Méthodes statistiques pour la mise en correspondance de descripteurs

Olivier Collier

Méthodes statistiques pour la mise en correspondance de descripteurs

Thèse de doctorat de l'Université Paris-Est

École doctorale MSTICS

Spécialité Mathématiques Appliquées

Méthodes statistiques pour la mise en correspondance de descripteurs

Olivier Collier

Thèse soutenue le 2 octobre 2013 devant le jury composé de :

M.	Arnak Dalalyan	Université Paris-Est	(Directeur)
M.	Jérémie Bigot	SupAéro, Toulouse	(Rapporteur)
M.	Vladimir Spokoiny	Institut Weierstrass, Berlin	(Rapporteur)
Mme	Cristina Butucea	Université Paris-Est	(Examinateur)
M.	Alexandre Tsybakov	CREST/GENES	(Examinateur)
M.	Nicolas Vayatis	ENS Cachan	(Examinateur)
M.	Jean-Philippe Vert	Mines de Paris	(Examinateur)

Cette thèse a été préparée au laboratoire IMAGINE :

IMAGINE / CERTIS
École des Ponts ParisTech
6, Av Blaise Pascal - Cité Descartes
Champs-sur-Marne
77455 Marne-la-Vallée cedex 2 - France

puis au Laboratoire de Statistiques du CREST

INSEE-CREST
3, avenue Pierre Larousse
92240 MALAKOFF.

Voici la liste des travaux terminés durant cette thèse :

Articles de revue :

– *Curve registration by nonparametric goodness-of-fit testing,* avec Arnak Dalalyan. Soumis.
– *Minimax hypothesis testing for curve registration.* Electron. J. Statist., 6 :1129–1154, 2012.
– *Permutation estimation and minimax matching thresholds,* avec Arnak Dalalyan. Soumis.

Conférences :

– Conférence AISTATS 2012, présentation du poster : *Wilks' phenomenon and penalized likelihood-ratio test for nonparametric curve registration,* avec Arnak Dalalyan. Journal of Machine Learning Research - Proceedings Track 22 : 264-272.
– Conférence AISTATS 2012, présentation du poster : *Minimax hypothesis testing for curve registration.* Journal of Machine Learning Research - Proceedings Track 22 : 236-245.
– Conférence AISTATS 2013, présentation orale d'un article récompensé du titre de Notable Paper : *Permutation estimation and minimax rates of identifiability,* avec Arnak Dalalyan. Journal of Machine Learning Research - Proceedings Track 31 : 10-19.

Méthodes statistiques pour la mise en correspondance de descripteurs

Résumé :

De nombreuses applications, en vision par ordinateur ou en médecine notamment, ont pour but d'identifier des similarités entre plusieurs images ou signaux. On peut alors détecter des objets, les suivre, ou recouper des prises de vue. Dans tous les cas, les procédures algorithmiques qui traitent les images utilisent une sélection de points-clefs qu'elles essayent ensuite de mettre en correspondance par paire. Elles calculent pour chaque point un descripteur qui le caractérise, le discrimine des autres. Parmi toutes les procédures possibles, la plus utilisée aujourd'hui est SIFT, qui sélectionne les points-clefs, calcule des descripteurs et propose un critère de mise en correspondance globale.

Dans une première partie, nous tentons d'améliorer cet algorithme en changeant le descripteur original qui nécessite de trouver l'argument du maximum d'un histogramme : en effet, son calcul est statistiquement instable. Nous devons alors également changer le critère de mise en correspondance de deux descripteurs. Il en résulte un problème de test non-paramétrique dans lequel à la fois l'hypothèse nulle et alternative sont composites, et même non-paramétriques. Nous utilisons le test du rapport de vraisemblance généralisé afin d'exhiber des procédures de test consistantes, et proposons une étude minimax du problème.

Dans une seconde partie, nous nous intéressons à l'optimalité d'une procédure globale de mise en correspondance. Nous énonçons un modèle statistique dans lequel des descripteurs sont présents dans un certain ordre dans une première image, et dans un autre dans une seconde image. La mise en correspondance revient alors à l'estimation d'une permutation. Nous donnons un critère d'optimalité au sens minimax pour les estimateurs. Nous utilisons en particulier la vraisemblance afin de trouver plusieurs estimateurs consistants, et même optimaux sous certaines conditions. Enfin, nous nous sommes intéressés à des aspects pratiques en montrant que nos estimateurs étaient calculables en temps raisonnable, ce qui nous a permis ensuite d'illustrer la hiérarchie de nos estimateurs par des simulations.

Mots-clefs : vision par ordinateur, test minimax, estimation minimax, problème non-paramétrique, hypothèse nulle composite, test du maximum de vraisemblance généralisé, phénomène de Wilks, détection de signal, estimation de permutation.

Statistical methods for descriptor matching

Summary:

Many applications, as in computer vision or medicine, aim at identifying the similarities between several images or signals. Thereafter, it is possible to detect objects, to follow them, or to overlap different pictures. In every case, the algorithmic procedures that treat the images use a selection of keypoints that they try to match by pairs. The most popular algorithm nowadays is SIFT, that performs keypoint selection, descriptor calculation, and provides a criterion for global descriptor matching.

In the first part, we aim at improving this procedure by changing the original descriptor, that requires to find the argument of the maximum of a histogram: its computation is indeed statistically unstable. So we also have to change the criterion to match two descriptors. This yields a nonparametric hypothesis testing problem, in which both the null and the alternative hypotheses are composite, even nonparametric. We use the generalized likelihood ratio test to get consistent testing procedures, and carry out a minimax study.

In the second part, we are interested in the optimality of the procedure of global matching. We give a statistical model in which some descriptors are present in a given order in a first image, and in another order in a second image. Descriptor matching is equivalent in this case to the estimation of a permutation. We give an optimality criterion for the estimators in the minimax sense. In particular, we use the likelihood to find several consistent estimators, which are even optimal under some conditions. Finally, we tackled some practical aspects and showed that our estimators are computable in reasonable time, so that we could then illustrate the hierarchy of our estimators by some simulations.

Keywords: computer vision, minimax testing, minimax estimation, nonparametric problem, composite null hypothesis, generalized likelihood ratio test, Wilks' property, signal detection, permutation estimation.

Table des matières

Exergue

C'est un grand charme ajouté à la vie dans une station balnéaire comme était Balbec, si le visage d'une jolie fille, une marchande de coquillages, de gâteaux ou de fleurs, peint en vives couleurs dans notre pensée, est quotidiennement pour nous dès le matin le but de chacune de ces journées oisives et lumineuses qu'on passe sur la plage. Elles sont alors, et par là, bien que désœuvrées, alertes comme des journées de travail, aiguillées, aimantées, soulevées légèrement vers un instant prochain, celui où tout en achetant des sablés, des roses, des ammonites, on se délectera à voir sur un visage féminin, les couleurs étalées aussi purement que sur une fleur. Mais au moins, ces petites marchandes, d'abord on peut leur parler, ce qui évite d'avoir à construire avec l'imagination les autres côtés que ceux que nous fournit la simple perception visuelle, et à recréer leur vie, à s'exagérer son charme, comme devant un portrait ; surtout, justement parce qu'on leur parle, on peut apprendre où, à quelles heures on peut les retrouver. Or il n'en était nullement ainsi pour moi en ce qui concernait les jeunes filles de la petite bande. Leurs habitudes m'étant inconnues, quand certains jours je ne les apercevais pas, ignorant la cause de leur absence, je cherchais si celle-ci était quelque chose de fixe, si on ne les voyait que tous les deux jours, ou quand il faisait tel temps, ou s'il y avait des jours où on ne les voyait jamais. Je me figurais d'avance ami avec elles et leur disant «Mais vous n'étiez pas là tel jour ?» «Ah ! oui, c'est parce que c'était un samedi, le samedi nous ne venons jamais parce que…» Encore si c'était aussi simple que de savoir que le triste samedi il est inutile de s'acharner, qu'on pourrait parcourir la plage en tous sens, s'asseoir à la devanture du pâtissier, faire semblant de manger un éclair, entrer chez le marchand de curiosités, attendre l'heure du bain, le concert, l'arrivée de la marée, le coucher du soleil, la nuit sans voir la petite bande désirée. Mais le jour fatal ne revenait peut-être pas une fois par semaine. Il ne tombait peut-être pas forcément un samedi. Peut-être certaines conditions atmosphériques influaient-elles sur lui ou lui étaient-elles entièrement étrangères. Combien d'observations patientes mais non point sereines, il faut recueillir sur les mouvements en apparence irréguliers de ces mondes inconnus avant de pouvoir être sûr qu'on ne s'est pas laissé abuser par des coïncidences, que nos prévisions ne seront pas trompées, avant de dégager les lois certaines, acquises au prix d'expériences cruelles, de cette astronomie passionnée. Me rappelant que je ne les avais pas vues le même jour qu'aujourd'hui, je me disais qu'elles ne viendraient pas, qu'il était inutile de rester sur la plage. Et justement je les apercevais. En revanche, un jour où, autant que j'avais pu supposer que des lois réglaient le retour de ces constellations j'avais calculé devoir être un jour faste, elles ne venaient pas. Mais à cette première incertitude si je les verrais ou non le jour même venait s'en ajouter une plus grave, si je les reverrais jamais, car j'ignorais en somme si elles ne devaient pas partir pour l'Amérique, ou rentrer à Paris. Cela suffisait pour me faire commencer à les aimer. On peut avoir du goût pour une personne. Mais pour déchaîner cette tristesse, ce sentiment de l'irréparable, ces angoisses, qui préparent l'amour, il faut – et il est peut-être ainsi, plutôt que ne l'est une personne, l'objet même que cherche anxieusement à étreindre la passion – le risque d'une impossibilité. Ainsi agissaient déjà ces influences qui se répètent au cours d'amours successives, pouvant du reste se produire mais alors plutôt dans l'existence des grandes villes au sujet d'ouvrières dont on ne sait pas les jours de congé et qu'on s'effraye de ne pas avoir vues à la sortie de l'atelier ou du moins qui se renouvelèrent au cours des miennes. Peut-être sont-elles inséparables de l'amour ; peut-être tout ce qui fut une particularité du premier vient-il s'ajouter aux suivants, par souvenir, suggestion, habitude et à travers les périodes successives de notre vie donner à ses aspects différents un caractère général.

Marcel PROUST, *A l'ombre des jeunes filles en fleur*.

Remerciements

Tout d'abord, je voudrais remercier Cristina Butucea, Alexandre Tsybakov, Nicolas Vayatis, Jean-Philippe Vert, et tout particulièrement les rapporteurs Jérémie Bigot et Vladimir Spokoiny, pour avoir accepté de participer à mon jury de thèse. Merci aux rapporteurs pour leur lecture attentive du manuscrit, et leurs remarques constructives !

Ensuite, je remercie mon directeur de thèse, Arnak Dalalyan, pour avoir partagé ses connaissances avec moi, pour m'avoir inculqué aussi un plus grand sens de la rigueur, pour son exigence et sa patience.

Je remercie les membres du laboratoire IMAGINE des Ponts et Chaussée, pour la bonne ambiance qu'ils font régner dans l'équipe, en particulier ceux que j'y ai longtemps côtoyés : Jean-Yves Audibert, Alexandre Boulc'h, Renaud Marlet, Pascal Monasse, Brigitte Mondou, Pierre Moulon, Bertrand Neveu, David Ok, Victoria Rudakova, Olivier Tournaire, Hiep Hoang Vu. Merci également aux membres du Laboratoire de Statistiques du CREST, en particulier aux nouveaux doctorants : Mathieu Gerber, Adélaïde Olivier, James Ridgway, Jean-Bernard Salomond.

Enfin, pour avoir patiemment écouté mes plaintes, pour s'être sans excès intéressé à ma thèse, pour m'avoir aidé à passer agréablement ces trois années, je voudrais remercier ma famille, mais aussi mes amis, mes confidents : Youssef Akel, Pauline Delannoy, François Devémy, Stanislas Dorey, Neşe Gökşen, Adeline Holt, Raluca Ilie, Adrian Iuga, Ekin Karapınar, Clotilde Lambert, Audrey Lefebvre, Alban Pétiau, Antoine Salomon, Jan-Frerk Stein, Oliver Troll, Christian Wenzel, et Blandine.

Plan de la thèse : Motivations et résultats de la thèse

En vision par ordinateur, il est fréquent de devoir identifier des correspondances entre deux images. Par exemple, on peut créer un panorama à partir de plusieurs prises de vue, à condition d'être capable de déterminer les zones de chevauchement. Il s'agit de la technique d'**assemblage de photos**, qui permet notamment d'obtenir une carte à partir de prises de vue par satellite. On peut également citer l'exemple de la **reconnaissance faciale**, dont le but est de détecter une éventuelle correspondance entre une image donnée d'un visage et un élément d'une base de donnée. Enfin, la technique du **traçage** est utilisée pour suivre le mouvement d'une cible dans une succession d'images. Il s'agit donc de repérer un objet ou une personne dans la première image et de déterminer sa position dans les suivantes.

Dans tous les cas, il n'est pas concevable en pratique d'essayer de faire correspondre deux images pixel par pixel ; il est nécessaire de sélectionner tout d'abord un sous-ensemble de **points-clefs** dans chaque image. Si ces points sont suffisamment peu nombreux, la complexité des calculs est ainsi nettement diminuée. Après cette sélection, on associe à chaque point-clef un **descripteur**, c'est-à-dire un objet mathématique caractérisant localement l'image autour du point. Le descripteur peut être assimilé à un code qui permet de le distinguer des autres.

Le plus célèbre exemple de ces descripteurs a été introduit par David Lowe en 2004 (*cf.* Lowe (2004)). Après avoir sélectionné les points qui maximisent un critère de stabilité locale, on associe à chaque point un descripteur dont voici, de manière très simplifiée, la construction :

– Calculer les gradients dans une petite zone autour du point.
– Construire l'histogramme de l'amplitude moyenne du gradient en fonction de son orientation.
– Recentrer l'histogramme afin que son maximum soit atteint en 0.

Avec ce choix de descripteur, un même point physique recevra le même descripteur quelque soit sa position dans l'image (**invariance par translation**), et même si on tourne l'appareil photographique autour de son axe (**invariance par rotation**). Ainsi, pour la plupart des transformations basiques, un même point physique recevra un descripteur proche, et on cherchera naturellement à associer deux points ayant des descripteurs comparables.

Enfin, une fois ce critère de mise en correspondance établi, Lowe propose une procédure globale permettant, étant donnés un grand nombre de points répartis dans deux images, de les associer par paires en écartant les correspondances non-pertinentes.

Le but principal de cette thèse est d'étudier d'un point de vue statistique le problème de la mise en correspondance. On a vu qu'il comportait trois aspects : la sélection des points-clefs, le choix des descripteurs, qui est associé à un critère de mise en correspondance, et la procédure de décision globale. Nous ne nous intéressons pas ici à la première étape ; nous supposons dans la suite que des points-clefs ont été sélectionnés, et que cette sélection a été bien faite en un certain sens. Par exemple, dans la deuxième partie, nous nous intéresserons principalement au cas où le même objet est représenté sur deux images, et où un algorithme a sélectionné des points correspondant au même ensemble de points physiques.

Dans la **première partie de cette thèse**, nous explorons le choix d'un autre descripteur : nous utiliserons l'histogramme du SIFT sans le recentrer. En effet, le recentrage nécessite le calcul de l'argument du maximum de l'histogramme, et on sait que cette recherche ne peut pas être effectuée de manière consistante. Utiliser l'histogramme non-recentré permet donc d'améliorer la procédure en évitant le calcul de l'argument du maximum, mais cela nécessite de modifier le critère de mise en correspondance : dorénavant, on cherchera à associer deux histogrammes qui sont proches à une translation près. Il en ressort un problème de test non-paramétrique : étant données deux fonctions bruités f et $f^\#$, observées à travers X et $X^\#$ vérifiant

$$\begin{cases} dX_t = f(t)dt + \sigma dW_t, \\ dX_t^\# = f^\#(t)dt + \sigma^\# dW_t^\#, \end{cases} \quad t \in [0, 2\pi],$$

où W et $W^\#$ sont deux mouvements browniens indépendants, on se donne les hypothèses

$$\begin{cases} H_0: & \exists\, \tau \in [0, 2\pi],\ f^\# = f_\tau, \\ H_1: & \forall\, \tau \in [0, 2\pi],\ \|f^\# - f_\tau\| > \rho_\sigma, \end{cases}$$

où f_τ désigne la translatée de f par τ.

Dans ce problème, à la fois l'hypothèse nulle et l'hypothèse alternative sont composites, et même non-paramétriques, si on suppose seulement que f et $f^\#$ appartiennent à une boule de Sobolev. Ce genre de problème a été peu étudié, puisque la plupart des résultats concernent des hypothèses nulles simples.

De plus, il est montré dans Fan *et al.* (2001) que, pour les problèmes de test nonparamétrique, l'utilisation du test du rapport de vraisemblance conduisait à des procédures non-optimales. Il faut alors utiliser la technique du rapport de vraisemblance généralisé ; la statistique de test a alors la forme d'un rapport de vraisemblance, mais les paramètres, qui sont normalement fixés pour maximiser le rapport, peuvent être choisis plus souplement. Dans le **premier chapitre de la première partie**, nous déterminons une procédure de test par la méthode du rapport de vraisemblance généralisé, et nous démontrons qu'elle possède la propriété de Wilks, à l'instar des tests du rapport de vraisemblance dans les cas simples, c'est-à-dire que la distribution asymptotique de la statistique de test est indépendante des paramètres de nuisance.

Dans le **deuxième chapitre de la première partie**, le problème est étudié au sens minimax. On définit les erreurs de première et seconde espèces pour tout test ψ :

$$\begin{cases} \alpha(\psi,\Theta_0) &=& \sup_{\Theta_0} \mathbf{P}_{f,f^\#}(\psi=1), \\ \beta(\psi,\Theta_1) &=& \sup_{\Theta_1} \mathbf{P}_{f,f^\#}(\psi=0), \end{cases}$$

où

$$\begin{cases} \Theta_0 = \{f, f^\# \in \mathscr{F}_{s,L} \,|\, \exists \tau \in \mathbb{R}, f^\# = f_\tau\}, \\ \Theta_1 = \{f, f^\# \in \mathscr{F}_{s,L} \,|\, \min_\tau \|f^\# - f_\tau\| \geq C\rho_\sigma\}, \end{cases}$$

avec $\mathscr{F}_{s,L}$ une boule de Sobolev de régularité s et de rayon L.

Avec ces notations, nous cherchons le paramètre de séparation ρ_σ minimal qui permette de trouver des procédures de test consistantes. Nous démontrons que le paramètre optimal est, à un facteur logarithmique près,

$$\rho_\sigma = \left(\sigma^2 \sqrt{\log \sigma^{-1}}\right)^{\frac{s}{4s+1}},$$

et nous définissons une procédure qui est minimax, ou proche de l'être. Enfin, nous donnons une procédure adaptative par rapport à la régularité de la boule de Sobolev considérée et nous montrons que la même performance est alors atteinte.

Dans une **deuxième partie**, nous nous intéressons à la procédure globale de mise en correspondance des points-clefs de deux images. Pour simplifier la présentation, nous choisissons d'associer les points dont les descripteurs sont exactement égaux, plutôt que translatés l'un par rapport à l'autre. De plus, nous nous plaçons dans le cas où les deux images contiennent exactement les mêmes points-clefs. Avec ces hypothèses, le problème revient à l'estimation d'une permutation π^* telle que

$$\begin{cases} X_i = \theta_i + \sigma_i \xi_i, \\ X_i^\# = \theta_{\pi^*(i)} + \sigma_i^\# \xi_i^\#, \end{cases} \quad i = 1, \ldots, n,$$

où

- n est le nombre de points-clefs,
- $\theta_1, \ldots, \theta_n$ sont les descripteurs dans $\mathbb{R}^d$,
- $X_1, \ldots, X_n, X_1^\#, \ldots, X_n^\#$ sont les observations bruitées,
- $\xi_1, \ldots, \xi_n, \xi_1^\#, \ldots, \xi_n^\# \overset{iid}{\sim} \mathcal{N}(0,1)$.

Nous nous focalisons sur plusieurs cas :

1. $\sigma_1 = \ldots, \sigma_n = \sigma_1^\#, \ldots, \sigma_n^\# = \sigma$ (**cas homoscédastique**) et σ est inconnu,
2. les variances sont a priori distinctes (**cas hétéroscédastique**) et connues,
3. les variances sont a priori distinctes (**cas hétéroscédastique**) et inconnues.

De plus, nous supposons toujours que

$$\sigma_{\pi^*(1)}^\# = \sigma_1, \ldots, \sigma_{\pi^*(n)}^\# = \sigma_n.$$

Dans le dernier cas, le plus général, l'estimateur du maximum de vraisemblance donne

$$\hat{\pi} = \arg\min_{\pi \in \mathfrak{S}_n} \sum_{i=1}^{n} \log \|X_{\pi(i)} - X_i^\#\|^2.$$

Si la quantité

$$\kappa \triangleq \min_{i \neq j} \|\theta_i - \theta_j\|$$

est nul, la permutation n'est pas identifiable, et on montre que si ce paramètre est plus petit, à une constante multiplicative près, que

$$\kappa^0 = \sigma \max\left(\sqrt{log n}, (d \log n)^{1/4}\right),$$

il n'existe aucun estimateur consistant. En revanche, s'il est plus grand que κ^0, alors l'estimateur ci-dessus est consistant. Enfin, on montre que cet estimateur est calculable en pratique en temps polynômial.

Introduction aux tests minimax non-paramétriques

Détection de signal

L'objet de cette thèse est un problème de test minimax non-paramétrique. Sa résolution est inspirée des travaux sur la détection de signal, que nous citons dans cette section.

Le problème de la détection de signal est le plus simple exemple de test non-paramétrique. Le champ de ses applications est très large, allant de la médecine à la robotique. Il se pose en effet à chaque fois que l'on doit prendre une décision basée sur la présence ou non d'une stimulation qui ne serait pas simplement due à l'appareil de détection. Sa difficulté est liée au rapport entre l'intensité du signal et l'intensité du bruit, c'est-à-dire de l'activité aléatoire du détecteur : si le signal est trop faible, il ne pourra être distingué du bruit.

On peut en donner la modélisation suivante :

$$dX_t = f(t)dt + \sigma dW_t\,, \quad t \in [0,1], \tag{1}$$

où

$$\begin{cases} f \text{ est le signal à détecter,} \\ X \text{ est le processus observé,} \\ W \text{ est un processus de Wiener représentant le bruit,} \\ \sigma \text{ est l'intensité du bruit.} \end{cases}$$

Ce modèle est appelé **modèle de bruit blanc gaussien**. Il est très particulier, puisqu'il suppose une certaine structure gaussienne du bruit, et que l'on a accès à toute observation X_t sur le segment $[0,1]$, ce qui n'est pas concevable en pratique.

Cependant, il a été démontré que, sous certaines conditions, le modèle de bruit blanc gaussien était équivalent au **modèle de régression non-paramétrique** (*cf.* Brown *et* Low (1996)), dans lequel on ne dispose que d'un nombre fini d'observations. Ces observations peuvent être aléatoirement réparties (*cf.* Reiß (2008)), le bruit peut ne pas être gaussien (*cf.* Grama *et* Nussbaum (1998) and Grama *et* Nussbaum (2002)). Rappelons qu'il est équivalent à d'autres modèles encore : **estimation de densité** (*cf.* Nussbaum (1996)) ou **diffusion ergodique** (*cf.* Dalalyan *et* Reiß (2006)).

D'autre part, le modèle de bruit blanc gaussien est équivalent au **modèle de suite gaussienne**, qui nous intéressera par la suite. En effet, en utilisant la transformation en série de Fourier, on obtient le modèle :

$$X_i = \theta_i + \sigma\xi_i\,, \quad i \in \mathbb{Z}, \tag{2}$$

où $\boldsymbol{\theta} \triangleq \{\theta_i\}_{i\in\mathbb{Z}}$ est une suite de nombres complexes dans $L^2(\mathbb{C})$, et $\{\xi_i\}_{i\in\mathbb{Z}}$ est une suite de variables aléatoires i.i.d. avec

$$\mathrm{Re}(\xi_i), \mathrm{Im}(\xi_i) \overset{iid}{\sim} \mathcal{N}(0,1).$$

Si nous nous plaçons dans ce modèle, le problème de la détection de signal revient à tester H_0 contre H_1, avec

$$\begin{cases} H_0 : & \boldsymbol{\theta} = 0, \\ H_1 : & N(\boldsymbol{\theta}) \geq C\rho, \end{cases}$$

où N est une norme ou une pseudo-norme, et $C, \rho > 0$. La valeur du produit $C\rho$ détermine l'existence de procédures de test capables de tester de manière consistante entre H_0 et H_1, c'est-à-dire, en particulier, mieux qu'on pourrait le faire avec une procédure aveugle. Si le produit est plus petit qu'une certaine valeur dépendante de σ, alors aucune procédure ne sera consistante. Pour rendre compte de cette dépendance, nous écrirons

$$"\rho = \rho_\sigma".$$

Maintenant, pour déterminer précisément la valeur limite du produit $C\rho_\sigma$, nous avons besoin de certaines définitions. Tout d'abord, nous notons Θ_0 et Θ_1 les ensembles correspondant aux hypothèses nulle et alternative :

$$\begin{cases} \Theta_0 = \{0\}, \\ \Theta_1 = \{\boldsymbol{\theta} \in \mathscr{F} \,|\, N(\boldsymbol{\theta}) \geq C\rho_\sigma\}, \end{cases}$$

où $\mathscr{F}$ est un certain ensemble de régularité dont nous traiterons plus loin. Nous pouvons alors définir les erreurs de première et de seconde espèce pour un test générique ψ :

$$\begin{cases} \alpha(\psi, \Theta_0) = \sup_{\Theta_0} \mathbf{P}_{\boldsymbol{\theta}}(\psi = 1), \\ \beta(\psi, \Theta_1) = \sup_{\Theta_1} \mathbf{P}_{\boldsymbol{\theta}}(\psi = 0), \end{cases}$$

où $\mathbf{P}_{\boldsymbol{\theta}}$ est la loi des observations lorsque $\boldsymbol{\theta}$ est la suite de coefficients correspondant au signal sous-jacent.

On dira alors, pour une paire de réels strictement positifs α et β fixée, que ρ_σ^* est la **vitesse de séparation minimax non-asymptotique** s'il existe un test ψ^* tel que, pour $C\rho_\sigma > \rho_\sigma^*$,

$$\begin{cases} \alpha(\psi^*, \Theta_0) \leq \alpha, \\ \beta(\psi^*, \Theta_1) \leq \beta, \end{cases}$$

et un tel test ne peut être trouvé pour $C\rho_\sigma < \rho_\sigma^*$.

Dans une étude asymptotique, on peut mettre en avant la contribution de σ dans le paramètre de séparation. Ainsi, on dira que ρ_σ^* est la **vitesse de séparation minimax asymptotique** s'il existe deux constantes $C_* < C^*$ telles que :

– Il existe un test ψ^* tel que, si $\rho_\sigma = \rho_\sigma^*$ et $C > C^*$,

$$\overline{\lim}\, \alpha(\psi^*, \Theta_0) \leq \alpha,$$
$$\overline{\lim}\, \beta(\psi^*, \Theta_1) \leq \beta,$$

– Si $\rho_\sigma = \rho_\sigma^*$ et $C < C_*$, alors aucun test ne satisfait les conditions précédentes.

Dans la pratique, l'unicité de la vitesse de séparation est assurée par une règle informelle de plus courte écriture. De plus, si

$$C_* = C^* = C,$$

on appelle C la **constante minimax asymptotique exacte**.

Cependant, la détection de signal serait impossible sans certaines hypothèses supplémentaires sur le signal. En effet, il a été démontré dans Ibragimov *et* Khasminskii (1977) que si

$$N(\cdot) = \|\cdot\|_2$$

alors $\rho_\sigma^* = +\infty$. Un résultat similaire a été trouvé dans Burnashev (1979) pour $p > 0$ et

$$N(\cdot) = \|\cdot\|_p.$$

On suppose donc habituellement une certaine régularité du signal afin de pouvoir obtenir des procédures de tests consistantes.

Retournons au modèle de bruit blanc gaussien pour énoncer ces conditions de régularité. On pourra supposer que le signal appartient

– à une classe de Sobolev

$$\mathcal{F}_{s,L} = \{f \in L^2([0,1]) \mid \|f^{(s)}\|_2 \leq L\}, \tag{3}$$

– à une classe de Hölder,
  - si $p \geq 1$,
$$\mathcal{F}_{s,p,L} = \{f \in L^2([0,1]) \mid \|f^{(s)}\|_p \leq L\}, \tag{4}$$
  - si $p = +\infty$, en appelant l la partie entière de s,

$$\mathcal{F}_{s,p,L} = \left\{f \in L^2([0,1]) \mid \sup_{x_1 \neq x_2} \frac{|f^{(l)}(x_1) - f^{(l)}(x_2)|}{|x_1 - x_2|^l} \leq L\right\}, \tag{5}$$

– à une classe de Besov (*cf.* Triebel (1992)), etc.

L'étude de la vitesse minimax de séparation a été étudié en premier lieu d'un point de vue asymptotique : dans Ingster (1982) lorsque $N(\cdot) = \|\cdot\|_2$ et quand le signal appartient à une ellipsoïde, et dans Ermakov (1990) and Ermakov (1996) pour les constances minimax exactes ; dans Ingster (1993), pour $N(\cdot) = \|\cdot\|_p$, $p \geq 2$ et quand le signal appartient à une classe de Sobolev ; dans Lepski *et* Tsybakov (2000), pour $N(\cdot) = \|\cdot\|_\infty$, et quand le signal appartient à une classe de Hölder ou de Sobolev ; dans Lepski *et* Spokoiny (1999), pour $N(\cdot) = \|\cdot\|_p$, $p < 2$, et quand le signal appartient à une classe de Besov. Une revue complète de ces résultats asymptotiques peut être trouvée dans Ingster *et* Suslina (2003).

Le problème de la détection de signal a également été étudié de manière non-asymptotique dans Baraud (2002) pour $N(\cdot) = \|\cdot\|_2$ et quand le signal appartient à une ellipsoïde de L^p, $p \leq 2$.

TEST DU RAPPORT DE VRAISEMBLANCE

Une partie importante de la recherche d'une vitesse minimax de séparation consiste donc à trouver un test performant ; une technique classique est alors d'utiliser le test du rapport de vraisemblance. Or, celui-ci se révèle insatisfaisant dans un contexte non-paramétrique comme en détection du signal, ou dans notre modèle. En effet, il est souvent non-optimal au sens minimax, en plus d'être difficilement calculable. La méthode du test du rapport de vraisemblance généralisé permet de résoudre ce problème en conservant la propriété de Wilks, qui est son principal atout dans le cas paramétrique. Dans cette section, nous exposons la propriété de Wilks, et nous montrons comment modifier le test du rapport de vraisemblance dans le cas non-paramétrique tout en conservant cette propriété.

Cas paramétrique

Définissons un modèle de test paramétrique général. On dispose des observations

$$X_1, \ldots, X_n \overset{iid}{\sim} \mathbf{P}_\theta \ ,$$

où, pour une certaine mesure μ,

$$\mathbf{P}_\theta \ll \mu, \quad \frac{d\mathbf{P}_\theta}{d\mu} = p_\theta,$$

et θ appartient à un ensemble paramétrique Θ que nous choisissons ici comme un sous-ensemble d'un espace $\mathbb{R}^m$ pour un certain entier m non-nul :

$$\theta \in \Theta \subset \mathbb{R}^m.$$

Les hypothèses du problème sont

$$\begin{cases} H_0 : & \theta \in \Theta_0, \\ H_1 : & \theta \in \Theta \setminus \Theta_0, \end{cases}$$

où Θ_0 est un sous-ensemble de Θ dans lequel r composantes sont fixées avec

$$1 \leq r \leq m.$$

On peut alors calculer la statistique

$$L_n(X_1, \ldots, X_n) = 2 \sup_{\theta \in \Theta} \Big\{ \sum_{i=1}^n \log p_\theta(X_i) \Big\} - 2 \sup_{\theta \in \Theta_0} \Big\{ \sum_{i=1}^n \log p_\theta(X_i) \Big\}.$$

On peut démontrer, sous certaines conditions de régularité, la convergence de cette statistique sous H_0 :

$$L_n(X_1, \ldots, X_n) \xrightarrow{\mathscr{L}} \chi^2(r).$$

Le fait que la loi limite est indépendante des éventuels paramètres de nuisance et de type χ^2 est connu sous le nom de **propriété de Wilks**. Cette propriété permet d'expliciter un test de H_0 contre H_1 au niveau désiré, à partir des quantiles de la loi du χ^2 au nombre de degrés de liberté adéquat.

Cas non-paramétrique

Dans le cas non-paramétrique, si le test du maximum de vraisemblance peut posséder la propriété de Wilks, l'optimalité au sens minimax n'est pas assurée. Ainsi, dans Fan *et al.* (2001), on considère le modèle 2 et le problème de détection de signal dans une classe de Sobolev 3. Dans ce cas, la statistique du rapport de vraisemblance est

$$\lambda_\sigma = \frac{n}{2}\sum_{j=1}^{+\infty}\left(1 - \frac{j^{4s}\hat{\xi}^2}{(1+j^{2s}\hat{\xi}))^2}\right),$$

avec $\hat{\xi}$ un multiplicateur de Lagrange vérifiant l'équation

$$\sum_{j=1}^{+\infty}\frac{j^{2s}}{(1+\hat{\xi}\,j^{2s})^2}Y_j^2 = L^2.$$

Les auteurs donnent ensuite deux résultats. Tout d'abord, ils donnent la répartition asymptotique de la statistique du rapport de vraisemblance.

Proposition 1. *Sous H_0,*

$$r_s\lambda_\sigma \xrightarrow{\mathscr{L}} \chi^2(a_\sigma)$$

avec

$$r_s = \frac{4s+2}{2s-1}, \quad a_\sigma = \frac{(2s+1)^2}{2s-1}\left[\frac{\pi}{4s^2\sin(\pi/2s)}\right]^{2s/2s+1}(L/\sigma)^{2/2s+1}.$$

Ceci prouve que le test du rapport de vraisemblance possède la propriété de Wilks dans ce cas précis. D'autre part, le test du rapport de vraisemblance

$$\mathbb{1}_{r_s\lambda_\sigma > a_\sigma + (2a_\sigma)^{-1}z_\alpha},$$

où z_a est le quantile d'ordre $1-\alpha$ de la loi gaussienne standard, est asymptotiquement de niveau α. Cependant, le résultat suivant en montre la non-optimalité.

Proposition 2. *Il existe un $\boldsymbol{\theta} \in \mathscr{F}_{s,L}$ avec*

$$\|\boldsymbol{\theta}\| = \sigma^{2(s+d)/(2s+1)}$$

et $d > 1/8$, tel que

$$\overline{\lim}\,\mathbf{P}_{\boldsymbol{\theta}}\big(r_s\lambda_\sigma > a_\sigma + (2a_\sigma)^{-1}z_\alpha\big) \le \alpha.$$

Ce résultat affirme que la puissance du test est bornée par α pour un certain choix de $\boldsymbol{\theta}$ situé sur la sphère de rayon $\sigma^{2(s+d)/(2s+1)}$. D'un point de vue minimax, on peut donc en déduire que la vitesse de séparation ne peut pas être plus petite que ce rayon. Or, la vitesse minimax prouvée dans Ingster (1993) est $\sigma^{4s/4s+1}$, donc le test du rapport de vraisemblance n'est pas minimax dès que $s > 1/4$.

La motivation du test du rapport de vraisemblance généralisé est alors de trouver un test, inspiré du test du rapport de vraisemblance classique, possédant la propriété de Wilks et optimal au sens minimax.

L'idée de la méthode est que le paramètre $\hat{\xi}$, fixé par l'optimisation dans les classes de Sobolev, n'est pas approprié, et qu'il est préférable de le choisir de manière flexible. En effet, en remplaçant $\hat{\xi}$ par

$$\xi_\sigma = \sigma^{8s/4s+1}$$

et en appelant λ_σ^* la statistique obtenue, on fait disparaître les défauts du test du rapport de vraisemblance.

Tout d'abord, le test du rapport de vraisemblance généralisé possède toujours la propriété de Wilks :

Proposition 3. *Sous H_0,*

$$r'_s \lambda_\sigma^* \xrightarrow{\mathscr{L}} \chi^2(a'_\sigma)$$

avec

$$\begin{cases} r'_s = \frac{4s+2}{2s-1}\frac{48s^2}{24s^2+14s+1}, \\ a'_\sigma = \frac{(2s+1)^2}{2s-1}\frac{24s^2}{24s^2+14s+1}\left[\frac{\pi}{4s^2\sin(\pi/2s)}\right]^{2s/2s+1}(L/\sigma)^{2/2s+1}. \end{cases}$$

Enfin, le test du rapport de vraisemblance généralisé est optimal au sens minimax :

Proposition 4. *Si $\rho_\sigma \gg \sigma^{4s/4s+1}$, alors*

$$\inf \mathbf{P}_{\boldsymbol{\theta}}\big(r'_s\lambda_\sigma^* > a'_\sigma + (2a'_\sigma)^{-1}z_\alpha\big) \to 1,$$

où l'infimum est pris sur tous les $\boldsymbol{\theta} \in \mathscr{F}_{s,L}$ tels que $\|\boldsymbol{\theta}\| \geq \rho_\sigma$.

Plusieurs modèles sont étudiés dans Fan *et al.* (2001), mais aucun théorème ne permet d'affirmer que le rapport de vraisemblance généralisé permettra toujours de trouver des procédures de tests possédant la propriété de Wilks et minimax à la fois. Ces propriétés doivent donc être vérifiées dans chaque cas étudié. Dans la première partie de ce travail, nous effectuerons cette vérification pour notre modèle.

Problèmes de test et hypothèses composites

La difficulté du problème de test traité dans notre première partie réside en particulier dans la forme de l'hypothèse nulle, qui est non-paramétrique. Dans cette section, nous passons en revue les articles qui abordent ce type de problème.

Dans un problème de test, l'hypothèse nulle joue un rôle particulier : elle représente une simplification, un cas favorable pour l'utilisateur, dans lequel il pourra faire un traitement particulier des données. Aussi est-elle toujours la plus simple possible, et rarement composite.

Cependant, il est naturel de vouloir s'assurer qu'un signal appartient à un modèle paramétrique plutôt qu'à un modèle non-paramétrique général. Ce cas est étudié dans le modèle de régression dans Horowitz *et* Spokoiny (2001), ainsi que dans de nombreux autres travaux cités dans cet article. Plus précisément, on teste l'hypothèse nulle

$$H_0 : \quad \exists\, \theta \in \Theta,\ f = F(\cdot,\theta),$$

avec Θ un espace de paramètres inclus dans un espace euclidien et F une certaine fonction suffisamment régulière.

L'idée générale est alors de comparer un estimateur paramétrique de la fonction de régression et un estimateur non-paramétrique. La puissance des tests qui en résultent est calculée pour des **alternatives locales**, c'est-à-dire de la forme

$$f_\sigma = F(\cdot, \theta_0) + \rho_\sigma g,$$

où
- θ_0 est un certain paramètre,
- g est une fonction donnée,
- ρ_σ une suite de réels positifs tendant vers 0.

Le test sera puissant s'il peut rejeter des hypothèses alternatives locales pour lesquelles ρ_σ décroît rapidement. La plupart des articles cités dans Horowitz *et* Spokoiny (2001) obtiennent des vitesses paramétriques, c'est-à-dire que ρ_σ peut être pris de l'ordre de σ.

Cependant, le problème n'a été étudié que dans Horowitz *et* Spokoiny (2001) d'un point de vue **minimax**, c'est-à-dire pour des hypothèses alternatives de la forme

$$H_1 : d(f, H_0) \geq \rho_\sigma,$$

où

$$d(f, H_0) \triangleq \inf_{\theta \in \Theta} \|f - F(\cdot, \theta)\|.$$

Dans ce cas, en supposant que le signal appartient à une classe de Sobolev, d'Hölder ou de Besov, les auteurs montrent que la vitesse minimax de séparation asymptotique est la même que dans le problème de détection de signal.

