k=1}^{K}\alpha_k\mathbf{c}_k, \quad \boldsymbol{\alpha} = \mathbf{G}^{-1}\mathbf{v}, \quad \mathbf{G}_{n,k} = \mathbf{c}_n^H\mathbf{c}_k \end{aligned} \quad . \quad (2)$$

Unfortunately, as shown hereinafter, if Θ contains a continuous subset of $\mathbb{R}$, then the norm minimization under a set of an infinite number of linear constraints (1) leads to an integral equation (8) with no analytical solution in general. Therefore, since the original work of Barankin [1], many studies [2, and references therein][3] have been dedicated to the derivation of "computable" lower bounds approximating the MSE of the locally-best unbiased estimator (BB). All these approximations derive from sets of discrete or integral linear transform of the "Barankin" constraint (1), and accordingly of the LR, and can be obtained using the following simple rationale.
Let $\boldsymbol{\theta}^N = \left(\theta^1, \ldots, \theta^N\right)^T \in \Theta^N$ be a vector of N test points, $\boldsymbol{\nu}\left(\mathbf{x};\boldsymbol{\theta}^N\right) = \left(\nu(\mathbf{x};\theta^1), \ldots, \nu(\mathbf{x};\theta^N)\right)^T$ be the vector of LR associated to $\boldsymbol{\theta}^N$, $\xi(\theta) = g(\theta) - g(\theta^0)$ and $\boldsymbol{\xi}\left(\boldsymbol{\theta}^N\right) = \left(\xi(\theta^1), \ldots, \xi(\theta^N)\right)^T$.
Any unbiased estimator $\widehat{g(\theta^0)}(\mathbf{x})$ satisfying (1) must comply with

$$E_{\theta^0}\left[\left(\widehat{g(\theta^0)}(\mathbf{x}) - g(\theta^0)\right)\boldsymbol{\nu}\left(\mathbf{x};\boldsymbol{\theta}^N\right)\right] = \boldsymbol{\xi}\left(\boldsymbol{\theta}^N\right), \quad (3)$$

and with any subsequent linear transformation of (3). Therefore, any given set of K ($K \le N$) independent linear transformations of (3):

$$E_{\theta^0}\left[\left(\widehat{g(\theta^0)}(\mathbf{x}) - g(\theta^0)\right)\mathbf{h}_k^T\boldsymbol{\nu}\left(\mathbf{x};\boldsymbol{\theta}^N\right)\right] = \mathbf{h}_k^T\boldsymbol{\xi}\left(\boldsymbol{\theta}^N\right), \quad (4)$$

$\mathbf{h}_k \in \mathbb{R}^N, 1 \le k \le K$, provides with a lower bound on the MSE (2):

$$MSE_{\theta^0}\left[\widehat{g(\theta^0)}\right] \ge \boldsymbol{\xi}\left(\boldsymbol{\theta}^N\right)^T \widetilde{\mathbf{G}}_{\mathbf{H}_K}\boldsymbol{\xi}\left(\boldsymbol{\theta}^N\right), \quad (5)$$

where $\widetilde{\mathbf{G}}_{\mathbf{H}_K} = \mathbf{H}_K\left(\mathbf{H}_K^T\mathbf{R}_{\boldsymbol{\nu}}\mathbf{H}_K\right)^{-1}\mathbf{H}_K^T$, $\mathbf{H}_K = [\mathbf{h}_1 \ldots \mathbf{h}_K]$ and $(\mathbf{R}_{\boldsymbol{\nu}})_{n,m} = E_{\theta^0}\left[\nu(\mathbf{x};\theta^n)\,\nu(\mathbf{x};\theta^m)\right]$. The BB is obtained by taking the supremum of (5) over all the existing degrees of freedom $\left(N, \boldsymbol{\theta}^N, K, \mathbf{H}_K\right)$. Moreover, for a given vector of test points $\boldsymbol{\theta}^N$, the lower bound (5) reaches its maximum iff the matrix $\mathbf{H}_K$ is invertible ($K = N$), which represents a bijective transformation of the set of the N initial constraints (3):

$$MSE_{\theta^0}\left[\widehat{g(\theta^0)}\right] \ge \boldsymbol{\xi}\left(\boldsymbol{\theta}^N\right)^T\widetilde{\mathbf{G}}_{\mathbf{I}_N}\boldsymbol{\xi}\left(\boldsymbol{\theta}^N\right) \ge \boldsymbol{\xi}\left(\boldsymbol{\theta}^N\right)^T\widetilde{\mathbf{G}}_{\mathbf{H}_K}\boldsymbol{\xi}\left(\boldsymbol{\theta}^N\right)$$

where $\mathbf{I}_N$ is the identity matrix with dimension N. All known bounds on the MSE deriving from the Barankin Bound is a particular implementation of (5), including the most general formalism introduced lately in [3]. Indeed, the limit of (4) where $N \to \infty$ and $\boldsymbol{\theta}^N$ uniformly samples Θ leads to the linear integral constraint:

$$\begin{aligned} &E_{\theta^0}\left[\left(\widehat{g(\theta^0)}(\mathbf{x}) - g(\theta^0)\right)\eta(\mathbf{x},\tau)\right] = \Gamma_h(\tau), \quad (6) \\ &\eta(\mathbf{x},\tau) = \int_\Theta h(\tau,\theta)\,\nu(\mathbf{x};\theta)\,d\theta, \quad \Gamma_h(\tau) = \int_\Theta h(\tau,\theta)\,\xi(\theta)\,d\theta, \end{aligned}$$

where each $\mathbf{h}_k = \left(h(\tau_k,\theta^1), \ldots, h(\tau_k,\theta^N)\right)^T$ is the vector of samples of a parametric function $h(\tau,\theta), \tau \in \Lambda \subset \mathbb{R}$, integrable over $\Theta, \forall\tau \in \Lambda$. Then, for any subset of K values of τ, $\{\tau_k\}_{1\le k\le K}$, the subset of the associated K linear integral constraints (6) leads to the following lower bound (2):

$$\left\{\begin{array}{l} MSE_{\theta^0}\left[\widehat{g(\theta^0)}(\mathbf{x})\right] \ge MSE_{\theta^0}\left[\widehat{g(\theta^0)}_{lmvu}(\mathbf{x})\right] \\ MSE_{\theta^0}\left[\widehat{g(\theta^0)}_{lmvu}(\mathbf{x})\right] = \boldsymbol{\Gamma}_h^T\mathbf{R}_\eta^{-1}\boldsymbol{\Gamma}_h = \boldsymbol{\Gamma}_h^T\left(\frac{\mathbf{a}}{\lambda}\right) \\ \widehat{g(\theta^0)}_{lmvu}(\mathbf{x}) - g(\theta^0) = \sum_{k=1}^{K}\frac{a_k}{\lambda}\eta(\mathbf{x},\tau_k) \\ \mathbf{R}_\eta\left(\frac{\mathbf{a}}{\lambda}\right) = \boldsymbol{\Gamma}_h \end{array}\right. \quad (7)$$

where $(\mathbf{R}_\eta)_{k,k'} = E_{\theta^0}\left[\eta(\mathbf{x},\tau_k)\,\eta(\mathbf{x},\tau_{k'})\right]$ and $(\boldsymbol{\Gamma}_h)_k = \Gamma(\tau_k)$. Therefore, when $K \to \infty$ and the set $\{\tau_k\}_{1\le k\le K}$ uniformly samples Λ, by setting $\frac{1}{\lambda} = d\tau = \tau_{k+1} - \tau_k$, $\boldsymbol{\beta} = \frac{\mathbf{a}}{\lambda}$, the integral form of the above lower bound appears straightforwardly:

$$\left\{\begin{array}{l} MSE_{\theta^0}\left[\widehat{g(\theta^0)}_{lmvu}(\mathbf{x})\right] = \int_\Lambda \Gamma_h(\tau)\,\beta(\tau)\,d\tau \\ \widehat{g(\theta^0)}_{lmvu}(\mathbf{x}) - g(\theta^0) = \int_\Lambda \eta(\mathbf{x},\tau)\,\beta(\tau)\,d\tau \\ \int_\Lambda K_h(\tau',\tau)\,\beta(\tau)\,d\tau = \Gamma_h(\tau') \end{array}\right. , \quad (8)$$

$$\begin{aligned} K_h(\tau,\tau') &= E_{\theta^0}\left[\eta(\mathbf{x},\tau)\,\eta(\mathbf{x},\tau')\right] \\ &= \iint_\Theta h(\tau,\theta)\,R_{\boldsymbol{\nu}}(\theta,\theta')\,h(\tau',\theta')\,d\theta d\theta', \end{aligned}$$

$$R_{\boldsymbol{\nu}}(\theta,\theta') = E_{\theta^0}\left[\frac{p(\mathbf{x};\theta)}{p(\mathbf{x};\theta^0)}\frac{p(\mathbf{x};\theta')}{p(\mathbf{x};\theta^0)}\right] = \int_\Omega \frac{p(\mathbf{x};\theta)\,p(\mathbf{x};\theta')}{p(\mathbf{x};\theta^0)}d\mathbf{x},$$

which is exactly the main result introduced in [3] and is a generalization of the Kiefer Bound [4] ($K = 2$). Note that if $h(\tau,\theta) = \delta(\tau-\theta)$ (limit case of $\mathbf{H}_N = \mathbf{I}_N$ where $N = K \to \infty$) then $K_h(\tau,\tau') = R_{\boldsymbol{\nu}}(\tau,\tau')$ and (8) becomes the simplest expression of the exact Barankin Bound [2, (10)]. As mentioned above, in most practical cases, it is impossible to find either the limit of (7) or an analytical solution of (8) to obtain an explicit form of the exact Barankin Bound on the MSE, which somewhat limits its interest.

Nevertheless this formalism allows to use discrete (4) or integral (6) linear transforms of the LR, possibly non-invertible, possibly optimized for a set of p.d.f. (such as the Fourier transform in [3]) in order to get a tight approximation of the BB.

2.2. Non-linear transformations of the unbiasedness constraint

Let us consider the set of estimation problems characterized by a p.d.f. for which there exists a real valued function t such that:

$$t\left(p\left(\mathbf{x};\theta\right)\right) = k\left(\theta,t\right)p\left(\mathbf{x};\gamma\left(\theta,t\right)\right), \quad k\left(\theta,t\right) = \int_{\Omega} t\left(p\left(\mathbf{x};\theta\right)\right)d\mathbf{x} \tag{9}$$

Then an unbiased estimator satisfying (1) satisfies as well [5],$\forall\theta\in\Theta$:

$$E_{\theta^0}\left[\left(\widehat{g\left(\theta^0\right)}\left(\mathbf{x}\right)-g\left(\theta^0\right)\right)\frac{t\left(p\left(\mathbf{x};\theta\right)\right)}{p\left(\mathbf{x};\theta^0\right)}\right] = k\left(\theta,t\right)\left[g\left(\gamma\left(\theta,t\right)\right)-g\left(\theta^0\right)\right].$$

Moreover, if there exists a set of functions t_θ satisfying (9), then we can update the definition of $\nu\left(\mathbf{x};\theta\right)$ and $\xi\left(\theta\right)$ in (6) according to:

$$\nu\left(\mathbf{x};\theta\right) = \frac{t_\theta\left(p\left(\mathbf{x};\theta\right)\right)}{p\left(\mathbf{x};\theta^0\right)}, \quad \xi\left(\theta\right) = k\left(\theta,t_\theta\right)\left[g\left(\gamma\left(\theta,t_\theta\right)\right)-g\left(\theta^0\right)\right], \tag{10}$$

and all the results released in the previous Section still hold, the linear integral transformation becoming a mixture of linear and non-linear integral transformations:

$$\begin{aligned}\eta\left(\mathbf{x},\tau\right) &= \int_{\Theta} h\left(\tau,\theta\right)\frac{t_\theta\left(p\left(\mathbf{x};\theta\right)\right)}{p\left(\mathbf{x};\theta^0\right)}d\theta,\\ \Gamma_h\left(\tau\right) &= \int_{\Theta} h\left(\tau,\theta\right)k\left(\theta,t_\theta\right)\left[g\left(\gamma\left(\theta,t_\theta\right)\right)-g\left(\theta^0\right)\right]d\theta.\end{aligned} \tag{11}$$

At first sight, the proposed rationale does not seem appealing, since a non-linear transformation of type (9) is unlikely to exist whatever the form of the p.d.f., although the linear transformation of the LR (6) is always possible. Fortunately, it is applicable to a subset of M-dimensional complex circular Gaussian p.d.f.:

$$p\left(\mathbf{x};\boldsymbol{\theta}\right) = p\left(\mathbf{x};\mathbf{m}\left(\boldsymbol{\theta}\right),\mathbf{C}\left(\boldsymbol{\theta}\right)\right) = \frac{e^{-\left(\mathbf{x}-\mathbf{m}\left(\boldsymbol{\theta}\right)\right)^H\mathbf{C}\left(\boldsymbol{\theta}\right)^{-1}\left(\mathbf{x}-\mathbf{m}\left(\boldsymbol{\theta}\right)\right)}}{\pi^M\left|\mathbf{C}\left(\boldsymbol{\theta}\right)\right|}$$

Indeed, the transformation $t_q\left(y\right) = y^q$ can be applied to the observation model resulting from a mixture of deterministic and stochastic signals in presence of Gaussian interference [5]. In this case $\mathbf{m}\left(\boldsymbol{\theta}\right) = \mathbf{m}\left(\boldsymbol{\varepsilon}\right)$, $\mathbf{C}\left(\boldsymbol{\theta}\right) = \boldsymbol{\Psi}\left(\boldsymbol{\zeta}\right)\mathbf{C_s}\boldsymbol{\Psi}\left(\boldsymbol{\zeta}\right)^H + \mathbf{C_n}$, $\boldsymbol{\theta} = \left[\boldsymbol{\varepsilon}^T,\boldsymbol{\zeta}^T,vec\left(\mathbf{C_s}\right)^T,vec\left(\mathbf{C_n}\right)^T\right]^T$.

3. CONJECTURAL LOWER BOUNDS

Although initially introduced by resorting to the covariance inequality, the Bayesian bounds of the Weiss-Weinstein family have been lately revisited by authors in [6] who have shown that these bounds are also solutions of a norm minimization under linear constraints (see [6]§III.B) analogous to the one introduce in section 2. Therefore any deterministic lower bounds have a corresponding Bayesian bound: Cramér-Rao bound, Bhattacharyya bound, Hammersley-Chapman-Robbins bound, ... Our idea is to argue from analogy from the Bayesian bounds towards the deterministic bounds to explore new possible bounds. As an example, in the case of a single

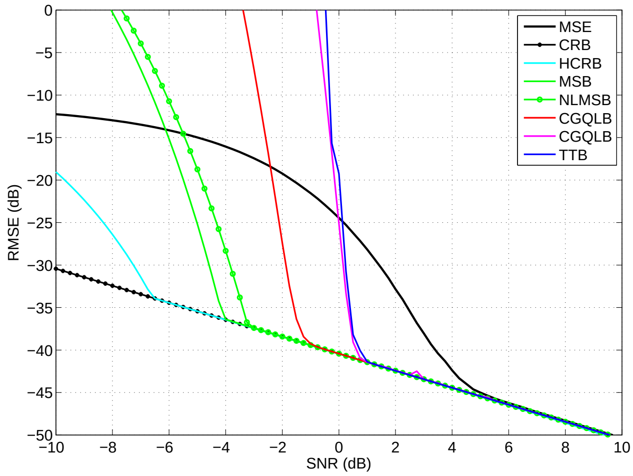


Fig. 1. Comparison of MSE lower bounds versus SNR $(M = 8, \theta = 0)$

unknown parameter θ - for sake of simplicity -, the Bayesian Weiss-Weinstein bound is associated with the linear constraint:

$$\begin{aligned}&\iint_{\Theta,\Omega}\left(\widehat{\theta}\left(\mathbf{x}\right)-\theta\right)\left[\frac{p\left(\mathbf{x},\theta+\delta\right)^q}{p\left(\mathbf{x},\theta\right)^q}-\frac{p\left(\mathbf{x},\theta-\delta\right)^{1-q}}{p\left(\mathbf{x},\theta\right)^{1-q}}\right]p\left(\mathbf{x},\theta\right)d\mathbf{x}d\theta\\ &\qquad = -\delta\iint_{\Theta,\Omega}\frac{p\left(\mathbf{x},\theta-\delta\right)^{1-q}}{p\left(\mathbf{x},\theta\right)^{1-q}}p\left(\mathbf{x},\theta\right)d\mathbf{x}d\theta\end{aligned}$$

where $p\left(\mathbf{x},\theta\right) = p\left(\mathbf{x}\mid\theta\right)p\left(\theta\right) = p\left(\mathbf{x};\theta\right)p\left(\theta\right)$, $q\in\left[0,1\right]$. The corresponding linear constraint for deterministic estimation is (drawn from examples in [6]§III.B):

$$\begin{aligned}&\int_{\Omega}\left(\widehat{\theta^0}\left(\mathbf{x}\right)-\theta^0\right)\left[\frac{p\left(\mathbf{x};\theta^0+\delta\right)^q}{p\left(\mathbf{x};\theta^0\right)^q}-\frac{p\left(\mathbf{x};\theta^0-\delta\right)^{1-q}}{p\left(\mathbf{x};\theta^0\right)^{1-q}}\right]p\left(\mathbf{x};\theta^0\right)d\mathbf{x}\\ &\qquad = -\delta\int_{\Omega}\frac{p\left(\mathbf{x};\theta^0-\delta\right)^{1-q}}{p\left(\mathbf{x};\theta^0\right)^{1-q}}p\left(\mathbf{x};\theta^0\right)d\mathbf{x}\end{aligned} \tag{12}$$

leading to the deterministic Weiss-Weinstein bound (WWB):

$$MSE_{\theta^0}\left[\widehat{\theta^0}\right] \geq \sup_{q,\delta}\left\{\frac{\delta^2 E_{\theta^0}\left[\frac{p\left(\mathbf{x};\theta^0-\delta\right)^{1-q}}{p\left(\mathbf{x};\theta^0\right)^{1-q}}\right]^2}{E_{\theta^0}\left[\left(\frac{p\left(\mathbf{x};\theta^0+\delta\right)^q}{p\left(\mathbf{x};\theta^0\right)^q}-\frac{p\left(\mathbf{x};\theta^0-\delta\right)^{1-q}}{p\left(\mathbf{x};\theta^0\right)^{1-q}}\right)^2\right]}\right\} \tag{13}$$

The WWB (13) is a bound for estimators satisfying (12). The conjecture is that unbiased estimators satisfy (12) as well. It is true where $q = 0$ or $q = 1$ since then (12) amounts to the Hammersley-Chapman-Robbins constraint. Unfortunately so far, we have not been able to prove that (12) derives from the mixture of integral linear and non-linear transformations of the unbiasedness constraint. And yet, but nevertheless simulations performed for the single tone threshold analysis clearly shows that the WWB (13) is a very tight bound for unbiased estimators; at least in this application case.

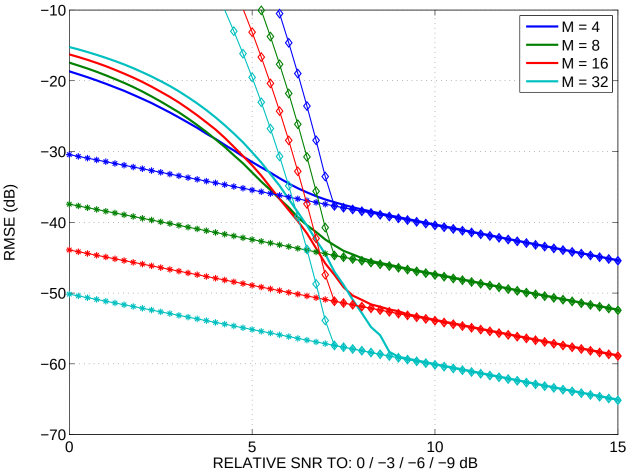


Fig. 2. MSE, CRB and WWB (Sup over $q \in [0,1]$) versus SNR

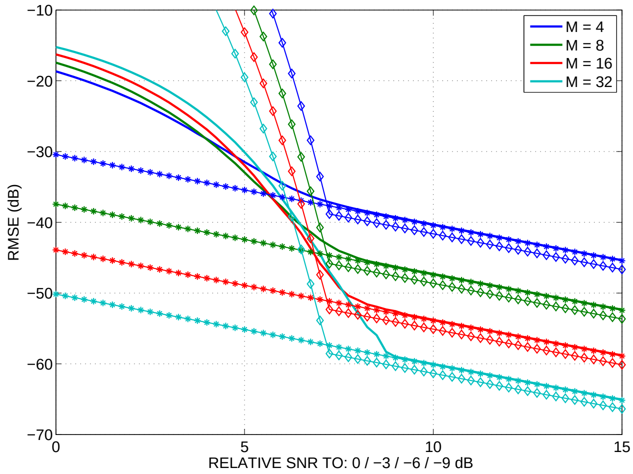


Fig. 3. MSE, CRB and WWB $\left(q = \frac{1}{2}\right)$ versus SNR

4. CONCLUSION: SINGLE TONE THRESHOLD ANALYSIS

A reference problem in threshold analysis is the estimation of a single tone $\theta \in \left]-0.5, 0.5\right[$ for a deterministic observation model:

$$\mathbf{x} = a\boldsymbol{\psi}\left(\theta^0\right) + \mathbf{n}, \quad \boldsymbol{\psi}(\theta) = \left[1, ..., e^{j(M-1)2\pi\theta}\right]^T \\ p\left(\mathbf{x};\theta\right) = \frac{e^{-\|\mathbf{x}-a\boldsymbol{\psi}(\theta)\|^2}}{\pi^M} \quad (14)$$

where a^2 is the known SNR $(a > 0)$ and $\mathbf{n}$ is a complex circular Gaussian noise, with zero mean and a known covariance matrix $\mathbf{C_n} = \mathbf{I}$. In the simulations:

- $\delta \in \left]0, 0.5\right[$, $\widehat{\theta}_{ML} = \max_{\theta} \left\{\text{Re}\left[\boldsymbol{\psi}(\theta)^H \mathbf{x}\right]\right\}$.
- the HCRB [2] is the simplest approximation of the BB (5) based on 2 test-points $\boldsymbol{\theta}^2 = \left(\theta^0, \theta^0 + \delta\right)^T$ + supremum on δ,
- the MSB [2] is the simplest approximation of the BB based on 3 test-points $\boldsymbol{\theta}^3 = \left(\theta^0, \theta^0 + \delta, \theta^0 - \delta\right)^T$ + supremum on δ,
- the NLMSB [5] is the nonlinear generalisation (10) of the MSB based on 3 test-points + supremum on δ and $q \in \left]0.5, 2\right[$,
- the CGQLB [2] is the generalization of the CRB based on 3 test-points $\boldsymbol{\theta}^3 = \left(\theta^0, \theta^0 + \delta, \theta^0 - \delta\right)^T$ + supremum on δ,
- the TTB [3] is the combination of CRB$\left(\theta^0\right)$ and of (5) where $N = 1024$, $K = 32$ and $\mathbf{H}_K$ is an ad hoc submatrix of FFT matrix of dimension N.