Donnons un autre exemple où il est naturel de considérer des hypothèses nulles composites : certains travaux s'intéressent à la monotonicité de la fonction de régression (*cf.* Bowman *et al.* (1998)). En effet, des contraintes de monotonicité peuvent améliorer les performances d'un estimateur si l'hypothèse est correcte (*cf.* Mammen (1991)), mais peuvent aussi les diminuer dans le cas contraire. Baraud *et al.* (2003), Baraud *et al.* (2005) et Juditsky *et* Nemirovski (2002) vont même plus loin en traitant le cas où le signal est positif, croissant ou convexe, où il vérifie une inéquation différentielle, ou bien quand l'échantillon du signal appartient à un espace vectoriel donné. Tous ces problèmes peuvent être écrit de la manière générique suivante :

$$\begin{cases} H_0 : & f \in \mathcal{M}, \\ H_1 : & d(f, H_0) \geq \rho_\sigma, \end{cases}$$

avec

$$d(f, H_0) \triangleq \inf_{g \in \mathcal{M}} \|f - g\|.$$

On démontre que les vitesses minimax pour ces problèmes sont les mêmes que pour le problème de détection de signal, à un possible facteur logarithmique près, bien que l'hypothèse nulle soit composite, et même de nature non-paramétrique.

Enfin, notons que Gayraud *et* Pouet (2001) et Gayraud *et* Pouet (2005) s'intéressent à un problème plus général, où l'ensemble auquel doit appartenir le signal sous H_0 n'est pas explicité, mais seulement connu à travers ses caractéristiques. En particulier, on contrôle l'entropie de cet ensemble. La vitesse minimax du problème de détection de signal s'applique également à ce problème, bien que, encore une fois, l'hypothèse nulle soit non-paramétrique.

Dans les exemples que nous avons cités, l'hypothèse nulle est composite. Cependant, le problème considéré est suffisamment proche du problème de détection de signal pour qu'on ne perde pas, ou peu, de performance. Cette remarque s'appliquera encore à notre modèle, comme on le verra dans le deuxième chapitre de la première partie.

ADAPTATION

Dans tous les problèmes de test que nous avons évoqués jusqu'ici, les procédures de test optimales nécessitent certaines informations sur le signal sous-jacent, en particulier, sur sa régularité. Ceci ne semble pas naturel ; dans cette section, nous passons en revue les résultats permettant de se libérer de cette dépendance.

Le premier article à proposer une solution à ce problème est Spokoiny (1996). On y considère le problème de la détection de signal dans le modèle 1, où le signal appartient à une classe de Besov $\mathscr{F}_\tau\big(\rho_\sigma(\tau)\big)$ dépendant du quadruplet de paramètres $\tau = (s, p, q, M)$ (*cf.* Triebel (1992) pour une définition précise). La vitesse minimax de séparation asymptotique $\rho^*_\sigma(\tau)$ dépend de ce quadruplet lorsque celui-ci est fixé.

On suppose alors que l'on sait seulement que τ appartient à un ensemble $\mathscr{P}$ régulier dans un certain sens, et on considère les erreurs de première et seconde espèces :

$$\begin{cases} \alpha(\psi) = \mathbf{P}_0(\psi = 1), \\ \beta(\psi, \mathscr{P}) = \sup_{\tau\in\mathscr{P}} \sup_f \ \mathbf{P}_f(\psi = 0), \end{cases}$$

où le dernier supremum est pris parmi tous les signaux de $\mathscr{F}_\tau\big(\rho_\sigma(\tau)\big)$.

Ainsi, un test adaptatif, c'est-à-dire ne possédant pas d'information sur le signal sous-jacent, doit pouvoir traiter des signaux de régularités diverses. On ne peut espérer de gain de perfomance pour une régularité fixée, car une procédure possède intuitivement moins d'information dans le cas adaptatif, si bien qu'il ne peut y avoir de procédure consistante si

$$\exists\, \tau \in \mathscr{P}, \quad \rho_\sigma(\tau) < \rho^*_\sigma(\tau).$$

On dira qu'on ne perd pas de performance quand il y a toujours égalité dans la précédente relation.

Cependant, le premier résultat de Spokoiny (1996) est que l'adaptation ne peut être réalisée sans perte de performance dans le modèle considéré. Plus précisément, si

$$\alpha + \beta < 1,$$

et

$$\rho_\sigma(\tau) = c\, t_\sigma \rho^*_\sigma(\tau)$$

avec

$$\begin{cases} c > 0, \\ t_\sigma \ll (\log\log\sigma^{-1})^{1/4}, \end{cases}$$

alors aucun test adaptatif ne peut vérifier

$$\begin{cases} \overline{\lim}_{\sigma\to 0}\, \alpha(\psi) \le \alpha, \\ \overline{\lim}_{\sigma\to 0}\, \beta(\psi, \mathscr{P}) \le \beta. \end{cases}$$

En d'autres termes, l'adaptation n'est pas possible sans perte de performance, et d'autre part, cette perte de performance, matérialisée par le facteur t_σ, est au moins de l'ordre de $(\log\log\sigma^{-1})^{1/4}$.

Le deuxième résultat montre qu'il existe une constante $c^* > 0$ et un test ψ^* telle que si

$$\rho_\sigma(\tau) = c^*(\log\log\sigma^{-1})^{1/4}\rho^*_\sigma(\tau),$$

alors

$$\begin{cases} \lim_{\sigma\to 0}\alpha(\psi^*) \to 0, \\ \lim_{\sigma\to 0}\beta(\psi^*,\mathscr{P}) \to 0. \end{cases}$$

Ainsi, le facteur optimal est effectivement de l'ordre de $(\log\log\sigma^{-1})^{1/4}$.

Ce premier résultat d'adaptation a ensuite été étendus à de nombreux autres problèmes, par exemple dans Baraud *et al.* (2005), Gayraud *et* Pouet (2005), Horowitz *et* Spokoiny (2001), etc.

Un autre principe d'adaptation peut également être mis en avant. Dans Lepski *et* Tsybakov (2000), les auteurs considèrent le problème de détection de signal dans le modèle 1 avec des classes de Sobolev ou de Hölder, et les hypothèses

$$\begin{cases} H_0 : \quad f(t_0) = 0, \\ H_1 : \quad |f(t_0)| \geq C\rho_\sigma, \end{cases}$$

où $t_0 \in (0,1)$ est fixé. Ils montrent alors que la vitesse minimax de séparation asymptotique est

$$\sigma^{2s/2s+1},$$

et que si on considère plutôt les hypothèses

$$\begin{cases} H_0 : \quad f = 0, \\ H_1 : \quad \|f\|_\infty \geq C\rho_\sigma, \end{cases}$$

la perte n'est que logarithmique, au sens où la vitesse minimax de séparation asymptotique est alors

$$(\sigma^2\sqrt{\log\sigma^{-1}})^{s/2s+1}.$$

Ce résultat peut être considéré comme une adaptation dans le sens où le second problème est similaire au premier, à la différence que, dans l'hypothèse alternative, aucune information n'est disponible sur t_0.

Ainsi, un manque d'information peut entraîner une perte de performance d'ordre $\log\log$ ou logarithmique. Nous utiliserons ce type de raisonnement dans le problème qui nous occupe dans la première partie de ce travail : dans notre cas, qui est une forme d'adaptation au paramètre de translation, nos résultats suggèrent que la perte est d'ordre logarithmique.

TEST MULTIPLE PAR ESTIMATION DE PERMUTATION

Dans la première partie de cette thèse, on énonce un problème de test correspondant à la mise en correspondance de deux descripteurs. Mais en pratique, c'est un grand nombre de descripteurs qu'il faut faire se correspondre par paire. Il s'agit donc d'effectuer le test individuel pour toutes les paires de descripteurs possibles, simultanément. Dans cette section, nous parlons de ces problèmes de test multiple et montrons qu'il est équivalent à l'estimation d'une permutation.

Supposons que nous disposons de deux images I et $I^\#$, auxquelles sont associés respectivement les ensembles de descripteurs

$$\begin{cases} I: & \theta_1,\dots,\theta_n, \\ I^\#: & \theta_1^\#,\dots,\theta_m^\#. \end{cases}$$

Le problème est alors de tester un ensemble $H_0^{i,j}$, $1 \le i \le n, 1 \le j \le m$, avec

$$H_0^{i,j}: \quad \theta_i = \theta_j^\#.$$

Une approche naïve serait de tester une à une les $n \times m$ hypothèses. Imaginons que l'on utilise une procédure (dite de Bonferroni) qui à chaque test commet une erreur de première espèce α, alors la procédure globale commet en moyenne $nm\alpha$ mauvais rejets. Dans les applications, ceci est prohibitif. D'autant que d'autres procédures, prenant en compte la structure particulière du problème, atteignent de meilleures performances.

Avant de parler de ces procédures, nous devons d'abord définir des notions précises d'erreur. Pour l'erreur de première espèce, on peut choisir par exemple

- la probabilité que le nombre de faux rejets dépasse un certain seuil fixé,
- la proportion moyenne de faux rejets,
- la probabilité que la proportion de faux rejets dépasse un certain seuil...

Roquain (2011) décrit des procédures de test adaptées à ces différents critères.

Pour le problème de la mise en correspondance globale de descripteurs, une procédure de test multiple, qui détermine les paires de points ayant le même descripteur, est présentée dans Lowe (2004). Supposons que l'on dispose d'observations bruitées $X_1,\dots,X_n,X_1^\#,\dots,X_m^\#$ de $\theta_1,\dots,\theta_n,\theta_1^\#,\dots,\theta_m^\#$. Cette procédure suit les étapes suivantes :

1. Déterminer les deux plus proches voisins de X_1 dans la deuxième image $I^\#$:

$$j_1 = \arg\min_k \|X_1 - X_k^\#\|_2,$$
$$j_2 = \arg\min_{k\neq j_1} \|X_1 - X_k^\#\|_2.$$

2. Si, pour un certain seuil δ fixé à l'avance,

$$\frac{\|X_1 - X_{j_1}^\#\|_2}{\|X_1 - X_{j_2}^\#\|_2} > \delta,$$

c'est-à-dire si $X_{j_1}^\#$ est significativement plus proche de X_1 que $X_{j_2}^\#$, alors on associe X_1 et $X_{j_1}^\#$.

3. Sinon, on abandonne X_1 et on ne lui associe aucun descripteur. On poursuit ainsi itérativement avec $X_2, X_3,\dots$

Cette procédure obtient de bons résultats en pratique, et l'algorithme a connu un grand succès dans le domaine de la vision par ordinateur. Cependant, sa performance n'a fait l'objet d'aucune étude statistique. A fortiori, on ne peut pas être sûr qu'il n'existe pas de meilleure procédure.

Dans la deuxième partie de notre travail, nous donnons un cadre mathématique précis au problème de la mise en correspondance de descripteurs et recherchons des solutions optimales. Plus particulièrement, nous nous intéressons au cas simple où les deux images représentent le même objet et contiennent les même descripteurs, c'est-à-dire

$$n = m \quad \text{et} \quad \{\theta_i\}_{i=1,\dots,n} = \{\theta_i^\#\}_{i=1,\dots,n}.$$

Alors, il existe (au moins) une permutation π^* telle que

$$\forall\ i = 1, \dots, n,\ \theta_i^\# = \theta_{\pi^*(i)},$$

et cette permutation est unique si et seulement si

$$\kappa \triangleq \min_{i \neq j} \|\theta_i - \theta_j\|_2 > 0.$$

Ainsi, le problème initial de test multiple peut être résolu par l'estimation d'une permutation quand $\kappa > 0$.

Nous montrons que cette quantité détermine aussi l'existence d'estimateurs consistants de la permutation π^* : il existe une valeur limite de κ à partir de laquelle il est possible d'effectuer la mise en correspondance de manière consistante. Le problème de la mise en correspondance de descripteurs est ainsi relié à la théorie des tests minimax, puisque κ s'apparente à une vitesse de séparation.

Première partie

Changement de descripteur et problème du test de décalage

CHAPTER

1 Detecting curve shifts with the generalized likelihood ratio

In this chapter, we explore the implications of a descriptor change in the SIFT algorithm. If we use the non-centered histograms, we need another matching criterion, *i.e.,* to match descriptors that are shifted from each other. This leads to a new statistical problem, for which we give a rigorous model.

But if descriptor matching inspired this work, it can also be applied to many other fields. So first, we review the literature on the related subject of curve registration (Section 1.1), and explain why this problem may be interesting for other applications. Then, we state the statistical model (Section 1.2), discuss it (Section 1.3) and we use the method of the generalized likelihood ratio to exhibit a well performing testing procedure (Section 1.4). In the next section, we establish the properties of this test, in particular Wilks' property (Section 1.5), confirming the validity of the generalized likelihood ratio approach. All our results are extended to the heteroscedastic case and to the multidimensional case in the next section (Section 1.6) Finally, we show some simulation results (Section 1.7). The proofs of the theorems are postponed to the end of this chapter (Section 1.8).

1.1 THE PROBLEM OF CURVE REGISTRATION

Recently, the problem of curve registration, and more particularly, some aspects of this problem related to nonparametric and semiparametric estimation, have received a lot of attention in statistical research. In the related papers, the model used for deriving statistical inference assumes that the input data consist of a finite collection of noisy signals possessing the following characteristic: each input signal is obtained from a given signal, termed mean template or structural pattern, by a parametric deformation and by adding a white noise. In what follows, we will refer to this as the "deformed mean template" model. The interest for this model is justified by the difficulties to develop statistical inference, caused by the nonlinearity of the deformations and the fact that not only the deformations but also the mean template that was used to generate the observed data are unknown.

Let us give the mathematical formulation of the "deformed mean template" model in full generality. We are given a sample of size n of noisy signals $\{Y_m; m = 1, \dots, n\}$ having common structural pattern f, *i.e.*, for every $x \in [0,1]^d$ and every $m = 1, \dots, n$,

$$dY_m(x) = f\big(\phi(x, \tau_m)\big)\, dx + \sigma_m\, dW_m(x), \tag{1.1}$$

with

$$\begin{cases} \phi \text{ a known function determining the type of the deformation,} \\ \tau_m \text{ is a finite-dimensional parameter allowing to instantiate the deformation.} \end{cases}$$

Typical examples are:

(a) the shifted curve model $\phi(x, \tau) = x - \tau$, where $\tau \in \mathbb{R}^d$ is the shift parameter,

(b) the periodic signal model $\phi(x, \tau) = \tau x$, where the signal f is a 1-periodic univariate function, x is 1-dimensional and $\tau \in \mathbb{R}$ is the period of the noise-free signal,

(c) the rigid deformation model $\phi(x, \tau) = s(\mathrm{R}x + t)$, where $\tau = (s, \mathrm{R}, t)$ with $s > 0$ being the scale, R being the rotation and $t \in \mathbb{R}^d$ being the translation.

Starting from Golubev (1988) and Kneip *et* Gasser (1992), semiparametric and nonparametric estimations of the mathematical objects appearing in different instances of problem (1.1) have been intensively investigated, see for instance Rønn (2001), Dalalyan *et al.* (2006), Gamboa *et al.* (2007), Dalalyan (2007), Castillo *et* Loubes (2009), Bigot *et* Gadat (2010), Trigano *et al.* (2011), Castillo (2012) for the shifted curve model ; Härdle (1990), Carroll (1992), Vimond (2010) for a slightly extended case of affine transforms of shifted curves ; Castillo *et al.* (2006), Castillo (2007) for the periodic signal model and Bigot *et al.* (2009) for the rigid deformation model. More general deformations have been considered in Ramsay *et* Li (1998), Reilly *et al.* (2004), Gervini *et* Gasser (2005), Juditsky *et al.* (2009), Bigot *et al.* (2012) with applications to image warping in Glasbey *et* Mardia (2001), Bigot *et al.* (2009).

But, if a collection $\{Y_m\ ;\ m = 1, \dots, n\}$ of sample curves is available, and if we want to estimate the common template, the deformations or any other object involved in (1.1) like it is done in the quoted papers, it appears natural to check the appropriateness of model (1.1) first.

For example, to test the appropriateness of the shifted curve model or the slightly more general translated curve model with $n = 2$, we assume that two functions Y and $Y^\#$ are observed such that

$$\begin{cases} dY(x) = f(x)dx + \sigma dW(x), \\ dY^\#(x) = f^\#(x)dx + \sigma^\# dW^\#(x), \end{cases} \qquad x \in [0,1], \tag{1.2}$$

where

$$\begin{cases} W, W^\# \text{ are two independent Brownian motions,} \\ f, f^\# \text{ are two unknown 1-periodic signals,} \\ \sigma, \sigma^\# > 0 \text{ are positive parameters representing the noise magnitude.} \end{cases}$$

Then, the hypothesis we wish to test is that the curves f and $f^\#$ coincide, up to a shift of the argument and to a vertical translation:

$$\mathbf{H_0}: \quad \forall\, x \in [0,1], \quad f(x) = f^\#(x + \tau^*) + b^*, \tag{1.3}$$

for some unknown $(b^*, \tau^*) \in \mathbb{R} \times [0,1]$. If the null hypothesis $\mathbf{H_0}$ is accepted, then we are in the setting of model (1.1) for the particular case of deformation given by a shift and a translation. This model seems to be a very narrow subclass of models given by (1.1), but it plays a central role in several applications, for example:

ECG interpretation: An electro-cardiogram (ECG) can be seen as a collection of replica of nearly the same signal, up to a time shift. This signal contains significant informations about heart malformations or diseases that can be extracted by aligning the curves and computing their mean value. But using improperly aligned curves would lead to a smoothed signal where small abnormalities can not be observed. For more details we refer to Trigano *et al.* (2011), where random shifts are considered, and they are estimated along with their common distribution in the asymptotic of a growing number of curves.

Road traffic forecast: In Loubes *et al.* (2006), a road traffic forecasting procedure is introduced. For this, archetypes of the different types of road trafficking behavior on the Parisian highway network are built, using a hierarchical classification method. In each obtained cluster, the curves all represent the same events, only randomly shifted in time. The mean of the unshifted curves is more significant of a given behavior than the mean of the shifted ones, and hence provides more efficient predictions.

So, this problem, that will be our concern in this chapter, possesses a vast field of possible applications, and our work is an original contribution to its resolution. Indeed, the problem of estimating the parameters of the deformation is a semiparametric one, since the deformation involves a finite number of parameters that have to be estimated by assuming that the unknown mean template is merely a nuisance parameter. But our testing problem of detecting shifts is nonparametric: both the null hypothesis and the alternative in the context of the present study are nonparametric, *i.e.*, the parameter describing the probability distribution of the observations is infinite-dimensional not only under the alternative but also under the null hypothesis.

The statistical literature on this type of testing problems is very scarce. If Gayraud *et* Pouet (2001) and Horowitz *et* Spokoiny (2001) analyze the optimality and the adaptivity of testing procedures in the setting of a parametric null hypothesis against a nonparametric alternative, to the best of our knowledge, the only papers concerned with nonparametric null hypotheses are Baraud *et al.* (2003; 2005) and Gayraud *et* Pouet (2005). Moreover, the results derived in Baraud *et al.* (2003; 2005) are inapplicable in our set-up since the null hypothesis in our problem is neither linear nor convex. The set-up of Gayraud *et* Pouet (2005) is closer to ours. However, their theoretical framework comprises a condition on the sup-norm-entropy of the null hypothesis, which is irrelevant in our set-up and may be violated.

1.2 Model and notation

In the following, we consider the curve registration problem in which the data $\{Y(x) : x \in [0,1]\}$ and $\{Y^\#(x) : x \in [0,1]\}$ are available, generated by the Gaussian white noise model

$$\begin{cases} dY(x) = f(x)\,dx + \sigma\,dW(x), \\ dY^\#(x) = f^\#(x)\,dx + \sigma^\#\,dW^\#(x), \end{cases} \qquad x \in [0,1], \tag{1.4}$$

where $(W, W^\#)$ is a two-dimensional Brownian motion. It is implicitly assumed that f and $f^\#$ are squared integrable, which makes model (1.4) sensible. Moreover, we consider that the noise levels σ and $\sigma^\#$ are known and focus on the hypotheses testing problem stemming from the curve registration set-up.

The Gaussian white noise model is equivalent to the Gaussian sequence model obtained by projecting the processes Y and $Y^\#$ onto the Fourier basis:

$$\begin{cases} Y_j = c_j + \sigma\epsilon_j, \\ Y_j^\# = c_j^\# + \sigma^\#\epsilon_j^\#, \end{cases} \qquad j \geq 0, \tag{1.5}$$

where

$$\begin{cases} c_j = \int_0^1 f(x)\,e^{2\mathrm{i}j\pi x}\,dx, \\ c_j^\# = \int_0^1 f^\#(x)\,e^{2\mathrm{i}j\pi x}\,dx, \end{cases} \qquad j \geq 0,$$

are the complex Fourier coefficients. The complex valued random variables ϵ_j, $\epsilon_j^\#$ are i.i.d. standard Gaussian:

$$\epsilon_j, \epsilon_j^\# \sim \mathcal{N}_{\mathbb{C}}(0,1),$$

which means that their real and imaginary parts are independent $\mathcal{N}(0,1)$ random variables.

We are interested in testing the hypothesis (1.3), which translates in the Fourier domain to

$$\mathbf{H_0}: \quad \forall\, j \geq 1, \quad c_j = e^{-\mathrm{i}j\tau^*} c_j^\#, \tag{1.6}$$

for some unknown $\tau^* \in [0, 2\pi[$. Indeed, one easily checks that the projection onto the functions $e^{2\mathrm{i}j\pi x}$ cancels the term b^* in (1.3), resulting in (1.6). Furthermore, if (1.6) is verified, then b^* can be recovered by the formula

$$b^* = c_0 - c_0^\#.$$

If no additional assumptions are imposed on the functions f and $f^{\#}$, or equivalently on their Fourier coefficients

$$\begin{cases} \boldsymbol{c} \triangleq (c_0, c_1, \cdots), \\ \boldsymbol{c}^{\#} \triangleq (c_0^{\#}, c_1^{\#}, \ldots), \end{cases}$$

the nonparametric testing problem has no consistent solution. A natural assumption widely used in nonparametric statistics is that $\boldsymbol{c}$ and $\boldsymbol{c}^{\#}$ belong to some Sobolev ball

$$\mathscr{F}_{s,L} = \Big\{ \boldsymbol{u} = (u_0, u_1, \ldots) : \sum_{j=0}^{+\infty} j^{2s} |u_j|^2 \leq L^2 \Big\},$$

where the positive real numbers s and L stand for the smoothness and the radius of the class $\mathscr{F}_{s,L}$.

Finally, our testing problem can be reformulated in the Fourier domain as follows:

$$\begin{cases} \mathbf{H_0}: & d(\boldsymbol{c}, \boldsymbol{c}^{\#}) = 0, \\ \mathbf{H_1}: & d(\boldsymbol{c}, \boldsymbol{c}^{\#}) \geq \rho, \end{cases} \tag{1.7}$$

for some $\rho > 0$, with

$$d^2(\boldsymbol{c}, \boldsymbol{c}^{\#}) \triangleq \min_{\tau \in \mathbb{R}} \sum_{j=1}^{+\infty} |c_j - e^{-\mathrm{i}j\tau} c_j^{\#}|^2.$$

In other words, under $\mathbf{H_0}$ the graph of the function $f^{\#}$ is obtained from that of f by a translation.

1.3 Discussion of the model

The choice of our model was inspired by practical considerations, and we intend to apply it to a problem in computer vision: that of keypoint matching. Accordingly, it is necessary to justify the realism of model (2.2).

First, our choice of the Gaussian sequence model is not restrictive, since this model is equivalent in Le Cam's sense to many other models, including Gaussian white noise, density estimation $\big($*cf.* Nussbaum (1996)$\big)$, nonparametric regression $\big($*cf.* Brown *et* Low (1996), in the case of random design in Reiß (2008), in the case of nonGaussian noise in Grama *et* Nussbaum (1998) and Grama *et* Nussbaum (2002)$\big)$, ergodic diffusion $\big($*cf.* Dalalyan *et* Reiß (2006)$\big)$. On the other hand, the Gaussian noise is accepted in computer vision as a good approximation of the Poisson noise, that is more natural in this context.

Variance

In particular, the procedure that we propose admits a simple counterpart in the regression model, at least in the case of deterministic equidistant design. According to the theory on the asymptotic equivalence, our results hold true for this model as well, provided that $s > 1/2$ $\big($*cf.* Rohde (2004)$\big)$. However, in the model of regression, it is not realistic to assume that the noise level is known in advance.

Nevertheless, one can compute a consistent estimator of the variance $\big($*cf.* Rice (1984)$\big)$ and plug this estimator in the testing procedure. In an analogous setup, it is proved in Gayraud *et* Pouet (2001) for example, that this plug-in strategy preserves the rate-optimality of the testing procedure. We believe that a similar result can be deduced in our set-up as well.

Symmetry of the model

In our modelization, the two parts corresponding in the Gaussian white noise model to two different functions are treated symmetrically: the same model, with the same variance and the same noise, applies to both. Indeed, in applications, the signals that we want to match with each other are thought to have the same nature. In addition, it seems that it is not meaningful to consider the case when the regularities of the Sobolev balls are different for the signals: under $\mathbf{H_0}$, the regularity has to be the same.

Besides, one could want to normalize both equations to get the same variance for both sides. But, this would also change the functions, which would not only differ from each other by a shift, but also by a dilatation. Therefore, the application of our methodology to this case is not straightforward.

In the heteroscedastic case, *i.e.,* when σ and $\sigma^{\#}$ are a priori different, the difficulty of testing is no longer only dependent on σ, but also on $\sigma^{\#}$. In fact, this difficulty is captured by the parameter

$$\sigma_* = \max(\sigma, \sigma^{\#}),$$

which is the biggest constraint exercised in the model. Theorems 1 and 2 still hold if we replace σ by σ_*.

1.4 Penalized Likelihood Ratio Test

Because of the Gaussian nature of the noise, the negative log-likelihood of the parameters $\boldsymbol{u}, \boldsymbol{u}^{\#}$ given the data $\boldsymbol{Y}, \boldsymbol{Y}^{\#}$ is

$$\ell(\boldsymbol{u}, \boldsymbol{u}^{\#}) = \frac{1}{2\sigma^2}\|\boldsymbol{Y} - \boldsymbol{u}\|_2^2 + \frac{1}{2(\sigma^{\#})^2}\|\boldsymbol{Y}^{\#} - \boldsymbol{u}^{\#}\|_2^2. \tag{1.8}$$

Hence, the classical likelihood ratio test is based on the test statistic

$$\Delta(\boldsymbol{Y}, \boldsymbol{Y}^{\#}) = \min_{\boldsymbol{u}\in\mathscr{F}_{s,L}} \min_{\tau\in\mathbb{R}} \ell(\boldsymbol{u}, e(\tau)\circ\boldsymbol{u}) - \min_{\boldsymbol{u},\boldsymbol{u}^{\#}\in\mathscr{F}_{s,L}} \ell(\boldsymbol{u}, \boldsymbol{u}^{\#}), \tag{1.9}$$

where $e(\tau) = (1, e^{-\mathrm{i}\tau}, e^{-2\mathrm{i}\tau}, \dots)$ and $\circ$ denotes the component-by-component multiplication.

The computations lead to a problem of optimization under a Sobolev constraint, which can be solved by the method of the Lagrange multipliers. For the second term in (1.9), we get

$$\min_{\boldsymbol{u},\boldsymbol{u}^{\#}\in\mathscr{F}_{s,L}} \ell(\boldsymbol{u}, \boldsymbol{u}^{\#}) = \frac{1}{2\sigma^2}\sum_{j\geq 1}\frac{\omega_{1,j}}{1+\omega_{1,j}}|Y_j|^2 + \frac{1}{2(\sigma^{\#})^2}\sum_{j\geq 1}\frac{\omega_{2,j}}{1+\omega_{2,j}}|Y_j^{\#}|^2, \tag{1.10}$$

where $\boldsymbol{\omega}_1$, $\boldsymbol{\omega}_2$ are functions of the Lagrange multipliers, that are fixed in the optimization process.

The idea of the generalized likelihood ratio is to take the liberty of choosing the parameters $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ differently from those that are fixed in the classical approach.

Now, we note that we can get a result similar to (1.10) by computing the penalized likelihood for a deterministic sequence $\boldsymbol{\omega}$ of weights:

$$p\ell(\boldsymbol{u},\boldsymbol{u}^{\#}) = \min_{\boldsymbol{u},\boldsymbol{u}^{\#}\in\mathscr{F}_{s,L}} \left\{ \frac{1}{2\sigma^2}\Big(\|\boldsymbol{Y}-\boldsymbol{u}\|_2^2+\sum_{j\geq 1}\omega_j|u_j|^2\Big) + \frac{1}{2(\sigma^{\#})^2}\Big(\|\boldsymbol{Y}^{\#}-\boldsymbol{u}^{\#}\|_2^2+\sum_{j\geq 1}\omega_j|u_j^{\#}|^2\Big)\right\}. \tag{1.11}$$

The only difference is that the random sequences $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ were both replaced by the deterministic sequence $\boldsymbol{\omega}$. A similar remark can be done for

$$\min_{\boldsymbol{u}\in\mathscr{F}_{s,L}} \ell(\boldsymbol{u}, e(\tau)\circ\boldsymbol{u}).$$

Hence, we can relax the definition of the classical likelihood ratio statistic by replacing $\ell(\boldsymbol{u},\boldsymbol{u}^{\#})$ by $p\ell(\boldsymbol{u},\boldsymbol{u}^{\#})$ in (1.9) for some deterministic sequence $\boldsymbol{\omega}$ to be chosen later. We get the generalized likelihood ratio statistic defined as

$$\tilde{\Delta}(\boldsymbol{Y},\boldsymbol{Y}^{\#}) = \min_{\boldsymbol{u}\in\mathscr{F}_{s,L}}\min_{\tau\in\mathbb{R}} p\ell(\boldsymbol{u},e(\tau)\circ\boldsymbol{u}) - \min_{\boldsymbol{u},\boldsymbol{u}^{\#}\in\mathscr{F}_{s,L}} p\ell(\boldsymbol{u},\boldsymbol{u}^{\#}). \tag{1.12}$$

The complete computations of the test statistic gives

$$\tilde{\Delta}(\boldsymbol{Y},\boldsymbol{Y}^{\#}) = \frac{\sigma^2+(\sigma^{\#})^2}{2(\sigma\sigma^{\#})^2}\min_{\tau\in[0,2\pi]}\sum_{j=1}^{+\infty}\frac{|Y_j - e^{-\mathrm{i}j\tau}Y_j^{\#}|^2}{1+\omega_j}. \tag{1.13}$$

$\tilde{\Delta}(\boldsymbol{Y},\boldsymbol{Y}^{\#})$ is always non-negative. Furthermore, it is small when $\mathbf{H_0}$ is satisfied and is large if $\mathbf{H_0}$ is violated.

From now on, it will be more convenient to use the notation

$$\nu_j = \frac{1}{1+\omega_j}.$$

The elements of the sequence $\boldsymbol{\nu} = \{\nu_j;\ j\geq 1\}$ are hereafter referred to as shrinkage weights. They are allowed to take any value between 0 and 1. Even the value 0 will be authorized, corresponding to the limit case when $w_j = +\infty$, or equivalently to our belief that the corresponding Fourier coefficient is 0. Finally, denoting

$$\|\boldsymbol{u}\|_{2,\boldsymbol{\nu}}^2 = \sum_{j=1}^{+\infty}\nu_j|u_j|^2,$$

we get the test statistic

$$\tilde{\Delta}(\boldsymbol{Y},\boldsymbol{Y}^{\#}) = \frac{\sigma^2+(\sigma^{\#})^2}{2(\sigma\sigma^{\#})^2}\min_{\tau\in[0,2\pi]}\|\boldsymbol{Y}-\boldsymbol{e}(\tau)\circ\boldsymbol{Y}^{\#}\|_{2,\boldsymbol{\nu}}^2, \tag{1.14}$$

and our goal is to find the asymptotic distribution of this quantity under the null hypothesis.

1.5 RESULTS

The test based on the generalized likelihood ratio statistic involves a sequence $\boldsymbol{\nu}$, which is completely modulable by the user. However, we are able to provide theoretical guarantees only under some conditions on these weights. This is no surprise, considering the interpretation of the penalization method that was applied: for instance, it is natural to think that $\boldsymbol{\nu}$ has to converge to 0, as $\boldsymbol{\omega}$ has to converge to $+\infty$.

For simplicity sake, we focus on the case

$$\sigma = \sigma^{\#}$$

and choose a positive integer

$$N = N_\sigma \geq 2,$$

which represents the number of Fourier coefficients involved in our testing procedure. In addition to requiring that

$$\forall\ j \geq 1, \quad 0 \leq \nu_j \leq 1,$$

we assume that:

$$\begin{aligned}
&\textbf{(A)} \quad \nu_1 = 1 \quad \text{and} \quad \mathbf{supp}(\boldsymbol{\nu}) \subset \{1, \dots, N_\sigma\},\\
&\textbf{(B)} \quad \exists\, \underline{c} > 0,\ \sum_{j\geq 1} \nu_j^2 \geq \underline{c}\, N_\sigma.
\end{aligned}$$

Moreover, we will use the following condition in the proof of the consistency of the test:

$$\textbf{(C)} \quad \exists\, \overline{c} > 0, \quad \min\{j \geq 0, \nu_j < \overline{c}\} \to +\infty, \quad \text{as } \sigma \to 0.$$

In simple words, this condition implies that the number of terms ν_j that are above a given strictly positive level goes to $+\infty$ as σ converges to 0. If $N_\sigma \to +\infty$ as $\sigma \to 0$, then all the aforementioned conditions are satisfied for the shrinkage weights $\boldsymbol{\nu}$ of the form

$$\nu_{j+1} = h(j/N_\sigma),$$

where $h : \mathbb{R} \to [0,1]$ is an integrable function, supported on $[0,1]$, continuous in 0 and satisfying $h(0) = 1$. The classical examples of shrinkage weights include:

$$\nu_j = \begin{cases} \mathbb{1}_{\{j \leq N_\sigma\}}, & \\ \left\{1 + \left(\frac{j}{\kappa N_\sigma}\right)^{\mu}\right\}^{-1} \mathbb{1}_{\{j \leq N_\sigma\}}, & \kappa > 0, \mu > 1, \\ \left\{1 - \left(\frac{j}{N_\sigma}\right)^{\mu}\right\}_+, & \mu > 0. \end{cases} \tag{1.15}$$

These examples correspond respectively to
- the projection weights,
- the Tikhonov weights,
- the Pinsker weights.

Note that condition **(C)** is satisfied in all these examples with $\overline{c} = 0.5$, or any other value in $(0,1)$. Here on, we write $\tilde{\Delta}_\sigma(\boldsymbol{Y}, \boldsymbol{Y}^{\#})$ instead of $\tilde{\Delta}(\boldsymbol{Y}, \boldsymbol{Y}^{\#})$ in order to stress its dependence on σ.

Theorem 1 (Wilks' property)**.** *Assume that*

- $\boldsymbol{c} \in \mathscr{F}_{1,L}$ *and* $|c_1| > 0$.
- $\boldsymbol{\nu}$ *satisfies conditions* ***(A)*** *and* ***(B)***,
- $N_\sigma \to +\infty$ *and* $\sigma^2 N_\sigma^{5/2} \log(N_\sigma) \to 0$.

Then, under the null hypothesis, the test statistic $\tilde{\Delta}_\sigma(\boldsymbol{Y}, \boldsymbol{Y}^\#)$ *is asymptotically distributed as a Gaussian random variable:*

$$\frac{\tilde{\Delta}_\sigma(\boldsymbol{Y}, \boldsymbol{Y}^\#) - 4\|\boldsymbol{\nu}\|_1}{4\|\boldsymbol{\nu}\|_2} \xrightarrow{\mathscr{L}} \mathscr{N}(0,1). \tag{1.16}$$

The main outcome of this result is a test of hypothesis $\mathbf{H_0}$ that is asymptotically of a prescribed significance level $\alpha \in (0,1)$. Indeed, let us define the test that rejects $\mathbf{H_0}$ if and only if

$$\tilde{\Delta}_\sigma(\boldsymbol{Y}, \boldsymbol{Y}^\#) \geq 4\|\boldsymbol{\nu}\|_1 + 4z_{1-\alpha}\|\boldsymbol{\nu}\|_2, \tag{1.17}$$

where $z_{1-\alpha}$ is the $(1-\alpha)$-quantile of the standard Gaussian distribution. We can derive the properties of this test.

Corollary 1. *The test of hypothesis* $\mathbf{H_0}$ *defined by the critical region (1.17) is asymptotically of significance level* α.

Remark 1. The signification of the assumption "$c_1 > 0$" in Theorem 1 is that in this case, there is only one possible shift (modulo 2π).

Remark 2. Let us consider the case of projection weights

$$\nu_j = \mathbb{1}(j \leq N_\sigma).$$

One can reformulate the asymptotic relation stated in Theorem 1 by claiming that

$$\frac{1}{2}\tilde{\Delta}_\sigma(\boldsymbol{Y}, \boldsymbol{Y}^\#) \overset{\mathscr{L}}{\approx} \mathscr{N}(2N_\sigma, 4N_\sigma).$$

Since the latter distribution approaches the chi-squared distribution, we get:

$$\frac{1}{2}\,\tilde{\Delta}_\sigma(\boldsymbol{Y}, \boldsymbol{Y}^\#) \overset{\mathscr{L}}{\approx} \chi^2_{2N_\sigma}, \qquad \text{as} \qquad \sigma \to 0.$$

In the case of general shrinkage weights satisfying the assumptions stated in the beginning of this section, an analogous relation holds as well:

$$\frac{\|\boldsymbol{\nu}\|_1}{2\|\boldsymbol{\nu}\|_2^2}\,\tilde{\Delta}_\sigma(\boldsymbol{Y}, \boldsymbol{Y}^\#) \overset{\mathscr{L}}{\approx} \chi^2_{2\|\boldsymbol{\nu}\|_1^2/\|\boldsymbol{\nu}\|_2^2}, \qquad \text{as} \qquad \sigma \to 0.$$

This type of results are often referred to as Wilks' property.

Remark 3. The p-value of the aforementioned test based on the Gaussian or chi-squared approximation can be used as a measure of the goodness-of-fit or, in other words, as a measure of alignment for the pair of curves under consideration. If the observed two noisy curves lead to the data $\boldsymbol{y}, \boldsymbol{y}^\#$, then the (asymptotic) p-value is defined as

$$\alpha^* = \Phi\Big(\frac{\tilde{\Delta}_\sigma(\boldsymbol{y}, \boldsymbol{y}^\#) - 4\|\boldsymbol{\nu}\|_1}{4\|\boldsymbol{\nu}\|_2}\Big),$$

where Φ stands for the c.d.f. of the standard Gaussian distribution.

The previous theorem was focused on the behavior of the test under the null without paying attention on what happens under the alternative. The next theorem establishes the consistency of the test defined by the critical region (1.17).

Theorem 2. *Assume that*
- $\boldsymbol{v}$ *satisfies condition* **(C)**,
- $\sigma^2 N_\sigma \to 0$ *as* $\sigma \to 0$.

Then the test statistic

$$T_\sigma = \frac{\tilde{\Delta}_\sigma(\boldsymbol{Y}, \boldsymbol{Y}^\#) - 4\|\boldsymbol{v}\|_1}{4\|\boldsymbol{v}\|_2}$$

diverges under $\mathbf{H_1}$, *i.e.,*

$$T_\sigma \xrightarrow{P} +\infty, \qquad as \qquad \sigma \to 0.$$

In other words, the result above claims that the power of the test defined via (1.17) is asymptotically equal to 1 as the noise level σ decreases to 0.

1.6 Extension to the multidimensional case

Our results also extend to the multidimensional case, *i.e.*, when the signals f and $f^\#$ take their values in $\mathbb{R}^d$, with $d \geq 1$. Let us now redefine the model. We dispose of the observations

$$\begin{cases} Y_j = c_j + \sigma\epsilon_j, \\ Y_j^\# = c_j^\# + \sigma^\#\epsilon_j^\#, \end{cases} \qquad j \geq 0,$$

where

$$\begin{cases} \forall\ j \geq 1, \quad c_j, c_j^\# \in \mathbb{C}^d, \\ \sigma, \sigma^\# \text{ are diagonal matrices of size } d \text{ with positive diagonal coefficients.} \end{cases}$$

and the random variables ϵ_j, $\epsilon_j^\#$ are such that their components are i.i.d. standard Gaussian:

$$\epsilon_{j,i}, \epsilon_{j,i}^\# \sim \mathcal{N}_{\mathbb{C}}(0,1),$$

which means that their real and imaginary parts are independent $\mathcal{N}(0,1)$ random variables.

We are interested in testing the hypothesis

$$\mathbf{H_0}: \quad \forall\ j \geq 1, \quad c_j = e^{-\mathrm{i}j\tau^*} c_j^\#, \tag{1.18}$$

for some unknown $\tau^* \in [0, 2\pi[$, and we assume that each component of $\boldsymbol{c}_i, \boldsymbol{c}_i^\#, i = 1, \dots, d$ belongs to a Sobolev class

$$\mathscr{F}_{s_i, L_i} = \Big\{\boldsymbol{u} = (u_0, u_1, \dots) : \sum_{j=0}^{+\infty} j^{2s_i}|u_j|^2 \leq L_i^2\Big\}.$$

The method of the generalized likelihood ratio leads, for a given weights sequence $\boldsymbol{v}$, to the statistic

$$\tilde{\Delta}_{\sigma,\sigma^\#}(\boldsymbol{Y}, \boldsymbol{Y}^\#) = \min_{\tau\in[0,2\pi]} \sum_{j=1}^{+\infty} \|(\sigma^2 + (\sigma^\#)^2)^{-1/2}(Y_j - e^{-\mathrm{i}j\tau} Y_j^\#)\|_{2,\nu}^2. \tag{1.19}$$

Now, denoting
$$\sigma_* = \max\left(\|\sigma\|_\infty, \|\sigma^\#\|_\infty\right),$$
we obtain similar theorems as in the 1-dimensional case.

Theorem 3. *Assume that*
- $\forall\ i = 1,\dots,d,\ \boldsymbol{c}_i \in \mathscr{F}_{1,L_i}$ *and* $|c_{i,1}| > 0$.
- $\boldsymbol{v}$ *satisfies conditions **(A)** and **(B)**,*
- $N_{\sigma_*} \to +\infty$ *and* $\sigma_*^2 N_{\sigma_*}^{5/2}\log(N_{\sigma_*}) \to 0$.

Then, under the null hypothesis, the test statistic $\tilde\Delta_{\sigma,\sigma^\#}(\boldsymbol{Y},\boldsymbol{Y}^\#)$ *is asymptotically distributed as a Gaussian random variable:*
$$\frac{\tilde\Delta_\sigma(\boldsymbol{Y},\boldsymbol{Y}^\#) - 4d\|\boldsymbol{v}\|_1}{4\sqrt{d}\|\boldsymbol{v}\|_2} \xrightarrow{\mathscr{L}} \mathcal{N}(0,1). \tag{1.20}$$

Under the alternative hypothesis:

Theorem 4. *Assume that*
- $\boldsymbol{v}$ *satisfies condition **(C)**,*
- $\sigma_*^2 N_{\sigma_*} \to 0$ *as* $\sigma_* \to 0$.

Then the test statistic $T_{\sigma,\sigma^\#} = \frac{\tilde\Delta_{\sigma,\sigma^\#}(\boldsymbol{Y},\boldsymbol{Y}^\#) - 4d\|\boldsymbol{v}\|_1}{4\sqrt{d}\|\boldsymbol{v}\|_2}$ *diverges under* $\mathbf{H_1}$*, i.e.,*
$$T_{\sigma,\sigma^\#} \xrightarrow{P} +\infty, \qquad as \qquad \sigma_* \to 0.$$

1.7 Numerical experiments

In this section, we carry out some numerical experiments on synthetic data. They prove that the methodology described in this chapter is applicable, and illustrate how the different characteristics of the testing procedure, such as the significance level, the power, etc, depend on the noise level σ and on the shrinkage weights $\boldsymbol{v}$. The testing procedure (1.17) was implemented in Matlab.

Convergence of the test under $\mathbf{H_0}$ and the influence of the shrinkage weights

In order to illustrate the convergence of the test (1.14) when σ tends to 0, we made the following experiment. We chose the function HeaviSine, considered as a benchmark in the signal processing community, and computed its complex Fourier coefficients $\{c_j; j = 0,\dots,10^6\}$. For each value of σ taken from the set $\{2^{-k/2}, k = 1,\dots,15\}$, we repeated 5000 times the following computations:
- following the assumptions required by our theoretical results, set $N_\sigma = 50\sigma^{-1/2}$,
- generate the noisy sequence $\{Y_j; j = 0,\dots,N_\sigma\}$ by adding to $\{c_j\}$ an i.i.d. $\mathcal{N}_\mathbb{C}(0,\sigma^2)$ sequence $\{\xi_j\}$,
- randomly choose a parameter τ^* from the uniform distribution on $[0,2\pi]$, independently of $\{\xi_j\}$,
- generate the shifted noisy sequence $\{Y_j^\#; j = 0,\dots,N_\sigma\}$ by adding to $\{e^{\mathrm{i}j\tau^*}c_j\}$ an i.i.d. $\mathcal{N}_\mathbb{C}(0,\sigma^2)$ sequence $\{\xi_j^\#\}$, independent of $\{\xi_j\}$ and of τ^*,
- compute the three values of the test statistic $\tilde\Delta_\sigma$ corresponding to the classical shrinkage weights defined by (1.15) and compare these values with the threshold for $\alpha = 5\%$.

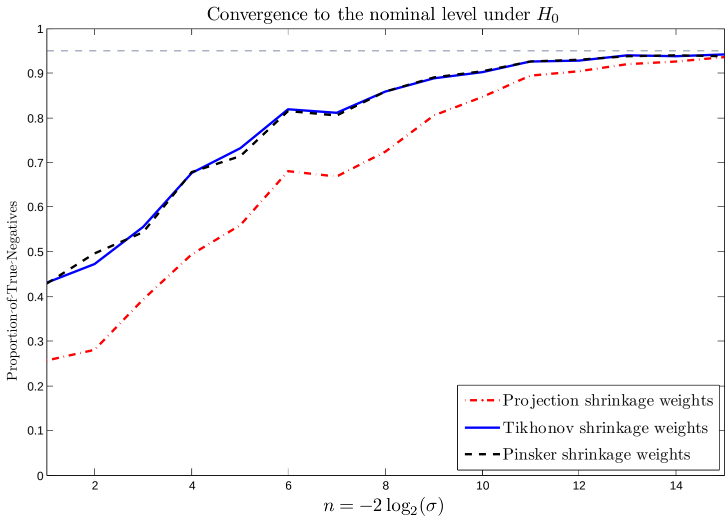

Figure 1.1 – *The proportion of true negatives in the experiment described in Section 1.7 as a function of* $\log_2 \sigma^{-2}$ *for three different shrinkage weights: projection (left), Tikhonov (middle) and Pinsker (right). One can observe that for* $\sigma \approx 5 \times 10^{-3}$, *the proportion of true negatives is almost equal to the nominal level* 0.95. *Another observation is that the Pinsker and the Tikhonov weights lead to a faster convergence to the nominal significance level.*

We denote by $p^{\mathrm{proj}}_{\mathrm{accept}}(\sigma)$, $p^{\mathrm{Tikh}}_{\mathrm{accept}}(\sigma)$ and $p^{\mathrm{Pinsk}}_{\mathrm{accept}}(\sigma)$ the proportion of experiments (among 10^3 that have been realized) leading to a value of the corresponding test statistic lower than the threshold, *i.e.,* the proportion of experiments leading to the acceptance of the null hypothesis. We plotted in Figure 1.1 the (linearly interpolated) curves

$$\begin{cases} k \longmapsto p^{\mathrm{proj}}_{\mathrm{accept}}(\sigma_k), \\ k \longmapsto p^{\mathrm{Tikh}}_{\mathrm{accept}}(\sigma_k, \\ k \longmapsto p^{\mathrm{Pinsk}}_{\mathrm{accept}}(\sigma_k), \end{cases}$$

with $\sigma_k = 2^{-k/2}$. It can be seen that for $\sigma = 2^{-7} \approx 8 \times 10^{-3}$, the proportion of true negatives is almost equal to the nominal level 0.95. It is also worth noting that the three curves are quite comparable, with a significant advantage for the curves corresponding to Pinsker's and Tikhonov's weights: these curves converge faster to the level $1 - \alpha = 95\%$ than the curve corresponding to the projection weights.

Power of the test

In the previous experiment, we illustrated the behavior of the penalized likelihood ratio test under the null hypothesis. The aim of the second experiment is to show what happens under the alternative. To this end, we still use the HeaviSine function as signal f and define

$$f^{\#} = f + \gamma\varphi,$$

where γ is a real parameter. Two cases are considered:

– $\varphi(t) = c\cos(4t)$,
– $\varphi(t) = c/(1+t^2)$,

where c is a constant ensuring that ϕ has an L^2 norm equal to 1. For each of these two pairs of functions $(f, f^{\#})$, we repeated 5000 times the following computations:

– set $\sigma = 1$ and $N_\sigma = 50\sigma^{-1/2}$,
– compute the complex Fourier coefficients $\{c_j; j = 0,\dots,10^6\}$ and $\{c_j^{\#}; j = 0,\dots,10^6\}$ of f and $f^{\#}$, respectively,
– generate the noisy sequence $\{Y_j; j = 0,\dots,N_\sigma\}$ by adding to $\{c_j\}$ an i.i.d. $\mathcal{N}_{\mathbb{C}}(0,\sigma^2)$ sequence $\{\xi_j\}$,
– generate the shifted noisy sequence $\{Y_j^{\#}; j = 0,\dots,N_\sigma\}$ by adding to $\{c_j^{\#}\}$ an i.i.d. $\mathcal{N}_{\mathbb{C}}(0,\sigma^2)$ sequence $\{\xi_j^{\#}\}$, independent of $\{\xi_j\}$,
– compute the value of the test statistic $\tilde{\Delta}_\sigma$ corresponding to the projection weights and compare this value with the threshold for $\alpha = 5\%$.

To show the behavior of the test under $\mathbf{H_1}$ when the distance between the null and the alternative changes, we computed for each γ the proportion of true positives, also called the empirical power, among the 5000 random samples we have simulated. The results, plotted in Figure 1.2 show that even for moderately small values of γ, the test succeeds in taking the correct decision. It is a bit surprising that the result for the case $\varphi(t) = c\cos(4t)$ is not better than that for $\varphi(t) = c/(1+t^2)$. Indeed, one can observe that the curve at the right panel approaches 1 much faster than the curve at the left panel.

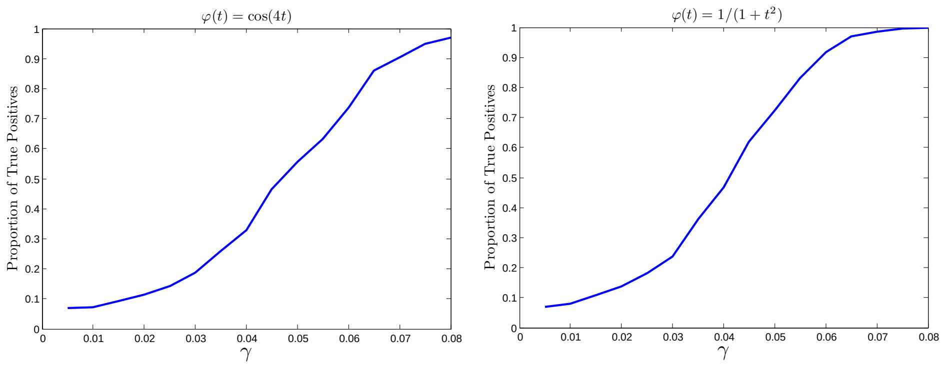

Figure 1.2 – *The proportion of true positives in the experiment described in Section 1.7 as a function of the parameter γ measuring the distance between the true parameter and the set of parameters characterizing the null hypothesis. The main observation is that both curves tend to* 1 *very fast.*

1.8 Proofs of the theorems

The proof of Wilks' property is divided into several parts. First we assume that $\mathbf{H_0}$ is true and study the convergence of the pseudo-estimator $\hat{\tau}$ (of the shift τ^*) defined as the maximizer of the log-likelihood over the interval $[\tau^* - \pi, \tau^* + \pi]$. Here, τ^* is an element of $[0, 2\pi[$ such that

$$\forall\ j \geq 1, \quad c_j = e^{-\mathrm{i}j\tau^*} c_j^{\#}.$$

Maximizer of the log-likelihood

Proposition 5. *Define $\hat{\tau}$ as the solution to the optimization problem*

$$\hat{\tau} = \arg \max_{|\tau-\tau^*|\leq\pi} M(\tau),$$

with

$$M(\tau) = \sum_{j\geq 1} \nu_j \operatorname{Re}(e^{ij\tau} Y_j \overline{Y_j^{\#}}).$$

Assume that

- $\boldsymbol{c} \in \mathscr{F}_{1,L}$ *and* $|c_1| > 0$,
- $\boldsymbol{\nu}$ *satisfies conditions* ***(A)*** *and* ***(B)***,

then, under $\mathbf{H_0}$, $\hat{\tau}$ *satisfies the asymptotic relation*

$$|\hat{\tau} - \tau^*| = \sigma\sqrt{\log N_\sigma}\big(1 + \sigma N_\sigma^{3/2}\big) O_P(1), \qquad as \quad \sigma \to 0.$$

Proof of Proposition 5. If we set

$$\begin{cases} \eta_j = e^{-ij\tau^*}\epsilon_j, \\ \eta_j^{\#} = \epsilon_j^{\#}, \end{cases}$$

we can write the decomposition

$$M(\tau) = \mathbf{E}[M(\tau)] + \sigma S(\tau) + \sigma^2 D(\tau + \tau^*),$$

where

$$\begin{cases} \mathbf{E}[M(\tau)] & = \sum_{j\geq 1} \nu_j |c_j|^2 \cos[j(\tau-\tau^*)], \\ S(\tau) & = \sum_{j\geq 1} \nu_j \operatorname{Re}\Big(e^{ij\tau}(\overline{c_j}\eta_j + c_j\overline{\eta_j^{\#}})\Big), \\ D(\tau) & = \sum_{j\geq 1} \nu_j \operatorname{Re}\big(e^{ij\tau}\eta_j\overline{\eta_j^{\#}}\big). \end{cases}$$

On the one hand, using the assumption $|c_1| > 0$ along with condition (A), we get that

$$\frac{\mathbf{E}\big[M(\tau)\big] - \mathbf{E}\big[M(\tau^*)\big]}{(\tau-\tau^*)^2} \leq -\nu_1 |c_1|^2 \frac{1-\cos(\tau-\tau^*)}{(\tau-\tau^*)^2} \leq -\frac{2|c_1|^2}{\pi^2}.$$

Denoting

$$C \triangleq \frac{2|c_1|^2}{\pi^2} > 0,$$

this yields

$$\begin{aligned} M(\tau) - M(\tau^*) &= \mathbf{E}[M(\tau)] - \mathbf{E}[M(\tau^*)] + \sigma\big[S(\tau) - S(\tau^*)\big] + \sigma^2\big[D(\tau) - D(\tau^*)\big] \\ &\leq -C|\tau-\tau^*|^2 + \sigma|\tau-\tau^*| \cdot \|S'\|_\infty + \sigma^2|\tau-\tau^*| \cdot \|D'\|_\infty \\ &= |\tau-\tau^*|\big\{\sigma\|S'\|_\infty + \sigma^2\|D'\|_\infty - C|\tau-\tau^*|\big\}. \end{aligned}$$

Using this result, we get for every $a > 0$

$$\begin{aligned} \mathbf{P}\big(|\hat{\tau} - \tau^*| > a\big) &\leq \mathbf{P}\Big\{\sup_{|\tau-\tau^*|>a} M(\tau) - M(\tau^*) \geq 0\Big\} \\ &\leq \mathbf{P}\Big\{\sup_{|\tau-\tau^*|>a} \big[\sigma\|S'\|_\infty + \sigma^2\|D'\|_\infty - C|\tau-\tau^*|\big] \geq 0\Big\} \\ &= \mathbf{P}\Big\{\sigma\|S'\|_\infty + \sigma^2\|D'\|_\infty \geq Ca\Big\}. \end{aligned}$$

Choosing

$$a = \sigma\sqrt{\log N_\sigma}\big(1+\sigma N_\sigma^{3/2}\big)z$$

with z some positive real number, it follows that

$$\mathbf{P}\big(|\hat\tau-\tau^*| > \sigma\sqrt{\log N_\sigma}\big(1+\sigma N_\sigma^{3/2}\big)z\big) \le \mathbf{P}\big(\|S'\|_\infty \ge Cz\sqrt{\log N_\sigma}\big) \\ + \mathbf{P}\big(\|D'\|_\infty \ge Cz\sqrt{N_\sigma^3\log N_\sigma}\big).$$

On the other hand, we can write

$$S'(t) = \sum_{j\ge1} j|c_j|v_j \operatorname{Re}\big(e^{\mathrm{i}j\tau}\zeta_j\big),$$

where ζ_j are i.i.d. complex valued random variable, whose real and imaginary parts are independent $\mathcal{N}(0,2)$ variables.

The large deviations of the sup-norm of S' can be controlled by using the following lemma.

Lemma 1. *Assume that*

$$F(t) = \sum_{j=0}^{K} s_j\{\cos(jt)\xi_j + \sin(jt)\xi'_j\},$$

where $\{\xi_j\}$ *and* $\{\xi'_j\}$ *are two independent sequences of i.i.d.* $\mathcal{N}(0,1)$ *random variables. Then*

$$\forall x>0,\quad \mathbf{P}(\|F\|_\infty \ge \|\boldsymbol{s}\|_2 x) \le (K+1)e^{-x^2/2}.$$

Using this lemma, the fact that $N_\sigma \ge 2$ and that

$$\sum_{j=1}^{+\infty} j^2|c_j|^2 \le L^2,$$

we get that

$$\mathbf{P}\big(\|S'\|_\infty \ge 2L\sqrt{2y\log N_\sigma}\big) \le 2N_\sigma^{1-y} \le 2^{2-y},$$

for every $y>1$. Finally, the large deviations of the term $\|D'\|_\infty$ are controlled by using Lemma 3 below.

Lemma 3. *Assume that*

- *N is some positive integer,*
- *$\eta_j, \eta_j^\#$, $j=1,\dots,N$ are independent complex valued random variables such that their real and imaginary parts are independent standard Gaussian variables,*
- *$\boldsymbol{s} = (s_1,\dots,s_N)$ is a vector of real numbers,*
- *F is defined by*

$$\forall\ t\in[0,2\pi],\ F(t) = \sum_{j=1}^{N} s_j \operatorname{Re}\big(e^{ijt}\eta_j\eta_j^\#\big),$$

- $\|F\|_\infty = \sup_{t\in[0,2\pi]}|S(t)|$.

Then,

$$\forall x,y>0,\quad \mathbf{P}\Big\{\|F\|_\infty > \sqrt{2}x\big(\|\boldsymbol{s}\|_2 + y\|\boldsymbol{s}\|_\infty\big)\Big\} \le (N+1)e^{-x^2/2} + e^{-y^2/2}.$$

Applied to D', this lemma yields

$$\forall\ x>0,\quad \mathbf{P}\big(\|D'\|_\infty > 4xN_\sigma^{3/2}\sqrt{\log N_\sigma}\big) \le 2N_\sigma^{1-x^2} + e^{-N_\sigma/2}.$$

Putting these inequalities together, we find that for any $\alpha \in (0,1)$, there exists $z>0$ such that

$$\mathbf{P}\big(|\hat\tau-\tau^*|>\sigma\sqrt{\log N_\sigma}\big(1+\sigma N_\sigma^{3/2}\big)z\big) \le \alpha.$$

In conclusion, we get that

$$\hat\tau-\tau^* = O_P\Big(\sigma\sqrt{\log N_\sigma}\big(1+\sigma N_\sigma^{3/2}\big)\Big).$$

□

Proof of Theorem 1

Under $\mathbf{H_0}$, there exists $\tau^* \in [0,2\pi[$ such that

$$\forall\ j\ge 1,\quad c_j = e^{-\mathrm{i}j\tau^*}c_j^{\#}.$$

We can decompose the test statistic as follows

$$\begin{aligned}\tilde\Delta_\sigma(\boldsymbol{Y},\boldsymbol{Y}^{\#}) &= \frac{1}{\sigma^2}\min_{\tau\in[0,2\pi[}\left[\sum_{j=1}^{+\infty}\nu_j|Y_j-e^{-\mathrm{i}j\tau}Y_j^{\#}|^2\right]\\ &= \frac{1}{\sigma^2}\big\{D_\sigma(\hat\tau)+2C_\sigma(\hat\tau)+P_\sigma(\hat\tau)\big\},\end{aligned}\tag{1.21}$$

where we have used the notation:

$$\begin{cases} D_\sigma(\tau) \;=\; \sum_{j=1}^{+\infty}\nu_j|c_j|^2|1-e^{-\mathrm{i}j(\tau-\tau^*)}|^2, \quad \text{(deterministic term)}\\ C_\sigma(\tau) \;=\; \sigma\sum_{j=1}^{+\infty}\nu_j\,\mathrm{Re}\big[c_j\big(1-e^{-\mathrm{i}j(\tau-\tau^*)}\big)\big(\overline{\epsilon_j-e^{-\mathrm{i}j\tau}\epsilon_j^{\#}}\big)\big], \quad \text{(cross term)}\\ P_\sigma(\tau) \;=\; \sigma^2\sum_{j=1}^{+\infty}\nu_j|\epsilon_j-e^{-\mathrm{i}j\tau}\epsilon_j^{\#}|^2. \quad \text{(principal term)}\end{cases}$$

We denote by $\hat\tau$ the pseudo-estimator of τ^* defined as the minimizer of the right-hand side of (1.21) and study the asymptotic behavior of the terms D_σ, C_σ and P_σ separately.

– For the deterministic term, it holds that

$$\begin{aligned}|D_\sigma(\hat\tau)| &\le \sum_{j=1}^{+\infty} j^2\nu_j|c_j|^2(\hat\tau-\tau^*)^2 \le (\hat\tau-\tau^*)^2\sum_{j=1}^{+\infty}j^2|c_j|^2\\ &\le L^2(\hat\tau-\tau^*)^2\\ &= \{\sigma^2(1+\sigma^2N_\sigma^3)\log N_\sigma\}\times O_p(1).\end{aligned}$$

– Let us turn now to the cross term. It holds that

$$\begin{aligned}C_\sigma(\tau) = \sigma\sum_{j=1}^{+\infty}\nu_j\Big\{&(1-\cos[j(\tau-\tau^*)])\,\mathrm{Re}\left[c_j\big(\overline{\epsilon_j-e^{-\mathrm{i}j\tau^*}\epsilon_j^{\#}}\big)\right]\\ &+\sin[j(\tau^*-\tau)]\,\mathrm{Im}\left[c_j\big(\overline{\epsilon_j+e^{-\mathrm{i}j\tau^*}\epsilon_j^{\#}}\big)\right]\Big\}.\end{aligned}$$

Thus, as $C_\sigma(\tau^*) = 0$, we have

$$|C_\sigma(\hat\tau)| \leq |\hat\tau - \tau^*| \cdot \|C'_\sigma\|_\infty.$$

Using again Lemma 1 and Lemma 3 like in the proof of Proposition 5, we check that

$$\|C'_\sigma\|_\infty = O_P\big(\sigma\sqrt{\log N_\sigma}\big).$$

Therefore, it holds that

$$|C_\sigma(\hat\tau)| = \{\sigma^2(1+\sigma N_\sigma^{3/2})\log N_\sigma\} \times O_p(1).$$

– Let us now study the last term,

$$P_\sigma(\tau) = \sigma^2 \sum_{j=1}^{+\infty} \nu_j |\epsilon_j - e^{-\mathrm{i}j\tau}\epsilon_j^\#|^2,$$

which will determine the asymptotic behavior of the test statistic. Now, denoting

$$\begin{cases} \eta_j = e^{\mathrm{i}j\tau^*}\epsilon_j, \\ \eta_j^\# = \epsilon_j^\#, \end{cases}$$

we can rewrite this term as

$$P_\sigma(\tau) = \sigma^2 \sum_{j=1}^{+\infty} \nu_j |\eta_j - e^{-\mathrm{i}j(\tau-\tau^*)}\eta_j^\#|^2.$$

We are going to prove that under $\mathbf{H_0}$, if
– $\boldsymbol{\nu}$ satisfies conditions **(A)** and **(B)**,
– $N_\sigma \to +\infty$ and $\sigma^2 N_\sigma^{5/2}\log(N_\sigma) \xrightarrow{P} 0$,
then

$$T_\sigma(\hat\tau) = \frac{P_\sigma(\hat\tau) - 4\sigma^2\sum_{k\geq 1}\nu_k}{4\sigma^2(\sum_{k\geq 1}\nu_k^2)^{1/2}} \xrightarrow{\mathscr{L}} \mathcal{N}(0,1).$$

To check this property, we decompose the principal term as follows:

$$T_\sigma(\hat\tau) = T_\sigma(\tau^*) + \underbrace{\frac{P_\sigma(\hat\tau) - P_\sigma(\tau^*)}{4\sigma^2(\sum_{k\geq 1}\nu_k^2)^{\frac{1}{2}}}}_{R_\sigma(\hat\tau)}.$$

We start by writing $T_\sigma(\tau^*)$ as

$$T_\sigma(\tau^*) = \sum_{j=1}^{N_\sigma} X_{j,\sigma},$$

with

$$X_{j,\sigma} = \frac{\nu_j(|\eta_j - \eta_j^\#|^2 - 4)}{4(\sum_{k\geq 1}\nu_k^2)^{\frac{1}{2}}},$$

and we apply the Berry-Esseen inequality (*cf.* Petrov (1995, Theorem 5.4)), which is possible since the $X_{j,\sigma}$'s are independent random variables with mean 0 and finite third moment. Furthermore, we have

$$B_\sigma \triangleq \sum_{j=1}^{N_\sigma} \mathbf{Var}\big(X_{j,\sigma}\big) = 1$$

and
$$L_\sigma \triangleq B_\sigma^{-\frac{3}{2}} \sum_{j=1}^{N_\sigma} \mathbf{E}|X_{j,\sigma}|^3 \le c\, N_\sigma^{-\frac{1}{2}},$$

for some positive constant c. Therefore, the Berry-Esseen inequality yields

$$\sup_x |F_\sigma(x) - \Phi(x)| \le K\, L_\sigma,$$

where
$$F_\sigma(x) = \mathbf{P}\big(B_\sigma^{-\frac{1}{2}} \sum_{j=1}^{N_\sigma} X_{j,\sigma} < x\big),$$

Φ is the c.d.f. of the standard Gaussian distribution and K is an absolute constant. Hence
$$T_\sigma(\tau^*) \xrightarrow{\mathscr{L}} \mathscr{N}(0,1).$$

It remains now to prove that
$$R_\sigma \xrightarrow{P} 0,$$

which, in view of Slutski's lemma, will be sufficient to complete the proof. It holds that

$$\begin{aligned}
R_\sigma(\tau) &= \sum_{j=1}^{+\infty} \frac{\nu_j}{2(\sum_{j=1}^{+\infty} \nu_j^2)^{\frac{1}{2}}} \operatorname{Re} \eta_j \overline{\eta_j^{\#}}\, (e^{\mathrm{i}j(\tau-\tau^*)} - 1) \\
&= \sum_{j=1}^{N_\sigma} \frac{j\nu_j(\tau-\tau^*)}{2(\sum_{j=1}^{+\infty} \nu_j^2)^{\frac{1}{2}}} \operatorname{Re}\big(e^{\mathrm{i}jt} \eta_j \overline{\eta_j^{\#}}\big),
\end{aligned}$$

with t some real number between τ and τ^*. Then, by virtue of Lemma 3,

$$\begin{aligned}
|R_\sigma(\hat\tau)| &\le \frac{|\hat\tau - \tau^*|}{2(\sum_{j=1}^{+\infty} \nu_j^2)^{\frac{1}{2}}} \sup_{t\in[0,2\pi]} \Big| \sum_{j=1}^{N_\sigma} j\nu_j \operatorname{Re}\big(e^{\mathrm{i}jt} \eta_j \overline{\eta_j^{\#}}\big)\Big| \\
&= \{\sigma(1+\sigma N_\sigma^{3/2}) N_\sigma \log N_\sigma\} \cdot O_P(1).
\end{aligned}$$

Hence, $R_\sigma(\hat\tau) \xrightarrow{P} 0$ and the desired result follows.

Proof of Theorem 2

Let us study the test statistic

$$T_\sigma = \frac{\tilde\Delta_\sigma(\boldsymbol{Y}, \boldsymbol{Y}^{\#}) - 4\|\boldsymbol{\nu}\|_1}{4\|\boldsymbol{\nu}\|_2},$$

and show that it tends to $+\infty$ in probability under $\mathbf{H_1}$. Actually, the hypothesis $\mathbf{H_1}$ will be supposed to be satisfied throughout this section. It holds true that:

$$\begin{aligned}
\tilde\Delta_\sigma(\boldsymbol{Y}, \boldsymbol{Y}^{\#}) &= \frac{1}{\sigma^2} \min_{\tau\in[0,2\pi]} \sum_{j=1}^{+\infty} \nu_j |Y_j - e^{-\mathrm{i}j\tau} Y_j^{\#}|^2 \\
&= \frac{1}{\sigma^2} \min_{\tau\in[0,2\pi]} \sum_{j\ge1} \nu_j \Big| (c_j - e^{-\mathrm{i}j\tau} c_j^{\#}) + \sigma(\epsilon_j - e^{-\mathrm{i}j\tau}\epsilon_j^{\#}) \Big|^2,
\end{aligned}$$

so that

$$\tilde{\Delta}_\sigma(\boldsymbol{Y},\boldsymbol{Y}^\#) \geq \frac{1}{\sigma^2}\min_{\tau\in[0,2\pi]}\Big\{\sum_{j\geq 1}\nu_j|c_j - e^{-\mathrm{i}j\tau}c_j^\#|^2\Big\}$$
$$-\frac{2}{\sigma}\max_{\tau\in[0,2\pi]}\Big\{\sum_{j\geq 1}\nu_j|c_j - e^{-\mathrm{i}j\tau}c_j^\#|\cdot|\epsilon_j - e^{-\mathrm{i}j\tau}\epsilon_j^\#|\Big\}.$$

Let us focus on the first term. Denoting

$$\delta_\sigma = \min\{j\geq 1, \nu_j < \overline{c}\},$$

we get by condition **(C)** that

$$\delta_\sigma \to +\infty.$$

which implies that

$$\begin{aligned}\min_{\tau\in[0,2\pi]}\sum_{j\geq 1}\nu_j|c_j - e^{-\mathrm{i}j\tau}c_j^\#|^2 &\geq \overline{c}\min_{\tau\in[0,2\pi]}\sum_{j=1}^{\delta_\sigma}|c_j - e^{-\mathrm{i}j\tau}c_j^\#|^2\\ &\geq \overline{c}\left(\min_{\tau\in[0,2\pi]}\sum_{j=1}^{+\infty}|c_j - e^{-\mathrm{i}j\tau}c_j^\#|^2 - 4L^2\delta_\sigma^{-2}\right)\\ &\geq \overline{c}\rho^2 + o_P(1).\end{aligned}$$

Now, the second term satisfies

$$\begin{aligned}\max_{\tau\in[0,2\pi]}\sum_{j\geq 1}\nu_j|c_j - e^{-\mathrm{i}j\tau}c_j^\#|\cdot|\epsilon_j - e^{-\mathrm{i}j\tau}\epsilon_j^\#| &\leq \max_{j=1,\ldots,N_\sigma}(|\epsilon_j|\vee|\epsilon_j^\#|)\sum_{j\geq 1}^{N_\sigma}\Big(|c_j| + |c_j^\#|\Big)\\ &= O_P(\sqrt{\log N_\sigma})\sum_{j\geq 1}^{N_\sigma}\Big(|c_j| + |c_j^\#|\Big),\end{aligned}$$

and since, if $\boldsymbol{u}\in\mathscr{F}_{1,L}$,

$$\sum_{j\geq 1}^{N_\sigma}|u_j| \leq \Big(\sum_{j\geq 1}^{N_\sigma}j^{-2}\Big)^{1/2}\Big(\sum_{j\geq 1}^{N_\sigma}j^2|u_j|^2\Big)^{1/2} \leq \frac{\pi L}{\sqrt{6}},$$

we get

$$\max_{\tau\in[0,2\pi]}\sum_{j\geq 1}\nu_j|c_j - e^{-\mathrm{i}j\tau}c_j^\#|\cdot|\epsilon_j - e^{-\mathrm{i}j\tau}\epsilon_j^\#| = O_P(\sqrt{\log N_\sigma}).$$

Putting it all together, we get

$$T_\sigma = \frac{\tilde{\Delta}_\sigma(\boldsymbol{Y},\boldsymbol{Y}^\#) - 4\|\boldsymbol{\nu}\|_1}{4\|\boldsymbol{\nu}\|_2} \geq \frac{\overline{c}\rho^2 + O_P\Big(\sigma\sqrt{\log N_\sigma}\Big) + O_P(\sigma^2 N_\sigma)}{4\sigma^2\sqrt{N_\sigma}} \xrightarrow{P} +\infty,$$

using the fact that

$$\sigma^2 N_\sigma \xrightarrow{P} 0.$$

1.9 Bounds for the maxima of random sums

In the proofs of the theorems, we needed some concentration inequalities, that we gather here with their proofs. First, the following proposition gives a control on the supremum of a continuum of Gaussian variables.

Proposition (Berman (1988))**.** *Assume that g_j are continuously differentiable functions satisfying*

$$\forall\ t, \quad \sum_{j=1}^{n} g_j(t)^2 = 1$$

and

$$\xi_j \overset{iid}{\sim} \mathcal{N}(0,1).$$

Then, for every $x > 0$,

$$\mathbf{P}\Big(\sup_{[a,b]} \sum_{j=1}^{n} g_j(t)\xi_j \geq x\Big) \leq \frac{L_0}{2\pi} e^{-\frac{x^2}{2}} + \int_x^{+\infty} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}\, dt,$$

with

$$L_0 = \int_a^b \Big[\sum_{j=1}^{n} g_j'(t)^2\Big]^{1/2} dt.$$

We used a direct consequence of this proposition:

Lemma 1. *Assume that*

$$F(t) = \sum_{j=0}^{K} s_j \{\cos(jt)\xi_j + \sin(jt)\xi_j'\},$$

where $\{\xi_j\}$ and $\{\xi_j'\}$ are two independent sequences of i.i.d. $\mathcal{N}(0,1)$ random variables. Then

$$\forall x > 0, \quad \mathbf{P}(\|F\|_\infty \geq \|\boldsymbol{s}\|_2 x) \leq (K+1)e^{-x^2/2}.$$

On the other hand, we needed the following fact about moderate deviations of the random variables that can be written as the sum of squares of independent centered Gaussian random variables.

Lemma 2. *Assume that*

- *N is some positive integer,*
- *$\eta_j^{\#}$, $j = 1, \ldots, N$ are independent complex valued random variables such that their real and imaginary parts are independent standard Gaussian variables,*
- *$\boldsymbol{s} = (s_1, \ldots, s_N)$ is a vector of real numbers.*

Then, for any $y \geq 0$,

$$\mathbf{P}\Big\{\sum_{j=1}^{N} s_j^2 |\eta_j^{\#}|^2 \geq 2\|\boldsymbol{s}\|_2^2 + 2\sqrt{2}\|\boldsymbol{s}\|_4^2 y + 2\|\boldsymbol{s}\|_\infty^2 y^2\Big\} \leq e^{-y^2/2},$$

with the standard notations

$$\begin{cases} \|\boldsymbol{s}\|_\infty = \max\limits_{j=1,\ldots,N} |s_j|, \\ \|\boldsymbol{s}\|_q^q = \sum_{j=1}^{N} |s_j|^q. \end{cases}$$

Proof of Lemma 2. This is a direct consequence of Laurent *et* Massart (2000, Lemma 1). □

Berman's proposition and the last lemma have a more elaborate consequence:

Lemma 3. *Assume that*

- *N is some positive integer,*
- *$\eta_j, \eta_j^\#$, $j = 1, \ldots, N$ are independent complex valued random variables such that their real and imaginary parts are independent standard Gaussian variables,*
- *$\boldsymbol{s} = (s_1, \ldots, s_N)$ is a vector of real numbers,*
- *F is defined by*

$$\forall\ t \in [0, 2\pi],\ F(t) = \sum_{j=1}^{N} s_j \operatorname{Re}\left(e^{ijt} \eta_j \eta_j^\#\right),$$

- $\|F\|_\infty = \sup_{t\in[0,2\pi]} |S(t)|$.

Then,

$$\forall x, y > 0, \quad \mathbf{P}\Big\{\|F\|_\infty > \sqrt{2}x\big(\|\boldsymbol{s}\|_2 + y\|\boldsymbol{s}\|_\infty\big)\Big\} \leq (N+1)e^{-x^2/2} + e^{-y^2/2}.$$

Proof of Lemma 3. First note that we can not directly use Berman's formula, since the summands are not Gaussian. However, they are conditionally Gaussian if the conditioning is done, for example, with respect to the sequence $\{\eta_j^\#\}$. Indeed,

$$\sum_{j=1}^{N} s_j \operatorname{Re}\left(e^{ij\tau}\eta_j\eta_j^\#\right) \Big| (\eta_j^\#) \sim \sum_{j=1}^{N} s_j |\eta_j^\#| \big(\cos(j\tau)\xi_j - \sin(j\tau)\xi'_j\big)$$

with

$$\xi_j, \xi'_j \overset{iid}{\sim} \mathcal{N}(0,1).$$

It follows by Lemma 1 that

$$\mathbf{P}\Big(\sup_{[0,2\pi]} \Big| \sum_{j=1}^{N} s_j \operatorname{Re}\left(e^{ij\tau}\eta_j\eta_j^\#\right)\Big| \geq x\Big(\sum_{j=1}^{N} s_j^2 |\eta_j^\#|^2\Big)^{\frac{1}{2}} \Big| (\eta_j^\#)\Big) \leq (N+1)\exp\Big(-\frac{x^2}{2}\Big).$$

Let us now denote by ζ the nonnegative random variable such that

$$\zeta^2 = \sum_{j=1}^{N} s_j^2 |\eta_j^\#|^2.$$

It holds that for all $a > 0$,

$$\begin{aligned}
\mathbf{P}(\|F\|_\infty \geq ax) &= \mathbf{P}(\|F\|_\infty \geq ax;\ \zeta \leq a) + \mathbf{P}(\|F\|_\infty \geq ax;\ \zeta > a)\\
&\leq \mathbf{P}(\|F\|_\infty \geq x\zeta) + \mathbf{P}(\zeta > a)\\
&\leq (N+1)e^{-x^2/2} + \mathbf{P}(\zeta > a).
\end{aligned}$$

To complete the proof, it suffices to replace a by

$$\sqrt{2}(\|\boldsymbol{s}\|_2 + y\|\boldsymbol{s}\|_\infty)$$

and to apply Lemma 2 along with the inequalities

$$\begin{aligned}
\|\boldsymbol{s}\|_2 + \|\boldsymbol{s}\|_\infty y &= (\|\boldsymbol{s}\|_2^2 + 2\|\boldsymbol{s}\|_\infty\|\boldsymbol{s}\|_2 y + \|\boldsymbol{s}\|_\infty^2 y^2)^{1/2}\\
&\geq (\|\boldsymbol{s}\|_2^2 + \sqrt{2}\|\boldsymbol{s}\|_4^2 y + \|\boldsymbol{s}\|_\infty^2 y^2)^{1/2}.
\end{aligned}$$

□

CONCLUSION OF THE CHAPTER

In this chapter, we considered the problem of shift testing, motivated by possible applications in computer vision. We found a rigorous statistical model and a hypothesis testing problem that seems to convey the main difficulties of shift testing: we assumed that the noise that corrupts the signals is additive and white Gaussian. In this context, we used the generalized likelihood ratio method which is appropriate in a nonparametric context, as pointed out in Fan *et al.* (2001). It was proven in this chapter that a weighted l^2-penalization of the likelihood leads to statistically relevant testing procedures, provided that the weights satisfy some mild assumptions. In particular, these tests possess Wilks' property, *i.e.,* their asymptotic distribution under the null hypothesis is independent of the nuisance parameters.

However, the question remains open as to determining the minimax rate of separation between the null hypothesis and the alternative. Our proofs suggest that this rate is not slower than $s^{1/2}(\log \sigma^{-1})^{1/4}$. However, it is very likely that this latter rate is suboptimal. There is a large body of literature on the topic of minimax rates of separation (*cf.* Ingster *et* Suslina (2003) and references therein), but they mainly concentrate on the case of a simple null hypothesis. At this stage, we expect that the composite character of the null hypothesis in our set-up will slow down the rate of convergence at least by a logarithmic factor. The adaptive choice of the tuning parameter N_σ is another central issue. These questions will be answered in the next chapter.

CHAPTER

2 Minimax study of the problem of shift testing

In the previous chapter, we proposed a model for testing shifts between curves, then a class of consistent testing procedures possessing the important Wilks' property. However, it is not possible to rank these procedures among all possible tests. That is why we wanted to give a minimax frame for the problem of shift testing, *i.e.,* a criterion to organize all testing procedures into a hierarchy and to look for an optimal test.