All these lower bounds are displayed on figure (1) and compared with the MSE of the MLE estimator (5×10^5 trials) for $M = 8$ and $\theta^0 = 0$. The first occurrence of the CGQLB is obtained for δ lying on a discretization of $\left]-0.5, 0.5\right[$ with a step of $1/1024$. The second one is obtained for a step of $1/\left(1024 * 128\right)$. The purpose of the 2 cases is to show that it is generally difficult to compare tightness of bounds which are based on subsets of constraints that are not included one in each other. For each bound, tightness may depend on specific optimization parameters.
Additionally, the tightness of CGQLB and TTB (or any existing bound) could be improved by updating their associated linear constraints with the non-linear transformation (10) as we did for the MSB, which is a topic for future work.

Finally, a much more impartial criterion could be the computation time (and possibly the memory load).
Nevertheless, all these bounds seem to provide a still too coarse prediction of the SNR threshold value (underestimated by at least 5 dB), an imperfection mostly compensated by the WWB (13) as shown on figure (2). This figure clearly shows that the WWB is not only a lower bound for unbiased estimators whatever the value of M (also checked for $M = 2, 64, 128$), but it is an extremely tight lower bounds, far tighter that all the existing ones.
Moreover, the deterministic WWB seems to share the same property as its Bayesian analogue one, i.e. to be nearly the tightest for $q = 0.5$ as shown on figure (3). Under that form, the WWB is as simple to implement as the HCRB.
Such a simple and tight bound really deserves to be derived!

5. REFERENCES

[1] E.W. Barankin, ”Locally best unbiased estimates”, Ann. Math. Stat., vol. 20, no. 4, pp. 477-501, 1949.

[2] E. Chaumette, J. Galy, A. Quinlan, P. Larzabal, ”A New Barankin Bound Approximation for the Prediction of the Threshold Region Performance of Maximum-Likelihood Estimators”, IEEE Trans. on SP, vol. 56, no. 11, pp. 5319-5333, Nov. 2008

[3] K. Todros and J. Tabrikian, ”A new lower bound on the mean-square error of unbiased estimators”, in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., pp. 3913-3916, 2008

[4] H.L. Van Trees, *Detection, Estimation and Modulation Theory, Part 1*. New York: Wiley, 1968.

[5] E. Chaumette, A. Renaux, P. Larzabal, ”Lower bounds on the mean square error derived from mixture of linear and non-linear transformations of the unbiasness definition”, in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., pp. 3045-3048, 2009

[6] A. Renaux, P. Forster, P. Larzabal, C.D. Richmond, A. Nehorai, ”A Fresh Look at the Bayesian Bounds of the Weiss-Weinstein Family”, IEEE Trans. on SP, vol. 56, no. 11, pp. 5334-5352, Nov. 2008

L. High resolution techniques for radar: myth or reality ? (EUSIPCO)
in Proc. Eurasip EUSIPCO, 2013.

HIGH RESOLUTION TECHNIQUES FOR RADAR: MYTH OR REALITY ?

Chengfang Ren [(1)], *Jerome Galy*[(2)], *Eric Chaumette*[(3,4)], *Pascal Larzabal*[(4)] *and Alexandre Renaux*[(1)]

[(1)] Universite Paris-Sud/LSS
3, Rue Joliot-Curie, 91192 Gif-sur-Yvette, France. Emails: cren@lss.supelec.fr, renaux@lss.supelec.fr
[(2)] Universite de Montpellier 2/LIRMM
161 rue Ada 34392 Montpellier Cedex 5, France. Email: galy@lirmm.fr
[(3)] ONERA/The French Aerospace Lab
Chemin de la Huniere Palaiseau Cedex 91123, France. Email: eric.chaumette@onera.fr
[(4)] Ecole Normale de Cachan/SATIE
61 av. du President Wilson, 94235 Cachan cedex, France. Email: pascal.larzabal@satie.ens-cachan.fr

ABSTRACT

We address the problem of effectiveness of the high resolution techniques applied to the conditional model. The rationale is based on a definition of the probability of resolution of maximum likelihood estimators which is computable in the asymptotic region of operation (in SNR and/or in large number of snapshots). The application case is the multiple tones estimation problem (Doppler frequencies estimation in radar).

Index Terms— high resolution techniques, maximum likelihood estimators, Cramer-Rao bound, multiple tones

1. INTRODUCTION

The resolvability of closely spaced signals, in terms of parameter of interest, for a given scenario (e.g., for a given Signal-to-Noise Ratio (SNR), for a given number of snapshots and/or for a given number of sensors) is a former and challenging problem which was recently updated by Smith [1, ref.12], Shahram and Milanfar [1, ref.13], Liu and Nehorai [1, ref.14], Amar and Weiss [1, ref.15] and El Korso et al [1]. Historically, the concept of Statistical Resolution Limit (SRL) has been introduced as the minimum distance between two closely spaced signals embedded in an additive noise that allows a correct resolvability/parameter estimation. The SRL is therefore an important statistical tool to quantify the ultimate performance for parametric estimation problems. Lately authors in [1] have generalized the concept of the SRL to the Multidimensional SRL (MSRL) applied to the multidimensional harmonic retrieval model. In that paper, they derive the SRL for the so-called multidimensional harmonic retrieval model by using a generalization of the previously introduced SRL concepts called Multidimensional SRL (MSRL). They first derive the MSRL using an hypothesis test approach (Liu and Nehorai). This statistical test is shown to be asymptotically an uniformly most powerful test which is the strongest optimality statement that one could expect to obtain. Second, they link the proposed asymptotic MSRL based on the hypothesis test approach to a new extension of the SRL based on the Cramér-Rao Bound approach (Smith). Thus, a closed-form expression of the asymptotic MSRL is given and analyzed in the framework of the multidimensional harmonic retrieval model. In the present paper we propose a different rationale to address the problem of resolvability of closely spaced signals, in terms of parameter of interest. It is based on a definition of the probability of resolution of maximum likelihood estimators (MLEs) which is computable in the asymptotic region of operation (in SNR and/or in large number of snapshots) for the conditional model. The results obtained with the proposed rationale must be regarded as an "upper bound" in terms of resolvability, in the sense it assumes that the number of source is known and that all the sources are present as well. The application case is the multiple tones estimation problem (Doppler frequencies estimation in radar).

This work has been partly funded by the European Network of excellence NEWCOM#

2. PROBABILITY OF RESOLUTION

Throughout the present paper, unless otherwise stated, $\mathbf{x}$ denotes the random observation vector of dimension N, Ω denotes the observations space and $L^2(\Omega)$ denotes the complex Hilbert space of square integrable functions over Ω. The probability density function (p.d.f.) of $\mathbf{x}$ is denoted $p(\mathbf{x};\mathbf{\Theta})$ and depends on a vector of $\overline{P}$ real parameters $\mathbf{\Theta} = (\theta_1, \ldots, \theta_{\overline{P}}) \in \Phi$, where Φ denotes the parameter space. The probability of an event $\mathcal{D} \subset \Omega$ is denoted $\mathcal{P}(\mathcal{D};\mathbf{\Theta})$. Let $\mathbf{\Theta}^0$ be a selected value of the parameter $\mathbf{\Theta}$, and $\widehat{\mathbf{g}(\mathbf{\Theta}^0)}(\mathbf{x})$ ($\widehat{\mathbf{g}(\mathbf{\Theta}^0)}$ in abbreviated form) an estimator of $\mathbf{g}(\mathbf{\Theta}^0)$ where $\mathbf{g}(\mathbf{\Theta}) = \left(g_1(\mathbf{\Theta}), \ldots, g_{\overline{Q}}(\mathbf{\Theta})\right)^T$ is a vector of $\overline{Q}$ real-valued (for the sake of simplicity) functions of $\mathbf{\Theta}$. For any selected value $\mathbf{\Theta}^0$, $\widehat{\mathbf{g}\left(\mathbf{\Theta}^0\right)}(\mathbf{x})$ stands for a mapping of the observation space Ω into an estimate of $\mathbf{g}\left(\mathbf{\Theta}^0\right)$.

2.1. Estimation precision and bounds

The quality (i.e. the precision) of an estimator $\widehat{\mathbf{g}\left(\mathbf{\Theta}^0\right)}$ can be measured using the following canonical objective function:

$$\mathcal{O}_{\mathbf{\Theta}^0}\left(\widehat{\mathbf{g}\left(\mathbf{\Theta}^0\right)}, \boldsymbol{\xi}^-, \boldsymbol{\xi}^+\right) = \mathcal{P}\left(\bigcap_{q=1}^{\overline{Q}}\left(\widehat{g_q\left(\mathbf{\Theta}^0\right)} \in \left]g_q\left(\mathbf{\Theta}^0\right) - \xi_q^-, g_q\left(\mathbf{\Theta}^0\right) + \xi_q^+\right[\right); \mathbf{\Theta}^0\right) \quad (1)$$

where $\boldsymbol{\xi}^- = \left(\xi_1^-, \ldots, \xi_{\overline{Q}}^-\right)^T$ and $\boldsymbol{\xi}^+ = \left(\xi_1^+, \ldots, \xi_{\overline{Q}}^+\right)^T$ define the (left and right) errors on the estimation of $\mathbf{g}\left(\mathbf{\Theta}^0\right)$ and $\mathcal{O}_{\mathbf{\Theta}^0}\left(\widehat{\mathbf{g}\left(\mathbf{\Theta}^0\right)}, \boldsymbol{\xi}^-, \boldsymbol{\xi}^+\right)$ is a measure of the probability that errors does not exceed $\boldsymbol{\xi}^-$ and $\boldsymbol{\xi}^+$. This objective function is identified as "canonical" since it is deduced naturally from the problem under study: the match between the observations of a random vector and a deterministic vector of interest. We also qualify it as *exhaustive*, in the sense that it incorporates all the available information on the problem, in other words the probabilities. Consequently, we consider that (1) defines the *exhaustive* precision (of estimation). Nevertheless, it is more fruitful practically to consider the *quasi-exhaustive* precision obtained when $\boldsymbol{\xi}^- = \boldsymbol{\xi}^+ = \boldsymbol{\xi}$, then $\mathcal{O}_{\mathbf{\Theta}^0}\left(\widehat{\mathbf{g}\left(\mathbf{\Theta}^0\right)}, \boldsymbol{\xi}^-, \boldsymbol{\xi}^+\right)$ (1) reduces to $\mathcal{O}_{\mathbf{\Theta}^0}\left(\widehat{\mathbf{g}\left(\mathbf{\Theta}^0\right)}, \boldsymbol{\xi}\right)$ defined as:

$$\mathcal{O}_{\mathbf{\Theta}^0}\left(\widehat{\mathbf{g}\left(\mathbf{\Theta}^0\right)}, \boldsymbol{\xi}\right) = \mathcal{P}\left(\bigcap_{q=1}^{\overline{Q}}\left(\left|\widehat{g_q\left(\mathbf{\Theta}^0\right)} - g_q\left(\mathbf{\Theta}^0\right)\right| < \xi_q\right); \mathbf{\Theta}^0\right) \quad (2)$$

and is bounded by:

$$\mathcal{P}\left(\sum_{q=1}^{\overline{Q}} \frac{\left(\widehat{g_q\left(\mathbf{\Theta}^0\right)} - g_q\left(\mathbf{\Theta}^0\right)\right)^2}{\xi_q^2} < 1; \mathbf{\Theta}^0\right) \leq \mathcal{O}_{\mathbf{\Theta}^0}\left(\widehat{\mathbf{g}\left(\mathbf{\Theta}^0\right)}, \boldsymbol{\xi}\right) \leq \mathcal{P}\left(\sum_{q=1}^{\overline{Q}} \frac{\left(\widehat{g_q\left(\mathbf{\Theta}^0\right)} - g_q\left(\mathbf{\Theta}^0\right)\right)^2}{\xi_q^2} < \overline{Q}; \mathbf{\Theta}^0\right) \quad (3)$$

where $\sum_{q=1}^{\overline{Q}} \frac{\left(\widehat{g_q\left(\mathbf{\Theta}^0\right)} - g_q\left(\mathbf{\Theta}^0\right)\right)^2}{\xi_q^2}$ is a weighted total square error. In the following, for the sake of legibility, we focus on the case where $\mathbf{g}\left(\mathbf{\Theta}\right) = \mathbf{\Theta}$, and $\mathbf{\Theta}^0$ (respectively $\widehat{\mathbf{\Theta}^0}$) is denoted $\mathbf{\Theta}$ (respectively $\widehat{\mathbf{\Theta}}$) wherever it is unambiguous.

2.2. Probability of resolution and bounds

Thus we consider a parameter estimation problem where the parameters of interest are the vectors $\{\boldsymbol{\theta}^m\}_{m=1}^M$, where $\boldsymbol{\theta} \in \mathbb{R}^P$ and $\boldsymbol{\theta}^m \neq \boldsymbol{\theta}^l$, $\forall l \neq m \in [1, M]$. Then $\mathbf{\Theta}^T = \left(\left(\boldsymbol{\theta}^1\right)^T, \ldots, \left(\boldsymbol{\theta}^M\right)^T\right)$ $(\overline{P} = PM, \overline{Q} = \overline{P})$, $\boldsymbol{\xi}^T = \left(\boldsymbol{\varepsilon}^1, \ldots, \boldsymbol{\varepsilon}^M\right)$ where $\boldsymbol{\varepsilon}^m = \left(\varepsilon_1^m, \ldots, \varepsilon_P^m\right)$. Let $\mathcal{C}^m$ be the hypercube with centre $\boldsymbol{\theta}^m$ defined by $\mathcal{C}^m\left(\boldsymbol{\varepsilon}^m\right) = \left\{\boldsymbol{\theta} : \bigcap_{p=1}^{P}\left|\theta_p - \theta_p^m\right| < \varepsilon_p^m\right\}$. We define the probability of resolvability (of vectors of multiple parameters $\boldsymbol{\theta}$) with precision $\boldsymbol{\xi}$ as the probability $\mathcal{O}_{\mathbf{\Theta}}\left(\widehat{\mathbf{\Theta}}, \boldsymbol{\xi}\right)$ (2) when $\widehat{\boldsymbol{\theta}^m} \in \mathcal{C}^m\left(\boldsymbol{\varepsilon}^m\right), \forall m \in [1, M]$, and the hypercubes are disjoint:

$$\mathcal{C}^m\left(\boldsymbol{\varepsilon}^m\right) \cap \mathcal{C}^l\left(\boldsymbol{\varepsilon}^l\right) = \varnothing, \ \forall l \neq m \in [1, M]. \quad (4)$$

In other words, we do not consider as successful a trial leading to at least one $\widehat{\boldsymbol{\theta}^m}$ outside $\mathcal{C}^m\left(\boldsymbol{\varepsilon}^m\right)$. The underlying idea is that estimates switch among hypercubes $\mathcal{C}^m\left(\boldsymbol{\varepsilon}^m\right)$ is not allowed. Parameters vector $\{\boldsymbol{\theta}^m\}_{m=1}^M$ will be said "resolved" by estimators $\widehat{\mathbf{\Theta}}$ if:

$$0.9 \leq \mathcal{O}_{\mathbf{\Theta}}\left(\widehat{\mathbf{\Theta}}, \boldsymbol{\xi}\right) \text{ s.t. (4)} \leq 0.99 \quad (5)$$

2.3. Gaussian p.d.f.

The (lower and upper) bounds on $\mathcal{O}_{\mathbf{\Theta}}\left(\widehat{\mathbf{\Theta}}, \boldsymbol{\xi}\right)$ given by (3) are particularly convenient when $\widehat{\mathbf{\Theta}}(\mathbf{x}) - \mathbf{\Theta} \sim \mathcal{N}\left(\mathbf{b}\left(\mathbf{\Theta}\right), \mathbf{C}\left(\mathbf{\Theta}\right)\right)$, that is $\widehat{\mathbf{\Theta}}(\mathbf{x})$ is a Gaussian estimator of $\mathbf{\Theta}$ with bias vector $\mathbf{b}\left(\mathbf{\Theta}\right)$ and covariance matrix $\mathbf{C}\left(\mathbf{\Theta}\right)$. Then a straightforward linear transformation of the Gaussian random vector yields that (3) is equivalent to :

$$\mathcal{P}\left(e\chi_Q^2\left(\delta\left(\mathbf{\Theta}\right), \boldsymbol{\sigma}^2\left(\mathbf{\Theta}\right)\right) < 1\right) \leq \mathcal{O}_{\mathbf{\Theta}}\left(\widehat{\mathbf{\Theta}}, \boldsymbol{\xi}\right) \leq \mathcal{P}\left(e\chi_Q^2\left(\delta\left(\mathbf{\Theta}\right), \boldsymbol{\sigma}^2\left(\mathbf{\Theta}\right)\right) < Q\right)$$

where $\delta\left(\mathbf{\Theta}\right) = \left\|\mathbf{M}^T\left(\mathbf{\Theta}\right)\mathbf{D}_{\boldsymbol{\varepsilon}}^{-1}\mathbf{b}\left(\mathbf{\Theta}\right)\right\|^2$, $\mathbf{D}_{\boldsymbol{\varepsilon}}^{-1}\mathbf{C}\left(\mathbf{\Theta}\right)\mathbf{D}_{\boldsymbol{\varepsilon}}^{-1} = \mathbf{M}\left(\mathbf{\Theta}\right)\mathbf{D}_{\boldsymbol{\sigma}^2\left(\mathbf{\Theta}\right)}\mathbf{M}^T\left(\mathbf{\Theta}\right)$, $\left(\mathbf{D}_{\boldsymbol{\alpha}}\right)_{q,p} = \alpha_q \delta_p^q$ and $e\chi_Q^2\left(\delta, \boldsymbol{\sigma}^2\right)$ is a non-central quadratic form [3], that is an extension of non-central chi-square with corresponding degrees of freedom in Q and positive noncentrality parameters in δ where the power of each component is not constant:

$$e\chi_Q^2\left(\delta = \sum_{q=1}^{Q}\delta_q, \boldsymbol{\sigma}^2\right) \sim \sum_{q=1}^{Q}\sigma_q^2\left|z_q + \sqrt{\delta_q}\right|^2$$
$$\boldsymbol{\sigma}^2 = \left(\sigma_1^2, \ldots, \sigma_Q^2\right)^T, \mathbf{z} = \left(z_1, \ldots, z_Q\right)^T \sim \mathcal{N}\left(\mathbf{0}, \mathbf{I}\right)$$

If $\delta = 0$ (unbiased estimates) then $u \sim e\chi_Q^2\left(0, \boldsymbol{\sigma}^2\right)$ and:

$$p\left(u\right) = \int \prod_{q=1}^{Q}\left(1 + j2\pi f\sigma_q^2\right)^{\frac{-1}{2}} e^{j2\pi f u} du = \frac{{}_0F^0\left(-\frac{1}{2}\mathbf{D}_{\boldsymbol{\sigma}^2}^{-1}, u\right)}{2^{\frac{Q}{2}}\Gamma\left(\frac{Q}{2}\right)\left|\mathbf{D}_{\boldsymbol{\sigma}^2}\right|}$$

where ${}_0F^0\left(\ \right)$ is a generalized hypergeometric function [3].

3. ASYMPTOTIC PERFORMANCE OF CONDITIONAL MODEL

Historically the first MSE lower bound for deterministic parameters to be derived was the Cramér-Rao Bound (CRB),

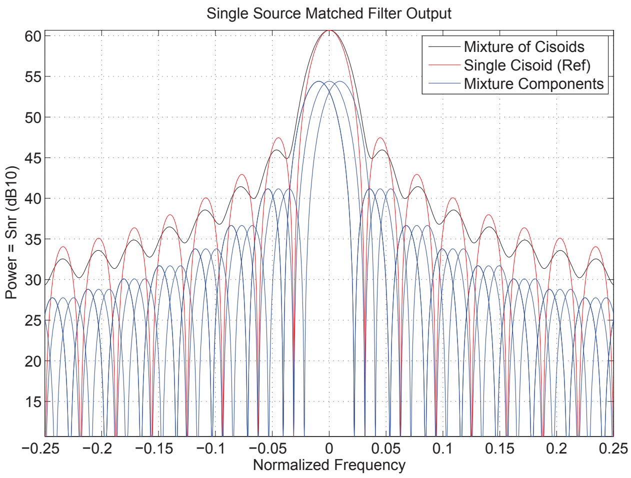


Fig. 1. Single source matched filter output: 3 cisoids separated by $\Delta\theta_{3dB} \times 0.3$