In this chapter, we use the minimax approach to measure the performance of the testing procedures, and the tool of the minimax rate of separation. We find the optimal minimax rate, up to a possible logarithmic factor, and at least nearly optimal testing procedures. Moreover, as these procedures depend on the parameters of the regularity set, which is unpractical, we also look for the optimal adaptive minimax rate and optimal adaptive procedures.

More precisely, we present our model, the minimax frame and the problem of adaptation in Section 2.1. Then, we exhibit testing procedures and their performances in the nonadaptive case (Section 2.2), and in the adaptive case (Section 2.3). The optimality of these procedures is discussed in Section 2.4. Finally, Section 2.5 and 2.6 gather the proofs of the theorems and of the lemmas.

2.1 Description of the problem

Shifted curve model

This paper deals with the shifted curve model, which we will state in a Gaussian sequence form, but which originally relates on two 2π-periodic functions

$$f, f^\# \in \mathbb{L}_2.$$

Expanding these functions in the complex Fourier basis, we get

$$\begin{cases} f(t) = \sum_{j=-\infty}^{+\infty} c_j(f) e^{\mathrm{i}jt}, \\ f^\#(t) = \sum_{j=-\infty}^{+\infty} c_j(f^\#) e^{\mathrm{i}jt}, \end{cases} \qquad t \in [0, 2\pi],$$

where

$$\begin{cases} c_j(f) = \frac{1}{2\pi}\int_0^{2\pi} f(t) e^{-\mathrm{i}jt}\, dt, \\ c_j(f^\#) = \frac{1}{2\pi}\int_0^{2\pi} f^\#(t) e^{-\mathrm{i}jt}\, dt. \end{cases}$$

With this notation, if f and $f^\#$ only differ from each other by a shift, then the Fourier coefficients satisfy

$$\forall\ j \neq 0, \quad c_j(f^\#) = e^{\mathrm{i}j\tau} c_j(f),$$

for some real τ in $[0, 2\pi]$. Hence, denoting

$$\begin{cases} \boldsymbol{c} = (c_0, c_1, \ldots), \\ \boldsymbol{c}^\# = (c_0^\#, c_1^\#, \ldots), \end{cases}$$

if we introduce the pseudo-distance d such that

$$d^2(\boldsymbol{c}, \boldsymbol{c}^\#) \triangleq \inf_\tau \sum_{j=1}^{+\infty} |c_j - e^{-\mathrm{i}j\tau} c_j^\#|^2, \tag{2.1}$$

and given that

$$\forall\ j \geq 1, \quad c_j(f) = \overline{c_{-j}(f)},$$

testing that f was shifted from $f^\#$ amounts to testing if

$$d(\boldsymbol{c}, \boldsymbol{c}^\#) = 0.$$

Now, if we assume that the observations are given by the white noise model

$$\begin{cases} dY(t) = f(t)\, dt + \sigma\, dW(t), \\ dY^\#(t) = f^\#(t)\, dt + \sigma\, dW^\#(t), \end{cases} \qquad t \in [0, 2\pi],$$

where $\sigma > 0$ and $W, W^\#$ are independent standard Wiener processes, we can state our model in a more convenient Gaussian sequence form:

$$\begin{cases} Y_j = c_j + \sigma \xi_j \\ Y_j^\# = c_j^\# + \sigma \xi_j^\# \end{cases}, \qquad j = 1, 2, \ldots, \tag{2.2}$$

where

– $\{\xi_j, \xi_j^\#; j = 1, 2, \dots\}$ is a family of independent complex random variables, whose real and imaginary parts are independent standard Gaussian variables,
– σ is assumed to be known.

Our problem amounts to testing $\mathbf{H_0}$ against $\mathbf{H_1}$ with

$$\begin{cases} \mathbf{H_0}: & d(\boldsymbol{c}, \boldsymbol{c}^\#) = 0, \\ \mathbf{H_1}: & d(\boldsymbol{c}, \boldsymbol{c}^\#) \geq C\rho_\sigma, \end{cases} \tag{2.3}$$

where

$$\begin{cases} C \text{ is a positive constant,} \\ \rho_\sigma \text{ is a sequence of positive real numbers.} \end{cases}$$

Working with Fourier coefficients, it seems natural to use the Sobolev regularity, and so we assume that $\boldsymbol{c}$ and $\boldsymbol{c}^\#$ belong under the alternative to a Sobolev ball

$$\mathscr{F}_{s,L} \triangleq \Big\{ \boldsymbol{u} = (u_1, u_2, \dots) \mid \|\boldsymbol{u}^{(s)}\|_2^2 \triangleq \sum_{j=1}^{\infty} j^{2s} |u_j|^2 \leq L^2 \Big\}, \tag{2.4}$$

with $s > 0$. With this notation, we denote Θ_0 and Θ_1 the parameter sets corresponding to the hypotheses $\mathbf{H_0}$ and $\mathbf{H_1}$, $\boldsymbol{Y}$ and $\boldsymbol{Y}^\#$ the sequences $(Y_1, Y_2, \dots)$ and $(Y_1^\#, Y_2^\#, \dots)$, and we call $\mathbf{P}_{\boldsymbol{c},\boldsymbol{c}^\#}$ the probability engendered by $(\boldsymbol{Y}, \boldsymbol{Y}^\#)$ when the underlying Fourier coefficients are $\boldsymbol{c}$ and $\boldsymbol{c}^\#$.

Minimax testing

A randomized test in our model is a random variable taking values in $[0, 1]$ and measurable with respect to the σ-algebra engendered by $(\boldsymbol{Y}, \boldsymbol{Y}^\#)$. In practice, the user simulates an independent random variable with a Bernoulli distribution of parameter the value of the test, which was computed from the data $(\boldsymbol{Y}, \boldsymbol{Y}^\#)$. The null hypothesis is accepted, respectively rejected, when the result of the simulation is 0 or 1. We say that a test is nonrandomized when it only takes the values 0 or 1.

To measure the performance of a test ψ, we choose the minimax point of view, in which the errors of first and second kinds are defined by

$$\begin{cases} \alpha(\psi, \Theta_0) = \sup_{\Theta_0} \mathbf{E}_{\boldsymbol{c},\boldsymbol{c}^\#}\big(\psi\big), \\ \beta(\psi, \Theta_1) = \sup_{\Theta_1} \mathbf{E}_{\boldsymbol{c},\boldsymbol{c}^\#}\big(1 - \psi\big). \end{cases} \tag{2.5}$$

Note that in the nonrandomized case,

$$\begin{cases} \alpha(\psi, \Theta_0) = \sup_{\Theta_0} \mathbf{P}_{\boldsymbol{c},\boldsymbol{c}^\#}(\psi = 1), \\ \beta(\psi, \Theta_1) = \sup_{\Theta_1} \mathbf{P}_{\boldsymbol{c},\boldsymbol{c}^\#}(\psi = 0). \end{cases}$$

We say that consistent testing in the asymptotic minimax sense is possible if for all $\alpha, \beta > 0$, there exists a test ψ_σ such that

$$\begin{cases} \overline{\lim\limits_{\sigma \to 0}}\, \alpha(\psi_\sigma, \Theta_0) \leq \alpha, \\ \overline{\lim\limits_{\sigma \to 0}}\, \beta(\psi_\sigma, \Theta_1) \leq \beta. \end{cases} \tag{2.6}$$

The distance between the null and the alternative hypotheses, $C\rho_\sigma$, determines the existence of such tests. Indeed, if $C\rho_\sigma$ is too small, no testing procedure is asymptotically better than a blind guess, for which

$$\alpha(\psi, \Theta_0) + \beta(\psi, \Theta_1) = 1.$$

For a fixed pair α, β, we call ρ_σ^* the asymptotic minimax separation rate if there are two positive constants C_* and C^* such that consistent testing is

– impossible for $\rho_\sigma = \rho_\sigma^*$ and $C < C_*$,
– and possible for $\rho_\sigma = \rho_\sigma^*$ and $C > C^*$.

The best constants C_* and C^* satisfying these conditions are called exact separation constants. Conventionally, one applies the informal minimal writing length rule to avoid nonuniqueness of the minimax separation rate and of these constants. Moreover, a test which is consistent when $\rho_\sigma = \rho_\sigma^*$ and for some $C > 0$ is called asymptotically minimax optimal.

Adaptive testing

A limitation of the minimax approach is that the optimal tests depend on the smoothness class to which the parameters belong. This is not convenient from a practical point of view, because the choice of the type of smoothness class and of the parameters of this class seems to be unnatural and arbitrary. To obtain handier procedures, we need an adaptive definition for hypothesis testing.

Prior to testing, some sets of smoothness parameters s, L must be chosen, over which adaptation is performed. Typically, these sets are taken as compact intervals $[s_1, s_2]$, $[L_1, L_2]$. To each couple of smoothness parameters (s, L), we associate the smoothness set $\mathscr{F}_{s,L}$, and we write $\Theta_0^{s,L}$ and $\Theta_1^{s,L}$ the sets corresponding to the null and alternative hypotheses. Note that, in our problem,

$$\Theta_0^{s,L} \equiv \Theta_0 = \{\boldsymbol{c}, \boldsymbol{c}^\# \in \mathbb{L}_2 \,|\, d(\boldsymbol{c}, \boldsymbol{c}^\#) = 0\},$$

is independent of the smoothness parameters, and that

$$\Theta_1^{s,L} = \{\boldsymbol{c}, \boldsymbol{c}^\# \in \mathscr{F}_{s,L} \,|\, d(\boldsymbol{c}, \boldsymbol{c}^\#) \geq C\rho_\sigma\}$$

depends on (s, L), not only because $\boldsymbol{c}$ and $\boldsymbol{c}^\#$ are in $\mathscr{F}_{s,L}$, but also since ρ_σ is allowed to be a function of s: as a matter of fact, $\Theta_1^{s,L}$ depends on the choice of the radius $C\rho_\sigma(s)$.

The easiest way to achieve adaptation is to use the test corresponding to the most constraining smoothness (s_1, L_2), but this entails a significant loss of efficiency if the underlying parameters are in fact smoother.

Thus, we prefer a more economical approach and we will say that consistent adaptive testing is possible uniformly over $s \in [s_1, s_2]$ and $L \in [L_1, L_2]$, if for all $\alpha, \beta > 0$, there is a test ψ_σ depending only on $s_1, s_2, L_1, L_2, \alpha$ and β such that

$$\begin{cases} \overline{\lim\limits_{\sigma \to 0}}\, \alpha(\psi_\sigma, \Theta_0) \leq \alpha, \\ \overline{\lim\limits_{\sigma \to 0}}\, \sup\limits_{s,L} \beta(\psi_\sigma, \Theta_1^{s,L}) \leq \beta. \end{cases} \tag{2.7}$$

However, adaptive testing is not always possible without loss of efficiency, *i.e.*, taking

$$\rho_\sigma(s) = \rho_\sigma^*(s)$$

for each s. That is why it was suggested in Spokoiny (1996) to replace σ by σd_σ in the expression of $\rho_\sigma^*(s)$, where d_σ is a sequence of positive real numbers, which can be seen as a necessary payment regarding the intensity of the noise to achieve adaptation.

Now, we say that $\rho^*_{\sigma d_\sigma}(s), s \in [s_1, s_2]$ is the adaptive asymptotic minimax separation rate if there are two positive constants C_* and C^* such that adaptive consistent testing is

- impossible for $\rho_\sigma(s) = \rho^*_{\sigma d_\sigma}(s)$ and $C < C_*$,
- possible for $\rho_\sigma(s) = \rho^*_{\sigma d_\sigma}(s)$ and $C > C^*$.

2.2 Nonadaptive testing procedure

Theorem

Here, we propose a testing procedure which is consistent when the separation rate is of order $(\sigma^2\sqrt{\log\sigma^{-1}})^{2s/4s+1}$. This rate will be proven to be minimax, up to a possible logarithmic factor. Indeed, no testing procedure is consistent for a separation rate smaller than $\sigma^{4s/4s+1}$, which is the rate of signal detection in the Gaussian sequence model when the signal to be detected belongs to a Sobolev ball and the separation from 0 is measured by the l_2-norm.

Our proposal is based on the generalized likelihood ratio tests presented in the previous chapter. The test statistic that we obtained was then

$$\frac{1}{\sigma^2}\min_{\tau\in[0,2\pi]}\sum_{j=1}^{+\infty}\nu_j|Y_j - e^{-\mathrm{i}j\tau}Y_j^\#|^2,$$

where $\boldsymbol{\nu}$ is a sequence of weights chosen by the user. In the followings, we choose these weights to be the projection weights

$$\forall\ j \le 1, \quad \nu_j = \mathbb{1}_{j\le N}.$$

Indeed, we believe that more complex sequences would only allow to win in the constant, but not in the minimax rate. And as we know this rate only up to a possible logarithmic factor, it makes little sense to optimize the constant. Furthermore, the first chapter gave only asymptotic bounds for N in order to get the desired Wilks' property. Hereafter, we will give precise values for N.

Hence, our test statistic, which can be interpreted as a standardized version of an estimator of $d(\boldsymbol{c}, \boldsymbol{c}^\#)$, is

$$\lambda_\sigma(N) = \frac{1}{4\sigma^2\sqrt{N}}\min_\tau\Big[\sum_{j=1}^{N}|Y_j - e^{-\mathrm{i}j\tau}Y_j^\#|^2\Big] - \sqrt{N}, \tag{2.8}$$

and

$$\psi_\sigma(N,q) = \mathbb{1}_{\{\lambda_\sigma(N)>q\}}, \tag{2.9}$$

for $N \in \mathbb{N}^*$ and $q \in \mathbb{R}$. Put into words, the test $\psi_\sigma(N,q)$ rejects the null hypothesis when the statistic $\lambda_\sigma(N)$ exceeds the threshold q and accepts it otherwise.

The following theorem establishes the minimax properties of this testing procedure for a proper choice of the tuning parameters.

Theorem 5. *Set*

$$\begin{cases} \Theta_0 = \Big\{(\boldsymbol{c}, \boldsymbol{c}^\#) \in l_2 \times l_2 \,|\, d(\boldsymbol{c}, \boldsymbol{c}^\#) = 0\Big\}, \\ \Theta_1 = \Big\{(\boldsymbol{c}, \boldsymbol{c}^\#) \in \mathscr{F}_{s,L} \times \mathscr{F}_{s,L} \,|\, d(\boldsymbol{c}, \boldsymbol{c}^\#) \ge C\rho_\sigma\Big\}, \end{cases} \tag{2.10}$$

with s and L some positive real numbers, and

$$\begin{cases} \rho_\sigma = \left(\sigma^2 \sqrt{\log \sigma^{-1}}\right)^{\frac{2s}{4s+1}}, \\ C^2 > 4L^2 c_{s,L}^{-2s} + \sqrt{\frac{256\, c_{s,L}}{4s+1}}, \\ c_{s,L} = (4sL^2\sqrt{4s+1})^{2/4s+1}. \end{cases}$$

Denote ψ_σ the test $\psi_\sigma(N,q)$ defined in (2.9) with

$$\begin{cases} N = N_\sigma(s,L) = [c_{s,L}\,\rho_\sigma^{-1/s}], \\ q = q_\alpha, \end{cases}$$

where q_α is the quantile of order $1-\alpha$ of the standard Gaussian distribution. Then

$$\begin{cases} \overline{\lim}_{\sigma\to 0}\, \alpha(\psi_\sigma, \Theta_0) \leq \alpha\,, \\ \lim_{\sigma\to 0} \beta(\psi_\sigma, \Theta_1) = 0\,. \end{cases}$$

Remark 4. In the rest of this section and in the proof, we skip the dependence of $N_\sigma(s,L)$ in s and L when no confusion is possible.

The proof of this result is given in Section 2.5. Let us now develop a brief heuristic describing how one could have guessed the optimal value of ρ_σ.

Heuristic for the performance of the nonadaptive procedure

Following the logic of the maximum likelihood ratio test, from where our procedure is originated, we accept the null hypothesis when $\lambda_\sigma(N_\sigma)$ is smaller than some quantity to be defined. Now, our proof will show that, under $\mathbf{H_0}$, $\lambda_\sigma(N_\sigma)$ is bounded from above in probability. Thus, we decide to take this quantity as a constant with respect to σ.

To get an intuition on the minimax rate, it remains to inspect the behavior of the statistic under the alternative hypothesis and give a condition on ρ_σ under which the test statistic is orders of magnitude larger than a constant, so that the procedure can have the desired power.

Developing the expression of $\lambda_\sigma(N_\sigma)$, we get the approximative lower bound

$$\begin{aligned} &\frac{1}{4\sqrt{N_\sigma}\sigma^2}\min_\tau \sum_{j=1}^{N_\sigma} |c_j - e^{-\mathrm{i}j\tau} c_j^\#|^2 && \text{(deterministic term)} \qquad (2.11)\\ &-\Big|\sum_{j=1}^{N_\sigma} \frac{|\xi_j|^2 + |\xi_j^\#|^2 - 4}{4\sqrt{N_\sigma}}\Big| && \text{(random term)} \\ &-\frac{1}{2\sqrt{N_\sigma}}\max_\tau \Big|\sum_{j=1}^{N_\sigma} \mathrm{Re}\left(e^{\mathrm{i}j\tau}\xi_j\overline{\xi_j^\#}\right)\Big| && \text{(perturbation term).} \end{aligned}$$

The first term, up to a $4\sqrt{N_\sigma}\sigma^2$ factor, is an approximation of the square of the pseudo-distance $d(\boldsymbol{c}, \boldsymbol{c}^\#)$. The remainder of the sum can be bounded from above, using the assumption that $\boldsymbol{c}$ and $\boldsymbol{c}^\#$ lie in a Sobolev class, by N_σ^{-2s}, up to a constant factor:

$$\min_\tau \sum_{j=1}^{N_\sigma} |c_j - e^{-\mathrm{i}j\tau} c_j^\#|^2 \approx d^2(\boldsymbol{c}, \boldsymbol{c}^\#) - O(N_\sigma^{-2s}).$$

The proof will establish that the second term is bounded in probability, while the third, that we call perturbative, is of order $\sqrt{\log N_\sigma}$. In a nutshell, we get the heuristical inequality

$$\lambda_\sigma(N_\sigma) \gtrsim \frac{d^2(\boldsymbol{c},\boldsymbol{c}^\#) - Cste \cdot N_\sigma^{-2s}}{\sqrt{N_\sigma}\sigma^2} - O_P(\sqrt{\log N_\sigma}).$$

Considering that ρ_σ^2 is the minimum of $d^2(\boldsymbol{c},\boldsymbol{c}^\#)$ for which $\mathbf{H_1}$ can be detected, this suggests that the minimax rate of separation satisfies

$$\begin{aligned}\rho_\sigma^2 &\gg \max\left(\sigma^2\sqrt{N_\sigma}, N_\sigma^{-2s}, \sigma^2\sqrt{N_\sigma \log N_\sigma}\right)\\ &\sim \left(\sigma^2\sqrt{\log \sigma^{-1}}\right)^{\frac{4s}{4s+1}}.\end{aligned}$$

Heuristic for the constant C

As we already mentioned, we are not interested by the minimax constant. But, once we have chosen our testing procedure which nearly reaches the minimax rate, we would like to optimize its parameters so that we can pull the best out of our theorem in term of performance.

Indeed, the previous optimization shows that the test achieves its best rate when N_σ is of the order of $\rho_\sigma^{*\,-1/s}$. Now, denoting $N_\sigma = [c\rho_\sigma^{*\,-1/s}]$, a similar heuristic can give an optimized constant C in the definition of Θ_1. First, Lemma 9 gives the more precise lower bound

$$\min_\tau \sum_{j=1}^{N_\sigma} |c_j - e^{-\mathrm{i}j\tau} c_j^\#|^2 \geq (C^2 - 4L^2 c^{-2s})\rho_\sigma^{*\,2}$$

for the sum in the first term, and we will prove the exact order of magnitude of the third to be

$$\frac{1}{2\sqrt{N_\sigma}} \max_\tau \left| \sum_{j=1}^{N_\sigma} \mathrm{Re}\left(e^{\mathrm{i}j\tau}\xi_j \overline{\xi_j^\#}\right)\right| \approx \sqrt{\frac{256\,c}{4s+1}\log N_\sigma}\,.$$

Thus

$$\lambda_\sigma(N_\sigma) \gtrsim \left(C^2 - 4L^2 c^{-2s} - \sqrt{\frac{256\,c}{4s+1}}\right)\sqrt{\log N_\sigma},$$

and this leads to a minimization problem determining the choice of c that we made in Theorem 5.

2.3 Adaptive testing procedure

The procedure given in the previous section possesses asymptotic minimax optimality properties thanks to an appropriate choice of the tuning parameter N_σ, but the practician needs to determine values of s and L to implement the test. As it seems arbitrary and nonintuitive to make assumptions on the smoothness of the signals, it is necessary to design testing procedures independent of s and L that are nearly as good, in the minimax sense, as the procedure proposed in the previous section.

From now on, we only assume that an interval $[s_1, s_2]$ is available such that $c, c^\# \in \mathscr{F}_{s,L}$ for some $s \in [s_1, s_2]$ and $L \in [L_1, L_2]$. We propose a testing procedure depending on s_1 and s_2 but independent of s and L, that achieves the same rate of separation, *i.e.*, $\left(\sigma^2\sqrt{\log\sigma^{-1}}\right)^{2s/4s+1}$, as the test based on the precise knowledge of s and L. Furthermore, this rate is achieved uniformly over the Sobolev classes $\mathscr{F}_{s,L}$ with $s \in [s_1, s_2]$ and L belonging to any compact interval $[L_1, L_2] \subset \mathbb{R}^+$.

Here is the idea of its construction. The nonadaptive testing procedure proposed above depends on s only via the tuning parameter $N_\sigma(s,L)$. In the followings, we will change the definition of $N_\sigma(s,L)$ to avoid the dependence on L and we will write $N_\sigma(s)$. Using a Bonferroni procedure like in Gayraud *et* Pouet (2005) or Horowitz *et* Spokoiny (2001), we consider the maximum of these tests for several values of $N_\sigma(s)$, more precisely, we consider tests of the form

$$\tilde{\psi}_\sigma(q) = \max_{N\in\mathcal{N}} \psi_\sigma(N,q).$$

For this kind of test, the next proposition gives bounds for the first and second type errors:

Proposition 6. *Let $\mathcal{N}$ be a set of positive integers and denote*

$$\tilde{\psi}_\sigma(q) = \max_{N\in\mathcal{N}} \psi_\sigma(N,q),$$

where ψ_σ is defined in 2.9, then

$$\begin{cases} \alpha(\tilde{\psi}_\sigma(q),\Theta_0) \le \sum_{N\in\mathcal{N}} \alpha(\psi_\sigma(N,q),\Theta_0), \\ \beta(\tilde{\psi}_\sigma(q),\Theta_1^{s,L}) \le \min_{N\in\mathcal{N}} \beta(\psi_\sigma(N,q),\Theta_1^{s,L}). \end{cases}$$

This stresses the conditions that a proper choice for $\mathcal{N}$ should satisfy:

– the set $\mathcal{N}$ has to be as small as possible, because a large $\mathcal{N}$ would lead to a large first kind error,
– the set $\mathcal{N}$ has to be rich enough to approximate the set of all $N_\sigma(s)$ for $s \in [s_1, s_2]$.

We will show in the proofs that each $N \in \mathcal{N}$ contributes to adaptation over all Sobolev balls of regularity s such that

$$S \le s \le S + \frac{1}{\log\sigma^{-1}}$$

with

$$N = N_\sigma(S).$$

This is the reason why we will define a grid of regularity parameters with a spacing of $1/\log\sigma^{-1}$. For every $s_2 > s_1 > 0$, define

$$\begin{cases} \Sigma(s_1,s_2) = \left\{s_1 + \frac{j}{\log\sigma^{-1}} \mid j \ge 0,\ s_1 + \frac{j}{\log\sigma^{-1}} \le s_2\right\}, \\ \mathcal{N}(s_1,s_2) = \left\{N_\sigma(s) = \left[\rho^*_\sigma(s)^{-1/s}\right] \mid s \in \Sigma(s_1,s_2)\right\}. \end{cases} \tag{2.12}$$

Theorem 6. *Set*

$$\begin{cases} \Theta_0 = \left\{(\boldsymbol{c},\boldsymbol{c}^\#) \in l_2 \times l_2 \mid d(\boldsymbol{c},\boldsymbol{c}^\#) = 0\right\}, \\ \Theta_1^{s,L} = \left\{(\boldsymbol{c},\boldsymbol{c}^\#) \in \mathscr{F}_{s,L} \times \mathscr{F}_{s,L} \mid d(\boldsymbol{c},\boldsymbol{c}^\#) \ge \rho_\sigma(s)\right\}, \end{cases} \tag{2.13}$$

with

$$\begin{cases} C > 0, \\ \rho_\sigma(s) = C\rho^*_\sigma(s), \\ \rho^*_\sigma(s) = \left(\sigma^2\sqrt{\log\sigma^{-1}}\right)^{\frac{2s}{4s+1}}. \end{cases}$$

Consider the test

$$\tilde{\psi}_\sigma = \max_{N \in \mathcal{N}(\sigma_1,\sigma_2)} \psi_\sigma\big(N, \sqrt{2\log\log\sigma^{-1}}\big),$$

where ψ_σ is defined in (2.9). Then, for the interval $[s_1, s_2]$ used in the construction of the test $\tilde{\psi}_\sigma$ and for any interval $[L_1, L_2]$ included in $\mathbb{R}_*^+$, there is a constant C such that

$$\begin{cases} \lim_{\sigma\to 0} \alpha\big(\tilde{\psi}_\sigma, \Theta_0\big) = 0, \\ \lim_{\sigma\to 0} \sup_{[L_1,L_2]} \sup_{[s_1,s_2]} \beta\big(\tilde{\psi}_\sigma, \Theta_1^{s,L}\big) = 0, \end{cases} \tag{2.14}$$

Remark 5. In the statement of this theorem, we observe that the constants L_1 and L_2 are not used in the definition of the test, while L was, in the definition of the nonadaptive procedure. Indeed, we optimized the separation constant C and gave an expression depending on L, while this optimization was not our matter in the second theorem.

Remark 6. The theorem claims that there exists a value of C for which the first and second type errors can be controlled. From the proof of the theorem, we see that it is sufficient that such a constant satisfies

$$\begin{cases} C^2 - 4L_2^2 e^{\frac{8}{(4s_1+1)^2}} - \frac{C}{2} > 0, \\ C > \frac{64}{\sqrt{4s_1+1}}, \end{cases}$$

which is verified when

$$C > \max\Big(\frac{64}{\sqrt{4s_1+1}}, \frac{1}{4} + \big(\frac{1}{16} + 4L_2^2 e^{\frac{8}{(4s_1+1)^2}}\big)^{1/2}\Big).$$

Heuristic for the performance of the adaptive procedure

Here we explain why our adaptive procedure achieves the same rate as the nonadaptive one. The heuristic of the previous section roughly holds:

$$\lambda_\sigma(N_\sigma) \gtrsim \frac{d^2(\boldsymbol{c}, \boldsymbol{c}^\#) - Cste \cdot N_\sigma^{-2s}}{\sqrt{N_\sigma}\sigma^2} - O_P(\sqrt{\log N_\sigma}),$$

with the difference that

$$\max_N \lambda_\sigma(N) = O(\log\log\sigma^{-1})$$

under the null hypothesis. But this is negligible in view of the perturbative term, so that the performance of the test does not deteriorate in the adaptive problem.

2.4 Lower bound for the minimax rate

As we wrote at the beginning of this chapter, the minimax frame allows us to discriminate between the different testing procedures, and it provides the tool of the minimax rate of separation. Here, we prove that our procedures are at least nearly minimax optimal, because their performance is close to a lower bound of the minimax rate.

Comparison to the classical signal detection

Our method is the following: we prove that the detection of a signal lying in a Sobolev ball when the separation from 0 is measured by the l_2-norm $\big($*cf.* (2.15) above for a precise definition$\big)$ is simpler than ours, in the sense that every lower bound result for this model is adaptable for our purpose. Indeed, the only difference between the two models is the shift: if you know exactly the shift, the test amounts to comparing the difference of the signals after registration, with 0.

Let us first introduce the classical signal detection problem, for which the minimax separation rate, and even the exact separation constants, are known:

$$\begin{cases} Y_j = c_j + \sigma \xi_j, \quad j = 1, 2, \dots, \\ \Theta_0^{\mathbf{class}} = \{0\}, \\ \Theta_1^{\mathbf{class}} = \Big\{ \boldsymbol{c} \in \mathscr{F}_{s,L} \mid \|\boldsymbol{c}\|_2 \geq C\rho_\sigma \Big\}. \end{cases} \tag{2.15}$$

For this model, we define the errors of first and second kind of a test $\psi^{\mathbf{class}}$ by

$$\begin{cases} \alpha^{\mathbf{class}}(\psi^{\mathbf{class}}, \Theta_0^{\mathbf{class}}) = \sup_{\Theta_0^{\mathbf{class}}} \mathbf{E}_{\boldsymbol{c}}\big(\psi^{\mathbf{class}}\big), \\ \beta^{\mathbf{class}}(\psi^{\mathbf{class}}, \Theta_1^{\mathbf{class}}) = \sup_{\Theta_1^{\mathbf{class}}} \mathbf{E}_{\boldsymbol{c}}\big(1 - \psi^{\mathbf{class}}\big), \end{cases} \tag{2.16}$$

where we denote $\mathbf{P}_{\boldsymbol{c}}$ the probability engendered by $\boldsymbol{Y} = (Y_1, Y_2, \dots)$ when $\boldsymbol{c} = (c_1, c_2, \dots)$.

Theorem 7. *Given the two models exposed in (2.2) and (2.15), we have*

$$\inf_{\psi_\alpha} \beta(\psi_\alpha, \Theta_1) \geq \inf_{\psi_\alpha^{class}} \beta^{class}(\psi_\alpha^{class}, \Theta_1^{class}), \tag{2.17}$$

where the infima are taken over all tests of level α respectively for our model and for the classical one.

Thus, our model can benefit from every lower bound result on model (2.15).

Theorem for the nonadaptive case

We choose to exploit the nonasymptotic results presented in Baraud (2002), Proposition 3. The following theorem shows that the asymptotic minimax separation rate for our problem is not smaller than

$$\sigma^{4s/4s+1}.$$

Corollary 2. *Assume that*

- $\alpha, \beta \in]0, 1]$,
- $\eta = 2(1 - \alpha - \beta)$,
- $\mathscr{L} = \log(1 + \eta^2)$
- $\rho^2 = \sup_{d \geq 1} \big[\sqrt{2\mathscr{L} d} \sigma^2 \wedge L^2 d^{-2s} \big]$.

Then

$$\rho_\sigma \leq \rho \quad \Rightarrow \quad \inf_{\psi^\alpha} \beta\Big(\psi^\alpha, \Theta_1\Big) \geq \beta, \tag{2.18}$$

where the infimum is taken over all tests of level α for the shifted curve model.

Remark 7. We can approximate ρ by computing

$$\sup_{x \in \mathbb{R}^+} \big[\sqrt{2\mathscr{L} x} \sigma^2 \wedge L^2 x^{-2s} \big] = L^{\frac{1}{4s+1}} \big(\sigma^2 \sqrt{2\mathscr{L}}\big)^{\frac{2s}{4s+1}}.$$

Theorem for the adaptive case

In the adaptive case, we exploit the results presented in Spokoiny (1996). Adapting Theorem 2.3 to our set-up, we get

Corollary 3. *Assume that*
- $s_1 < s_2$,
- $d_\sigma = o\big((\log\log\sigma^{-1})^{1/4}\big)$,
- $\rho^*_\sigma = \sigma^{4s/4s+1}$.

Then, for any $c > 0$ *and for any test* ϕ,

$$\mathbf{P}_0(\phi = 1) + \sup_{[s_1,s_2]} \sup_{[L_1,L_2]} \sup_{\boldsymbol{c}} \mathbf{P}_{\boldsymbol{c}}(\phi = 0) \to 1,$$

where the last supremum is taken over the set

$$\{\boldsymbol{c} \in \mathscr{F}_{s,L} \mid \|\boldsymbol{c}\|_2 \geq c\rho^*_{d_\sigma \sigma}\}.$$

The consequence of this theorem is that the adaptive minimax rate is not smaller than

$$\left(\sigma^2\sqrt{\log\log\sigma^{-1}}\right)^{2s/4s+1}.$$

Discussion

Strictly speaking, our results only prove that the minimax rate of separation satisfies

$$\sigma^{4s/4s+1} \lesssim \rho^*_\sigma \lesssim \left(\sigma^2\sqrt{\log\sigma^{-1}}\right)^{2s/4s+1}.$$

It could be argued that the upper bound is suboptimal, and that the minimax separation rate for the shifted curve model does not contain our logarithmic factor. Indeed, one could have expected the adaptive minimax rate to be different from the nonadaptive minimax rate for a loglog factor. The fact that adaptation is achieved here without loss of efficiency may be interpreted as a sign that the minimax separation rate is strictly smaller than our upper bound.

On the other hand, the problem considered in the present work is qualitatively different from the aforementioned works on the minimax separation rate, since our null hypothesis is not only composite but also nonparametric. As already stressed out, shift testing only differs from signal detection by the fact that no information is available on the shift. Now, it seems that the finite-dimensional parameter cannot be uniformly consistently estimated, which contrasts with the situation of Horowitz *et* Spokoiny (2001), in particular because in a minimax context, no lower bound is available on $|c_1|$: in the alternative case, one could get a consistent estimator of the shift parameter, and apply the classical signal detection methods to the sequence

$$\left\{Y_j - e^{\mathrm{i}j\hat{\tau}} Y^\#_j\right\}_{j\geq 1}.$$

The shift may even not be identifiable. Hence, the study of the perturbative term (*cf.* first heuristic after Theorem 5) is unavoidable, in order to take into account every possible shift. That is why we believe that the logarithmic term is unavoidable.

Moreover, the problem of testing the goodness-of-fit of the shifted curve model can be regarded as an adaptation to the unknown shift parameter. As a matter of fact, if adaptation to the unknown smoothness typically entails a loglog factor, other types of adaptation can bring simple logarithmic ones: it is proved in Lepski *et* Tsybakov (2000) that the asymptotic minimax separation rate for signal detection when the signal to be detected belongs to a Sobolev or Hölder ball and the separation from 0 is measured by the sup-norm is $(\sigma^2\sqrt{\log\sigma^{-1}})^{s/2s+1}$, while it is $\sigma^{2s/2s+1}$ when the separation from 0 is measured by the value of the signal at a fixed point. The logarithmic factor can be interpreted as a payment for the adaptation of the problem of testing at one point when this point is unknown. Furthermore, note that the same logarithmic factor appears in Fromont *et* Lévy-Leduc (2006), where upper bounds on the minimax separation rate are established in the problem of periodic signal detection with unknown period.

2.5 Proofs

Proof of Theorem 5

First kind error

Here, we prove that the asymptotic first kind error of the test ψ_σ does not exceed the prescribed level α:

$$\overline{\lim_{\sigma\to 0}}\sup_{\Theta_0}\mathbf{P}_{\boldsymbol{c},\boldsymbol{c}^\#}(\psi_\sigma = 1)\leq\alpha.$$

To this end, denote τ^* a real number such that, under $\mathbf{H_0}$,

$$\forall j\geq 1,\quad c_j^\# = e^{\mathrm{i}j\tau^*}c_j.$$

For simplicity sake, we skip the dependence of τ^* on $\boldsymbol{c}$ and $\boldsymbol{c}^\#$. Using the inequality

$$\begin{aligned}\min_\tau\sum_{j=1}^{N_\sigma}|Y_j - e^{-\mathrm{i}j\tau}Y_j^\#|^2 &\leq \sum_{j=1}^{N_\sigma}|Y_j - e^{-\mathrm{i}j\tau^*}Y_j^\#|^2\\ &= \sigma^2\sum_{j=1}^{N_\sigma}|\xi_j - e^{-\mathrm{i}j\tau^*}\xi_j^\#|^2,\end{aligned}$$

we get

$$\begin{aligned}\alpha(\psi_\sigma,\Theta_0) &= \sup_{\Theta_0}\mathbf{P}_{\boldsymbol{c},\boldsymbol{c}^\#}\Big(\frac{1}{4\sigma^2\sqrt{N_\sigma}}\min_\tau\sum_{j=1}^{N_\sigma}|Y_j - e^{-\mathrm{i}j\tau}Y_j^\#|^2 - \sqrt{N_\sigma} > q_\alpha\Big)\\ &\leq \mathbf{P}\Big(\frac{1}{4\sqrt{N_\sigma}}\sum_{j=1}^{N_\sigma}\big(\eta_j^2+\tilde\eta_j^2 - 4\big) > q_\alpha\Big),\end{aligned}$$

where

$$\begin{cases}\eta_j = \mathrm{Re}(\xi_j - e^{-\mathrm{i}j\tau^*}\xi_j^\#),\\ \tilde\eta_j = \mathrm{Im}(\xi_j - e^{-\mathrm{i}j\tau^*}\xi_j^\#),\end{cases}$$

and

$$\eta_j,\tilde\eta_j \overset{iid}{\sim}\mathcal{N}(0,2).$$

Finally, using Berry-Esseen's inequality (*cf.* Theorem 8), we get

$$\alpha(\psi_\sigma,\Theta_0)\leq\alpha+\frac{1}{\sqrt{2\pi N_\sigma}},$$

and this gives the desired asymptotic level.