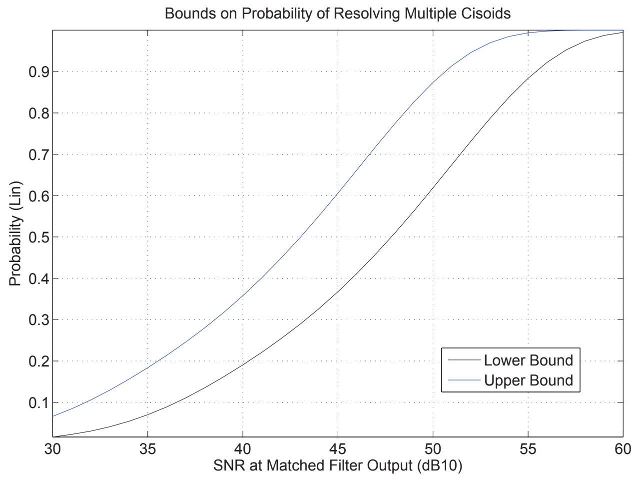


Fig. 2. Bounds on probability of resolving multiple cisoids: 3 cisoids separated by $\Delta\theta_{3dB} \times 0.3, T = 1$

which was introduced to investigate fundamental limits of a parameter estimation problem or to assess the relative performance of a specific estimator (efficiency) [2]. It has since become the most popular lower bound due to its simplicity of calculation, the fact that in many cases it can be achieved asymptotically (high SNR [4] and/or large number of snapshots [2]) by Maximum Likelihood Estimators (MLE). This initial characterization of locally unbiased estimators has been significantly generalized by Barankin work [5], who established the general form of the highest lower bound on MSE (BB) for uniformly unbiased estimates, but unfortunately with a generally incomputable analytic solution. Therefore, since then, numerous works detailed in [5] have been devoted to deriving computable approximations of the BB and have shown that the CRB and the BB can be regarded as key representatives of two general classes of bounds, respectively the Small-Error bounds and the Large-Error bounds. These works have also shown that in non-linear estimation problems three distinct regions of operation can be observed. In the asymptotic region, the MSE is small and, in many cases, close to the Small-Error bounds. In the *a priori* performance region where the number of independent snapshots and/or the SNR are very low, the observations provide little information and the MSE is close to that obtained from the prior knowledge about the problem. Between these two extremes, there is an additional ambiguity region, also called the transition region. In this region, the MSE of MLEs usually deteriorates rapidly with respect to Small-Error bounds and exhibits a threshold behaviour corresponding to a ”performance breakdown”. The nature of this phenomenon is specified by a complicated non-smooth behaviour of the likelihood function in the ”threshold” area where it tends to generate outliers [2]. Small-Error bound such as the CRB are not able to handle the threshold phenomena, whereas it is revealed by Large-Error bounds that can be used to predict the threshold value. Therefore, provided that one keeps in mind the CRB limitations, that is, to become an excessively optimistic lower bound when the observation conditions degrade (low SNR and/or low number of snapshots), the CRB is still a lower bound of great interest for system analysis and design in the asymptotic region.

3.1. Asymptotic performance of radar conditional model

The choice of focusing on the (Gaussian) conditional model comes from our primary interest for active systems such as radar (or sonar) where a known waveform is transmitted, and the signals scattered from the targets of interest are used to estimate their parameters. Typically, the received signals are modelled as scaled, delayed, and Doppler-shifted versions of the transmitted signal. Estimation of the time delay and Doppler shift provides information about the range and radial velocity of the targets. The use of spatial diversity, i.e. antenna arrays, compared with a single sensor, guarantees more accurate range and velocity estimation and allows estimation of the targets direction. Last, but no least, waveform diversity may be used to improve the estimation of all targets parameters. In an active system, as the waveform parametric model is known and deterministic (in opposition with a passive system where a probabilistic modelling of the waveform is generally considered), the most accurate statistical prediction for an observation will be obtained when considering the signal amplitudes as deterministic (since it is well known that the complex Gaussian amplitude modelling provide an average unconditional CRB higher that the corresponding conditional CRB [2]). The asymptotic (in SNR and/or in large number of

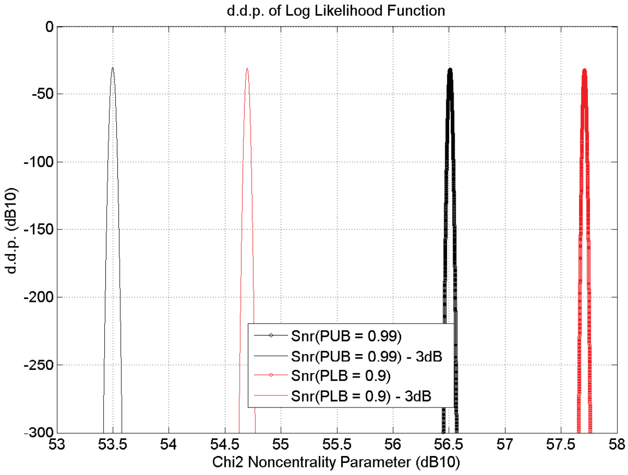


Fig. 3. Loglikelihood p.d.f. at limits of SNR interval allowing resolution of: 3 cisoids separated by $\Delta\theta_{3dB} \times 0.3, T = 1$

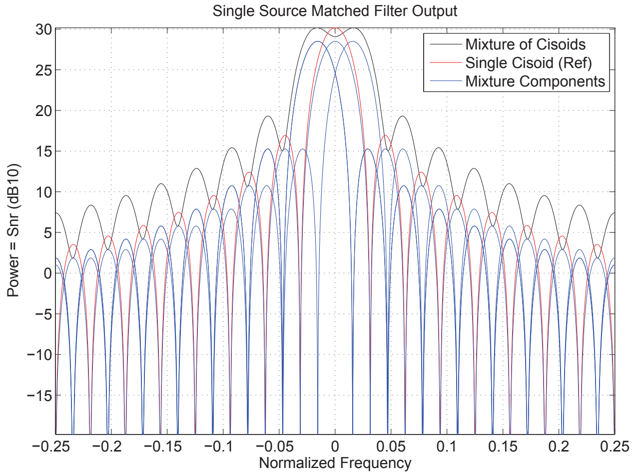


Fig. 4. Single source matched filter output: 3 cisoids separated by $\Delta\theta_{3dB} \times 0.5$

snapshots) Gaussianity and efficiency of CMLEs (conditional MLEs) in the multiple parameters case has been proved under the assumption that the maximum of the (reduced) log likelihood function belongs to its main lobe. As an example, let us consider the general linear observation model:

$$\mathbf{x}_t\left(\mathbf{\Theta}^0\right) = \mathbf{A}\left(\mathbf{\Theta}^0\right)\mathbf{s}_t + \mathbf{n}_t, \quad t \in [1, T]$$

where T is the number of independent observation, M is the number of signal sources, $\mathbf{s}_t = (s_{t,1}, \ldots, s_{t,M})^T$ is the vector of complex amplitudes of the M sources for the t^{th} observation, $\mathbf{A}(\mathbf{\Theta}) = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_M)]$ where $\mathbf{\Theta} = (\theta_1, \ldots, \theta_M)^T$ and $\mathbf{a}(\,)$ is a vector of N parametric functions depending on a single parameter θ (for sake of simplicity), $\mathbf{n}_t$ are Gaussian complex circular independent noises with spatially white covariance matrix: $\mathbf{C_n} = \sigma_\mathbf{n}^2 \mathbf{I}_N$, independent from the M sources. Then the reduced log likelihood function $L\left(\mathbf{\Theta}; \mathbf{\Theta}^0\right)$ is given by [2]:

$$L\left(\mathbf{\Theta};\mathbf{\Theta}^0\right) = \frac{\sum_{t=1}^{T}\left\|\mathbf{\Pi}_{\mathbf{A}(\mathbf{\Theta})}\mathbf{x}_t\left(\mathbf{\Theta}^0\right)\right\|^2}{TM} \sim \mathcal{CX}_{MT}^2\left(F\left(\mathbf{\Theta};\mathbf{\Theta}^0\right), \frac{\sigma^2}{TM}\right)$$

$$F\left(\mathbf{\Theta};\mathbf{\Theta}^0\right) = \frac{\sum_{t=1}^{T}\left\|\mathbf{\Pi}_{\mathbf{A}(\mathbf{\Theta})}\mathbf{A}\left(\mathbf{\Theta}^0\right)\mathbf{s}_t\right\|^2}{TM}$$

where $F\left(\mathbf{\Theta}; \mathbf{\Theta}^0\right)$ is a generalized correlation function (aka generalized matched filter) and $\mathcal{CX}_K^2\left(\delta, \sigma^2\right)$ denotes a noncentral complex (circular) chi-square with corresponding degrees of freedom in K and positive noncentrality parameters in δ. Let $\widehat{\mathbf{\Theta}} \triangleq \widehat{\mathbf{\Theta}}(\mathbf{x}) = \arg\max\left\{L\left(\mathbf{\Theta}; \mathbf{\Theta}^0\right)\right\}$ denote the CMLE of $\mathbf{\Theta}$ and let $\Upsilon_{\mathbf{\Theta}^0}(\alpha) = \left\{\mathbf{\Theta} : \frac{F\left(\mathbf{\Theta};\mathbf{\Theta}^0\right)}{F\left(\mathbf{\Theta}^0;\mathbf{\Theta}^0\right)} > \alpha\right\}$ denote the main lobe at α $(0 < \alpha < 1)$. Then a condition of asymptotic region of operation for CMLE can be :

$$P\left(\mathbf{\Theta} \in image\left(\widehat{\mathbf{\Theta}}\right) \mid \mathbf{\Theta} \notin \Upsilon_{\mathbf{\Theta}^0}\left(\frac{1}{2}\right)\right) \approx 0 \qquad (6)$$

where $\Upsilon_{\mathbf{\Theta}^0}\left(\frac{1}{2}\right)$ is the usual main lobe at $-3dB$. The quasi-nullity of the probability of an outlier (6) can be demonstrated by computing the p.d.f. of $L\left(\mathbf{\Theta}^0; \mathbf{\Theta}^0\right)$ and $L\left(\mathbf{\Theta}; \mathbf{\Theta}^0\right)$ where $F\left(\mathbf{\Theta}; \mathbf{\Theta}^0\right) = \frac{1}{2}$ and by checking that their supports do not overlap above a certain p.d.f. threshold value, as small as possible (10^{-30} in the present paper). As the p.d.f. of $\mathcal{CX}_{MT}^2\left(F\left(\mathbf{\Theta}; \mathbf{\Theta}^0\right), \frac{\sigma^2}{TM}\right)$ is an increasing function in $F\left(\mathbf{\Theta}; \mathbf{\Theta}^0\right)$, it is sufficient to check that p.d.f. of $L\left(\mathbf{\Theta}^0; \mathbf{\Theta}^0\right)$ and $L\left(\mathbf{\Theta}; \mathbf{\Theta}^0\right)$ where $F\left(\mathbf{\Theta}; \mathbf{\Theta}^0\right) = \frac{1}{2}$ do not overlap to ensure that this property is valid for any $\mathbf{\Theta} \notin \Upsilon_{\mathbf{\Theta}^0}\left(\frac{1}{2}\right)$, what proves that $image\left(\widehat{\mathbf{\Theta}}\right) \subset \Upsilon_{\mathbf{\Theta}^0}\left(\frac{1}{2}\right)$. Then, in the asymptotic region [6]:

$$\widehat{\mathbf{\Theta}}(\mathbf{x}) \sim \mathcal{N}\left(\mathbf{\Theta}, \frac{\sigma_\mathbf{n}^2}{2T}\operatorname{Re}\left\{\mathbf{H}(\mathbf{\Theta}) \odot \widehat{\mathbf{R}}_\mathbf{s}^T\right\}^{-1}\right)$$

$$\widehat{\mathbf{R}}_\mathbf{s} = \sum_{t=1}^{T}\frac{\mathbf{s}_t\mathbf{s}_t^H}{T}, \mathbf{H}(\mathbf{\Theta}) = \frac{\partial\mathbf{A}(\mathbf{\Theta})}{\partial\theta}^H \mathbf{\Pi}_{\mathbf{A}(\mathbf{\Theta})}^\perp \frac{\partial\mathbf{A}(\mathbf{\Theta})}{\partial\theta}$$

and it has been proved that for each source [6]: the highest (worst) variance is obtained when the sources amplitudes are fully correlated and the lowest (best) variance is obtained when the sources amplitudes are uncorrelated.