Second kind error

It remains to study the second kind error of the test, and to show that it tends to 0:

$$\sup_{\Theta_1} \mathbf{P}_{\boldsymbol{c},\boldsymbol{c}^\#}(\psi_\sigma = 0) \to 0.$$

Our proof is based on the heuristic given earlier in Section 2.2: we decompose $\lambda_\sigma(N_\sigma)$ into several terms, and make use of their respective orders of magnitude. Decomposing $4\sigma^2\sqrt{N_\sigma}\lambda_\sigma(N_\sigma)$, we get the lower bound

$$\min_\tau \Big\{ \sum_{j=1}^{N_\sigma} |c_j - e^{-\mathrm{i}j\tau} c_j^\#|^2 + 2\sigma \sum_{j=1}^{N_\sigma} \mathrm{Re}\big((c_j - e^{-\mathrm{i}j\tau} c_j^\#)\overline{(\xi_j - e^{-\mathrm{i}j\tau}\xi_j^\#)}\big)\Big\} \tag{2.19}$$
$$- \sigma^2\sqrt{N_\sigma}\Big|\sum_{j=1}^{N_\sigma} \frac{|\xi_j|^2 + |\xi_j^\#|^2 - 4}{\sqrt{N_\sigma}}\Big| - 2\sigma^2 \max_\tau \Big|\sum_{j=1}^{N_\sigma} \mathrm{Re}\big(e^{\mathrm{i}j\tau}\xi_j\overline{\xi_j^\#}\big)\Big|.$$

For simplicity sake, we introduce some notation:

$$\begin{cases} D_\sigma(\boldsymbol{c},\boldsymbol{c}^\#) = \min_\tau \Big\{\sum_{j=1}^{N_\sigma} |c_j - e^{-\mathrm{i}j\tau}c_j^\#|^2 \\ + 2\sigma \sum_{j=1}^{N_\sigma} \mathrm{Re}\big((c_j - e^{-\mathrm{i}j\tau}c_j^\#)\overline{(\xi_j - e^{-\mathrm{i}j\tau}\xi_j^\#)}\big)\Big\}, \\ A_\sigma = \Big|\sum_{j=1}^{N_\sigma} \frac{|\xi_j|^2 + |\xi_j^\#|^2 - 4}{\sqrt{N_\sigma}}\Big|, \\ B_\sigma = \max_\tau \Big|\sum_{j=1}^{N_\sigma} \mathrm{Re}\big(e^{\mathrm{i}j\tau}\xi_j\overline{\xi_j^\#}\big)\Big|, \end{cases}$$

which, combined with (2.19), leads to:

$$\beta(\psi_\sigma,\Theta_1) \le \sup_{\Theta_1} \mathbf{P}_{\boldsymbol{c},\boldsymbol{c}^\#}\Big(D_\sigma(\boldsymbol{c},\boldsymbol{c}^\#) - \sigma^2\sqrt{N_\sigma}A_\sigma - 2\sigma^2 B_\sigma \le 4q_\alpha\sigma^2\sqrt{N_\sigma}\Big).$$

In addition to $c_{s,L}$, introduced in the definition of N_σ,

$$N_\sigma = [c_{s,L}\,\rho_\sigma^{-1/s}],$$

we will need the constant c' and ϵ, defined as

$$\begin{cases} c' = \sqrt{\frac{256\, c_{s,L}}{4s+1}}, \\ \epsilon = \frac{1}{2}\big(C^2 - 4L^2 c_{s,L}^{-2s} - \sqrt{\frac{256\, c_{s,L}}{4s+1}}\big). \end{cases}$$

Separating the different terms to study them independently, we write

$$\begin{aligned} \beta(\psi_\sigma,\Theta_1) \le \sup_{\Theta_1} &\mathbf{P}_{\boldsymbol{c},\boldsymbol{c}^\#}\Big(D_\sigma(\boldsymbol{c},\boldsymbol{c}^\#) \le (c' + \epsilon + \frac{4q_\alpha\sqrt{c_{s,L}}}{\sqrt{\log\sigma^{-1}}})\rho_\sigma^2\Big) \\ &+ \mathbf{P}\Big(\sigma^2\sqrt{N_\sigma}A_\sigma > \epsilon\rho_\sigma^2\Big) \\ &+ \mathbf{P}\Big(2\sigma^2 B_\sigma > c'\rho_\sigma^2\Big). \end{aligned}$$

– Let us first study

$$\sup_{\Theta_1} \mathbf{P}_{\boldsymbol{c},\boldsymbol{c}^\#}\big(D_\sigma(\boldsymbol{c},\boldsymbol{c}^\#) \le (c' + \epsilon + \frac{4q_\alpha\sqrt{c_{s,L}}}{\sqrt{\log\sigma^{-1}}})\rho_\sigma^2\big),$$

which contains the dominant term when ρ_σ is too large. Denoting

$$\delta = \sqrt{C^2 - 4L^2 c_{s,L}^{-2s}},$$

Lemma 4 allows to apply Lemma 5 with

$$\begin{cases} x_0 = \delta\rho_\sigma, \\ M = (c' + \epsilon + \frac{4q_\alpha\sqrt{c_{s,L}}}{\sqrt{\log\sigma^{-1}}})\rho_\sigma^2. \end{cases}$$

The choice of the parameters yields for σ small enough

$$\Big(\frac{\delta}{4} - \frac{c'+\epsilon}{4\delta} - \frac{q_\alpha\sqrt{c_{s,L}}}{\delta\sqrt{\log\sigma^{-1}}}\Big)\rho_\sigma > 0\,,$$

so that the second part of Lemma 5 holds: the term we are studying is smaller than

$$2\Big(1 + \delta^{-1}L\rho_\sigma^{-1}\max\{1, N_\sigma^{1-s}\}\Big)$$
$$\times\left[\exp\Big\{-\Big(\delta^2 - c' - \epsilon - \frac{4q_\alpha\sqrt{c_{s,L}}}{\sqrt{\log\sigma^{-1}}}\Big)^2\frac{\rho_\sigma^2}{32\delta^2\sigma^2}\Big\} + \exp\Big\{-\frac{\rho_\sigma^2\delta^2}{8\sigma^2}\Big\}\right]$$

and this upper bound converges to 0 as σ goes to 0, since

$$\frac{\rho_\sigma}{\sigma} \to +\infty.$$

– Let us now turn to

$$\mathbf{P}\big(\sigma^2\sqrt{N_\sigma}A_\sigma > \epsilon\rho_\sigma^2\big).$$

Prior to using Berry-Esseen's inequality (*cf.* Theorem 8), we derive

$$\frac{\epsilon\rho_\sigma^2}{4\sigma^2\sqrt{N_\sigma}} \geq \frac{\epsilon}{4\sqrt{c_{s,L}}}\sqrt{\log\sigma^{-1}},$$

so that, putting

$$x = \frac{\epsilon}{4\sqrt{c_{s,L}}}\sqrt{\log\sigma^{-1}}$$

into the formula of the theorem and using the bound

$$\forall\ x > 0, \quad 1 - \Phi(x) \leq \frac{e^{-\frac{x^2}{2}}}{x\sqrt{2\pi}},$$

we get

$$\mathbf{P}\Big(\sigma^2\sqrt{N_\sigma}A_\sigma > \epsilon\rho_\sigma^2\Big) \leq \sqrt{\frac{2}{\pi N_\sigma}} + \sqrt{\frac{32c_{s,L}}{\pi\epsilon^2}}\frac{\sigma^{\frac{\epsilon^2}{32c}}}{\sqrt{\log\sigma^{-1}}} \to 0.$$

– Finally, it remains to control

$$\mathbf{P}\big(2\sigma^2 B_\sigma > c'\rho_\sigma^2\big).$$

We apply Lemma 6:

$$\begin{aligned}\mathbf{P}\Big(2\sigma^2 B_\sigma > c'\rho_\sigma^2\Big) &\leq 2c(\log\sigma^{-1})^{\frac{-1}{4s+1}}\sigma^{\frac{c'^2}{64c} - \frac{4}{4s+1}} + e^{-N_\sigma/2} \\ &\leq 2c(\log\sigma^{-1})^{\frac{-1}{4s+1}} + e^{-N_\sigma/2} \to 0.\end{aligned}$$

Proof of Theorem 6

Proposition 1

Let $\mathcal{N}$ be a set of positive integers and denote

$$\tilde{\psi}_\sigma(q) = \max_{N\in\mathcal{N}} \psi_\sigma(N,q),$$

where ψ_σ is defined in 2.9.

– Concerning the first kind error:

$$\begin{aligned}\alpha(\tilde{\psi}_\sigma(q),\Theta_0) &= \sup_{(\boldsymbol{c},\boldsymbol{c}^\#)\in\Theta_0} \mathbf{P}_{\boldsymbol{c},\boldsymbol{c}^\#}\Big(\max_{N\in\mathcal{N}}\min_\tau \sum_{j=1}^N |Y_j - e^{-\mathrm{i}j\tau}Y_j^\#|^2 > q\Big)\\ &\leq \sum_{N\in\mathcal{N}} \sup_{(\boldsymbol{c},\boldsymbol{c}^\#)\in\Theta_0} \mathbf{P}_{\boldsymbol{c},\boldsymbol{c}^\#}\Big(\min_\tau \sum_{j=1}^N |Y_j - e^{-\mathrm{i}j\tau}Y_j^\#|^2 > q\Big),\end{aligned}$$

and the last right side term is exactly

$$\sum_{N\in\mathcal{N}} \alpha(\psi_\sigma(N,q),\Theta_0).$$

– Concerning the second kind error:

$$\begin{aligned}\beta(\tilde{\psi}_\sigma(q),\Theta_1^{s,L}) &= \sup_{\Theta_1^{s,L}} \mathbf{P}_{\boldsymbol{c},\boldsymbol{c}^\#}\Big(\max_{N\in\mathcal{N}}\min_\tau \sum_{j=1}^N |Y_j - e^{-\mathrm{i}j\tau}Y_j^\#|^2 \leq q\Big)\\ &\leq \sup_{\Theta_1^{s,L}} \mathbf{P}_{\boldsymbol{c},\boldsymbol{c}^\#}\Big(\min_\tau \sum_{j=1}^N |Y_j - e^{-\mathrm{i}j\tau}Y_j^\#|^2 \leq q\Big),\end{aligned}$$

for any $N\in\mathcal{N}$, so that

$$\beta(\tilde{\psi}_\sigma(q),\Theta_1^{s,L}) \leq \min_{N\in\mathcal{N}} \beta(\psi_\sigma(N,q),\Theta_1^{s,L}).$$

First kind error

Here, we prove that the first kind error of the test $\tilde{\psi}_\sigma$ converges to 0:

$$\sup_{\Theta_0} \mathbf{P}_{\boldsymbol{c},\boldsymbol{c}^\#}(\tilde{\psi}_\sigma = 1) \to 0.$$

To this end, denote τ^* a real number such that, under $\mathbf{H_0}$,

$$\forall j\geq 1, \quad c_j^\# = e^{\mathrm{i}j\tau^*}c_j.$$

We skip the dependence of τ^* on $\boldsymbol{c}$ and $\boldsymbol{c}^\#$. Using the inequality

$$\begin{aligned}\min_\tau \sum_{j=1}^{N_\sigma} |Y_j - e^{-\mathrm{i}j\tau}Y_j^\#|^2 &\leq \sum_{j=1}^{N_\sigma} |Y_j - e^{-\mathrm{i}j\tau^*}Y_j^\#|^2\\ &= \sigma^2\sum_{j=1}^{N_\sigma} |\xi_j - e^{-\mathrm{i}j\tau^*}\xi_j^\#|^2,\end{aligned}$$

we get

$$\alpha\big(\tilde{\psi}_\sigma,\Theta_0\big) \leq \sum_{N\in\mathcal{N}(s_1,s_2)} \mathbf{P}\Big(\frac{1}{4\sqrt{N}}\sum_{j=1}^N(\eta_j^2+\tilde{\eta}_j^2-4) > \sqrt{2\log\log\sigma^{-1}}\Big),$$

where

$$\begin{cases} \eta_j = \mathrm{Re}(\xi_j - e^{-\mathrm{i}j\tau^*}\xi_j^\#), \\ \tilde{\eta}_j = \mathrm{Im}(\xi_j - e^{-\mathrm{i}j\tau^*}\xi_j^\#), \end{cases}$$

and

$$\eta_j, \tilde{\eta}_j \overset{iid}{\sim} \mathcal{N}(0,2).$$

Thus, using Berry-Esseen's inequality (*cf.* Theorem 8) with

$$x = \sqrt{2 \ \log\log\sigma^{-1}}$$

and the bound

$$\forall\ x > 0, \quad 1 - \Phi(x) \le \frac{e^{-\frac{x^2}{2}}}{x\sqrt{2\pi}},$$

we get

$$\begin{aligned} \alpha\big(\tilde{\psi}_\sigma, \Theta_0\big) &\le \sum_{N \in \mathcal{N}(s_1,s_2)} \Big\{ \frac{1}{\sqrt{2\pi N}} + \frac{\exp(-\log\log\sigma^{-1})}{\sqrt{4\pi \log\log\sigma^{-1}}} \Big\} \\ &\le \frac{1}{\sqrt{2\pi}} \frac{\mathbf{Card}\,\mathcal{N}(s_1,s_2)}{\sqrt{N_\sigma(s_2)}} + \frac{1}{\sqrt{4\pi}} \frac{\mathbf{Card}\,\mathcal{N}(s_1,s_2)}{\log\sigma^{-1}\sqrt{\log\log\sigma^{-1}}}. \end{aligned}$$

Finally, as

$$\mathbf{Card}\,\mathcal{N}(s_1,s_2) = 1 + \big[\,(s_2 - s_1)\log\sigma^{-1}\big]$$

is of logarithmic order, this implies that $\alpha\big(\tilde{\psi}_\sigma, \Theta_0\big) \to 0$.

Second kind error

Finally, we study the second kind error and prove that it goes to 0.

$$\sup_{s,L} \sup_{\Theta_1^{s,L}} \mathbf{P}_{\boldsymbol{c},\boldsymbol{c}^\#}(\tilde{\psi}_\sigma = 0) \to 0.$$

For $s \in [s_1, s_2]$, define

$$S = \max\big\{t \in \Sigma(s_1,s_2) \mid t \le s\big\},$$

where we omit the dependence of S in s for simplicity sake. Note that

$$0 \le s - S \le \frac{1}{\log\sigma^{-1}}.$$

The regularity S is an approximation of s which will be sufficient for our purpose according to Lemma 9.

We introduce the notation

$$\begin{cases} D_\sigma^s(\boldsymbol{c},\boldsymbol{c}^\#) = \min_\tau \Big\{ \sum_{j=1}^{N_\sigma(s)} |c_j - e^{-\mathrm{i}j\tau} c_j^\#|^2 \\ \quad + 2\sigma \sum_{j=1}^{N_\sigma(s)} \mathrm{Re}\big((c_j - e^{-\mathrm{i}j\tau}c_j^\#)(\overline{\xi_j - e^{-\mathrm{i}j\tau}\xi_j^\#})\big)\Big\}, \\ A_\sigma^s = \Big| \sum_{j=1}^{N_\sigma(s)} \frac{|\xi_j|^2 + |\xi_j^\#|^2 - 4}{\sqrt{N_\sigma(s)}} \Big|, \\ B_\sigma^s = \max_\tau \Big| \sum_{j=1}^{N_\sigma(s)} \mathrm{Re}\big(e^{\mathrm{i}j\tau}\xi_j \overline{\xi_j^\#}\big)\Big|. \end{cases}$$

and computations similar to those of the previous section yield that the second kind error is smaller than

$$\sup_{s,L}\sup_{\Theta_1^{s,L}} \mathbf{P}_{\boldsymbol{c},\boldsymbol{c}^\#}\Big(D_\sigma^S(\boldsymbol{c},\boldsymbol{c}^\#) \le \sigma^2\sqrt{32\,N_\sigma(S)\,\log\log\sigma^{-1}} + \frac{C}{2}\rho_\sigma^2(S)\Big)$$
$$+\sum_{s\in\Sigma}\mathbf{P}\Big(\sigma^2\sqrt{N_\sigma(s)}A_\sigma^s > \frac{C}{4}\rho_\sigma^2(s)\Big) + \sum_{s\in\Sigma}\mathbf{P}\Big(2\sigma^2 B_\sigma^s > \frac{C}{4}\rho_\sigma^2(s)\Big).$$

– Let us study

$$\sup_{s,L}\sup_{\Theta_1^{s,L}} \mathbf{P}_{\boldsymbol{c},\boldsymbol{c}^\#}\big(D_\sigma^S(\boldsymbol{c},\boldsymbol{c}^\#) \le \sigma^2\sqrt{32\,N_\sigma(S)\,\log\log\sigma^{-1}} + \frac{C}{2}\rho_\sigma^2(S)\big).$$

Lemma 9 implies

$$\big(N_\sigma(S)+1\big)^{-2s} \le \rho_\sigma^*(S)^2 \le e^{\frac{8}{(4s_1+1)^2}}\,\rho_\sigma^*(s)^2,$$

so that, denoting

$$\delta^2 = C^2 - 4L^2 e^{\frac{8}{(4s_1+1)^2}},$$

Lemma 4 allows to apply Lemma 5 with

$$\begin{cases} x_0 = \delta\rho_\sigma^*(s), \\ M = \sigma^2\sqrt{32\,N_\sigma(S)\log\log\sigma^{-1}} + \frac{C}{2}\rho_\sigma^2(s). \end{cases}$$

On the other hand, the choice of δ entails that for C large and σ small enough

$$\forall\, s\in[s_1,s_2],\quad \Big(\frac{\delta}{4}-\frac{C}{8\delta}\Big)\rho_\sigma^*(s) - \frac{\sigma^2\sqrt{2\,N_\sigma(S)\log\log\sigma^{-1}}}{\delta\rho_\sigma^*(s)} > 0.$$

Hence, applying the second part of Lemma 8, we get an upper bound for the term that we are studying

$$2\Big(1+\delta^{-1}L\rho_\sigma(s_2)^{-1}\max\{1, N_\sigma(s_1)^{1-s_1}\}\Big)$$
$$\times\Big[\exp\Big\{-\Big((\delta^2-\frac{C}{2})\rho_\sigma^2(s_1) - \sqrt{32\,N_\sigma(s_1)\log\log\sigma^{-1}}\Big)^2/32\delta^2\rho_\sigma^2(s_1)\sigma^2\Big\}$$
$$+\exp\Big\{-\frac{\rho_\sigma^2(s_2)\delta^2}{8\sigma^2}\Big\}\Big],$$

and this upper bound converges to 0.

– Consider the second term. Berry-Esseen's theorem (*cf.* Theorem 8) implies the following inequality, where the right-hand side converges to 0 as σ tends to 0:

$$\sum_{s\in\Sigma}\mathbf{P}\Big(\sigma^2\sqrt{N_\sigma(s)}A_\sigma^s > \frac{C}{4}\rho_\sigma^2(s)\Big)$$
$$\le \mathbf{Card}\,\mathcal{N}(s_1,s_2)\cdot\Big[\sqrt{\frac{2}{\pi N_\sigma(s_2)}} + \sqrt{\frac{128}{\pi C}}\,\frac{\sigma^{\frac{C}{128}}}{\sqrt{\log\sigma^{-1}}}\Big].$$

– Let us turn to the third term. We apply Lemma 6 and get an inequality where once again the right-hand side converges to 0 as σ tends to 0:

$$\sum_{s\in\Sigma}\mathbf{P}\Big(2\sigma^2 B_\sigma^s > \frac{C}{4}\rho_\sigma^2(s)\Big)$$
$$\le \mathbf{Card}\,\mathcal{N}(s_1,s_2)\cdot\Big[2(\log\sigma^{-1})^{\frac{-1}{4s_2+1}}\sigma^{\frac{C^2}{1024}-\frac{4}{4s_1+1}} + e^{-N_\sigma/2}\Big].$$

Proof of Theorem 7

Consider a randomized test ψ in the shifted curve model. We will define a corresponding test in the classical model with smaller first and second kind errors, and this is sufficient to establish the result.

First note that there is a measurable function f with respect to the σ-algebra engendered by the sequences $\boldsymbol{Y}$ and $\boldsymbol{Y}^{\#}$ and with values in $[0,1]$ such that

$$\psi = f(\boldsymbol{Y}, \boldsymbol{Y}^{\#}).$$

Denoting $\boldsymbol{\epsilon}$ a sequence of random variables such that

$$\epsilon_j \overset{iid}{\sim} \mathcal{N}(0,\sigma^2)$$

and independent from $\boldsymbol{Y}$, we define

$$\psi^{\mathbf{class}} = \mathbf{E}_{\boldsymbol{\epsilon}}\big(f(\boldsymbol{Y},\boldsymbol{\epsilon})|\boldsymbol{Y}\big),$$

where $\mathbf{E}_{\boldsymbol{\epsilon}}$ is the integration with respect to the probability engendered by $\boldsymbol{\epsilon}$. $\psi^{\mathbf{class}}$ is $\sigma(\boldsymbol{Y})$-measurable and thus constitutes a test for the classical model.

This testing procedure can be interpreted as a test in the shifted curve model when $\boldsymbol{c}^{\#} = 0$. Indeed,

$$d(\boldsymbol{c}, \boldsymbol{c}^{\#}) = \|\boldsymbol{c}\|_2$$

when $\boldsymbol{c}^{\#} = 0$, so that

$$\begin{cases} \Theta_0^{\mathbf{class}} \times 0 \subseteq \Theta_0, \\ \Theta_1^{\mathbf{class}} \times 0 \subseteq \Theta_1. \end{cases}$$

By Tonelli-Fubini's theorem, $\psi^{\mathbf{class}}$ satisfies

$$\begin{aligned} \alpha^{\mathbf{class}}(\psi^{\mathbf{class}}, \Theta_0^{\mathbf{class}}) &= \sup_{\Theta_0^{\mathbf{class}}} \mathbf{E}_{\boldsymbol{c}}\big(\psi^{\mathbf{class}}\big) \\ &= \sup_{\Theta_0^{\mathbf{class}}} \mathbf{E}_{\boldsymbol{c},0}\big(f(\boldsymbol{Y}, \boldsymbol{Y}^{\#})\big) \\ &\le \alpha(\psi, \Theta_0). \end{aligned}$$

A similar inequality holds concerning the second kind error.

2.6 Lemmas

Lemma 4. *Assume that*
- $\boldsymbol{c}, \tilde{\boldsymbol{c}} \in \mathscr{F}_{s,L}$ *with* $s > 0$,
- $d(\boldsymbol{c}, \tilde{\boldsymbol{c}}) \ge C\rho$,
- $N + 1 \ge c\rho^{-1/s}$.

Then

$$\min_{\tau} \sum_{j=1}^{N} |c_j - e^{-ij\tau}\tilde{c}_j|^2 \ge (C^2 - 4L^2 c^{-2s})\rho^2.$$

Proof of Lemma 4. Since both $\boldsymbol{c}$ and $\tilde{\boldsymbol{c}}$ belong to $\mathscr{F}_{s,L}$, it holds that

$$\begin{aligned} \sum_{j>N} |c_j - e^{-ij\tau}\tilde{c}_j|^2 &\le \sum_{j>N} \big(2|c_j|^2 + 2|\tilde{c}_j|^2\big) \\ &\le 2(N+1)^{-2s} \sum_{j>N} j^{2s}\big(|c_j|^2 + |\tilde{c}_j|^2\big) \\ &\le 4L^2(N+1)^{-2s}. \end{aligned}$$

Consequently, taking into account that

$$\sum_{j=1}^{\infty} |c_j - e^{-ij\tau}\tilde{c}_j|^2 \geq d^2(c,\tilde{c}) \geq C^2\rho^2,$$

we get

$$\begin{aligned}\sum_{j=1}^{N} |c_j - e^{-ij\tau}\tilde{c}_j|^2 &= \sum_{j=1}^{\infty} |c_j - e^{-ij\tau}\tilde{c}_j|^2 - \sum_{j>N} |c_j - e^{-ij\tau}\tilde{c}_j|^2 \\ &\geq C^2\rho^2 - 4L^2(N+1)^{-2s},\end{aligned}$$

and the result follows in view of $N+1 \geq c\rho^{-1/s}$. □

Lemma 5. *Assume that*

- $N \in \mathbb{N}^*$,
- $\xi_j, \tilde{\xi}_j$, $j = 1,\dots,N$ *are independent complex valued random variables such that their real and imaginary parts are independent standard Gaussian variables,*
- $\boldsymbol{c}, \tilde{\boldsymbol{c}} \in \mathbb{C}^N$.

Denote $\boldsymbol{\xi} = (\xi_1,\dots,\xi_N)$, $\tilde{\boldsymbol{\xi}} = (\tilde{\xi}_1,\dots,\tilde{\xi}_N)$ *and*

$$\begin{cases} D_{\sigma,N}(\boldsymbol{c},\tilde{\boldsymbol{c}}) = \min_\tau \Big\{\sum_{j=1}^N |c_j - e^{-ij\tau}\tilde{c}_j|^2 \\ +2\sigma\sum_{j=1}^N \mathrm{Re}\big((c_j - e^{-ij\tau}\tilde{c}_j)\overline{(\xi_j - e^{-ij\tau}\tilde{\xi}_j)}\big)\Big\}, \\ d_{N,\tau}(\boldsymbol{c},\tilde{\boldsymbol{c}}) = \sqrt{\sum_{j=1}^N |c_j - e^{-ij\tau}\tilde{c}_j|^2}, \\ u_N(\boldsymbol{\xi},\boldsymbol{c},\tilde{\boldsymbol{c}}) = \sup_\tau \left|\sum_{j=1}^N \frac{\mathrm{Re}\left[\xi_j\overline{(c_j - e^{-ij\tau}\tilde{c}_j)}\right]}{d_{N,\tau}(\boldsymbol{c},\tilde{\boldsymbol{c}})}\right|. \end{cases}$$

If $x_0 \leq \min_\tau d_{N,\tau}(\boldsymbol{c},\tilde{\boldsymbol{c}})$, *then*

$$\begin{aligned}\forall M \in \mathbb{R},\quad &\mathbf{P}\Big(D_{\sigma,N}(\boldsymbol{c},\tilde{\boldsymbol{c}}) \leq M\Big) \\ &\leq 2\mathbf{P}\Big(\sigma u_N(\boldsymbol{\xi},\boldsymbol{c},\tilde{\boldsymbol{c}}) \geq \frac{x_0}{4} - \frac{M}{4x_0}\Big) + 2\mathbf{P}\Big(\frac{x_0}{2} < \sigma u_N(\boldsymbol{\xi},\boldsymbol{c},\tilde{\boldsymbol{c}})\Big).\end{aligned}$$

Assume further that

- $\boldsymbol{c}, \tilde{\boldsymbol{c}} \in \mathscr{F}_{s,L}$,
- $\frac{x_0}{4} - \frac{M}{4x_0} > 0$,

then combining the last result with Lemma 8,

$$\begin{aligned}\mathbf{P}\Big(D_{\sigma,N}(\boldsymbol{c},\tilde{\boldsymbol{c}}) \leq M\Big) &\leq 2\Big(1 + x_0^{-1}L\max\{1, N^{1-s}\}\Big) \\ &\times \Big(\exp\big\{-(x_0^2 - M)^2/32x_0^2\sigma^2\big\} + \exp\big\{-x_0^2/8\sigma^2\big\}\Big).\end{aligned}$$

Proof of Lemma 5. Using the notation introduced in the statement of this lemma, we can write

$$\begin{aligned}&\sum_{j=1}^{N} |c_j - e^{-\mathrm{i}j\tau}\tilde{c}_j|^2 + 2\sigma\sum_{j=1}^{N} \mathrm{Re}\Big((c_j - e^{-\mathrm{i}j\tau}\tilde{c}_j)\overline{(\xi_j - e^{-\mathrm{i}j\tau}\tilde{\xi}_j)}\Big) \\ &= d_{N,\tau}^2(\boldsymbol{c},\tilde{\boldsymbol{c}}) + 2\sigma d_{N,\tau}(\boldsymbol{c},\tilde{\boldsymbol{c}})\sum_{j=1}^{N} \frac{\mathrm{Re}\left[\xi_j\overline{(c_j - e^{-\mathrm{i}j\tau}\tilde{c}_j)}\right]}{d_{N,\tau}(\boldsymbol{c},\tilde{\boldsymbol{c}})} \\ &+ 2\sigma d_{N,\tau}(\boldsymbol{c},\tilde{\boldsymbol{c}})\sum_{j=1}^{N} \frac{\mathrm{Re}\left[\tilde{\xi}_j\overline{(e^{\mathrm{i}j\tau}c_j - \tilde{c}_j)}\right]}{d_{N,\tau}(\boldsymbol{c},\tilde{\boldsymbol{c}})},\end{aligned}$$

and obtain the following lower bound for this quantity

$$d_{N,\tau}^2(\boldsymbol{c},\tilde{\boldsymbol{c}}) - 2\sigma d_{N,\tau}(\boldsymbol{c},\tilde{\boldsymbol{c}}) \sup_\tau \Big| \sum_{j=1}^N \frac{\operatorname{Re}\big[\xi_j \overline{(c_j - e^{-\mathrm{i}j\tau}\tilde{c}_j)}\big]}{d_{N,\tau}(\boldsymbol{c},\tilde{\boldsymbol{c}})} \Big|$$
$$- 2\sigma d_{N,\tau}(\boldsymbol{c},\tilde{\boldsymbol{c}}) \sup_\tau \Big| \sum_{j=1}^N \frac{\operatorname{Re}\big[\tilde{\xi}_j \overline{(e^{\mathrm{i}j\tau}c_j - \tilde{c}_j)}\big]}{d_{N,\tau}(\boldsymbol{c},\tilde{\boldsymbol{c}})} \Big|.$$

With the notation

$$u_N(\boldsymbol{\xi},\boldsymbol{c},\tilde{\boldsymbol{c}}) = \sup_\tau \Big| \sum_{j=1}^N \frac{\operatorname{Re}\big[\xi_j \overline{(c_j - e^{-\mathrm{i}j\tau}\tilde{c}_j)}\big]}{d_{N,\tau}(\boldsymbol{c},\tilde{\boldsymbol{c}})} \Big|,$$

we obtain

$$D_{\sigma,N}(\boldsymbol{c},\tilde{\boldsymbol{c}}) \geq \min_{x\geq x_0}(x^2 - ax),$$

with

$$a = 2\sigma u_N(\boldsymbol{\xi},\boldsymbol{c},\tilde{\boldsymbol{c}}) + 2\sigma u_N(\tilde{\boldsymbol{\xi}},\tilde{\boldsymbol{c}},\boldsymbol{c}).$$

Now, using the fact that if $x_0 \geq \frac{a}{2}$,

$$\min_{x\geq x_0}(x^2 - ax) = x_0^2 - ax_0,$$

we get

$$\begin{aligned}
\mathbf{P}\Big(D_{\sigma,N}(\boldsymbol{c},\tilde{\boldsymbol{c}}) \leq M\Big) &\leq \mathbf{P}\Big(x_0^2 - 2x_0\sigma u_N(\boldsymbol{\xi},\boldsymbol{c},\tilde{\boldsymbol{c}}) - 2x_0\sigma u_N(\tilde{\boldsymbol{\xi}},\tilde{\boldsymbol{c}},\boldsymbol{c}) \leq M\Big) \\
&\quad + \mathbf{P}\Big(x_0 < \sigma u_N(\boldsymbol{\xi},\boldsymbol{c},\tilde{\boldsymbol{c}}) + \sigma u_N(\tilde{\boldsymbol{\xi}},\tilde{\boldsymbol{c}},\boldsymbol{c}))\Big) \\
&\leq 2\mathbf{P}\Big(\sigma u_N(\boldsymbol{\xi},\boldsymbol{c},\tilde{\boldsymbol{c}}) \geq \frac{x_0}{4} - \frac{M}{4x_0}\Big) \\
&\quad + 2\mathbf{P}\Big(\frac{x_0}{2} < \sigma u_N(\boldsymbol{\xi},\boldsymbol{c},\tilde{\boldsymbol{c}})\Big),
\end{aligned}$$

since $u_N(\boldsymbol{\xi},\boldsymbol{c},\tilde{\boldsymbol{c}})$ and $u_N(\tilde{\boldsymbol{\xi}},\tilde{\boldsymbol{c}},\boldsymbol{c})$ have the same distribution. This terminates the proof. □

Lemma 6. *Assume that*

- *$\xi_j, \tilde{\xi}_j$ are independent complex valued random variables such that their real and imaginary parts are independent standard Gaussian variables,*
- *c, s and σ are some positive real numbers.*

Denote

$$\begin{cases} \rho_\sigma = (\sigma^2\sqrt{\log\sigma^{-1}})^{\frac{2s}{4s+1}}, \\ N_\sigma = [c\rho_\sigma^{-1/s}], \\ B_\sigma = \max_\tau \Big|\sum_{j=1}^{N_\sigma} \operatorname{Re}\big(e^{ij\tau}\xi_j\tilde{\xi}_j\big)\Big|. \end{cases}$$

Then, for σ small enough and for every positive c',

$$\mathbf{P}\Big(2\sigma^2 B_\sigma > c'\rho_\sigma^2\Big) \leq 2c(\log\sigma^{-1})^{\frac{-1}{4s+1}}\sigma^{\frac{c'^2}{64c} - \frac{4}{4s+1}} + e^{-N_\sigma/2}.$$

Proof of Lemma 6. Applying Lemma 7, we state that, for σ small enough,

$$\mathbf{P}\Big(B_\sigma > 4x\sqrt{N_\sigma\log(\sigma^{-1})}\Big) \leq 2c(\log\sigma^{-1})^{\frac{-1}{4s+1}}\sigma^{x^2 - \frac{4}{4s+1}} + e^{-N_\sigma/2},$$

from which follows that

$$\mathbf{P}\Big(B_\sigma > 4x\rho_\sigma^{-1/2s}\sqrt{c\log(\sigma^{-1})}\Big) \le 2c(\log\sigma^{-1})^{\frac{-1}{4s+1}}\sigma^{x^2-\frac{4}{4s+1}} + e^{-N_\sigma/2}.$$

We conclude, observing that

$$4x\rho_\sigma^{-1/2s}\sqrt{c\log(\sigma^{-1})} = \frac{8x\rho_\sigma^2\sqrt{c}}{2\sigma^2}.$$

□

Lemma 7. *Assume that*

- $N \in \mathbb{N}^*$,
- $\xi_j, \tilde{\xi}_j$, $j = 1,\dots,N$, *are independent complex valued random variables such that their real and imaginary parts are independent standard Gaussian variables,*
- $\boldsymbol{u} \in \mathbb{R}^N$.

Denote

$$\forall\ t \in [0,2\pi],\quad S(t) = \sum_{j=1}^N u_j \operatorname{Re}\left(e^{ijt}\xi_j\tilde{\xi}_j\right)$$

and

$$\|S\|_\infty = \sup_{t\in[0,2\pi]} |S(t)|.$$

Then

$$\forall x, y > 0,\quad \mathbf{P}\Big(\|S\|_\infty > \sqrt{2}x\big(\|\boldsymbol{u}\|_2 + y\|\boldsymbol{u}\|_\infty\big)\Big) \le (N+1)e^{-x^2/2} + e^{-y^2/2}.$$

Proof of Lemma 7. We refer to the first chapter, Lemma 3, for a proof of this lemma. □

Lemma 8. *Assume that*

- $\boldsymbol{c}, \tilde{\boldsymbol{c}} \in \mathscr{F}_{s,L}$ *with* $s > 0$,
- $\eta_j, \tilde{\eta}_j \overset{iid}{\sim} \mathcal{N}(0,1)$,
- $N \in \mathbb{N}^*$.

Define

$$\forall\ t \in [0,2\pi],\quad S(t) = \sum_{j=1}^N \frac{\eta_j \operatorname{Re}(c_j - e^{-ijt}\tilde{c}_j) + \tilde{\eta}_j \operatorname{Im}(c_j - e^{-ijt}\tilde{c}_j)}{\sqrt{\sum_{j=1}^N |c_j - e^{-ijt}\tilde{c}_j|^2}}.$$

Then

$$\mathbf{P}\Big(\|S\|_\infty \ge x\Big) \le \Big(\frac{L\cdot\max\{1, N^{1-s}\}}{\sqrt{\min_\tau \sum_{j=1}^N |c_j - e^{-ij\tau}\tilde{c}_j|^2}} + 1\Big)e^{-\frac{x^2}{2}}.$$

First recall Berman's formula, that we will need in the proof.

Theorem (Berman (1988))**.** *Assume that*

- $N \in \mathbb{N}^*$,
- $a < b$,
- g_j, $j = 1,\dots,N$ *are continuously differentiable functions on* $[a,b]$ *satisfying*

$$\forall\ t \in \mathbb{R},\quad \sum_{j=1}^N g_j(t)^2 = 1,$$

– η_j, $j = 1, \ldots, N$, *are some independent standard Gaussian variables.*

Then

$$\mathbf{P}\Big(\sup_{[a,b]} \sum_{j=1}^{N} g_j(t)\eta_j \geq x\Big) \leq \frac{I}{2\pi} e^{-\frac{x^2}{2}} + \int_x^\infty \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}\, dt$$

with

$$I = \int_a^b \Big[\sum_{j=1}^{N} g_j'(t)^2\Big]^{1/2} dt.$$

Proof of Lemma 8. Denote

$$\begin{cases} f_j(t) = \frac{\operatorname{Re}(c_j - e^{-\mathrm{i}jt}\tilde{c}_j)}{\sqrt{\sum_{k=1}^N |c_k - e^{-\mathrm{i}kt}\tilde{c}_k|^2}}, \\ g_j(t) = \frac{\operatorname{Im}(c_j - e^{-\mathrm{i}jt}\tilde{c}_j)}{\sqrt{\sum_{k=1}^N |c_k - e^{-\mathrm{i}kt}\tilde{c}_k|^2}}. \end{cases}$$

We compute the derivatives of these functions:

$$\begin{aligned} f_j'(t) &= \frac{-\operatorname{Im}(je^{-\mathrm{i}jt}\tilde{c}_j)}{\sqrt{\sum_{k=1}^N |c_k - e^{-\mathrm{i}kt}\tilde{c}_k|^2}} \\ &\quad + \frac{\operatorname{Re}(c_j - e^{-\mathrm{i}jt}\tilde{c}_j)}{\big(\sum_{k=1}^N |c_k - e^{-\mathrm{i}kt}\tilde{c}_k|^2\big)^{\frac{3}{2}}} \sum_{k=1}^N \operatorname{Im}(k\bar{c}_k\tilde{c}_k e^{-\mathrm{i}kt}) \\ \text{and} \quad g_j'(t) &= \frac{\operatorname{Re}(je^{-\mathrm{i}jt}\tilde{c}_j)}{\sqrt{\sum_{k=1}^N |c_k - e^{-\mathrm{i}kt}\tilde{c}_k|^2}} \\ &\quad + \frac{\operatorname{Im}(c_j - e^{-\mathrm{i}jt}\tilde{c}_j)}{\big(\sum_{k=1}^N |c_k - e^{-\mathrm{i}kt}\tilde{c}_k|^2\big)^{\frac{3}{2}}} \sum_{k=1}^N \operatorname{Im}(k\bar{c}_k\tilde{c}_k e^{-\mathrm{i}kt}), \end{aligned}$$

whence

$$\begin{aligned} \sum_{j=1}^N \big(f_j'(t)^2 + g_j'(t)^2\big) &= \frac{\sum_{j=1}^N j^2|\tilde{c}_j|^2}{\sum_{k=1}^N |c_k - e^{-\mathrm{i}kt}\tilde{c}_k|^2} - \Big(\frac{\sum_{k=1}^N \operatorname{Im}(k\bar{c}_k\tilde{c}_k e^{-\mathrm{i}kt})}{\sum_{k=1}^N |c_k - e^{-\mathrm{i}kt}\tilde{c}_k|^2}\Big)^2 \\ &\leq \frac{L^2 \max\{1, N^{2-2s}\}}{\min_t \sum_{k=1}^N |c_k - e^{-\mathrm{i}kt}\tilde{c}_k|^2}. \end{aligned}$$

The conclusion follows from Berman's formula. □

Lemma 9. *Assume that*

– $\sigma > 0$,

– s, S *in* $[s_1, s_2] \subseteq \mathbb{R}_*^+$ *are such that*

$$0 \leq s - S \leq \frac{1}{\log \sigma^{-1}}.$$

Denote

$$\rho_\sigma^*(s) = \big(\sigma^2\sqrt{\log\sigma^{-1}}\big)^{\frac{2s}{4s+1}}.$$

Then, for σ *small enough,*

$$\frac{\rho_\sigma^*(S)}{\rho_\sigma^*(s)} \leq e^{\frac{4}{(4s_1+1)^2}}.$$

Proof of Lemma 9. By the definition of $\rho^*_\sigma(s)$, we have

$$\frac{\rho^*_\sigma(S)}{\rho^*_\sigma(s)} = \left(\sigma^2\sqrt{\log(\sigma^{-1})}\right)^{\frac{2(S-s)}{(4s+1)(4S+1)}},$$

which, when σ is so small that

$$\sigma^2\sqrt{\log\sigma^{-1}} \leq 1,$$

leads, with the hypothesis on s and S,

$$\frac{\rho^*_\sigma(S)}{\rho^*_\sigma(s)} \leq \left(\sigma^2\sqrt{\log(\sigma^{-1})}\right)^{\frac{-2}{(4s_1+1)^2\log\sigma^{-1}}}.$$

Then, we compute

$$\begin{aligned}
&\left(\sigma^2\sqrt{\log(\sigma^{-1})}\right)^{\frac{-2}{(4s_1+1)^2\log\sigma^{-1}}}\\
&= \exp\left\{\frac{-2}{(4s_1+1)^2\log\sigma^{-1}}(2\log\sigma + \frac{1}{2}\log\log\sigma^{-1})\right\}\\
&= \exp\left\{\frac{4}{(4s_1+1)^2}(1-\frac{\log\log\sigma^{-1}}{4\log\sigma^{-1}})\right\}\\
&\leq e^{\frac{4}{(4s_1+1)^2}},
\end{aligned}$$

and this concludes the proof. □

Finally, we recall here Berry-Esseen's inequality, in a simpler version than Theorem 5.4 of Petrov (1995).

Theorem 8 (Berry-Esseen's inequality)**.** *Assume that*

- $N \in \mathbb{N}^*$
- $X_1, \ldots, X_N \overset{iid}{\sim} X$ *are such that*

$$\begin{cases} \mathbf{E}(X) = 0, \\ \mathbf{Var}(X) = \gamma^2, \\ \mathbf{E}|X|^3 = m^3 < +\infty. \end{cases}$$

Denote

$$F_N(x) = \mathbf{P}\Big(\frac{1}{\sqrt{N}\gamma}\sum_{j=1}^{N} X_j < x\Big)$$

and Φ *the distribution function of the standard Gaussian variable. Then*

$$\sup_x |F_N(x) - \Phi(x)| \leq \frac{Am^3}{\gamma^3}\frac{1}{\sqrt{N}},$$

for an absolute constant number A. Moreover, in the case when

$$X = Y^2 - 1$$

and Y *has a standard Gaussian distribution, and using the majoration*

$$A \leq \frac{1}{2},$$

we get

$$\sup_x |F_N(x) - \Phi(x)| \leq \frac{1}{\sqrt{2\pi N}}.$$

Conclusion of the chapter

This chapter provided a minimax frame for the problem of shit testing, and we showed that the testing procedures coming from the generalized likelihood ratio approach are at least nearly optimal. So this closes an aspect of the problem that we have tackled in this work: it is possible to use the non-centered histograms as descriptors and efficient procedures to match pairs of descriptors corresponding to translated signals, avoiding the unstable problem of finding the argument of the maximum of an histogram.

But we can not directly use our testing procedure to match the descriptors of an image pair by pair. Indeed, it would be highly under-optimal to carry out a test for each possible pair of descriptors, because the individual errors would add up. Now, we need to investigate the problem of global matching.