3.2. Doppler frequency (multiple tones) estimation

As an example for radar, we consider the problem of Doppler frequency estimation which is a particular application case of the very general multiple tones estimation problem where :

$$\mathbf{a}(\theta) = \left[1, \ldots, e^{j2\pi n\theta}, \ldots, e^{j2\pi(N-1)\theta}\right]^T \qquad (7)$$

For sake of simplicity but without loss of generality, we consider only scenarios where the Doppler frequencies are equispaced with a step $d\theta = \Delta\theta_{3dB} \times \beta$ in order to take into

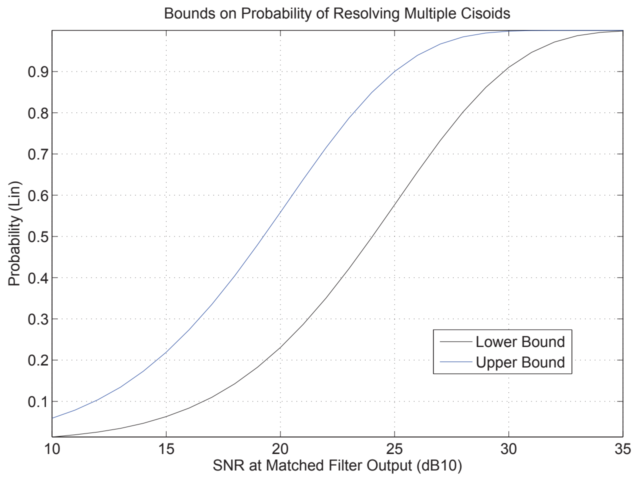


Fig. 5. Bounds on probability of resolving multiple cisoids: 3 cisoids separated by $\Delta\theta_{3dB} \times 0.5, T = 2$

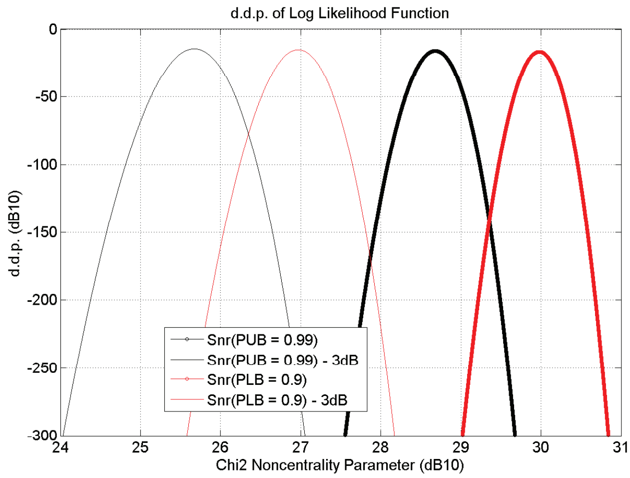


Fig. 6. Loglikelihood p.d.f. at limits of SNR interval allowing resolution of: 3 cisoids separated by $\Delta\theta_{3dB} \times 0.5, T = 2$

account an isotropic estimation error $\boldsymbol{\xi} = \frac{d\theta}{2}\mathbf{1}_M$ in the definition of probability of resolution (5), where $\mathbf{1}_M$ is a M-dimensional vector with components equal to 1. Additionally in all scenarios: $N = 32$ and the target amplitude are equal and therefore fully correlated $\mathbf{s}_t = \sqrt{\frac{SNR}{N}}\mathbf{1}_M$ (but it may not be the worst correlation case [6]). The main resolution features of each scenario are described with 3 figures:

• the output of the single source matched filter $\left(\frac{1}{N}\left\|\mathbf{a}\left(\boldsymbol{\Theta}\right)^H \mathbf{x}_t\left(\boldsymbol{\Theta}^0\right)\right\|^2\right)$ which could be the first step in a practical implementation of the CMLE (Clean algorithm, Alternating Projection algorithm).

• the probability (lower and upper) bounds (PLB and PUB) defined by (3) under (4) where $\boldsymbol{\xi} = \frac{d\theta}{2}\mathbf{1}_M$, as a function of the SNR computed at output of the single source matched filter. These bounds allow to determine the SNR interval containing the SNR from which the sources are resolved according to (5): SNR_{res}. Indeed : $SNR\left(PUB = 0.99\right) \leq SNR_{res} \leq SNR\left(PLB = 0.9\right)$.

• the p.d.f. of $L\left(\boldsymbol{\Theta}^0; \boldsymbol{\Theta}^0\right)$ and $L\left(\boldsymbol{\Theta}^0_{3dB}; \boldsymbol{\Theta}^0\right)$ for $SNR\left(PUB = 0.99\right)$ and $SNR\left(PLB = 0.9\right)$ to prove that within $\left[SNR\left(PUB = 0.99\right), SNR\left(PLB = 0.9\right)\right]$ the condition of asymptotic region of operation for CMLE is valid.

4. CONCLUSION

In the first scenario $SNR_{res} \in [54.4, 55.6]$ dB, which is a quite high required value to resolved a non demanding high resolution scenario of 3 targets ($d\theta = \Delta\theta_{3dB} \times 0.3$). This result suggests that high resolution techniques in operational radar system with a limited transmitted power will be rather a myth. On the other hand the addition of a second observation ($T = 2$) in the second scenario coupled with more spaced parameters values ($d\theta = \Delta\theta_{3dB} \times 0.5$) allows to decrease $SNR_{res} \in [28.5, 29.8]$. This result suggests that high resolution techniques in operational radar system can be a reality in some not too demanding scenarios provided a relevant waveform is transmitted.

5. REFERENCES

[1] M. N. El Korso, R. Boyer, A. Renaux and S. Marcos, "Statistical resolution limit for the multidimensional harmonic retrieval model: hypothesis test and Cramér-Rao Bound approaches", Journal on ASP, Special Issue, 2011

[2] H.L. Van Trees, *Optimum Array Processing*, New-York, Wiley-Interscience, 2002

[3] A.T. James, "Distributions of Matrix Variates and Latent Roots Derived from Normal Samples", Ann. Math. Statist. 35(2): 475-501, 1964

[4] A. Renaux, P. Forster, E. Chaumette, P. Larzabal, "On the High RSB CML Estimator Full Statistical Characterization", IEEE Trans. on SP, vol 54(12): 4840-4843, 2006

[5] K. Todros and J. Tabrikian, "General Classes of Performance Lower Bounds for Parameter Estimation-Part I: Non-Bayesian Bounds for Unbiased Estimators", IEEE Trans. on IT, vol. 56, no. 10, Oct. 2010.

[6] S. F. Yau and Y. Bresler, "Worst Case Cramér-Rao Bounds for Parametric Estimation of Superimposed Signals", IEEE Trans. on SP, vol. 40. no. 12, pp 2973-2986, Dec. 1992

REFERENCES

[Abe93] J.S. Abel, ”A bound on mean-square-estimate error”, IEEE Trans. on IT, 39(5): 1675-1680, 1993

[Alb73] J.P. Albuquerque, ”The Barankin Bound : a Geometric Interpretation”, IEEE Trans. on IT, 19(4): 559-561, 1973

[AW08] A. Amar and A. Weiss, “Fundamental limitations on the résolution of deterministic signals”, IEEE Trans. on SP, 56(11): 5309-5318, 2008

[Ath05] F. Athley, ”Threshold region performance of Maximum{Likelihood direction of arrival estimators”, IEEE Trans. on SP, 53(4): 1359-1373, 2005

[Bat46] A. Bhattacharyya, ”On some analogues of the amount of information and their use in statistical estimation,” Sankhya Indian J. Stat, 8: 1-14, 1946

[Bar49] E.W. Barankin, ”Locally best unbiased estimates”, Ann. Math. Stat., 20(4): 477-501, 1949

[BHE09] Z. Ben-Haim, Y. C. Eldar, ”On the Constrained Cramér-Rao Bound With a Singular Fisher Information Matrix”, IEEE SP Letters, 16(6): 453-456, 2009

[BHE10] Z. Ben-Haim, Y. C. Eldar, ”The Cramér-Rao Bound for Estimating a Sparse Parameter Vector”, IEEE Trans. on SP, 58(6): 3384-3389, 2010

[Ber99] N. Bertaux, P. Larzabal, C. Adnet et E. Chaumette; ”Maximum de vraisemblance paramétrée (PML) : application à la localisation spatio-temporelle radar”, Revue Traitement du Signal, 16(3): 187-201, 1999

[BTM69] W. R. Blischke, A. J. Truelove and P. B. Mundle, ”On non-regular estimation. I.Variance bounds for estimators of location parameters”, Journal of American Stat. Assoc., 64(327): 1056-1072, 1969

[Bou74] N. Bourbaki, *Elements of mathematics. Algebra: Algebraic structures. Linear algebra, 1*, Addison-Wesley (1974)

[BFL04-D] E. Boyer and P. Forster and P. Larzabal,”Non asymptotic performances analysis of beamforming for deterministic signals”, IEEE SP Letters, 11(1): 20-22, 2004

[BFL04-S] E. Boyer and P. Forster and P. Larzabal, ”Non asymptotic performances analysis of beamforming with stochastic signals”, IEEE SP Letters, 11(1): 23-25, 2004

[Bly74] C. R. Blyth, ”Necessary and sufficient conditions for inequalities of Cramér-Rao type,” Ann. Math. Stat., 2(3): 464-473, 1974

[Can13] C. Candan, ”Analysis and Further Improvement of Fine Resolution Frequency Estimation Method From Three DFT Samples”, IEEE SP Letters, 20(9): 913-916, 2013

[CR51] D.G. Chapman, H. Robbins, ”Minimum variance estimation without regularity assumptions”, Ann. Math. Stat., 22: 581-586, 1951

[Chau04] E. Chaumette, ”Contribution to the performance characterization of the joint Detection-Estimation problem”, PhD Thesis, Ecole Normale Supérieure de Cachan, France, Dec. 2004

[CLF05] E. Chaumette, P. Larzabal, P. Forster, ”On the Influence of a Detection Step on Lower Bounds for Deterministic Parameters Estimation”, IEEE Trans. on SP, 53(11): 4080-4090, 2005

[CL07C] E. Chaumette, P. Larzabal, ”Cramér-Rao Bound Conditioned by the Energy Detector”, IEEE SP Letters, 14(7): 477-480, 2007

[CL07M] E. Chaumette, P. Larzabal, ”The Statistical Prediction of monopulse Angle Measurement”, IET Proc. RSN, 1(5): 377-387, 2007

[CGVRL07] E. Chaumette, J. Galy, F. Vincent, A. Renaux and P. Larzabal, ”Mse lower bounds conditioned by the energy detector”, in Proc. Eurasip EUSIPCO, 2007