Deuxième partie

Procédure globale de mise en correspondance des descripteurs

CHAPTER

3 Matching by permutation estimation

In the previous part, we inspected a possible improvement of the descriptor SIFT: indeed, the computation of the argument of the maximum of a histogram is not stable. We showed that this can be avoided, if we adapt the matching criterion. In this part, we tackle another possible improvement of the SIFT algorithm, more precisely the procedure that finds pairs of matching descriptors in two images. Our goal is to give a statistical approach for this problem and to find optimal procedures.

First, we introduce a model for this problem in Section 3.1, in which matching is equivalent to the estimation of a permutation. We derive several estimators of this permutation (Section 3.2). Then in Section 3.3, we introduce the matching threshold to measure the performance of the estimators and use it in Sections 3.4 and 3.5 to assess the performance of our estimators. Finally, we show in Section 3.6 how to compute our estimators in practice, which allowed us to conduct some experiments illustrating our results in Section 3.7. Sections 3.8, 3.9 and 3.10 are devoted to the statement of some possible extensions of our results and the discussion of the different assumptions that were used in the statements in the theorems. The proofs of the theorems and of the lemmas are postponed to Sections 3.11 and 3.12.

3.1 MODEL

We now consider the task of matching the descriptors of two images. That means that we observe two sets of descriptors $\{X_1, \dots, X_n\}$ and $\{X_1^\#, \dots, X_m^\#\}$ with $n, m \geq 2$. We propose the following model for the observations: from the model

$$\begin{cases} X_i = \theta_i + \sigma_i \xi_i\ , \\ X_j^\# = \theta_j^\# + \sigma_j^\# \xi_j^\#, \end{cases} \qquad i = 1, \dots, n \text{ and } j = 1, \dots, m \tag{3.1}$$

where

- $\boldsymbol{\theta} = \{\theta_1, \dots, \theta_n\}$ and $\boldsymbol{\theta}^\# = \{\theta_1^\#, \dots, \theta_m^\#\}$ are some collections of distinct vectors from $\mathbb{R}^d$, corresponding to the original features, which are unavailable,
- $\sigma_1, \dots, \sigma_n, \sigma_1^\#, \dots, \sigma_m^\#$ are positive real numbers corresponding to the levels of noise contaminating each feature,
- $\xi_1, \dots, \xi_n$ and $\xi_1^\#, \dots, \xi_m^\#$ are two independent sets of i.i.d. random vectors drawn from the Gaussian distribution with zero mean and identity covariance matrix.

The task of descriptor matching consists in finding a bijection π^* between the largest possible subsets S_1 and S_2 of $\{1, \dots, n\}$ and $\{1, \dots, m\}$ respectively, such that

$$\forall\ i \in S_2, \quad \theta_i^\# \equiv \theta_{\pi^*(i)},$$

where $\equiv$ is an equivalence relation that we call **matching criterion**. In the first part of this work, we chose an elaborate case of matching criterion. But now, for the sake of simplicity, we will go back to the simpler case:

Assumption 1. *In this part, the equivalence relation "$\equiv$" stands for the usual equality "$=$".*

The features that do not belong to S_1 or S_2 are called **outliers** ; they have no match in the other image. We expect our task to be more complicated when a lot of outliers are present. Hence, an important simplification is to assume that there is no outlier at all. As we will see in the followings, the study of this simple case is still quite involved, so that it seems that this simplification was necessary at first. More precisely, we make the following assumption:

Assumption 2. *We assume that*

$$\begin{aligned} & m = n \\ \textit{and} \quad & S_1 = S_2 = \{1, \dots, n\}. \end{aligned}$$

In this formulation, there are only four sets of unknown parameters: $\boldsymbol{\theta}$, $\boldsymbol{\sigma} = (\sigma_i, \dots, \sigma_n)$, $\boldsymbol{\sigma}^\# = (\sigma_i^\#, \dots, \sigma_m^\#)$ and π^*. However, we will focus our attention on the problem of estimating the parameter π^* only, considering $\boldsymbol{\theta}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}^\#$ as nuisance parameters. In what follows, we denote by $\mathbf{P}_{\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\sigma}^\#,\pi^*}$ the probability distribution of the vector $(X_1, \dots, X_n, X_1^\#, \dots, X_n^\#)$ defined by (3.10). The set of all permutations of $\{1, \dots, n\}$ will be denoted by $\mathfrak{S}_n$.

3.2 Some estimators

Greedy estimator

Maybe the simplest procedure that you can think of is what we called the greedy algorithm and that we denoted by π^{gr}. It is defined as follows:

$$\pi^{\mathrm{gr}}(1) = \arg \min_{j \in \{1,\dots,n\}} \|X_j - X_1^{\#}\|$$

and recursively, for every $i \in \{2,\dots,n\}$,

$$\pi^{\mathrm{gr}}(i) = \arg \min_{j \notin \{\pi^{\mathrm{gr}}(1),\dots,\pi^{\mathrm{gr}}(i-1)\}} \|X_j - X_i^{\#}\|. \tag{3.2}$$

A drawback of this estimator is that it is not symmetric: the resulting permutation depends on the initial numbering of the features. Let us give an example when $d = 1$ and $n = 2$. Assume that

- $X_1 = \theta_1 = 0$ and $X_2 = \theta_2 = 1$,
- $\theta_1^{\#} = \theta_1$ and $\theta_2^{\#} = \theta_2$ so that $\pi^* = id$,
- $X_1^{\#} = -1$ and $X_2^{\#} = 0.5$.

In this case, the greedy procedure will select the transposition (1 2), but it would have chosen the correct permutation *id* if we had ordered the descriptors differently. This problem also occurs in larger dimension, and it can not be solved by randomly choosing the ordering: there always exists some situations that make the procedure often select incorrect permutations.

However, we will show that this estimator possesses nice optimality properties in the homoscedastic setting .

Maximum likelihood estimator in the homoscedastic case

To avoid this problem of ordering, we need estimators that take all observations into account at the same time. This is satisfied by the estimators of the maximum likelihood. In the homoscedastic case, which is for the sake of simplicity the first setting that we studied, the computations lead to the following expression:

$$\pi^{\mathrm{LSS}} = \arg \min_{\pi \in \mathfrak{S}_n} \sum_{i=1}^{n} \|X_{\pi(i)} - X_i^{\#}\|^2. \tag{3.3}$$

In the homoscedastic setting, for which this estimator was designed, the **Least Sum of Squares** possesses optimality properties as well.

But in the heteroscedastic case, it may not perform as well. Indeed, this estimator only takes the distance between the observations into account, without considering the noise levels. Yet, in the heteroscedastic case, the distance between the observations is not as relevant as the signal-to-noise ratio

$$\frac{\|\theta_i - \theta_j\|^2}{\sigma_i^2 + \sigma_j^2}:$$

when the noise levels are small, a large distance will be more significant than if the noise levels are large.

Maximum likelihood estimator in the heteroscedastic case with known noise levels

The necessity of considering the signal-to-noise ratio is confirmed by the computation of the maximum likelihood estimator in the heteroscedastic case with known noise levels. In this setting, we obtain the **Least Sum of Normalized Squares**:

$$\pi^{\mathrm{LSNS}} = \arg\min_{\pi \in \mathfrak{S}_n} \sum_{i=1}^{n} \frac{\| X_{\pi(i)} - X_i^{\#} \|^2}{\sigma_{\pi(i)}^2 + {\sigma_i^{\#}}^2}. \tag{3.4}$$

In the homoscedastic case, the LSNS equals the LSS. But in the heteroscedastic case, we can prove that the LSNS is optimal, while the LSS is not.

Maximum likelihood estimator in the heteroscedastic case

In the general setting, when no information on the noise levels is available, the maximum likelihood approach requires to compute the maximum over all nuisance parameters (descriptors, noise levels and the underlying permutation) of the likelihood. But this problem is underconstrained, and the result of this maximization is $+\infty$.

This can be circumvented by assuming a proper relation between the noise levels. We chose the following assumption:

Assumption 3. *We assume that*

$$\forall\ i \in \{1, \dots, n\}, \quad \sigma_i^{\#} = \sigma_{\pi^*(i)}.$$

In this chapter, we will always assume that this assumption is satisfied in the heteroscedastic case. Then, the maximum likelihood estimator is the **Least Sum of Logarithms** defined as

$$\pi^{\mathrm{LSL}} = \arg\min_{\pi \in \mathfrak{S}_n} \sum_{i=1}^{n} \log \| X_{\pi(i)} - X_i^{\#} \|^2. \tag{3.5}$$

We will prove that this estimator is optimal both in the homoscedastic and in the heteroscedastic case.

3.3 Notion of matching threshold

In this section, we give a criterion to evaluate the performance of an estimator. The estimation of the permutation is indeed more difficult when the descriptors are close to each other. So we introduce the inner distance $\kappa(\boldsymbol{\theta})$, which measures the separation of the descriptors compared to the magnitude of the corresponding noise, following our remark that the distance between the descriptors is not as relevant as the signal-to-noise ratio:

$$\kappa^2(\boldsymbol{\theta}) \triangleq \min_{i \neq j} \frac{\|\theta_i - \theta_j\|^2}{\sigma_i^2 + \sigma_j^2}. \tag{3.6}$$

It holds that if $\kappa(\boldsymbol{\theta}) = 0$, then the parameter π^* is nonidentifiable, in the sense that there exist two different permutations π_1^* and π_2^* such that the distributions

$\mathbf{P}_{\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\sigma}^{\#},\pi_1^*}$ and $\mathbf{P}_{\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\sigma}^{\#},\pi_2^*}$ coincide. Therefore, the condition $\kappa(\boldsymbol{\theta}) > 0$ is necessary for the existence of consistent estimators of π^*.

Furthermore, good estimators are those consistently estimating π^* even if $\kappa(\boldsymbol{\theta})$ is small. To give a precise sense to these considerations, let $\alpha \in (0,1)$ be a prescribed tolerance level and let us call **matching threshold** of a given estimation procedure $\hat{\pi}$ the quantity

$$\kappa_\alpha(\hat{\pi}) = \inf\Big\{\kappa > 0 \,|\, \max_{\pi\in\mathfrak{S}_n} \sup_{\kappa(\boldsymbol{\theta})>\kappa} \mathbf{P}_{\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\sigma}^{\#},\pi}(\hat{\pi} \neq \pi) \leq \alpha\Big\},$$

where we skip the dependence on n, d, $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}^{\#}$.

We then define the **minimax matching threshold** as

$$\kappa_\alpha = \inf_{\hat{\pi}} \kappa_\alpha(\hat{\pi}),$$

where the infimum is taken over all possible estimators of π^*.

3.4 Performance of the estimators

The purpose of this section is to present conditions for the consistency of our estimators in the form of upper bounds of their matching threshold, and to give lower bounds on the performance of an estimator.

Homoscedastic setup

We start by considering the homoscedastic case, in which upper and lower bounds matching up to a constant are obtained.

Theorem 9. *Assume that*

- $\alpha \in (0,1)$ *is a tolerance level,*
- *the noise levels are equal to the constant* σ*,*
- $\hat{\pi}$ *denotes any one of the estimators (3.2)-(3.5).*

Then we have

$$\kappa_\alpha^2(\hat{\pi}) \leq 16\max\Big\{2\log\frac{8n^2}{\alpha}, \Big(d\log\frac{4n^2}{\alpha}\Big)^{1/2}\Big\}.$$

Remark 8. An equivalent way of stating this result is that if

$$\kappa^2 = 16\max\Big\{2\log\frac{8n^2}{\alpha}, \Big(d\log\frac{4n^2}{\alpha}\Big)^{1/2}\Big\}$$

and Θ_κ is the set of all $\boldsymbol{\theta} \in \mathbb{R}^{n\times d}$ such that $\kappa(\boldsymbol{\theta}) \geq \kappa$, then

$$\max_{\pi^*\in\mathfrak{S}_n} \sup_{\boldsymbol{\theta}\in\Theta_\kappa} \mathbf{P}_{\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\sigma}^{\#},\pi^*}(\hat{\pi} \neq \pi^*) \leq \alpha$$

for all the estimators defined in Section 3.2.

Remark 9. Note that this result is nonasymptotic. Roughly speaking, it tells us that the matching threshold of the procedures under consideration is at most of the order of

$$\max\Big\{(\log n)^{1/2}, (d\log n)^{1/4}\Big\}. \tag{3.7}$$

However, this result does not allow us to deduce any hierarchy between the four estimators, since it provides the same upper bound for all of them. On the other hand, as stated in the next theorem, this bound is optimal up to a multiplicative constant.

Theorem 10. *Assume that*

– $n \geq 4$,
– Θ_κ *is the set of all* $\boldsymbol{\theta} \in \mathbb{R}^{n\times d}$ *such that* $\kappa(\boldsymbol{\theta}) \geq \kappa$.

Then, there exist two absolute constants $c, C > 0$ *such that:*

$$\kappa \leq \frac{1}{4}\max\Big\{\sqrt{\log n}, c(d\log n)^{1/4}\Big\},$$

implies that

$$\inf_{\hat{\pi}} \max_{\pi^* \in \mathfrak{S}_n} \sup_{\boldsymbol{\theta}\in\Theta_\kappa} \mathbf{P}_{\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\sigma}^\#,\pi^*}\big(\hat{\pi} \neq \pi^*\big) > C,$$

where the infimum is taken over all permutation estimators.

Remark 10. We can prove that the constant C in the previous theorem can be chosen larger than 15% .

So the performance of our estimators can not be improved in the homoscedastic case.

Heteroscedastic setup

We switch now to the heteroscedastic setting, which will allow us to discriminate between the four procedures. Note that the greedy algorithm, the LSS and the LSL have an advantage over the LSNS since they can be computed without knowing the noise levels $\boldsymbol{\sigma}$.

Theorem 11. *Assume that*

– $\alpha \in (0,1)$,
– $\forall\ i \in \{1,\ldots,n\}, \quad \sigma_i^\# = \sigma_{\pi^*(i)}$,
– $\hat{\pi}$ *is either* π^{LSNS} *(if the noise levels* $\sigma_i, \sigma_i^\#$ *are known) or* π^{LSL} *(when the noise levels are unknown).*

Then

$$\kappa_\alpha^2(\hat{\pi}) \leq 16\max\Big\{2\log\frac{8n^2}{\alpha}, \Big(d\log\frac{4n^2}{\alpha}\Big)^{1/2}\Big\}$$

Now, can the performance of the LSNS or LSL be improved ? In full generality, the answer to that question is yes. Indeed, under Assumption 3, it is possible to estimate the permutation π^* by only considering the noise levels, without any assumption on the distance between the descriptors.

Suppose that we know $\boldsymbol{\sigma}$, then we can define an estimator $\hat{\pi}$ as follows:

$$\hat{\pi}(1) = \arg\min_{j\in\{1,\ldots,n\}} \arg\min_{i=1,\ldots,n} \left|\frac{\|X_1 - X_i^\#\|^2}{2d} - \sigma_1^2\right|$$

and recursively, for every $i \in \{2,\ldots,n\}$,

$$\hat{\pi}(i) = \arg\min_{j\notin\{\pi^{\mathrm{gr}}(1),\ldots,\pi^{\mathrm{gr}}(i-1)\}} \left|\frac{\|X_i - X_j^\#\|^2}{2d} - \sigma_i^2\right|.$$

This estimator will be efficient as soon as the noise levels are different enough from each other, and more precisely, when

$$\forall i \in \{1,\dots,n\}, \quad \frac{\|X_i - X_i^{\#}\|^2}{2d} - \sigma_i^2 \ll \max_{j\neq i} \frac{\|X_i - X_j^{\#}\|^2}{2d} - \sigma_i^2.$$

Now, the left-hand side is of the order $\sigma_i^2\sqrt{d}$ while the right-hand side is of the order $\max_j(\sigma_j^2 - \sigma_i^2)/\sqrt{\log n}$, so that we can consistently identify the permutation when

$$\forall (i,j) \in \{1,\dots,n\}^2, \quad \frac{\sigma_j^2}{\sigma_i^2} - 1 \gg \sqrt{\frac{\log n}{d}}.$$

So in full generality, Assumption 3 allows to improve the performance of our estimators. However, this gain seems a bit unnatural, since it strongly depends on this assumption. Hence, to investigate the optimality of our estimators, we will avoid pure noise-level-based estimation by using the following assumption.

Assumption 4. *We assume that*

$$\forall\ (i,j) \in \{1,\dots,n\}^2, \quad \frac{\sigma_i^2}{\sigma_j^2} - 1 \le c_1 \sqrt{\frac{\log n}{d}}.$$

With this assumption, the noise levels can not be too different from each other. Furthermore, in this case, our estimators can not be improved, as shown by the following theorem.

Theorem 12. *Assume that*

- $n \ge 4$,
- Θ_κ *is the set of all* $\boldsymbol{\theta} \in \mathbb{R}^{n\times d}$ *such that* $\kappa(\boldsymbol{\theta}) \ge \kappa$,
- $\forall\ (i,j) \in \{1,\dots,n\}^2$,

$$\frac{\sigma_i^2}{\sigma_j^2} - 1 \le c_1 \sqrt{\frac{\log n}{d}},$$

with $c_1 \le 1/4$.

Then, there exist two constants $c_2, C > 0$ *such that*

$$\kappa < \frac{1}{8}\max\{\sqrt{\log n}, c_2(d\log n)^{1/4}\},$$

implies that

$$\inf_{\hat\pi} \max_{\pi^*\in\mathfrak{S}_n} \sup_{\boldsymbol{\theta}\in\Theta_\kappa} \mathbf{P}_{\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\sigma}^{\#},\pi^*}(\hat\pi \neq \pi^*) > C,$$

where the infimum is taken over all permutation estimators.

Remark 11. We can prove that the constant C in the previous theorem can be chosen larger than 8.75% .

Note that Theorem 11 does not tell anything about the theoretical properties of the greedy algorithm and the LSS under heteroscedasticity. In fact, the matching thresholds of these two procedures are significantly worse than those of the LSNS and the LSL especially for large dimensions d. We state the corresponding result for the greedy algorithm, a similar conclusion being true for the LSS as well. The superiority of the LSNS and LSL is also confirmed by numerical simulations presented in Section 3.7 below.

Theorem 13. *Assume that*
- $d \geq 225 \log 6$,
- $n = 2$,
- $\sigma_1^2 = 3$ *and* $\sigma_2^2 = 1$.

Then

$$\kappa > 0.1(2d)^{1/2}$$

implies that

$$\sup_{\boldsymbol{\theta} \in \Theta_\kappa} \mathbf{P}_{\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\sigma}^\#,id}(\pi^{gr} \neq id) \geq 1/2.$$

This theorem shows that if d is large, the necessary condition for π^{gr} to be consistent is much stronger than the one obtained for π^{LSL} in Theorem 11. Indeed, for the consistency of π^{gr}, κ needs to be at least of the order of $d^{1/2}$, whereas $d^{1/4}$ is sufficient for the consistency of π^{LSL}. Hence, our estimators of the maximum likelihood are, as expected, more interesting than the simple greedy estimator.

3.5 Extension: mean error proportion

In the previous sections, the error of estimation is measured by the risk

$$\mathbf{P}(\hat{\pi} \neq \pi),$$

which may be considered as to restrictive. Indeed, one could be interested in an estimate having a couple of incorrect matches, more than in an estimate with only one half of correct matches. In applications, there is a lot of incorrect permutations with a low proportion of errors that still allow to pertinently estimate the deformation of an object between two pictures. On the other hand, it is much more useful to use the mean error proportion in simulations, to illustrate the different behaviors of the estimators.

So we will use the risk of the mean error proportion, defined as

$$\max_{\pi^* \in \mathfrak{S}_n} \sup_{\boldsymbol{\theta}} \mathbf{E}_{\boldsymbol{\theta},\pi^*}\Big(\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\hat{\pi}(i) \neq \pi^*(i)}\Big).$$

With this risk, the matching threshold of $\hat{\pi}$ becomes

$$\bar{\kappa}_\alpha(\hat{\pi}) = \inf\Big\{\kappa > 0 \,|\, \max_{\pi \in \mathfrak{S}_n} \sup_{\kappa(\boldsymbol{\theta}) > \kappa} \mathbf{E}_{\boldsymbol{\theta},\pi^*}\Big(\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\hat{\pi}(i) \neq \pi^*(i)}\Big) \leq \alpha\Big\}.$$

We prove that our estimators do not lose any efficiency with this new criterion, since

$$\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\hat{\pi}(i) \neq \pi^*(i)} \leq \mathbb{1}_{\hat{\pi} \neq \pi^*}.$$

This translates into the following theorems. First, in the homoscedastic case:

Theorem 14. *Assume that*
- $\alpha \in (0,1)$ *is a tolerance level,*
- *the noise levels are equal to the constant* σ.
- $\hat{\pi}$ *denotes either one of the estimators (3.2)-(3.5).*

Then we have

$$\bar{\kappa}_\alpha^2(\hat{\pi}) \le 16 \max\Big\{2\log\frac{8n^2}{\alpha}, \Big(d\log\frac{4n^2}{\alpha}\Big)^{1/2}\Big\}.$$

Then, in the heteroscedastic case:

Theorem 15. *Assume that*
- $\alpha \in (0,1)$,
- $\forall\, i \in \{1,\dots,n\}, \quad \sigma_i^\# = \sigma_{\pi^*(i)}$,
- $\hat{\pi}$ *is either* π^{LSNS} *(if the noise levels* $\sigma_i, \sigma_i^\#$ *are known) or* π^{LSL} *(when the noise levels are unknown).*

Then

$$\bar{\kappa}_\alpha^2(\hat{\pi}) \le 16 \max\Big\{2\log\frac{8n^2}{\alpha}, \Big(d\log\frac{4n^2}{\alpha}\Big)^{1/2}\Big\}.$$

This was only a direct translation of Theorems 9 and 11, but the lower bound is not so obvious. Nevertheless, we proved in the homoscedastic case, that the performance of our estimators can not be improved. So the minimax matching threshold has the same order of magnitude when the error criterion is the probability of error and when it is the mean error proportion. This goes from the following theorem.

Theorem 16. *Assume that*
- $n \ge 25$,
- Θ_κ *is the set of all* $\boldsymbol{\theta} \in \mathbb{R}^{n\times d}$ *such that* $\kappa(\boldsymbol{\theta}) \ge \kappa$.

Then, there exist two absolute constants $c, C > 0$ *such that:*

$$\kappa \le \frac{1}{4}\max\Big\{(\log n)^{1/2}, c(d\log n)^{1/4}\Big\}$$

implies that

$$\inf_{\hat{\pi}} \max_{\pi^*\in\mathfrak{S}_n} \sup_{\boldsymbol{\theta}\in\Theta_\kappa} \mathbf{E}_{\boldsymbol{\theta},\pi}\Big(\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\hat{\pi}(i)\ne\pi^*(i)}\Big) > C,$$

where the infimum is taken over all permutation estimators.

3.6 Computational aspects

Now, there is a major obstacle to the use of the maximum likelihood estimators, in simulations as in general practice. Indeed, at first sight, the computation of the estimators (3.3)-(3.5) requires to perform an exhaustive search over the set of all possible permutations, the number of which, $n!$, is prohibitively large. This computation is thus impossible in practice as soon as $n \ge 20$.

In this section, we show how to compute these maximum likelihood estimators in polynomial time using linear programming, as it is done for instance in Jebara (2003). Let us consider the LSS estimator

$$\pi^{\mathrm{LSS}} = \arg\min_{\pi\in\mathfrak{S}_n} \sum_{i=1}^n \|X_{\pi(i)} - X_i^\#\|^2.$$

For every permutation π, we denote by P^π the $n\times n$ permutation matrix with coefficients

$$P_{ij}^\pi = \mathbb{1}_{\{j=\pi(i)\}}.$$

Then we can give the equivalent formulation

$$\pi^{\text{LSS}} = \arg\min_{\pi\in\mathfrak{S}_n} \mathbf{tr}\big(MP^{\pi}\big), \tag{3.8}$$

where M is the matrix with coefficient $\|X_i - X_j^{\#}\|^2$ at the i^{th} row and j^{th} column. The cornerstone of our next argument is the Birkhoff-von Neumann theorem stated below, which can be found for example in Budish *et al.* (2009).

Theorem 17 (Birkhoff-von Neumann Theorem)**.** *Assume that $\mathscr{P}$ is the set of all doubly stochastic matrices of size n, i.e., the matrices whose entries are nonnegative and sum up to* 1 *in every row and every column.*
Then, every matrix in $\mathscr{P}$ is a convex combination of matrices $\{P^{\pi} : \pi \in \mathfrak{S}_n\}$. Furthermore, permutation matrices are the vertices of the simplex $\mathscr{P}$.

In view of this result, the combinatorial optimization problem (3.8) is equivalent to the following problem of continuous optimization:

$$P^{\text{LSS}} = \arg\min_{P\in\mathscr{P}} \mathbf{tr}\big(MP\big), \tag{3.9}$$

in the sense that π is a solution to (3.8) if and only if P^{π} is a solution to (3.9). To prove this claim, let us remark that for every $P \in \mathscr{P}$, there exist coefficients

$$\alpha_1, \dots, \alpha_{n!} \in (0,1)$$

such that

$$P = \sum_{i=1}^{n!} \alpha_i P^{\pi_i} \text{ and } \sum_{i=1}^{n!} \alpha_i = 1.$$

Therefore,

$$\mathbf{tr}\big(MP\big) = \sum_{i=1}^{n!} \alpha_i \mathbf{tr}\big(MP^{\pi_i}\big) \geq \min_{\pi\in\mathfrak{S}_n} \mathbf{tr}\big(MP^{\pi}\big)$$

and

$$\mathbf{tr}\big(MP^{\text{LSS}}\big) \geq \mathbf{tr}\big(MP^{\pi^{\text{LSS}}}\big).$$

The great advantage of (3.9) is that it concerns the minimization of a linear function under linear constraints and, therefore, is a problem of linear programming that can be efficiently solved even for large values of n. The same arguments apply to the estimators π^{LSNS} and π^{LSL} (only the matrix M needs to be changed).
Now, there is another way to compute the estimators in practice. Indeed, the computation of one of our maximum likelihood estimators is a particular case of the assignment problem, *i.e.,* finding a minimum weight matching in a weighted bipartite graph, where the matrix of the costs is the matrix M from above, which means that the cost of assigning the i^{th} descriptor of the first image to the j^{th} descriptor of the second image is either

- the square distance between the corresponding observations $\|X_i - X_j^{\#}\|^2$,
- or the normalized square distance $\|X_i - X_j^{\#}\|^2/(\sigma_i^2 + (\sigma_j^{\#})^2)$,
- or the logarithm of the square distance $\log\|X_i - X_j^{\#}\|^2$.

The so-called Hungarian algorithm presented in Kuhn (1955) solves the assignment problem in time $0(n^3)$.

3.7 EXPERIMENTAL RESULTS

Using this computational trick, we have access to the values of the maximum likelihood estimators for reasonably large values of n. We have implemented all the procedures in Matlab and carried out a certain number of numerical experiments on synthetic data. To simplify, we have used the general-purpose solver SeDuMi (*cf.* Sturm (1998)) for solving linear programs. We believe that it is possible to speed-up the computations by using more adapted first-order optimization algorithms, such as coordinate gradient descent. However, even with this simple implementation, the running times are reasonable: for a problem with $n = 500$ features, it takes about six seconds to compute a solution to (3.9) on a standard PC.

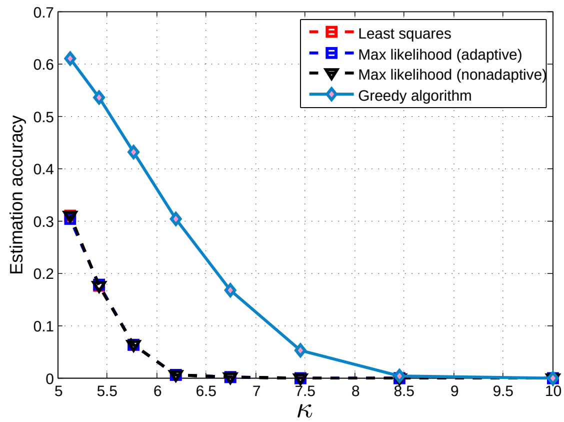

Figure 3.1 – *Average error rate of the four estimating procedures in the experiment with homoscedastic noise as a function of the minimal distance κ between distinct features. One can observe that the LSS, LSNS and LSL procedures are indistinguishable and perform much better than the greedy algorithm.*

Homoscedastic noise Here is the procedure that we followed:

- Choose $n = d = 200$.
- Randomly generate a $n \times d$ matrix $\boldsymbol{\theta}$ with i.i.d. entries uniformly distributed on $[0,\tau]$, with several values of τ varying between 1.4 and 3.5.
- Randomly choose a permutation π^* uniformly over $\mathfrak{S}_n$.
- Generate the sets $\{X_i\}$ and $\{X_i^\#\}$ according to (3.10) with $\sigma_i = \sigma_i^\# = 1$.
- Compute the four estimators of π^* and evaluate their mean error proportion
$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(\hat{\pi}(i) \neq \pi^*(i)).$$

Note that the three estimators originating from the maximum likelihood methodology lead to the same estimators, while the greedy algorithm provides an estimator which is much worse than the others when the parameter κ is small.

Heteroscedastic noise This experiment is similar to the previous one, but the noise levels are different from each other. We followed this procedure

- Choose $n = d = 200$.
- Choose $\boldsymbol{\theta} = \tau I_d$, where I_d is the identity matrix and τ varies between 4 and 10.
- Randomly choose a permutation π^* uniformly over $\mathfrak{S}_n$.

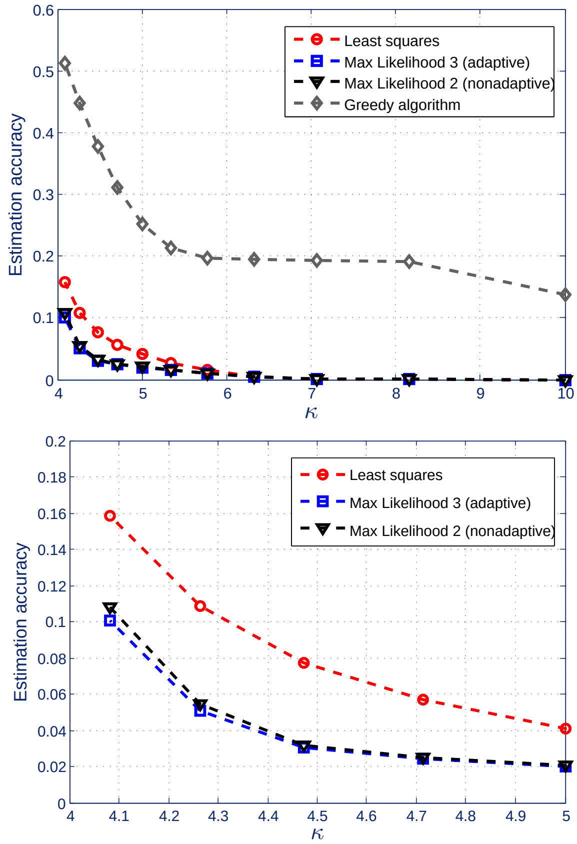

Figure 3.2 – *Top: Mean error proportion of the four estimating procedures in the experiment with heteroscedastic noise as a function of the minimal distance κ between distinct features. Bottom: zoom on the same plots. One can observe that the LSNS and LSL are almost indistinguishable and, as predicted by the theory, perform better than the LSS and the greedy algorithm.*

- Generate the sets $\{X_i\}$ and $\{X_i^\#\}$ according to (3.10) with $\sigma_i = \sigma^\#_{\pi^*(i)} = 1$ for 10 randomly chosen values of i and $\sigma_i = \sigma^\#_{\pi^*(i)} = 0.5$ for the others.
- Compute the four estimators of π^* and evaluated the mean error proportion

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(\hat{\pi}(i) \neq \pi^*(i)).$$

The result, averaged over 500 independent trials, is plotted in Fig. 3.2. Note that among the noise-level-adaptive estimators, LSL outperforms the two others and is as accurate as, and even slightly better than the LSNS pseudo-estimator.

3.8 Extension: other matching criteria

In the previous sections, we considered the case when two vectors θ_i and $\theta_j^\#$ are matched if $\theta_i \equiv \theta_j^\#$, and $\equiv$ is the usual equality $=$. This choice is appropriate when for example the classical SIFT descriptors have to be matched.
In this section, we show that our results can be extended to more general matching criteria, defined as follows: we write $a \equiv_A b$, where A is a given projection matrix, when

$$A(a-b) = 0.$$

Note that the previous case of $=$ is obtained with $\equiv_I$, where I is the identity matrix of size d. Now, denoting

$$A = \frac{1}{d}\begin{pmatrix} d-1 & -1 & \dots & -1 \\ -1 & d-1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & -1 \\ -1 & \dots & -1 & d-1 \end{pmatrix},$$

we define another criterion that matches θ_i and $\theta_j^{\#}$ when they are translated from each other. This could be useful in the case of an illumination change and another descriptor than SIFT for example.
On the other hand, if you consider a descriptor that includes the position of the keypoint (which the SIFT does not: it is invariant by translation), and with our class of matching criteria, we can take geometric constraints into account, by penalizing matches between distant points. This is pertinent when we know that the two images represent the same object or scene taken at two close moments.
In the rest of this section, we show that we can adapt our estimators to define efficient permutation estimators. Our theorem proves that there is no loss of efficiency in this more general case, when d is replaced by **rank**(A). We believe that their performance is also optimal, which would confirm the relevance of our model and the generality of our results.

Now, assume that we have the model

$$\begin{cases} X_i = \theta_i + \sigma_i \xi_i\,, \\ X_i^{\#} = \theta_i^{\#} + \sigma_i^{\#}\xi_i^{\#}, \end{cases} \qquad i = 1,\dots,n \tag{3.10}$$

and for a given permutation π^*, we have

$$\forall\ i \in \{1,\dots,n\}, \quad \theta_{\pi^*(i)} \equiv_A \theta_i^{\#}.$$

We still assume that

$$\forall\ i \in \{1,\dots,n\}, \quad \sigma_i^{\#} = \sigma_{\pi^*(i)},$$

and we define a distance between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^{\#}$ by

$$\kappa_A^2(\boldsymbol{\theta},\boldsymbol{\theta}^{\#}) \triangleq \min_{j\neq\pi^*(i)} \frac{\left\| A(\theta_j - \theta_i^{\#})\right\|^2}{\sigma_j^2 + \sigma_i^2}.$$

The next theorem shows that consistent estimation of the permutation π^* can be obtained in the heteroscedastic case with the estimator inspired by our previous maximum likelihood computations:

$$\pi_A^{\mathrm{LSL}} = \arg\min_{\pi\in\mathfrak{S}_n} \sum_{i=1}^{n} \log\left\| A(X_{\pi(i)} - X_i^{\#})\right\|^2.$$

Theorem 18. *Assume that*
- $\alpha \in (0,1)$,
- $\forall\ i \in \{1,\dots,n\}, \quad \sigma_i^{\#} = \sigma_{\pi^*(i)}$,
- $\hat\pi$ *is either* π^{LSNS} *(if the noise levels* $\sigma_i, \sigma_i^{\#}$ *are known) or* π^{LSL} *(when the noise levels are unknown).*

Then

$$\kappa_A^2 \le 16\max\Big\{2\log\frac{8n^2}{\alpha}, \Big(\mathbf{rank}(A)\log\frac{4n^2}{\alpha}\Big)^{1/2}\Big\}.$$

Remark 12. We could also define equivalents of the estimators π^{LSS} and π^{LSNS} and get the same result.

Remark 13. The only difference between the quantity κ in this theorem and κ in Theorem 11 is that d was replaced by $\mathbf{rank}(A) \le d$. Moreover, for the first example of matrix we gave, $\mathbf{rank}(A) = d$, while for the second, $\mathbf{rank}(A) = d-1$. The rank of A can be seen as the number of error sources.

3.9 Extension: rectangular case

An interesting extension concerns the case of the estimation of a general arrangement, *i.e.,* the case when m and n are not necessarily identical. In such a situation, without loss of generality, one can assume that $n \le m$ and look for an injective function

$$\pi^* : \{1,\dots,n\} \to \{1,\dots,m\}.$$

All the estimators presented in Section 3.2 admit natural counterparts in this rectangular setting. Furthermore, the computational tricks described in the previous section are valid in this setting as well, and are justified by the extension of the Birkhoff-von Neumann theorem recently proved by Budish *et al.* (2009). In this case, the minimization should be carried out over the set of all matrices P of size (n,m) such that

$$P_{i,j} \ge 0,$$

and

$$\begin{cases} \sum_{i=1}^n P_{i,j} \le 1 \\ \sum_{j=1}^m P_{i,j} = 1 \end{cases}, \quad (i,j) \in \{1,\dots,n\}\times\{1,\dots,m\}.$$

This extension is equivalent to allowing the images to contain outliers. We wish to prove the robustness of the LSL with respect to the presence of outliers. The detailed exploration of this problem was not treated in this work, but we would like to underline that the LSL seems well suited for such this situation because of the robustness of the logarithmic function: indeed, the correct matches are strongly rewarded because $\log(0) = -\infty$ and the outliers do not interfere too much with the estimation of the arrangement thanks to the slow growth of log in $+\infty$.

3.10 Extension: sparsity

In the last sections, we proved that the minimax matching threshold was of the order of

$$\max\{\sqrt{\log n}, (d\log n)^{1/4}\}.$$

So the difficulty of the estimation of the underlying permutation is mostly determined by the dimension of the descriptor, but does not depend much on the number of descriptors in the image. That the difficulty grows quickly with d as the total amount of noise in one descriptor grows.

But in the sparse case, when we know that most of the components of the descriptors are zero, the major part of the noise has not to be considered, so that the problem is greatly simplified. Moreover, if we suppose that for every $i \in \{1, \dots, n\}$,

$$\|\theta_i\|, \|\theta_i^{\#}\| \leq s,$$

and if the supports of the descriptors are exactly known, then the minimax matching threshold is obtained from the general one by replacing d by s, and it is thus of the order of

$$\max\left\{\sqrt{\log n}, (s \log n)^{1/4}\right\}.$$

But when no a priori information on the supports is available, then it is not clear wheter this rate can be reached or not. Maybe one has to pay to compensate this lack of information.
Looking for an upper bound result for this problem, we considered the penalized maximum likelihood

$$\min_{\pi,\boldsymbol{\theta}}\left\{\frac{1}{\sigma^2}\sum_{i=1}^n \|X_i - \theta_i\|^2 + \frac{1}{\sigma^2}\sum_{i=1}^n \|X_i^{\#} - \theta_{\pi(i)}\|^2 + \lambda \sum_{i=1}^n \|\theta_i\|_0\right\},$$

which leads to the penalized maximum likelihood estimator

$$\pi^{spar} = \arg\min_{\pi} \sum_{i=1}^n \sum_{j=1}^d \left[(X_{ij} + X^{\#}_{\pi^{-1}(i)j})^2 - 4\sigma^2\lambda\right]_+,$$

where

$$[x]_+ = \max(x, 0).$$

Unfortunately, this estimator does not seem to reach a better rate than $\sqrt{s \log n}$. This could be linked with the difficulty of estimating the square norm of a sparse vector in the Gaussian model. We wish to investigate this problem in the future.

3.11 Proofs of the theorems

In this section we collect the proofs of the theorems. We start with the proof of Theorem 11, since it concerns the more general setting and the proof of Theorem 9 can be deduced from that of Theorem 11 by simple arguments. We then prove the other theorems in the usual order and postpone the proofs of some technical lemmas to the next section.