[CGVL07] E. Chaumette, J. Galy, F. Vincent and P. Larzabal, ”Computable Lower Bounds for Deterministic Parameter Estimation”, in Proc. IEEE CAMPSAP, 2007

[CGQL08] E. Chaumette, J. Galy, A. Quinlan, P. Larzabal, ”A New Barankin Bound Approximation for the Prediction of the Threshold Region Performance of Maximum-Likelihood Estimators”, IEEE Trans. on SP, 56(11): 5319-5333, 2008

[CRL09] E. Chaumette; A. Renaux; P. Larzabal, ”Lower bounds on the mean square error derived from mixture of linear and non-linear transformations of the unbiasness definition”, in Proc. IEEE ICASSP, 2009

[CRL10] E. Chaumette, A. Renaux, P.Larzabal, ”New trends in deterministic lower bounds and SNR threshold estimation: From derivable bounds to conjectural bounds”, in Proc. IEEE SAM, 2010

[Cra46] H. Cramér, *Mathematical Methods of Statistics*. Princeton, NJ: Princeton Univ. Press, 1946

[CS81] S.K. Chow, P.M. Schultheiss, ”Delay estimation using narrow-band processes”, IEEE Trans. on ASSP, 29(3): 478-484, 1981

[Dar45] G. Darmois, ”Sur les lois limites de la dispersion de certaines estimations”, Rev. Int. Stat, 13: 9-15, 1945

[DNYSF09] A. De Maio, S. De Nicola, H. Yongwei, Z. Shuzhong and A. Farina, ”Adaptive Detection and Estimation in the Presence of Useful Signal and Interference Mismatches”, IEEE Trans. SP, 57(2): 436-450, 2009

[EM77] B. Efron and C. Morris, ”Stein’s paradox in statistics”, Scientific American, 236(5): 119-127, 1977

[Eld06] Y.C. Eldar, “Uniformly improving the Cramer-Rao bound and maximum likelihood estimation”, IEEE Trans. on SP, 54(8): 2943-2956, 2006

[EKBRM11] M. N. El Korso, R. Boyer, A.Renaux and S. Marcos, ”Statistical résolution limit for the multidimensional harmonic retrieval model: hypothesis test and Cramér-Rao Bound approaches”, Journal on ASP, Special Issue, June 2011

[EKRBM13] M. N. El Korso, A.Renaux, R. Boyer and S. Marcos,”Deterministic Performance Bounds on the Mean Square Error for Near Field Source Localization”, IEEE Trans. SP, 61(4): 871-877, 2013

[FLV06] A. Féréol, P. Larzabal and M. Viberg, ” On the Asymptotic Performance Analysis of Subspace DOA Estimation in the Presence of Modeling Errors: Case of MUSIC”, IEEE Trans. on SP, 54(3): 907-920, 2006

[Fer11] A. Féréol, ”Traitement d’Antennes : Goniométrie - Séparation de sources”, Habilitation à Diriger des Recherches, Ecole Normale Supérieure de Cachan, 2011

[FRT02] A. Farina, B. Ristic, L. Timmoneri, ”Cramer-Rao bound for nonlinear filtering with Pd $<$ 1 and its application to target tracking”, IEEE Trans. on SP, 50(8): 1916-1924, 2002

[FLB04] P. Forster, P. Larzabal, E. Boyer, "Threshold Performance Analysis of Maximum Likelihood DOA Estimation", IEEE Trans. on SP, 52(11): 3183-3191, 2004

[FBL04] P. Forster, E. Boyer, P. Larzabal, "Nonefficiency of stochastic beamforming bearing estimates at high SNR and finite number of samples", IEEE SP Letters, 11(5): 509-512, 2004

[FL02] P. Forster, P. Larzabal, "On lower Bounds For Deterministic Parameter Estimation", in Proc. IEEE ICASSP, 2002

[FG52] D.A.S. Fraser and I. Guttman, "Bhattacharyya bounds without regularity assumptions," Ann. Math. Stat., 23(4): 629-632, 1952

[Fre43] M. Fréchet, "Sur l'extension de certaines évaluations statistiques au cas de petits échantillons", Rev. Int. Stat, 11: 182-205, 1943

[Fis21] R.A. Fisher, "On an absolute criterion for fitting frequency curves", Mess. of Math., 41: 155-160, 1912

[GCL10] J. Galy, E. Chaumette, P. Larzabal, "Joint Detection Estimation Problem of Monopulse Angle Measurement", IEEE Trans. on AES, 46(1): 397-413, 2010

[GL01] V. Girardin, N. Limnios, *Probabilités*, Vuibert "Les grands cours", Paris, mai 2001

[Gla72] F.E. Glave, "A new look at the Barankin Lower Bound", IEEE Trans. on IT, 18(3): 349-356, 1972

[GO90] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints", IEEE Trans. on IT, 36(6): 1285-1301, 1990

[Ham50] J.M. Hammersley, "Estimating Restricted Parameters", Journal of Roy. Stat. Soc., 12(2): 192-240, 1950

[Haz02] M. Hazewinkel, *Encyclopaedia of Mathematics*, Springer-Verlag Berlin Heidelberg New York, 2002

[HRFT04] M. Hernandez, B. Ristic, A. Farina and L. Timmoneri, "A Comparison of Two Cramér–Rao Bounds for Nonlinear Filtering with pd less than 1", IEEE Trans. on SP, 52(9): 2361-2370, 2004

[HF94] A. O. Hero and J. Fessler, "A recursive algorithm for computing Cramér-Rao-type bounds on estimator covariance", IEEE Trans. on IT, 40(4): 1205-1210, 1994

[HFU96] A. O. Hero, J. A. Fessler, and M. Usman, "Exploring estimator biasvariancetradeoffs using the uniform CR bound", IEEE Trans. on SP, 44(8): 2026-2041, 1996

[HJ99] R.A. Horn, C.R. Johnson, *Matrix Analysis*. Cambridge University Press, 1999

[JR04] A.K. Jagannatham, B.D. Rao, "Cramér-Rao Lower Bound for Constrained Complex Parameters", IEEE SP letters, 11(11): 875-878, 2004

[JAW95] B. James, B. D. O. Anderson, and R. C. Williamson, "Characterization of threshold for single tone maximum likelihood frequency estimation", IEEE Trans. on SP, 43(4): 817-821, 1995

[Jam64] A.T. James, "Distributions of Matrix Variates and Latent Roots Derived from Normal Samples", Ann. Math. Statist. 35(2): 475-501, 1964

[JS61] W. James, and C. Stein, "Estimation with Quadratic Loss", Proc. Fourth Berkeley Symp. on Math. Statist. and Prob., vol. 1, pp361-379, 1961

[Kay98] S.M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1998

[Kay93] S.M. Kay, *Fundamentals of Statistical Signal Processing: estimation theory*, Prentice-Hall, 1993

[KE08] S. Kay and Y. C. Eldar, "Rethinking biased estimation", IEEE SP Mag., 25(3): 133-136, 2008

[Kie52] J. Kiefer, "On minimum variance estimators", Ann. Math. Stat., 23(4): 627–629, 1952

[Kno97] L. Knockaert, "The Barankin bound and threshold behaviour in frequency estimation", IEEE Trans. on SP, 45(9): 2398-2401, 1997

[Lee92] H. B. Lee, "The Cramér-Rao bound on frequency estimates of signals closely spaced in frequency", IEEE Trans. on SP, 40(6): 1507-1517, 1992

[Lev88] N. Levanon, *Radar Principles*, Wiley-Interscience, Wiley, 1988

[LC98] E. L. Lehmann and G. Casella, *Theory of Point Estimation (2nd ed.)*. Springer, 1998

[LS96] J. LI, and P. SToica, "Efficient Mixed-Spectrum Estimation with Applications to Target Feature Extraction", IEEE Trans. on SP, 44(2): 281-295, 1996

[LTN11] T. Li, J. Tabrikian, and A. Nehorai, "A Barankin-Type Bound on Direction Estimation Using Acoustic Sensor Arrays", IEEE Trans. on SP, 59(1): 431-435, 2011

[LN07] Z. Liu and A. Nehorai, "Statistical angular résolution limit for point sources", IEEE Trans. on SP, 55(11): 5521-5527, 2007

[Mar93] T.L. Marzetta, "A simple derivation of the constrained multiple parameter Cramér-Rao bound", IEEE Trans. on SP, 41(6): 2247-2249, 1993

[Mar97] T.L. Marzetta, "Computing the Barankin bound, by solving an unconstrained quadratic optimization problem", in Proc. IEEE ICASSP, 1997

[MS69] R. McAulay, L.P. Seidman, "A useful form of the Barankin Lower Bound and its application to PPM threshold analysis", IEEE Trans. on IT, 15(2): 273-279, 1969

[MH71] R. McAulay, E.M. Hofstetter, " Barankin Bounds on parameter estimation", IEEE Trans. on IT, 17(6): 669-676, 1971

[MCLB12] T. Menni, E. Chaumette, P. Larzabal, J.P. Barbot,"New Results on Deterministic Cramér–Rao Bounds for Real and Complex Parameters", IEEE Trans. on SP, 60(3): 1032-1049, 2012

[Men12] T. Menni, "Borne de Cramér-Rao déterministe pour l'analyse des performances asymptotiques en estimation d'un radar actif", PhD Thesis, Ecole Normale Supérieure de Cachan, France, Sept. 2012

[MGCL12] T. Menni, J. Galy, E. Chaumette and P. Larzabal, "On the versatility of constrained Cramér-Rao bound for estimation performance analysis and design of a system of measurement", IEEE Trans. on AES, accepté le 08/09/2013

[MVL13] X. Mestre, P. Vallet, P. Loubaton, "On the résolution probability of conditional and unconditional maximum likelihood doa estimation", in Proc. Eurasip EUSIPCO, 2013

[Mor83] C.N. Morris, "Natural exponential families with quadratic variance functions: statistical theory", Ann. Statist., 11(2): 515-529, 1983

[Mos69] E. Mosca, "Angle Estimation in Amplitude Comparison monopulse Systems", IEEE Trans. on AES, 5(2): 205-212, 1969

[MKS07] T.J. Moore, R.J. Kozick, B.M. Sadler, "The Constrained Cramér–Rao Bound From the Perspective of Fitting a Model", IEEE SP Letters, 14(8): 564-567, 2007

[MSK08] T.J. Moore, B.M. Sadler, R.J. Kozick, "Maximum-Likelihood Estimation, the Cramér-Rao Bound, and the Method of Scoring With Parameter Constraints", IEEE Trans. on SP, 56(3): 895-908, 2008

[NBL04] L. Najjar-Atallah, J.P. Barbot and P. Larzabal, "SNR threshold indicator in Data aided frequency synchronization", IEEE SP Letters, 11(8): 252-254, 2004

[NLF05] L. Najjar-Atallah, P. Larzabal and P. Forster, "Threshold region determination of ML estimation in known phase Data aided frequency synchronization", IEEE SP Letters, 12(9): 605-608, 2005

[NFBL09] L. Najjar Atallah, P. Forster, J. P. Barbot, P. Larzabal, "Tractable forms of some deterministic MSE lower bounds for biased estimation", in Proc. IEEE ISPA, 2009

[OVSN93] Ottersten and M. Viberg and P. Stoica and A. Nehorai: "Exact and Large Sample Maximum Likelihood Techniques for Parameter Estimation and Detection in Array Processing", *Radar Array Processing*, §4, Springer-Verlag, 1993

[Ova11] J.P Ovarlez, "Apport de Techniques d'Analyse, de Détection et d'Estimation en Radar et Imagerie SAR", Habilitation a Diriger des Recherches, Université Paris-Sud 11, 2011

[Pom03] D. Pommeret, "A construction of the UMVU estimator for simple quadratic natural exponential families", Journal of Multivariate Analysis, 85: 217-233, 2003

[QK94] B. G. Quinn and P. J. Kootsookos, "Threshold behavior of the maximum likelihood estimator of frequency", IEEE Trans. on SP, 42(11): 3291-3294, 1994

[Rao45] C.R. Rao, "Information and accuracy attainable in the estimation of statistical parameters", Bull. Calcutta Math. Soc., 37: 81-91, 1945

[RGCLR13] C. Ren, J. Galy, E. Chaumette, P. Larzabal, A. Renaux, "High résolution techniques for radar: myth or reality ?", in Proc. Eurasip EUSIPCO, 2013

[REKGCLR13] C. Ren, M. N. El Korso, J. Galy, E. Chaumette, P. Larzabal and A. Renaux, "Limitations on Resolution: a probabilistic approach in the asymptotic region of operation", soumis à IEEE Trans. on SP, en cours de révision pour re-soumission ((R) Reject and resubmit as a regular paper).