Proof of Theorem 11

To ease notation and without loss of generality, we assume that π^* is the identity permutation denoted by *id*. Furthermore, since there is no risk of confusion, we write $\mathbf{P}$ instead of $\mathbf{P}_{\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\sigma}^{\#},\pi^*}$. We wish to bound the probability of the event

$$\Omega = \{\hat{\pi} \neq id\}.$$

Let us first denote by $\hat{\pi}$ the maximum likelihood estimator π^{LSL} defined by (3.5). We have

$$\Omega = \bigcup_{\pi \neq id} \Omega_\pi,$$

where

$$\Omega_\pi = \Big\{ \sum_{i=1}^n \log \frac{\|X_i - X_i^\#\|^2}{\|X_{\pi(i)} - X_i^\#\|^2} > 0 \Big\}$$
$$= \Big\{ \sum_{i:\pi(i)\neq i} \log \frac{\|X_i - X_i^\#\|^2}{\|X_{\pi(i)} - X_i^\#\|^2} > 0 \Big\}.$$

On the one hand, for every permutation π,

$$\sum_{\pi(i)\neq i} \log\big(\tfrac{2\sigma_i^2}{\sigma_i^2+\sigma_{\pi(i)}^2}\big) = \sum_{i=1}^n \big(\log(2\sigma_i^2) - \log(\sigma_i^2 + \sigma_{\pi(i)}^2)\big)$$
$$= \sum_{i=1}^n \frac{\log(2\sigma_i^2) + \log(2\sigma_{\pi(i)}^2)}{2} - \log(\sigma_i^2 + \sigma_{\pi(i)}^2)$$

so, using the concavity of the logarithm, this quantity is nonpositive. Therefore,

$$\Omega_\pi \subset \Big\{ \sum_{i:\pi(i)\neq i} \log \frac{\|X_i - X_i^\#\|^2/(2\sigma_i^2)}{\|X_{\pi(i)} - X_i^\#\|^2/(\sigma_i^2+\sigma_{\pi(i)}^2)} > 0 \Big\}$$
$$\subset \bigcup_{i=1}^n \bigcup_{i\neq j} \Big\{ \frac{\|X_i - X_i^\#\|^2}{2\sigma_i^2} > \frac{\|X_j - X_i^\#\|^2}{\sigma_j^2 + \sigma_i^2} \Big\}.$$

This readily yields $\Omega \subset \bar\Omega$, where

$$\bar\Omega = \bigcup_{i=1}^n \bigcup_{i\neq j} \Big\{ \frac{\|X_i - X_i^\#\|^2}{2\sigma_i^2} \geq \frac{\|X_j - X_i^\#\|^2}{\sigma_j^2 + \sigma_i^2} \Big\}. \tag{3.11}$$

Furthermore, the same inclusion is true for the LSNS estimator as well. Therefore, the rest of the proof is common for the estimators LSNS and LSL.

Let us denote

$$\zeta_1 = \max_{i\neq j} \Big| \frac{(\theta_i - \theta_j)^\top(\sigma_i\xi_i - \sigma_j\xi_j^\#)}{\|\theta_i - \theta_j\|\sqrt{\sigma_i^2+\sigma_j^2}} \Big|,$$
$$\zeta_2 = d^{-1/2} \max_{i,j} \Big| \Big\| \frac{\sigma_i\xi_i - \sigma_j\xi_j^\#}{\sqrt{\sigma_i^2+\sigma_j^2}} \Big\|^2 - d \Big|.$$

Since $\pi^* = id$, it holds that for every $i \in \{1,\dots,n\}$,

$$\|X_i - X_i^\#\|^2 = \sigma_i^2 \|\xi_i - \xi_i^\#\|^2 \leq 2\sigma_i^2(d + \sqrt{d}\zeta_2).$$

Similarly, for every $j \neq i$,

$$\|X_j - X_i^\#\|^2 = \|\theta_j - \theta_i\|^2 + \|\sigma_j\xi_j - \sigma_i\xi_i^\#\|^2$$
$$+ 2(\theta_j - \theta_i)^\top(\sigma_j\xi_j - \sigma_i\xi_i^\#).$$

Therefore,

$$\frac{\|X_j - X_i^\#\|^2}{\sigma_i^2 + \sigma_j^2} \geq \frac{\|\theta_j - \theta_i\|^2}{\sigma_i^2 + \sigma_j^2} + d - \sqrt{d}\zeta_2$$
$$- 2\frac{\|\theta_j - \theta_i\|}{\sqrt{\sigma_i^2 + \sigma_j^2}}\zeta_1.$$

But, as we know that

$$\|\theta_j - \theta_i\| / \sqrt{\sigma_i^2 + \sigma_j^2} \geq \kappa(\boldsymbol{\theta}),$$

we infer that

$$\frac{\|X_j - X_i^\#\|^2}{\sigma_i^2 + \sigma_j^2} \geq \min_{a \geq \kappa(\boldsymbol{\theta})} \Big\{a^2 - 2a\zeta_1\Big\} + d - \sqrt{d}\zeta_2.$$

Now, we know from the study of polynomials of the second degree that

$$\min_{a \geq \kappa(\boldsymbol{\theta})} \Big\{a^2 - 2a\zeta_1\Big\} = \kappa^2(\boldsymbol{\theta}) - 2\kappa(\boldsymbol{\theta})\zeta_1$$

when $\kappa(\boldsymbol{\theta}) \geq \zeta_1$, so that on the event $\Omega_1 = \{\kappa(\boldsymbol{\theta}) \geq \zeta_1\}$, it holds that

$$\frac{\|X_j - X_i^\#\|^2}{\sigma_i^2 + \sigma_j^2} \geq \kappa(\boldsymbol{\theta})^2 - 2\kappa(\boldsymbol{\theta})\zeta_1 + d - \sqrt{d}\zeta_2.$$

Combining these bounds, we get that

$$\Omega \cap \Omega_1 \subset \Big\{d + \sqrt{d}\zeta_2 > \kappa(\boldsymbol{\theta})^2 - 2\kappa(\boldsymbol{\theta})\zeta_1 + d - \sqrt{d}\zeta_2\Big\},$$

which implies that

$$\begin{aligned}
\mathbf{P}(\Omega) &\leq \mathbf{P}(\Omega_1^\complement) + \mathbf{P}\big(\Omega \cap \Omega_1\big) \\
&\leq \mathbf{P}\big(\zeta_1 \geq \kappa(\boldsymbol{\theta})\big) + \mathbf{P}(2\sqrt{d}\zeta_2 + 2\kappa(\boldsymbol{\theta})\zeta_1 > \kappa(\boldsymbol{\theta})^2) \\
&\leq 2\mathbf{P}\big(\zeta_1 \geq \tfrac{\kappa(\boldsymbol{\theta})}{4}\big) + \mathbf{P}\big(\zeta_2 > \tfrac{\kappa(\boldsymbol{\theta})^2}{4\sqrt{d}}\big). \qquad (3.12)
\end{aligned}$$

Finally, one easily checks that for suitably chosen random variables $\zeta_{i,j}$ drawn from the standard Gaussian distribution, it holds that

$$\zeta_1 = \max_{i \neq j} |\zeta_{i,j}|.$$

Therefore, using the well-known tail bound for the standard Gaussian distribution in conjunction with the union bound, we get

$$\begin{aligned}
\mathbf{P}\big(\zeta_1 \geq \tfrac{1}{4}\kappa(\boldsymbol{\theta})\big) &\leq \sum_{i \neq j} \mathbf{P}\big(|\zeta_{i,j}| \geq \tfrac{1}{4}\kappa(\boldsymbol{\theta})\big) \\
&\leq 2n^2 e^{-\frac{1}{32}\kappa(\boldsymbol{\theta})^2}. \qquad (3.13)
\end{aligned}$$

Similarly, using the tail bound of Corollary 5, we get

$$\mathbf{P}\big(\zeta_2 > \tfrac{\kappa(\boldsymbol{\theta})^2}{4\sqrt{d}}\big) \leq 2n^2 e^{-\frac{(\kappa(\boldsymbol{\theta})/16)^2}{d}(\kappa^2(\boldsymbol{\theta}) \wedge 8d)}. \qquad (3.14)$$

Combining inequalities (3.12)-(3.14), we obtain that as soon as

$$\kappa(\boldsymbol{\theta}) \geq 4\Big(\sqrt{2\log(8n^2/\alpha)} \vee \big(d\log(4n^2/\alpha)\big)^{1/4}\Big),$$

we have

$$\mathbf{P}(\hat{\pi} \neq \pi^*) = \mathbf{P}(\Omega) \leq \alpha.$$

Proof of Theorem 9

It holds that, on the event

$$\mathscr{A} = \bigcap_{i=1}^{n} \bigcap_{j \neq i} \Big\{ \|X_{\pi^*(i)} - X_i^\#\| < \|X_{\pi^*(j)} - X_i^\#\| \Big\},$$

all the four estimators coincide with the true permutation π^*. Therefore, we have

$$\{\hat{\pi} \neq \pi^*\} \subseteq \bigcup_{i=1}^{n} \bigcup_{j \neq i} \Big\{ \|X_{\pi^*(i)} - X_i^\#\| > \|X_{\pi^*(j)} - X_i^\#\| \Big\}.$$

The latter event coincides with $\bar{\Omega}$ at the right-hand side of (3.11), the probability of which has been already evaluated in the previous proof. This completes the proof.

Proof of Theorem 10

For two probability measures $\mathbf{P}$ and $\mathbf{Q}$ such that $\mathbf{P}$ is absolutely continuous with respect to $\mathbf{Q}$, we denote by $K(\mathbf{P},\mathbf{Q})$ the Kullback-Leibler divergence between $\mathbf{P}$ and $\mathbf{Q}$ defined by

$$K(\mathbf{P},\mathbf{Q}) = \int \log \frac{d\mathbf{P}}{d\mathbf{Q}}\, d\mathbf{P}.$$

In our proof, we decided to separate the cases when

$$\max\Big\{ (\log n)^{1/2}, c(d\log n)^{1/4} \Big\} = (\log n)^{1/2}$$

and when

$$\max\Big\{ (\log n)^{1/2}, c(d\log n)^{1/4} \Big\} = c(d\log n)^{1/4}.$$

Besides, we will repeatedly use the fact that for $n \geq 4$,

$$\log(n/2) \geq \frac{1}{2}\log n$$

and, to ease notation, we will omit the subscripts $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}^\#$ in $\mathbf{P}_{\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\sigma}^\#,\pi^*}$.

First part: $\kappa \leq \frac{1}{4}\sqrt{\log n}$

Notice that in this first case $\kappa \leq \sqrt{\frac{1}{8}\log(n/2)}$, which is the bound we will use from now on.

We choose $\boldsymbol{\theta}_0 \in \mathbb{R}^{n\times d}$ such that

$$\theta_{0,i} = (i \times \sqrt{2}\sigma\kappa, 0, \ldots, 0)^\top \in \mathbb{R}^d$$

for every $i \in \{1,\ldots,n\}$. We reduce our problem from considering Θ_κ to considering only $\boldsymbol{\theta}_0 \in \Theta_\kappa$:

$$\begin{aligned}&\inf_{\hat{\pi}} \max_{\pi^* \in \mathfrak{S}_n} \sup_{\boldsymbol{\theta}\in\Theta_\kappa} \mathbf{P}_{\boldsymbol{\theta},\pi^*}\big(\hat{\pi} \neq \pi^*\big)\\ &\geq \inf_{\hat{\pi}} \max_{\pi^* \in \mathfrak{S}_n} \mathbf{P}_{\boldsymbol{\theta}_0,\pi^*}\big(\hat{\pi} \neq \pi^*\big).\end{aligned}$$

A lower bound for this quantity can be obtained from the following lemma:

Lemma 10 (Tsybakov (2008)). *Assume that*

- *M is an integer larger than* 2*,*
- *there exist distinct permutations $\pi_0, \dots, \pi_M \in \mathfrak{S}_n$ and mutually absolutely continuous probability measures $\mathbf{Q}_0, \dots, \mathbf{Q}_M$ such that*

$$\frac{1}{M}\sum_{j=1}^{M} K(\mathbf{Q}_j, \mathbf{Q}_0) \le \frac{1}{8}\log M.$$

Then

$$\inf_{\tilde{\pi}} \max_{j=0,\dots,M} \mathbf{Q}_j(\tilde{\pi} \neq \pi_j) \ge \frac{\sqrt{M}}{\sqrt{M}+1}\Big(\frac{3}{4} - \frac{1}{2\sqrt{\log M}}\Big),$$

where the infimum is taken over all permutation estimators.

We will apply this lemma with

$$M = n-1$$

and, for $i = 1, \dots, n$,

$$\pi_i = (i\ i+1),$$

the permutation which leaves all $j \in \{1, \dots, n\} \setminus \{i, i+1\}$ invariant and switches i and $i+1$. Then, we have

$$K\big(\mathbf{Q}_i, \mathbf{Q}_0\big) = \frac{1}{2\sigma^2}\sum_{k=1}^{n} \|\theta_{0,\pi_i(k)} - \theta_{0,k}\|^2 = \kappa^2.$$

Thus, if $n \ge 4$ and $\kappa \le \sqrt{\frac{1}{8}\log(n/2)}$, then

$$\frac{1}{M}\sum_{i=1}^{M} K\big(\mathbf{Q}_i, \mathbf{Q}_0\big) \le \frac{1}{8}\log(n/2) \le \frac{1}{8}\log M$$

and Lemma 10 yields the desired result with

$$\forall\ 0 \in \{, \dots, M\},\ \mathbf{Q}_i = \mathbf{P}_{\boldsymbol{\theta}_0,\pi_i}.$$

Second part: $\kappa \le \frac{c}{4}(d\log n)^{1/4}$
In this second part, we suppose that

$$c \le 1,$$

so that

$$\kappa \le \frac{1}{4}\big(d\log n\big)^{1/4},$$

which is the bound we will use in this proof. Furthermore, the hypothesis made on the maximum implies that

$$d \ge \frac{1}{c^4}\log n \ge \log n.$$

Now, let μ be a prior probability measure on $\mathbb{R}^{n\times d}$. Define the posterior probability

$$\mathbf{P}_{\mu,\pi} = \int_{\mathbb{R}^{nd}} \mathbf{P}_{\boldsymbol{\theta},\pi}\ \mu(d\boldsymbol{\theta}).$$

It holds that for every $M > 0$, and every permutations $\pi_0, \ldots, \pi_M \in \mathfrak{S}_n$,

$$\begin{aligned}
&\max_{\pi^* \in \mathfrak{S}_n} \sup_{\boldsymbol{\theta} \in \Theta_\kappa} \mathbf{P}_{\boldsymbol{\theta},\pi^*}\left(\hat{\pi} \neq \pi^*\right) \\
&\geq \max_{\pi_0,\ldots,\pi_M} \int_{\Theta_\kappa} \mathbf{P}_{\boldsymbol{\theta},\pi^*}\left(\hat{\pi} \neq \pi\right) \frac{\mu(d\boldsymbol{\theta})}{\mu(\Theta_\kappa)} \\
&\geq \max_{\pi_0,\ldots,\pi_M} \mathbf{P}_{\mu,\pi^*}\left(\hat{\pi} \neq \pi^*\right) - \mu(\mathbb{R}^{nd} \backslash \Theta_\kappa).
\end{aligned}$$

We will use Lemma 10 again with

$$\begin{cases} M = n, \\ \pi_0 = id \text{ and } \pi_1, \ldots, \pi_M \ M \text{ distinct transpositions,} \\ \forall i \in \{0, \ldots, M\},\ \mathbf{Q}_i = \mathbf{P}_{\mu,\pi_i}. \end{cases}$$

To this end, we state the following lemma, that allows us to bound the Kullback-Leibler divergence from above when

$$\mu \text{ is the uniform distribution on } \{\pm\epsilon\}^{n\times d}.$$

Lemma 11. *Assume that*

- *ϵ is a positive real number with $\epsilon \leq \sigma/2$,*
- *μ is the uniform distribution on $\{\pm\epsilon\}^{n\times d}$.*

Then, for any transposition π, we have

$$K(\mathbf{P}_{\mu,\pi}, \mathbf{P}_{\mu,id}) \leq \frac{8d\epsilon^4}{\sigma^4}.$$

Furthermore, if

$$\epsilon = \frac{\sqrt{2}\sigma\kappa}{\sqrt{d}},$$

then

$$\mu(\mathbb{R}^{nd} \backslash \Theta_\kappa) \leq \frac{n(n-1)}{2}\, e^{-d/8}.$$

Using the prior μ and the value of $\epsilon \leq \sigma/2$ defined in the previous lemma, we get

$$\frac{1}{M}\sum_{i=1}^{M} K(\mathbf{Q}_i, \mathbf{Q}_0) \leq \frac{8d\epsilon^4}{\sigma^4} = \frac{32\kappa^4}{d} \leq \frac{1}{8}\log n.$$

This implies that the minimum risk is larger than

$$\frac{\sqrt{3}}{\sqrt{3}+1}\left(\frac{3}{4} - \frac{1}{2\sqrt{\log 3}}\right) - \frac{n^2}{2} e^{-d/8}.$$

Finally, remembering that $d \geq \frac{1}{c^4}\log n$, we have

$$(n^2/2)e^{-d/8} \leq \frac{1}{2} n^{2-1/8c^4}.$$

Taking c small enough, we get the desired result.

Proof of Theorem 12

In our proof, we decide to separate the cases when

$$\max\left\{(\log n)^{1/2}, c_2(d\log n)^{1/4}\right\} = (\log n)^{1/2}$$

and when

$$\max\left\{(\log n)^{1/2}, c_2(d\log n)^{1/4}\right\} = c_2(d\log n)^{1/4}.$$

First part: $\kappa \le \frac{1}{8}\sqrt{\log n}$
Denote m the largest integer such that $2m \le n$. As $n \ge 4$, we have

$$\kappa \le \frac{1}{4}\sqrt{\log m},$$

which is the bound we will use from now on. Moreover, we assume without loss of generality that the noise levels are ranked in increasing order:

$$\sigma_1 \le \ldots \le \sigma_n.$$

Then, we construct a least favorable set of vectors for the estimation of the permutation.

Lemma 12. *Assume that m is the largest integer such that $2m \le n$. Then there is a set of vectors $\boldsymbol{\theta}$ such that*

$$\frac{\|\theta_1 - \theta_2\|}{\sqrt{\sigma_1^2 + \sigma_2^2}} = \ldots = \frac{\|\theta_{2m-1} - \theta_{2m}\|}{\sqrt{\sigma_{2m-1}^2 + \sigma_{2m}^2}} = \kappa,$$

and

$$\{i,j\} \neq \{1,2\}, \ldots, \{2m-1, 2m\} \Rightarrow \frac{\|\theta_i - \theta_j\|}{\sqrt{\sigma_i^2 + \sigma_j^2}} > \kappa + \frac{\max_{k=1,\ldots,m} |\eta_k|}{\sqrt{\sigma_i^2 + \sigma_j^2}},$$

where for every $k = 1, \ldots, m$,

$$\eta_k = \frac{\sigma_{2k}^2 - \sigma_{2k-1}^2}{\sigma_{2k}^2 + \sigma_{2k-1}^2}(\theta_{2k-1} - \theta_{2k}).$$

Denote $\boldsymbol{\theta}^0$ the constructed set. We define for every $i \in \{1, \ldots, m\}$

$$\boldsymbol{\theta}^i = \boldsymbol{\theta}^0 + (0, \ldots, \eta_i, \eta_i, \ldots, 0),$$

where only the $(2i-1)^{\text{th}}$ and $2i^{\text{th}}$ components are nonzero, and η_i is defined as in Lemma 12. It follows from Lemma 12 that

$$\boldsymbol{\theta}^0, \ldots, \boldsymbol{\theta}^m \text{ belong to } \Theta_\kappa,$$

so that, denoting for every $i \in \{1, \ldots, m\}$

$$\pi_i = (2i-1\ 2i)$$

the transposition of $\mathfrak{S}_n$ that only permutes $2i-1$ and $2i$, and

$$\pi_0 = id,$$

we get the following lower bound for the risk:

$$\inf_{\hat{\pi}} \max_{\pi^* \in \mathfrak{S}_n} \sup_{\boldsymbol{\theta} \in \Theta_\kappa} \mathbf{P}_{\pi^*, \boldsymbol{\theta}}(\hat{\pi} \neq \pi^*) \ge \inf_{\hat{\pi}} \sup_{i=0,\ldots,m} \mathbf{P}_{\boldsymbol{\theta}^i, \pi_i}(\hat{\pi} \neq \pi).$$

Now, in order to use Lemma 10, we compute for every $k \in \{1, \ldots, m\}$

$$\begin{aligned}
2K(\mathbf{P}_{\boldsymbol{\theta}^k, \pi_k}, \mathbf{P}_{\boldsymbol{\theta}^0, \pi_0}) = {} & \frac{\|\theta_{2k-1} + \eta_k - \theta_{2k}\|^2}{\sigma_{2k}^2} + d\Big(\frac{\sigma_{2k-1}^2}{\sigma_{2k}^2} - 1\Big) - d\log\frac{\sigma_{2k-1}^2}{\sigma_{2k}^2} \\
& + \frac{\|\theta_{2k} + \eta_k - \theta_{2k-1}\|^2}{\sigma_{2k-1}^2} + d\Big(\frac{\sigma_{2k}^2}{\sigma_{2k-1}^2} - 1\Big) - d\log\frac{\sigma_{2k}^2}{\sigma_{2k-1}^2}.
\end{aligned}$$

Using the definition of η_k, we get

$$2K(\mathbf{P}_{\boldsymbol{\theta}^k,\pi_k},\mathbf{P}_{\boldsymbol{\theta}^0,\pi_0}) = 4\frac{\|\theta_{2k}-\theta_{2k-1}\|^2}{\sigma_{2k}^2+\sigma_{2k-1}^2} + d\frac{\sigma_{2k-1}^2}{\sigma_{2k}^2}\big(1-\frac{\sigma_{2k}^2}{\sigma_{2k-1}^2}\big)^2$$
$$\leq 4\kappa^2 + c_1^2\log n.$$

Then, since $\log n \leq 2\log m$ for $n \geq 4$, and since we assumed that $c_1 \leq \frac{1}{4}$,

$$K(\mathbf{P}_{\boldsymbol{\theta}^k,\pi_k},\mathbf{P}_{\boldsymbol{\theta}^0,\pi_0}) \leq 2\kappa^2 + c_1^2\log m \leq \frac{1}{8}\log m.$$

We conclude by Lemma 10 that

$$\inf_{\hat\pi}\max_{\pi^*\in\mathfrak{S}_n}\sup_{\boldsymbol{\theta}\in\Theta_\kappa}\mathbf{P}_{\pi,\boldsymbol{\theta}}(\hat\pi\neq\pi) \geq \frac{\sqrt{2}}{\sqrt{2}+1}\Big(\frac{3}{4}-\frac{1}{2\sqrt{\log 2}}\Big).$$

Second part: $\kappa \leq \frac{c_2}{8}(d\log n)^{1/4}$
In this case, we have

$$d \geq \frac{1}{c_2^4}\log n \geq 2\log n$$

if we suppose that $c_2 \leq 2^{-1/4}$. Let

$$\pi_{i,j} = (i\ j),$$

the transposition that only permutes i and j, with $i<j$. Assume that

$$\boldsymbol{\theta} = (\theta_1,\ldots,\theta_n) \sim \mu,$$

where μ is the uniform distribution on $\prod_{i=1}^n\{\pm\epsilon_i\}^d$ and $\epsilon_1,\ldots,\epsilon_n$ are some given positive real numbers. Now, we state the following lemma

Lemma 13. *Assume that*
- $\epsilon_1,\ldots,\epsilon_n$ *are real numbers defined by* $\epsilon_k = \sqrt{2/d}\,\kappa\sigma_k$,
- μ *is the uniform distribution on* $\{\pm\epsilon_1\}^d\times\ldots\times\{\pm\epsilon_n\}^d$,
- $\pi = (i\ j)$ *is the transposition that only permutes* i *and* j,
- $\sigma_1 \leq \ldots \leq \sigma_n$.

Then

$$K(\mathbf{P}_{\mu,\pi},\mathbf{P}_{\mu,id}) \leq 4\kappa^2\Big(1-\frac{\sigma_i^2}{\sigma_j^2}\Big) + \frac{8\kappa^4}{d}\Big(2+\big(1+(2/d)\kappa^2\big)^2\frac{\sigma_j^4}{\sigma_i^4}\Big)$$
$$+\frac{1}{2}\big(d+2\kappa^2\big)\Big(\frac{\sigma_j^2}{\sigma_i^2}-1\Big)^2$$

and

$$\mu(\mathbb{R}^{nd}\backslash\Theta_\kappa) \leq (n(n-1)/2)\ e^{-d/8}.$$

The assumption on the noise levels entails that

$$1 \leq \frac{\sigma_j^2}{\sigma_i^2} \leq 1+\frac{1}{4}\sqrt{\frac{\log n}{d}},$$

and consequently,

$$\Big(\frac{\sigma_j^2}{\sigma_i^2}-1\Big)^2 \le \frac{1}{16}\Big(\frac{\log n}{d}\Big).$$

Furthermore,

$$\frac{\kappa^2}{d} \le \frac{c}{64} \le \frac{1}{64}$$

since $c \le 1$. Finally, it holds that

$$\begin{aligned} K(\mathbf{P}_{\mu,\pi_{i,j}}, \mathbf{P}_{\mu,id}) &\le \kappa^2\sqrt{\frac{\log n}{d}} + \frac{8\kappa^4}{d}\Big(2+\frac{33^2}{32^2}\big(1+0.25\big)^2\Big) + \frac{33d}{64}\times\frac{\log n}{16d} \\ &\le \frac{\log n}{8} \le \frac{\log n(n-1)/2}{8}, \end{aligned}$$

for c_2 small enough. Applying Lemma 10 with

- $M = n(n-1)/2$,
- $\mathbf{Q}_0 = \mathbf{P}_{\mu,id}$,
- $\{\mathbf{Q}_j\}_{j=1,\ldots,M} = \{\mathbf{P}_{\mu,\pi_{i,j}}\}_{j\neq i}\}$,

we obtain

$$\begin{aligned} \max_{\pi^*\in\mathfrak{S}_n}\sup_{\boldsymbol{\theta}\in\Theta_\kappa} \mathbf{P}_{\boldsymbol{\theta},\pi^*}\big(\hat\pi\neq\pi^*\big) &\ge \max_{\pi^*\in\{id\}\cup\{\pi\}}\int_{\Theta_\kappa}\mathbf{P}_{\boldsymbol{\theta},\pi^*}\big(\hat\pi\neq\pi^*\big)\frac{\mu(d\boldsymbol{\theta})}{\mu(\Theta_\kappa)} \\ &\ge \max_{\pi^*\in\{id\}\cup\{\pi\}}\mathbf{P}_{\mu,\pi^*}\big(\hat\pi\neq\pi^*\big) - \mu(\mathbb{R}^{nd}\backslash\Theta_\kappa) \\ &\ge \frac{\sqrt{15}}{\sqrt{15}+1}\Big(\frac{3}{4}-\frac{1}{2\sqrt{\log 15}}\Big) - \frac{1}{2}n^{2-1/8c_2^4}. \end{aligned}$$

The result follows by taking c_2 small enough.

Proof of Theorem 13

It holds that $\{\pi^{\mathrm{gr}}\neq\pi^*\}$ includes the smaller set

$$\Omega_2 = \{\|X_1 - X_1^\#\|^2 > \|X_1 - X_2^\#\|^2\}.$$

Thus, it is sufficient to give a lower bound of the probability of the event Ω_2. Then, we choose any $\boldsymbol{\theta}\in\mathbb{R}^{n\times d}$ satisfying

$$\|\theta_1-\theta_2\| = 2\tilde\kappa.$$

Now, it holds, for suitably chosen random variables

$$\begin{cases} \eta_1 \sim \chi_d^2, \\ \eta_2 \sim \chi_d^2, \\ \eta_3 \sim \mathcal{N}(0,1), \end{cases}$$

that

$$\|X_1 - X_1^\#\|^2 - \|X_1 - X_2^\#\|^2 = 6\eta_1 - 4\kappa^2 - 8\kappa\eta_3 - 4\eta_2.$$

These variables can be controlled using the following proposition and its corollary:

Proposition 7 (Laurent *et* Massart (2000))**.** *Assume that $Y \sim \chi^2(D)$, where $D \in \mathbb{N}^*$. Then, for every $x > 0$,*

$$\begin{cases} \mathbf{P}\big(Y - D \le -2\sqrt{Dx}\big) \le e^{-x}, \\ \mathbf{P}\big(Y - D \ge 2\sqrt{Dx} + 2x\big) \le e^{-x}. \end{cases}$$

Corollary 4. *Assume that $Y \sim \chi^2(D)$, where $D \in \mathbb{N}^*$. Then, for every $y > 0$,*

$$\mathbf{P}\big(D^{-1/2}|Y - D| \ge y\big) \le 2\exp\Big\{-\tfrac{1}{8}y(y \wedge \sqrt{D})\Big\}.$$

Hence, for every $x > 0$, each one of the following three inequalities holds true with probability at least $1 - e^{-x^2}$:

$$\begin{aligned} \eta_1 &\ge d - 2\sqrt{d}x, \\ \eta_2 &\le d + 2\sqrt{d}x + 2x^2, \\ \eta_3 &\le \sqrt{2}x. \end{aligned}$$

This implies that with probability at least $1 - 3e^{-x^2}$, we have

$$\|X_1 - X_1^\#\|^2 - \|X_1 - X_2^\#\|^2 \ge 2d - 20\sqrt{d}x - 4(\kappa + \sqrt{2}x)^2.$$

If $x = \sqrt{\log 6}$, then the conditions imposed on κ and d ensure that the right-hand side of the last inequality is positive. Therefore,

$$\mathbf{P}(\bar{\Omega}) \ge 1 - 3e^{-x^2} = 1/2.$$

Proof of Theorem 16

In our proof, we decided to separate the cases when

$$\max\Big\{(\log n)^{1/2}, c(d\log n)^{1/4}\Big\} = (\log n)^{1/2}$$

and when

$$\max\Big\{(\log n)^{1/2}, c(d\log n)^{1/4}\Big\} = c(d\log n)^{1/4}.$$

Besides, we will repeatedly use the fact that for $n \ge 4$,

$$\log(n/2) \ge \frac{1}{2}\log n$$

and, to ease notation, we will omit the subscripts $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}^\#$ in $\mathbf{P}_{\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\sigma}^\#,\pi^*}$.

First part: $\kappa \le \frac{1}{8}\sqrt{\log n}$
Notice that in this first case, as $n \ge 25$,

$$\kappa \le \frac{1}{16}\sqrt{\log n/8},$$

which is the bound we will use from now on. We choose $\boldsymbol{\theta}_0 \in \mathbb{R}^{n\times d}$ such that for every $i \in \{1,\ldots,n\}$,

$$\theta_{0,i} = (i \times \sqrt{2}\sigma\kappa, 0, \ldots, 0)^\top \in \mathbb{R}^d.$$

We reduce our problem from considering Θ_κ to considering only $\boldsymbol{\theta}_0 \in \Theta_\kappa$:

$$\begin{aligned}
&\inf_{\hat{\pi}} \max_{\pi^* \in \mathfrak{S}_n} \sup_{\boldsymbol{\theta} \in \Theta_\kappa} \mathbf{E}_{\boldsymbol{\theta},\pi^*}\Big(\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\hat{\pi}(i)\neq\pi^*(i)}\Big) \\
&\geq \inf_{\hat{\pi}} \max_{\pi^* \in \mathfrak{S}_n} \mathbf{E}_{\boldsymbol{\theta}_0,\pi^*}\Big(\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\hat{\pi}(i)\neq\pi^*(i)}\Big).
\end{aligned}$$

Now, using Markov's inequality, we get

$$\mathbf{E}_{\boldsymbol{\theta}_0,\pi^*}\Big(\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\hat{\pi}(i)\neq\pi^*(i)}\Big) \geq \frac{1}{4}\,\mathbf{P}_{\boldsymbol{\theta}_0,\pi}\Big(\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\hat{\pi}(i)\neq\pi^*(i)} \geq 1/4\Big).$$

A lower bound for this quantity can be obtained from the following lemma:

Lemma 14 (Tsybakov (2008))**.** *Assume that*

- *M is an integer larger than 2,*
- *there exist $\pi_0,\dots,\pi_M$ in $\mathfrak{S}_n$, mutually absolutely continuous probability measures $\mathbf{Q}_0,\dots,\mathbf{Q}_n$ and a pseudo-distance δ such that*

$$\begin{cases} \exists\, s>0,\ \forall\, i\neq j,\ \delta(\pi_i,\pi_j)\geq 2s, \\ \frac{1}{M}\sum_{j=1}^M K(\mathbf{Q}_j,\mathbf{Q}_0) \leq \frac{1}{8}\log M. \end{cases}$$

Then

$$\inf_{\hat{\pi}} \max_{j=0,\dots,n} \mathbf{Q}_j\big(\delta(\hat{\pi},\pi_j)\geq s\big) \geq \frac{\sqrt{M}}{\sqrt{M}+1}\Big(\frac{3}{4} - \frac{1}{2\sqrt{\log M}}\Big),$$

where the infimum is taken over all permutation estimators.

We now have to choose a proper M and $\pi_0,\dots,\pi_M$, which we take as in the next lemma.

Lemma 15. *Assume that n is an integer larger than 8. Then there exist some permutations $\pi_0,\pi_1,\dots,\pi_M$ with*

$$\pi_0 = id,\ M \geq (n/8)^{n/2}$$

and

$$\forall\, i,j\in\{0,\dots,M\},\ i\neq j \quad \Rightarrow \quad \frac{1}{n}\sum_{k=1}^n \mathbb{1}_{\pi_i(k)\neq\pi_j(k)} \geq \frac{1}{2}.$$

Denoting $\mathbf{Q}_i = \mathbf{P}_{\boldsymbol{\theta}_0,\pi_i}$, we compute

$$\begin{aligned}
K\big(\mathbf{Q}_i,\mathbf{Q}_0\big) &= \frac{1}{2\sigma^2}\sum_{k=1}^n \|\theta_{0,\pi_i(k)} - \theta_{0,k}\|^2 \\
&= \frac{1}{2\sigma^2}\sum_{k=1}^n \|\theta_{0,\pi_i(k)} - \theta_{0,k}\|^2 \mathbb{1}_{\pi_i(k)\neq k} \\
&\leq \frac{1}{2\sigma^2}(\sqrt{2}\sigma\kappa)^2 \sum_{k=1}^n \mathbb{1}_{\pi_i(k)\neq k} \\
&\leq \frac{n\kappa^2}{2},
\end{aligned}$$

and

$$\frac{n\kappa^2}{2} \leq \frac{1}{8}\log M,$$

by assumption on κ and M. Now, we have

$$M \geq 2,$$

so Lemma 14 yields that

$$\inf_{\hat{\pi}} \max_{\pi^* \in \mathfrak{S}_n} \sup_{\boldsymbol{\theta} \in \Theta_\kappa} \mathbf{E}_{\boldsymbol{\theta},\pi^*}\Big(\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\hat{\pi}(i)\neq\pi^*(i)}\Big) > \frac{\sqrt{2}}{\sqrt{2}+1}\Big(\frac{3}{4} - \frac{1}{2\sqrt{\log 2}}\Big).$$

Second part: $\kappa \leq \frac{c}{8}(d \log n)^{1/4}$
In this second part, we suppose that $c \leq 1$, so that

$$\kappa \leq \frac{1}{8}\big(d \log n\big)^{1/4}.$$

Furthermore, the hypothesis made on the maximum implies that

$$d \geq \frac{1}{c^4}\log n \geq \log n.$$

By Markov's inequality,

$$\begin{aligned}
&\sup_{\boldsymbol{\theta} \in \Theta_\kappa} \mathbf{E}_{\boldsymbol{\theta},\pi^*}\Big(\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\hat{\pi}(i)\neq\pi^*(i)}\Big) \\
&\geq \frac{3}{16}\sup_{\boldsymbol{\theta} \in \Theta_\kappa} \mathbf{P}_{\boldsymbol{\theta},\pi^*}\big(\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\hat{\pi}(i)\neq\pi^*(i)} \geq 3/16\big) \\
&\geq \frac{3}{16}\Big(\mathbf{P}_{\mu,\pi^*}\big(\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\hat{\pi}(i)\neq\pi^*(i)} \geq 3/16\big) - \mu(\mathbb{R}^{nd}\backslash\Theta_\kappa)\Big),
\end{aligned}$$

where μ is the uniform distribution on $\{\pm\epsilon\}^{n\times d}$ with

$$\epsilon = \frac{\sqrt{2}\sigma\kappa}{\sqrt{d}}.$$

Now, choose M and $\pi_0, \dots, \pi_M$ as in the following lemma.

Lemma 16. *Assume that n is an integer larger than 25. Then there exist permutations $\pi_0, \dots, \pi_M$ with*

$$\pi_0 = id, \ M \geq (n/24)^{n/6},$$

and for every $i \neq j \in \{0, \dots, M\}$,

$$\begin{cases} \pi_i \text{ is the composition of distinct transpositions,} \\ \frac{1}{n}\sum_{k=1}^n \mathbb{1}_{\pi_i(k)\neq\pi_j(k)} \geq \frac{3}{8}. \end{cases}$$

So, with this notation, we get

$$\begin{aligned}
&\max_{\pi^* \in \mathfrak{S}_n} \sup_{\boldsymbol{\theta} \in \Theta_\kappa} \mathbf{E}_{\boldsymbol{\theta},\pi^*}\Big(\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\hat{\pi}(i)\neq\pi^*(i)}\Big) \\
&\geq \frac{3}{16}\Big(\max_{i=0,\dots,M} \mathbf{P}_{\mu,\pi_i}\big(\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\hat{\pi}(i)\neq\pi^*(i)} \geq 3/16\big) - \mu(\mathbb{R}^{nd}\backslash\Theta_\kappa)\Big).
\end{aligned}$$

As π_i is a product of transpositions, the Kullback-Leibler divergence between $\mathbf{P}_{\mu,\pi_i}$ and $\mathbf{P}_{\mu,\pi_0}$ can be computed by independence thank to Lemma 11:

$$\frac{1}{M}\sum_{i=1}^M K(\mathbf{P}_{\mu,\pi_i}, \mathbf{P}_{\mu,\pi_0}) \leq \frac{n}{2} \times \frac{8d\epsilon^4}{\sigma^4} = \frac{16n\kappa^4}{d}.$$

We can now apply Lemma 14. It holds that $M \geq 2$, so that

$$\inf_{\hat\pi} \max_{\pi^* \in \mathfrak{S}_n} \sup_{\boldsymbol{\theta} \in \Theta_\kappa} \mathbf{E}_{\boldsymbol{\theta},\pi^*}\Big(\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\hat\pi(i)\neq\pi^*(i)}\Big) > \frac{3}{16} \times \frac{\sqrt{2}}{\sqrt{2}+1}\Big(\frac{3}{4} - \frac{1}{2\sqrt{\log 2}}\Big) - \frac{n^2}{2}e^{-d/8}.$$

Finally, remembering that $d \geq \frac{1}{c^4}\log n$, we have

$$(n^2/2)e^{-d/8} \leq \frac{1}{2}n^{2-1/8c^4}.$$

Taking c small enough, we get the desired result.

Proof of Theorem 18

This proof is formally identical to the proof of Theorem 11. Let us denote $\hat\pi = \pi_A^{\mathrm{LSL}}$ and assume that $\pi^* = id$. We have

$$\{\hat\pi \neq id\} \subset \bigcup_{i=1}^n \bigcup_{i\neq j} \Big\{\Big\| A\,\frac{X_i - X_i^\#}{\sqrt{2\sigma_i^2}}\Big\|^2 > \Big\| A\,\frac{X_j - X_i^\#}{\sqrt{\sigma_j^2+\sigma_i^2}}\Big\|^2\Big\},$$

and denoting

$$\zeta_1 = \max_{i\neq j}\Big|\Big(\tfrac{A(\theta_j-\theta_i^\#)}{\|A(\theta_j-\theta_i^\#)\|}\Big)^\top\Big(A\,\tfrac{\sigma_j\xi_j-\sigma_i\xi_i^\#}{\sqrt{\sigma_j^2+\sigma_i^2}}\Big)\Big|,$$
$$\zeta_2 = \mathbf{rank}(A)^{-1/2}\max_{i,j}\Big|\Big\|A\,\tfrac{\sigma_j\xi_j-\sigma_i\xi_i^\#}{\sqrt{\sigma_j^2+\sigma_i^2}}\Big\|^2 - \mathbf{rank}(A)\Big|,$$

and

$$\Omega = \{\kappa_A(\boldsymbol{\theta},\boldsymbol{\theta}^\#) \geq \zeta_1\},$$

we get

$$\{\hat\pi \neq id\}\cap\Omega \subset \Big\{\mathbf{rank}(A) + \sqrt{\mathbf{rank}(A)}\,\zeta_2 > \kappa_A^2(\boldsymbol{\theta},\boldsymbol{\theta}^\#)$$
$$-2\kappa_A(\boldsymbol{\theta},\boldsymbol{\theta}^\#)\,\zeta_1 + \mathbf{rank}(A) - \sqrt{\mathbf{rank}(A)}\,\zeta_2\Big\},$$

which implies that

$$\mathbf{P}(\hat\pi \neq id) \leq 2\,\mathbf{P}\big(\zeta_1 \geq \tfrac{\kappa_A(\boldsymbol{\theta},\boldsymbol{\theta}^\#)}{4}\big) + \mathbf{P}\big(\zeta_2 > \tfrac{\kappa_A^2(\boldsymbol{\theta},\boldsymbol{\theta}^\#)}{4\sqrt{\mathbf{rank}(A)}}\big).$$

The result follows.