[RP49] H. Robbins and E.J.G. Pitman, "Application of the method of mixtures to quadratic forms in normal variates", Ann. Math. Stat., 20: 55-560, 1949

[Ren06] A. Renaux, "Contribution à l'analyse des performances d'estimation en traitement statistique du signal", PhD Thesis, Ecole Normale Supérieure de Cachan, France, 2006

[RFCL06] A. Renaux, P. Forster, E. Chaumette, P. Larzabal, "On the High SNR CML Estimator Full Statistical Characterization", IEEE Trans. on SP, 54(12): 4840-4843, 2006

[RFBL07] A. Renaux, P. Forster, E. Boyer, P. Larzabal, "Unconditional Maximum Likelihood Performance at Finite Number of Samples and High Signal-to-Noise Ratio", IEEE Trans. on SP, 55(5): 2358-2364, 2007

[RFLRN08] A. Renaux, P. Forster, P. Larzabal, C.D. Richmond, A. Nehorai, "A Fresh Look at the Bayesian Bounds of the Weiss-Weinstein Family", IEEE Trans. on SP, 56(11): 5334-5352, 2008

[Ric05] C. D. Richmond, "Capon Algorithm Mean Squared Error Threshold SNR Prediction and Probability of Resolution", IEEE Trans. on SP, 53(8): 2748-2764, 2005

[RH12] C. D. Richmond and L. L. Horowitz, "Aspects of threshold region mean squared error prediction - Method of interval errors, bounds, Taylor's theorem and extensions", in Proc. ASILOMAR, 2012

[RB74] D.C. Rife, R.R. Boorstyn, "Single tone parameter estimation from discrete-time observations", IEEE Trans. on IT, 20(5): 591-598, 1974

[SKM01] B. M. Sadler, R. J. Kozick, and T. Moore, "Bounds on bearing and symbol estimation with side information" IEEE Trans. on SP, 49(4): 822-834, 2001

[Sch91] L.L. Scharf, *Detection, Estimation and Time Series Analysis*, Adison-Wesley, 1991

[SM05] M. Shahram and P. Milanfar, "On the resolvability of sinusoids with nearby frequencies in the presence of noise", IEEE Trans. on SP, 53(7): 2579-2585, 2005

[She84] S. M. Sherman, *Monopulse Principles and Techniques*, Dedham, MA: Artech House, 1984

[Smi05] S. T. Smith, "Statistical résolution limits and the complexified Cramér Rao bound", IEEE Trans. on SP, 53(5): 1597-1609, 2005

[Ste56] C. Stein, "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution", Proc. Third Berkeley Symp. on Math. Statist. and Prob., (1): 197-206, 1956

[SB85] A. O. Steinhardt and C. Bretherton, "Thresholds in frequency estimation" in Proc. ICASSP, 1985

[SN90] P. Stoica, A. Nehorai, "Performances study of conditional and unconditional direction of arrival estimation", IEEE Trans. on SP, 38(10): 1783-1795, 1990

[SM01] P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular information matrices", IEEE Trans. on SP, 49(1): 87-90, 2001

[SNg98] P. Stoica and B. C. Ng, "On the Cramér–Rao bound under parametric constraints", IEEE SP Letters, 5(7): 177-179, 1998

[SM97] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Prentice Hall, Upper Saddle River, NJ, 1997

[Swe60] P. Swerling, "Probability of detection for fluctuating targets", IEEE Trans. on IT, 6(2): 269-308, 1960

[TK99] J. Tabrikian, J.L. Krolik, "Barankin bounds for source localization in an uncertain ocean environnment", IEEE Trans. on SP, 47(11): 2917-2927, 1999

[TST94] J. K. Thomas and L. L. Scharf and D. W. Tufts, "The probability of a subspace swap in the SVD", IEEE Trans. on SP, 43(11): 730-736, 1994

[TT08] K. Todros and J. Tabrikian, "A new lower bound on the mean-square error of unbiased estimators", in Proc. IEEE ICASSP, 2008

[TT10-I] K. Todros and J. Tabrikian, ”General Classes of Performance Lower Bounds for Parameter Estimation-Part I: Non-Bayesian Bounds for Unbiased Estimators”, IEEE Trans. on IT, 56(10): 5064-5082, 2010

[TT10-ISIT] K. Todros and J. Tabrikian, ”On order relations between lower bounds on the MSE of unbiased estimators”, in Proc. IEEE ISIT, 2010

[TT10-II] K. Todros and J. Tabrikian, ” Uniformly Best Biased Estimators in Non-Bayesian Parameter Estimation”, IEEE Trans. on IT, 57(11): 7635-7647, 2011

[TS88] J.Tsao, B.D. Steinberg, ”Reduction of Sidelobe and Speckle Artifacts in Microwave Imaging: the CLEAN Technique”, IEEE Trans. on AP, 36(4): 543-556, 1988

[Tun12] P. Tune, ”Computing Constrained Cramer-Rao Bounds”, IEEE Trans. on SP, 60(10): 5543-5548, 2012

[VDB94] A. Van Den Bos, ”A Cramér-Rao Lower Bound for Complex Parameters”, IEEE Trans. on SP, 42(10): 2859, 1994

[VT68] H.L. Van Trees, *Detection, Estimation and Modulation Theory, Part 1*, New York, Wiley, 1968

[VT02] H.L. Van Trees, *Optimum Array Processing*, New-York, Wiley-Interscience, 2002

[VBC13] F. Vincent, O. Besson and E. Chaumette, ”Computationally efficient approximate maximum likelihood estimation of two closely spaced sources”, revision submitted to Signal Processing (Elsevier)

[YG05] Y. Yao, G. B. Giannakis, ”On Regularity and Identifiability of Blind Source Separation Under Constant-Modulus Constraints”, IEEE Trans. on SP, 53(4):1272-1281, 2005

[YB92] S. F. Yau and Y. Bresler, ”A Compact Cramér-Rao Bound Expression for Parametric Estimation of Superimposed Signals”, IEEE Trans. on SP, 40(5): 1226-1230, 1992

[YB92W] S. F. Yau and Y. Bresler, ”Worst Case Cramér-Rao Bounds for Parametric Estimation of Superimposed Signals”, IEEE Trans. on SP, 40(12): 2973-2986, 1992

[ZS93] A. Zeira, M.Schultheiss, ”Realizable Lower Bounds for Time Delay Estimation”, IEEE Trans. on SP, 41(11): 3102-3113, 1993

[ZJLJ11] S. Zhiguang, Z. Jianxiong, H. Lei, L. Jicheng, ”A New Derivation of Constrained Cramér-Rao Bound via Norm Minimization”, IEEE Trans. on SP, 59(4): 1879-1882, 2011

[ZW88] I. Ziskind and M. Wax, ”Maximum likelihood localization of multiple sources by alternating projection”, IEEE Trans. on ASSP, 36(10): 1553-1560, 1988

Caractérisation des problèmes conjoints de détection et d'estimation

Le thème général de ma recherche est la caractérisation des problèmes conjoints détection-estimation les plus fréquemment rencontrés en écoute active ou passive (radar, télécoms, sonar, ...) : l'estimation des paramètres déterministes (non aléatoires) d'un signal d'intérêt intermittent en présence d'un environnement permanent. Ce problème peut être modélisé par un test d'hypothèses binaires : H0 (environnement seul) et H1 (environnement et signal d'intérêt). L'intérêt premier pour l'estimation des paramètres déterministes provient de mon domaine d'application professionnel : le radar actif où l'approche à paramètres déterministe est privilégiée.
Dans le cadre de paramètres déterministes, cette recherche peut s'aborder graduellement (en terme de difficulté théorique et calculatoire) sous deux axes :
• l'étude des performances en estimation non conditionnelle (sans test
de détection préalable) par le biais des bornes de performance en estimation. Dans ce cas il n'y a qu'un seul modèle d'observation H1.
• l'étude des performances en estimation conditionnelle (avec test de détection préalable), c'est à dire la caractérisation des problèmes conjoints détection-estimation. Cette étude est réalisée par le biais des bornes de performance conditionnelle appliquées à deux modèles d'observation particuliers (l'antenne monopulse et le modèle d'observation gaussien déterministe) pour lesquels certains calculs analytiques sont accessibles.

Mots-clés : DETECTION ; ESTIMATION ; BORNE ; ECARTOMETRIE

Characterization of the joint detection-estimation problem

The general theme of my research deals with the characterization of the joint detection-estimation problem arising frequently both in passive and active systems of measurement (radar, telecoms, sonar): the estimation of the deterministic (non random) parameters of an intermittent source of signal of interest embedded in a permanent noisy environment. This problem can be modelled as a binary hypothesis testing: H0 (noise signal only) and H1 (noise signal and signal of interest). My primary interest for deterministic parameters originates in my involvement in numerous studies dedicated to active radar estimation performance where the deterministic parametric model is the privileged model.
Under the framework of deterministic parametric modelling, the problem under consideration can be addressed progressively, in terms of theoretical and computational complexity, in two steps:
• the assessment of unconditional estimation performance, that is without a prior detection test, by resorting to lower bounds on estimation performance. In this first step, there is a single observation model H1.
• the assessment of conditional estimation performance, that is including a prior detection test and leading to the characterization of the joint detection-estimation problem. This far more complex problem is investigated by studying the influence of a detection step on lower bounds on estimation performance when applied to two models of observations (the monopulse antenna and the Gaussian conditional observation model) for which some analytical expressions exist.

Keywords : DETECTION ; ESTIMATION ; LOWER BOUNDS ON PERFORMANCE ESTIMATION

Office National d'Études et de Recherches Aérospatiales
BP 72 - 29 avenue de la Division Leclerc
92322 CHATILLON CEDEX
Tél. : +33 1 46 73 40 40 - Fax : +33 1 46 73 41 41
http://www.onera.fr