3.12 Lemmas

Proof of Corollary 5

Recall the proposition borrowed from Laurent *et* Massart (2000) and the statement of the corollary:

Proposition 8 (Laurent *et* Massart (2000))**.** *Assume that $Y \sim \chi^2(D)$, where $D \in \mathbb{N}^*$. Then, for every $x > 0$,*

$$\begin{cases} \mathbf{P}\big(Y - D \leq -2\sqrt{Dx}\big) \leq e^{-x}, \\ \mathbf{P}\big(Y - D \geq 2\sqrt{Dx} + 2x\big) \leq e^{-x}. \end{cases}$$

Corollary 5. *Assume that $Y \sim \chi^2(D)$, where $D \in \mathbb{N}^*$. Then, for every $y > 0$,*

$$\mathbf{P}\big(D^{-1/2}|Y-D| \geq y\big) \leq 2\exp\Big\{-\tfrac{1}{8}y(y \wedge \sqrt{D})\Big\}.$$

We have

$$\begin{aligned}\mathbf{P}\big(D^{-1/2}|Y-D| \geq y\big) &= \mathbf{P}\big(Y-D \leq -\sqrt{D}y\big)\\ &\quad + \mathbf{P}\big(Y-D \geq \sqrt{D}y\big).\end{aligned}$$

The result follows if we prove that both probabilities in the right hand side are smaller than

$$\exp\Big\{-\frac{1}{8}y(y \wedge \sqrt{D})\Big\},$$

which holds if

$$\begin{cases}\sqrt{D}y \geq 2\sqrt{Dx_0}\\ \sqrt{D}y \geq 2\sqrt{Dx_0}+2x_0\end{cases} \qquad \text{with } x_0 = \frac{1}{8}y(y \wedge \sqrt{D}).$$

To conclude, we note that

$$\begin{aligned}2\sqrt{Dx_0}+2x_0 &\leq y\sqrt{D/2}+y\sqrt{D}/4\\ &\leq y\sqrt{D}.\end{aligned}$$

Proof of Lemma 11

We recall the statement of this lemma:

Lemma. *Assume that*
- *ϵ is a positive real number with $\epsilon \leq \sigma/2$,*
- *μ is the uniform distribution on $\{\pm\epsilon\}^{n\times d}$.*

Then, for any transposition π, we have

$$K(\mathbf{P}_{\mu,\pi}, \mathbf{P}_{\mu,id}) \leq \frac{8d\epsilon^4}{\sigma^4}.$$

Furthermore, if

$$\epsilon = \frac{\sqrt{2}\sigma\kappa}{\sqrt{d}},$$

then

$$\mu(\mathbb{R}^{nd}\backslash\Theta_\kappa) \leq \frac{n(n-1)}{2}\, e^{-d/8}.$$

Let us write

$$\pi = (i\ \ j).$$

If you randomly choose

$$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n) \sim \mu,$$

we define $\boldsymbol{\theta}'$ by

$$\begin{cases}\theta'_i = \theta'_j = 0,\\ k \not\in \{i,j\} \Rightarrow \theta'_k = \theta_k.\end{cases}$$

Let us denote by $\tilde{\mu}$ the probability distribution of $\boldsymbol{\theta}'$ and set

$$\mathbf{P}_{\tilde{\mu}} = \int_{\Theta} \mathbf{P}_{\boldsymbol{\theta},\pi}\, \tilde{\mu}(d\boldsymbol{\theta}).$$

We first compute the likelihood ratio of $\mathbf{P}_{\mu,\pi}$ and $\mathbf{P}_{\mu,id}$. We get, for every $(X,Y) = (X_1,\dots,X_n,Y_1,\dots,Y_n)$ in $\mathbb{R}^{2nd}$,

$$\begin{aligned}
\frac{d\mathbf{P}_{\mu,\pi}}{d\mathbf{P}_{\mu,id}}(X,Y) &= \frac{d\mathbf{P}_{\mu,\pi}}{d\mathbf{P}_{\tilde{\mu}}}(X,Y) \times \Big(\frac{d\mathbf{P}_{\mu,id}}{d\mathbf{P}_{\tilde{\mu}}}(X,Y)\Big)^{-1}\\
&= \mathbf{E}_\mu\Big[\frac{d\mathbf{P}_{\theta_i}}{d\mathbf{P}_0}(X_i)\frac{d\mathbf{P}_{\theta_j}}{d\mathbf{P}_0}(X_j)\frac{d\mathbf{P}_{\theta_j}}{d\mathbf{P}_0}(Y_i)\frac{d\mathbf{P}_{\theta_i}}{d\mathbf{P}_0}(Y_j)\Big]\\
&\times \mathbf{E}_\mu^{-1}\Big[\frac{d\mathbf{P}_{\theta_i}}{d\mathbf{P}_0}(X_i)\frac{d\mathbf{P}_{\theta_j}}{d\mathbf{P}_0}(X_j)\frac{d\mathbf{P}_{\theta_i}}{d\mathbf{P}_0}(Y_i)\frac{d\mathbf{P}_{\theta_j}}{d\mathbf{P}_0}(Y_j)\Big],
\end{aligned}$$

where $\mathbf{P}_\theta$ is the d-dimensional Gaussian distribution of mean θ and of variance matrix $\sigma^2 I$. Now, reminding that

$$\frac{d\mathbf{P}_\theta}{d\mathbf{P}_0}(X_i) = e^{-\frac{\epsilon^2 d}{2\sigma^2}+\frac{1}{\sigma^2}\langle X_i,\theta\rangle},$$

we get that

$$\begin{aligned}
\frac{d\mathbf{P}_{\mu,\pi}}{d\mathbf{P}_{\mu,id}}(X,Y) = \prod_{k=1}^{d}\frac{\cosh\big(\frac{\epsilon}{\sigma^2}(X_i^{(k)}+Y_j^{(k)})\big)}{\cosh\big(\frac{\epsilon}{\sigma^2}(X_i^{(k)}+Y_i^{(k)})\big)}\\
\times\prod_{k=1}^{d}\frac{\cosh\big(\frac{\epsilon}{\sigma^2}(X_j^{(k)}+Y_i^{(k)})\big)}{\cosh\big(\frac{\epsilon}{\sigma^2}(X_j^{(k)}+Y_j^{(k)})\big)}.
\end{aligned}$$

Then, we compute the Kullback-Leibler divergence,

$$\begin{aligned}
K(\mathbf{P}_{\mu,\pi},\mathbf{P}_{\mu,id}) &= \int \log\Big(\frac{d\mathbf{P}_{\mu,\pi}}{d\mathbf{P}_{\mu,id}}\Big)d\mathbf{P}_{\mu,\pi}\\
&= 2\sum_{k=1}^{d}\mathbf{E}_\mu\Big[\int \log\cosh\frac{\epsilon}{\sigma^2}(2\theta_{i,k}+\sigma\sqrt{2}X)d\mathbf{Q}\Big]\\
&-2\sum_{k=1}^{d}\mathbf{E}_\mu\Big[\int \log\cosh\frac{\epsilon}{\sigma^2}(\theta_{i,k}+\theta_{j,k}+\sigma\sqrt{2}X)d\mathbf{Q}\Big],
\end{aligned}$$

where $\mathbf{Q}$ is a standard Gaussian distribution and we used the fact that for any $i,j\in\{1,\dots,n\}$,

$$\mathbf{E}_\mu\Big[\int \log\cosh\frac{\epsilon}{\sigma^2}(2\theta_{i,k}+\sigma\sqrt{2}X)\,d\mathbf{Q}\Big] = \mathbf{E}_\mu\Big[\int \log\cosh\frac{\epsilon}{\sigma^2}(2\theta_{j,k}+\sigma\sqrt{2}X)\,d\mathbf{Q}\Big].$$

We control this divergence using the following lemma:

Lemma 17. *For every $x\in\mathbb{R}$,*

$$\frac{x^2}{2}-\frac{x^4}{12}\le \log\cosh(x)\le\frac{x^2}{2}.$$

We get that the general term in the first sum is smaller than

$$\frac{\epsilon^2}{\sigma^2}+2\frac{\epsilon^4}{\sigma^4},$$

while the second general term is larger than

$$\frac{\epsilon^2}{\sigma^2}-2\frac{\epsilon^6}{\sigma^6}-\frac{2}{3}\frac{\epsilon^8}{\sigma^8},$$

whence
$$K(\mathbf{P}_{\mu,\pi},\mathbf{P}_{\mu,id}) \le 4d\frac{\epsilon^4}{\sigma^4} + 4d\frac{\epsilon^6}{\sigma^6} + \frac{4d}{3}\frac{\epsilon^8}{\sigma^8} \le \frac{8d\epsilon^4}{\sigma^4}.$$
Finally, to upper bound $\mu(\mathbb{R}^{nd}\backslash\Theta_\kappa)$, we notice that
$$\begin{aligned}\mu(\mathbb{R}^{nd}\backslash\Theta_\kappa) &\le \tfrac{n(n-1)}{2}\mu\Big(\frac{\|\theta_1-\theta_2\|^2}{2\sigma^2} < \kappa^2\Big)\\ &\le \tfrac{n(n-1)}{2}\mathbf{P}\Big(\sum_{j=1}^{d}(\zeta_1^{(j)}-\zeta_2^{(j)})^2 < \tfrac{2\sigma^2\kappa^2}{\epsilon^2} = d\Big),\end{aligned}$$
where $\zeta_1^{(1)},\ldots,\zeta_1^{(d)},\zeta_2^{(1)},\ldots,\zeta_2^{(d)}$ are i.i.d. variables with Rademacher distribution. The Hoeffding inequality completes the proof.

Proof of Lemma 12

We recall the statement of this lemma:

Lemma. *Assume that m is the largest integer such that $2m \le n$. Then there is a set of vectors $\boldsymbol{\theta}$ such that*
$$\frac{\|\theta_1-\theta_2\|}{\sqrt{\sigma_1^2+\sigma_2^2}} = \ldots = \frac{\|\theta_{2m-1}-\theta_{2m}\|}{\sqrt{\sigma_{2m-1}^2+\sigma_{2m}^2}} = \kappa,$$
and
$$\{i,j\} \ne \{1,2\},\ldots,\{2m-1,2m\} \Rightarrow \frac{\|\theta_i-\theta_j\|}{\sqrt{\sigma_i^2+\sigma_j^2}} > \kappa + \frac{\max_{k=1,\ldots,m}|\eta_k|}{\sqrt{\sigma_i^2+\sigma_j^2}},$$
where for every $k=1,\ldots,m$,
$$\eta_k = \frac{\sigma_{2k}^2-\sigma_{2k-1}^2}{\sigma_{2k}^2+\sigma_{2k-1}^2}(\theta_{2k-1}-\theta_{2k}).$$

The proof is iterative. We first take θ_1,θ_2 distant of $\sqrt{\sigma_1^2+\sigma_2^2}\kappa$, then we choose θ_3,θ_4 satisfying a similar condition and distant enough from θ_1 and θ_2. The result follows.

Proof of Lemma 13

We recall the statement of this lemma:

Lemma. *Assume that*
- $\epsilon_1,\ldots,\epsilon_n$ *are real numbers defined by* $\epsilon_k = \sqrt{2/d}\,\kappa\sigma_k$,
- μ *is the uniform distribution on* $\{\pm\epsilon_1\}^d\times\ldots\times\{\pm\epsilon_n\}^d$,
- $\pi = (i\ j)$ *is the transposition that only permutes i and j,*
- $\sigma_1 \le \ldots \le \sigma_n$.

Then
$$\begin{aligned}K(\mathbf{P}_{\mu,\pi},\mathbf{P}_{\mu,id}) \le 4\kappa^2\Big(1-\frac{\sigma_i^2}{\sigma_j^2}\Big) + \frac{8\kappa^4}{d}\Big(2+\big(1+(2/d)\kappa^2\big)^2\frac{\sigma_j^4}{\sigma_i^4}\Big)\\ + \frac{1}{2}\big(d+2\kappa^2\big)\Big(\frac{\sigma_j^2}{\sigma_i^2}-1\Big)^2\end{aligned}$$
and
$$\mu(\mathbb{R}^{nd}\backslash\Theta_\kappa) \le (n(n-1)/2)\ e^{-d/8}.$$

Without loss of generality, we assume hereafter that $\pi = (1\ 2) \in \mathfrak{S}_n$. Let us introduce a probability mesure $\tilde{\mu}$ on $\mathbb{R}^{n\times d}$ defined as

$$\tilde{\mu} = \delta_{\mathbf{0}} \otimes \delta_{\mathbf{0}} \otimes \{\pm\epsilon_1\}^d \times \ldots \times \{\pm\epsilon_n\}^d$$

with $\delta_{\mathbf{0}}$ the Dirac measure at $\mathbf{0} \in \mathbb{R}^d$. We set

$$\mathbf{P}_{\tilde{\mu},id}(\cdot) = \int_\Theta \mathbf{P}_{\boldsymbol{\theta},id}(\cdot)\tilde{\mu}(d\boldsymbol{\theta}).$$

We first compute the density of $\mathbf{P}_{\mu,\pi}$ with relation to $\mathbf{P}_{\mu,id}$, which can be written as

$$\frac{d\mathbf{P}_{\mu,\pi}}{d\mathbf{P}_{\mu,id}}(\boldsymbol{X},\boldsymbol{X}^\#) = \frac{d\mathbf{P}_{\mu,\pi}}{d\mathbf{P}_{\tilde{\mu},id}}(\boldsymbol{X},\boldsymbol{X}^\#)\Big/\Big(\frac{d\mathbf{P}_{\mu,id}}{d\mathbf{P}_{\tilde{\mu},id}}(\boldsymbol{X},\boldsymbol{X}^\#)\Big).$$

We have

$$\begin{aligned}
&\frac{d\mathbf{P}_{\mu,\pi}}{d\mathbf{P}_{\tilde{\mu},id}}(\boldsymbol{X},\boldsymbol{X}^\#) = \mathbf{E}_\mu\Big[\frac{d\mathbf{P}_{\theta_1,\sigma_1}}{d\mathbf{P}_{0,\sigma_1}}(X_1)\frac{d\mathbf{P}_{\theta_2,\sigma_2}}{d\mathbf{P}_{0,\sigma_2}}(X_2)\frac{d\mathbf{P}_{\theta_2,\sigma_2}}{d\mathbf{P}_{0,\sigma_1}}(X_1^\#)\frac{d\mathbf{P}_{\theta_1,\sigma_1}}{d\mathbf{P}_{0,\sigma_2}}(X_2^\#)\Big]\\
&= \mathbf{E}_\mu\Big[\frac{d\mathbf{P}_{\theta_1,\sigma_1}}{d\mathbf{P}_{0,\sigma_1}}(X_1)\frac{d\mathbf{P}_{\theta_1,\sigma_1}}{d\mathbf{P}_{0,\sigma_2}}(X_2^\#)\Big]\times\mathbf{E}_\mu\Big[\frac{d\mathbf{P}_{\theta_2,\sigma_2}}{d\mathbf{P}_{0,\sigma_2}}(X_2)\frac{d\mathbf{P}_{\theta_2,\sigma_2}}{d\mathbf{P}_{0,\sigma_1}}(X_1^\#)\Big]\\
&= \prod_{k=1}^d \cosh\Big(\frac{\epsilon_1}{\sigma_1^2}(X_{1,k}+X_{2,k}^\#)\Big)\cosh\Big(\frac{\epsilon_2}{\sigma_2^2}(X_{2,k}+X_{1,k}^\#)\Big)\\
&\quad\times\exp\Big\{-\frac{1}{2}(\|X_1^\#\|^2-\|X_2^\#\|^2)(\sigma_2^{-2}-\sigma_1^{-2})\Big\}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\frac{d\mathbf{P}_{\mu,id}}{d\mathbf{P}_{\tilde{\mu},id}}(\boldsymbol{X},\boldsymbol{X}^\#) &= \mathbf{E}_\mu\Big[\frac{d\mathbf{P}_{\theta_1,\sigma_1}}{d\mathbf{P}_{0,\sigma_1}}(X_1)\frac{d\mathbf{P}_{\theta_2,\sigma_2}}{d\mathbf{P}_{0,\sigma_2}}(X_2)\frac{d\mathbf{P}_{\theta_1,\sigma_1}}{d\mathbf{P}_{0,\sigma_1}}(X_1^\#)\frac{d\mathbf{P}_{\theta_2,\sigma_2}}{d\mathbf{P}_{0,\sigma_2}}(X_2^\#)\Big]\\
&= \mathbf{E}_\mu\Big[\frac{d\mathbf{P}_{\theta_1,\sigma_1}}{d\mathbf{P}_{0,\sigma_1}}(X_1)\frac{d\mathbf{P}_{\theta_1,\sigma_1}}{d\mathbf{P}_{0,\sigma_1}}(X_1^\#)\Big]\times\mathbf{E}_\mu\Big[\frac{d\mathbf{P}_{\theta_2,\sigma_2}}{d\mathbf{P}_{0,\sigma_2}}(X_2)\frac{d\mathbf{P}_{\theta_2,\sigma_2}}{d\mathbf{P}_{0,\sigma_2}}(X_2^\#)\Big]\\
&= \prod_{k=1}^d \cosh\Big(\frac{\epsilon_1}{\sigma_1^2}(X_{1,k}+X_{1,k}^\#)\Big)\cosh\Big(\frac{\epsilon_2}{\sigma_2^2}(X_{2,k}+X_{2,k}^\#)\Big).
\end{aligned}$$

Thus, we get that

$$\begin{aligned}
\frac{d\mathbf{P}_{\mu,\pi}}{d\mathbf{P}_{\mu,id}}(\boldsymbol{X},\boldsymbol{X}^\#) = \prod_{k=1}^d \frac{\cosh\big(\frac{\epsilon_1}{\sigma_1^2}(X_{1,k}+X_{2,k}^\#)\big)}{\cosh\big(\frac{\epsilon_1}{\sigma_1^2}(X_{1,k}+X_{1,k}^\#)\big)}\times\prod_{k=1}^d\frac{\cosh\big(\frac{\epsilon_2}{\sigma_2^2}(X_{2,k}+X_{1,k}^\#)\big)}{\cosh\big(\frac{\epsilon_2}{\sigma_2^2}(X_{2,k}+X_{2,k}^\#)\big)}\\
\times\exp\Big\{-\frac{1}{2}(\|X_1^\#\|^2-\|X_2^\#\|^2)(\sigma_2^{-2}-\sigma_1^{-2})\Big\}.
\end{aligned}$$

Then, we compute the Kullback-Leibler divergence,

$$\begin{aligned}
&K(\mathbf{P}_{\mu,\pi},\mathbf{P}_{\mu,id}) = \int\log\Big(\frac{d\mathbf{P}_{\mu,\pi}}{d\mathbf{P}_{\mu,id}}(\boldsymbol{X},\boldsymbol{X}^\#)\Big)d\mathbf{P}_{\mu,\pi}(\boldsymbol{X},\boldsymbol{X}^\#)\\
&= \sum_{k=1}^d\sum_{j=1}^2\Big\{\mathbf{E}_\mu\Big[\int\log\cosh\Big[\frac{\epsilon_j}{\sigma_j^2}(2\theta_{j,k}+\sigma_j\sqrt{2}x)\Big]\varphi(x)dx\Big]\\
&\qquad-\mathbf{E}_\mu\Big[\int\log\cosh\Big[\frac{\epsilon_j}{\sigma_j^2}(\theta_{1,k}+\theta_{2,k}+\sigma_{12}x)\Big]\varphi(x)dx\Big]\Big\}\\
&\qquad+\frac{d}{2}\mathbf{E}_\mu\Big[\int_{\mathbb{R}}((\theta_{1,1}+\sigma_1x)^2-(\theta_{2,1}+\sigma_2x)^2)\varphi(x)\,dx\Big](\sigma_2^{-2}-\sigma_1^{-2}),
\end{aligned}$$

where φ is the density function of the standard Gaussian distribution. We evaluate the first two terms using Lemma 17, while for the third term the exact computation yields:

$$\begin{aligned}\mathbf{E}_\mu\Big[\int_{\mathbb{R}}((\theta_{1,1}+\sigma_1 x)^2-(\theta_{2,1}+\sigma_2 x)^2)\varphi(x)\,dx\Big] &= \epsilon_1^2+\sigma_1^2-\epsilon_2^2-\sigma_2^2\\ &= (\sigma_1^2-\sigma_2^2)(1+(2/d)\kappa^2).\end{aligned}$$

In conjunction with the facts that $\epsilon_1/\sigma_1=\epsilon_2/\sigma_2$, $\sigma_1\le\sigma_2$ and $\epsilon_1\le\epsilon_2$, this leads to

$$\begin{aligned}&\mathbf{E}_\mu\Big[\int\log\cosh\Big[\frac{\epsilon_j}{\sigma_j^2}(2\theta_{j,k}+\sigma_j\sqrt{2}x)\Big]\varphi(x)dx\le\frac{\epsilon_j^2}{\sigma_j^2}+2\frac{\epsilon_j^4}{\sigma_j^4}=\frac{\epsilon_2^2}{\sigma_2^2}+2\frac{\epsilon_2^4}{\sigma_2^4},\\ &\mathbf{E}_\mu\Big[\int\log\cosh\Big[\frac{\epsilon_j}{\sigma_j^2}(\theta_{1,k}+\theta_{2,k}+\sigma_{1,2}x)\Big]\varphi(x)dx\Big]\\ &\quad\ge\frac{\epsilon_j^2}{2\sigma_j^4}\Big(\epsilon_1^2+\epsilon_2^2+\sigma_1^2+\sigma_2^2\Big)\\ &\qquad-\frac{\epsilon_j^4}{12\sigma_j^8}\Big(\epsilon_1^4+\epsilon_2^4+3(\sigma_1^2+\sigma_2^2)^2+6\epsilon_1^2\epsilon_2^2+6(\sigma_1^2+\sigma_2^2)(\epsilon_1^2+\epsilon_2^2)\Big)\\ &\quad\ge\frac{\epsilon_2^2(\epsilon_1^2+\sigma_1^2)}{\sigma_2^4}-\frac{\epsilon_1^4(\epsilon_2^2+\sigma_2^2)^2}{\sigma_1^8}.\end{aligned}$$

Thus, we get that

$$\begin{aligned}(1/d)K(\mathbf{P}_{\mu,\pi},\mathbf{P}_{\mu,id}) &\le\frac{2\epsilon_2^2}{\sigma_2^2}+\frac{4\epsilon_2^4}{\sigma_2^4}-\frac{2\epsilon_2^2(\epsilon_1^2+\sigma_1^2)}{\sigma_2^4}+\frac{2\epsilon_1^4(\epsilon_2^2+\sigma_2^2)^2}{\sigma_1^8}\\ &\quad+\frac{1}{2}\big(1+(2/d)\kappa^2\big)(\sigma_1^2-\sigma_2^2)(\sigma_2^{-2}-\sigma_1^{-2})\\ &\le\frac{4\kappa^2}{d}\Big(1-\frac{\sigma_1^2}{\sigma_2^2}\Big)+\frac{16\kappa^4}{d^2}+\frac{8\kappa^4}{d^2}\big(1+(2/d)\kappa^2\big)^2\frac{\sigma_2^4}{\sigma_1^4}\\ &\quad+\frac{1}{2}\big(1+(2/d)\kappa^2\big)\Big(\frac{\sigma_2^2}{\sigma_1^2}-1\Big)^2.\end{aligned}$$

To complete the proof, we need to evaluate $\mu(\mathbb{R}^{nd}\setminus\bar\Theta_\kappa)$. We note that in view of the union bound,

$$\begin{aligned}\mu(\mathbb{R}^{nd}\setminus\bar\Theta_\kappa) &=\mu\Big(\bigcup_{k=1}^n\bigcup_{k'\ne k}\{\boldsymbol{\theta}:\|\theta_k-\theta_{k'}\|<\kappa\sigma_{k,k'}\}\Big)\\ &\le\frac{n(n-1)}{2}\max_{k\ne k'}\mu\big(\{\boldsymbol{\theta}:\|\theta_k-\theta_{k'}\|^2<\kappa^2\sigma_{k,k'}\}\big)\\ &=\frac{n(n-1)}{2}\max_{k\ne k'}\mathbf{P}\big(d\epsilon_k^2+d\epsilon_{k'}^2-2d\epsilon_k\epsilon_{k'}\bar\zeta<\kappa^2\sigma_{k,k'}^2\big),\end{aligned}$$

where $\bar\zeta=\frac{1}{d}\sum_{j=1}^d\zeta_j$ with $\zeta_1,\dots,\zeta_d$ being i.i.d. Rademacher random variables. One can check that

$$\frac{d\epsilon_k^2+d\epsilon_{k'}^2-\kappa^2\sigma_{k,k'}^2}{2d\epsilon_k\epsilon_{k'}}=\frac{2\sigma_k^2+2\sigma_{k'}^2-(\sigma_k^2+\sigma_{k'}^2)}{4\sigma_k\sigma_{k'}}\ge\frac{1}{2}.$$

Therefore, using the Hoeffding inequality, we get

$$\mu(\mathbb{R}^{nd}\setminus\bar\Theta_\kappa)\le\frac{1}{2}n(n-1)\mathbf{P}\big(\bar\zeta>1/2\big)\le\frac{1}{2}n(n-1)e^{-d/8}.$$

Proof of Lemma 16

We recall the statement of this lemma:

Lemma. *Assume that n is an integer larger than 25. Then there exist permutations $\pi_0, \dots, \pi_M$ with*

$$\pi_0 = id,\ M \geq (n/24)^{n/6},$$

and for every $i \neq j \in \{0, \dots, M\}$,

$$\begin{cases} \pi_i \text{ is the composition of distinct transpositions,} \\ \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\pi_i(k)\neq\pi_j(k)} \geq \frac{3}{8}. \end{cases}$$

Denote m the largest integer such that $2m \leq n$, and choose

$$\tilde{\pi}_0, \tilde{\pi}_1, \dots, \tilde{\pi}_M \in \mathfrak{S}_m \text{ with } M \leq (m/8)^{m/2}$$

as in Lemma 15, so that for every $i \neq j \in \{1, \dots, M\}$,

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\pi_i(k)\neq\pi_j(k)} \geq \frac{1}{2}.$$

We use these permutations of $\mathfrak{S}_m$ to construct other permutations of $\mathfrak{S}_n$. The idea of the construction is as follows: the permutations we are looking for are products of m transpositions of distinct supports, and each transposition permutes a even integer with an odd one.
Now, denote

$$\pi_0 = id$$

and for every i in $\{1, \dots, M\}$,

$$\pi_i = \big(1\ 2\tilde{\pi}_i(1)\big) \circ \big(3\ 2\tilde{\pi}_i(2)\big) \circ \dots \circ \big(2m-1\ 2\tilde{\pi}_i(m)\big) \in \mathfrak{S}_n.$$

With these choices, the number of differences between π_i and π_j, *i.e.*,

$$\sum_{k=1}^{n} \mathbb{1}_{\pi_i(k)\neq\pi_j(k)}$$

is twice as much as the number of differences between $\tilde{\pi}_i$ and $\tilde{\pi}_j$, *i.e.*,

$$\sum_{k=1}^{m} \mathbb{1}_{\tilde{\pi}_i(k)\neq\tilde{\pi}_j(k)}.$$

This entails that for every $i \neq j \in \{0, \dots, M\}$,

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\pi_i(k)\neq\pi_j(k)} \geq \frac{2m}{n} \times \frac{1}{m}\sum_{k=1}^{m} \mathbb{1}_{\tilde{\pi}_i(k)\neq\tilde{\pi}_j(k)} \geq \frac{3}{8}.$$

To complete the proof, we note that $m \geq n/3$, since m is the integer part of $n/2$.

Proof of Lemma ??

Counting the permutations such that

$$\sum_{i=1}^{n} \mathbb{1}_{\pi(i)\neq i} = k, \quad k \in \{m, \dots, n\},$$

we get

$$\#E_{id} = N_n + \binom{n}{1} N_{n-1} + \ldots + \binom{n}{m} N_{n-m},$$

where N_k is the number of derangements, the permutations such that none of the elements appear in their original position, in $\mathfrak{S}_k$ for $k \geq 1$. We know that

$$\forall\ k \geq 1,\ N_k = k! \times \sum_{j=0}^{k} \frac{(-1)^j}{j!},$$

which, using the alternating series test, gives

$$\forall\ k \geq 1,\ N_k \geq k! \times \left(e^{-1} - \frac{1}{(k+1)!}\right).$$

It follows that

$$\begin{aligned} \#E_{id} &\geq n! \times \left(e^{-1} - \frac{1}{(n-m+1)!}\right) \times \left(1 + \frac{1}{1!} + \ldots + \frac{1}{m!}\right) \\ &\geq n! \times \left(e^{-1} - \frac{1}{(n-m+1)!}\right) \times \left(e - \frac{e}{(m+1)!}\right), \end{aligned}$$

so that

$$\#E_{id}^{\complement} \leq n! \times \left(\frac{e}{(n-m+1)!} + \frac{1}{(m+1)!}\right) \leq \frac{4n!}{m!}.$$

Conclusion of the chapter

In this work, we have made a first step in solving the problem of descriptor matching. We have considered the simplest example of matching criterion when no outliers are present. We gave a complete minimax study of this problem, measuring the performance of an estimator using its matching threshold, and finding computable estimators whose performance can not be improved. More importantly, we showed that the results also carry over to more general cases, with different matching criteria or risk definitions.

The next step is to extend our work to the case of general images, representing maybe totally different images. An efficient procedure should be able to reject matches that are identified as incorrect.

Bibliographie

Y. BARAUD. Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, 8(5):577–606, 2002. (Cité pages 3 et 49.)

Y. BARAUD, S. HUET, AND B. LAURENT. Adaptive tests of linear hypotheses by model selection. *Ann. Statist.*, 31(1):225–251, 2003. (Cité pages 7 et 19.)

Y. BARAUD, S. HUET, AND B. LAURENT. Testing convex hypotheses on the mean of a Gaussian vector. Application to testing qualitative hypotheses on a regression function. *Ann. Statist.*, 33(1):214–257, 2005. (Cité pages 7, 9 et 19.)

S. BERMAN. Sojourns and extremes of a stochastic process defined as a random linear combination of arbitrary functions. *Comm. Statist. Stochastic Models*, 4(1):1–43, 1988. (Cité pages 35 et 60.)

J. BIGOT AND S. GADAT. A deconvolution approach to estimation of a common shape in a shifted curves model. *Ann. Statist.*, 38(4):2422–2464, 2010. (Cité page 17.)

J. BIGOT AND S. GADAT AND J.-M. LOUBES. Statistical M-estimation and consistency in large deformable models for image warping. *J. Math. Imaging Vision*, 34(3):270–290, 2009. (Cité page 17.)

J. BIGOT AND F. GAMBOA AND M. VIMOND. Estimation of translation, rotation, and scaling between noisy images using the Fourier-Mellin transform. *SIAM J. Imaging Sci.*, 2(2):614–645, 2009. (Cité page 17.)

J. BIGOT AND J.-M. LOUBES AND M. VIMOND. Semiparametric estimation of shifts on compact Lie groups for image registration. *Probab. Theory Related Fields*, 152(3-4):425–473, 2012. (Cité page 17.)

A.W. BOWMAN, M. JONES AND I. GIJBELS. Testing monotonivity of regression. *J. Comput. Graph. Stat.*, 7.4:489–500, 1998. (Cité page 7.)

L. BROWN AND M. LOW. Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.*, 24(6):2384–2398, 1996. (Cité pages 1 et 20.)

E. BUDISH ,Y. CHE, F. KOJIMA AND P. MILGROM. Implementing random assignments: A generalization of the Birkhoff-von Neumann Theorem. *Technical report*, 2009. (Cité pages 77 et 81.)

M. BURNASHEV. On the minimax detection of an inaccurately known signal in a Gaussian noise background. *Theory Prob. Appl.*, 24,107-119, 1977. (Cité page 3.)

R. CARROLL AND P. HALL. Semiparametric comparison of regression curves via normal likelihoods. *Austral. J. Statist.*, 34(3):471-487, 1992. (Cité page 17.)

I. CASTILLO. Semiparametric second-order efficient estimation of the period of a signal. *Bernoulli*, 13(4):910–932, 2007. (Cité page 17.)

I. CASTILLO. A semiparametric Bernstein-von Mises theorem for Gaussian process priors. *Probab. Theory Related Fields*, 152(1-2),53–99, 2012. (Cité page 17.)

I. CASTILLO AND C. LÉVY-LEDUC AND C. MATIAS. Exact adaptive estimation of the shape of a periodic function with unknown period corrupted by white noise. *Math. Methods Statist.*, 15(2):146–175, 2006. (Cité page 17.)

I. CASTILLO AND J.-M. LOUBES. Estimation of the distribution of random shifts deformation. *Math. Methods Statist.*, 18(1):21–42, 2009. (Cité page 17.)

A. DALALYAN. Stein shrinkage and second-order efficiency for semiparametric estimation of the shift. *Math. Methods Statist.*, 16(1):42–62, 2007. (Cité page 17.)

A. DALALYAN AND G. GOLUBEV AND A. TSYBAKOV. Penalized maximum likelihood and semiparametric second-order efficiency. *Ann. Statist.*, 34(1):169–201, 2006. (Cité page 17.)

A. DALALYAN AND M. REISS. Asymptotic statistical equivalence for scalar ergodic diffusions. *Probab. Theory Related Fields*, 134(2): 248–282, 2006. (Cité pages 1 et 20.)

M. ERMAKOV. Minimax detection of a signal in Gaussian white noise. *Teor. Veroyatnost. i Primenen.*, 35(4): 704–715, 1990. (Cité page 3.)

M. ERMAKOV. Asymptotically minimax criteria for testing complex nonparametric hypotheses. *Problems Inform. Transmission*, 33:184–196, 1996. (Cité page 3.)

J. FAN AND J. JIANG. Nonparametric inference with generalized likelihood ratio tests. *TEST*, 16(3):409–444, 2007.

J. FAN, C. ZHANG, AND J. ZHANG. Generalized likelihood ratio statistics and wilks phenomenon. *Ann. Statist.*, 29(1):153–193, 2001. (Cité pages xvi, 5, 6 et 37.)

M. FROMONT AND C. LÉVY-LEDUC. Adaptive tests for periodic signal detection with applications to laser vibrometry. *ESAIM P. S.*, 10:46–75 (electronic), 2006. (Cité page 51.)

F. GAMBOA AND J.-M. LOUBES AND E. MAZA Semi-parametric estimation of shifts. *Electron. J. Stat.*,1:616–640, 2007. (Cité page 17.)

G. GAYRAUD AND C. POUET. Minimax testing composite null hypotheses in the discrete regression scheme. *Math. Methods Statist.*, 10(4):375–394 (2002), 2001. Meeting on Mathematical Statistics (Marseille, 2000). (Cité pages 7, 19 et 21.)

G. GAYRAUD AND C. POUET. Adaptive minimax testing in the discrete regression scheme. *Probab. Theory Relat. Fields*, 133(4): 531–558, 2005. (Cité pages 7, 9, 19 et 47.)

D. GERVINI AND T. GASSER. Nonparametric maximum likelihood estimation of the structural mean of a sample of curves. *Biometrika*, 92(4):801–820, 2005. (Cité page 17.)

C. GLASBEY AND K. MARDIA. A penalized likelihood approach to image warping. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 63(3):465–514, 2001. (Cité page 17.)

G. GOLUBEV. Estimation of the period of a signal with an unknown form against a white noise background. *Problems Inform. Transmission*, 24(4):288–299, 1988 (Cité page 17.)

I. GRAMA AND M. NUSSBAUM. Asymptotic equivalence for nonparametric generalized linear models. *Probability Theory and Related Fields*, 111 (2):167–214, 1998. (Cité pages 1 et 20.)

I. GRAMA AND M. NUSSBAUM. Asymptotic equivalence for nonparametric regression. *Math. Methods Statist.*, 11(1):1–36, 2002. (Cité pages 1 et 20.)

W. HÄRDLE AND J. MARRON. Semiparametric comparison of regression curves. *Ann. Statist.*, 18(1):63–89, 1990. (Cité page 17.)

J. HOROWITZ AND V. SPOKOINY. An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, 69(3):599–631, 2001. (Cité pages 6, 7, 9, 19, 47 et 50.)

I. IBRAGIMOV AND R. KHASMINSKII. One problem of statistical estimation in Gaussian white noise. *Soviet Math. Dokl.*, 1977. (Cité page 3.)

Y. INGSTER. Minimax nonparametric detection of signals in white Gaussian noise. *Problems of Information Transmission*, 18:130–140, 1982. (Cité page 3.)

Y. INGSTER. Asymptotically minimax hypothesis testing for nonparametric alternatives. I,II,III. *Math. Methods Statist.*, 2(2):85–114, 1993. (Cité pages 3 et 5.)

Y. INGSTER AND I. SUSLINA. *Nonparametric goodness-of-fit testing under Gaussian models*. Springer Verlag, 2003. (Cité pages 3 et 37.)

T. JEBARA. Images as Bags of Pixels. *International Conference on Computer Vision*, 2003. (Cité page 76.)

A. JUDITSKY AND O. LEPSKI AND A. TSYBAKOV. Nonparametric estimation of composite functions. *Ann. Statist.*, 37(3):1360–1404, 2009. (Cité page 17.)

A. JUDITSKY, A. NEMIROVSKI. On nonparametric tests of positivity/monotonicity/convexity. *Ann. Statist.*, 30(2):498–527, 2002. (Cité page 7.)

A. KNEIP AND T. GASSER. Statistical tools to analyze data representing a sample of curves. *Ann. Statist.*, 20(3):1266–1305 (Cité page 17.)

H. KUHN. The Hungarian Method for the assignment problem. *Nav. Res. Logist. Q.*, 2:83–97 (Cité page 77.)

B. LAURENT AND P. MASSART. Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.*, 28(5):1302–1338, 2000. (Cité pages 36, 91 et 94.)

O. LEPSKI AND V. SPOKOINY. Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative. *Bernoulli*, 5(2):333–358, 1999. (Cité page 3.)

O. LEPSKI AND A. TSYBAKOV. Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probability Theory and Related Fields*, 117 (1):17–48, 2000. (Cité pages 3, 9 et 51.)

J.-M. LOUBES AND E. MAZA AND M. LAVIELLE AND L. RODRÍGUEZ. Road trafficking description and short term travel time forecasting, with a classification method. *Canad. J. Statist.*, 34(3):475–491, 2006. (Cité page 18.)

D. LOWE. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60 (2):91–110, 2004. (Cité pages xv et 10.)

E. MAMMEN. Estimating a smooth monotone regression function. *Ann. Statist.*, 19(2):724–740, 1991. (Cité page 7.)

M. NUSSBAUM. Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.*, 24(6):2399–2430, 1996. (Cité pages 1 et 20.)

V. PETROV. Limit theorems of probability theory. *Oxford Studies in Probability*, The Clarendon Press Oxford University Press, New York, 1995. (Cité pages 32 et 62.)

J. RAMSAY AND X. LI. Curve registration. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 60(2):351–363, 1998. (Cité page 17.)

C. REILLY AND P. PRICE AND A. GELMAN AND S. SANDGATHE. Using image and curve registration for measuring the goodness of fit of spatial and temporal predictions. *Biometrics*, 60(4):954–964, 2004. (Cité page 17.)

M. REISS. Asymptotic equivalence for nonparametric regression with multivariate and random design. *Ann. Statist.*, 36(4):1957–1982, 2008. (Cité pages 1 et 20.)

J. RICE. Bandwidth choice for nonparametric regression. *The Annals of Statistics*, 12(4): 1215–1230, 1984. (Cité page 21.)

A. ROHDE. On the asymptotic equivalence and rate of convergence of nonparametric regression and gaussian white noise. *Statistics & Decisions, International mathematical journal for stochastic methods and models*, 22(3/2004):235–243, 2004. (Cité page 20.)

E. ROQUAIN. Type I error rate control for testing many hypotheses: a survey with proofs. *J. SFdS*, 152:3–38, 2011. (Cité page 10.)

B. RØNN. Nonparametric maximum likelihood estimation for shifted curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 63(2):243–259, 2001. (Cité page 17.)

J. STURM. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. 1998. (Cité page 78.)

V. SPOKOINY. Adaptive hypothesis testing using wavelets. *Ann. Statist.*, 24(6):2477–2498, 1996. (Cité pages 8, 43 et 50.)

H. TRIEBEL. Theory of Function Spaces II. *Birkhäuser, Basel.* (Cité pages 3 et 8.)

U. ISSERLES AND Y. RITOV AND T. TRIGANO. Semiparametric curve alignment and shift density estimation for biological data. *IEEE Trans. Signal Process.*,59(5):1970–1984, 2011. (Cité pages 17 et 18.)

A. TSYBAKOV. Introduction to Nonparametric Estimation. *Springer,* 2008. (Cité pages 86 et 92.)

M. VIMOND. Efficient estimation for a subclass of shape invariant models. *Ann. Statist.*, 38(3):1885–1912, 2010. (Cité page 17.)