



HAL
open science

Les méthodes de régularisation optimale et leurs applications en tomographie : nouveaux algorithmes performants de reconstruction d'images

Didier A. Girard

► To cite this version:

Didier A. Girard. Les méthodes de régularisation optimale et leurs applications en tomographie : nouveaux algorithmes performants de reconstruction d'images. Modélisation et simulation. Institut National Polytechnique de Grenoble - INPG, 1984. Français. NNT: . tel-00311758

HAL Id: tel-00311758

<https://theses.hal.science/tel-00311758>

Submitted on 21 Aug 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE

présentée à

l'Institut National Polytechnique de Grenoble

pour obtenir le grade de
DOCTEUR-INGENIEUR
«Mathématiques Appliquées»

par

Didier GIRARD



**LES METHODES DE REGULARISATION OPTIMALE ET
LEURS APPLICATIONS EN TOMOGRAPHIE:
NOUVEAUX ALGORITHMES PERFORMANTS DE
RECONSTRUCTION D'IMAGES.**



Thèse soutenue le 29 octobre 1984 devant la commission d'examen.

F. ROBERT	Président
C. CARASSO	} Examineurs
Ph. GARDERET	
P.J. LAURENT	
F. NATTERER	

INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

Président: Daniel BLOCH

Vice-Présidents: René CARRE
Hervé CHERADAME
Jean-Pierre LONGQUEUE

Année universitaire 1983-1984

Professeur des Universités

ANCEAU	François	E.N.S.I.M.A.G	JOUBERT	Jean-Claude	E.N.S.I.E.G
BARIBAUD	Michel	E.N.S.E.R.G	JOURDAIN	Geneviève	E.N.S.I.E.G
BARRAUD	Alain	E.N.S.I.E.G	LACOUME	Jean-Louis	E.N.S.I.E.G
BAUDELET	Bernard	E.N.S.I.E.G	LATOMBE	Jean-Claude	E.N.S.I.M.A.G
BESSON	Jean	E.N.S.E.E.G	LESIEUR	Marcel	E.N.S.H.G
BLIMAN	Samuel	E.N.S.E.R.G	LESPINARD	Georges	E.N.S.H.G
BLOCH	Daniel	E.N.S.I.E.G	LONGQUEUE	Jean-Pierre	E.N.S.I.E.G
BOIS	Philippe	E.N.S.H.G	LOUCHET	François	E.N.S.E.E.G
BONNETAIN	Lucien	E.N.S.E.E.G	MASSELOT	Christian	E.N.S.I.E.G
BONNIER	Etienne	E.N.S.E.E.G	MAZARE	Guy	E.N.S.I.M.A.G
BOUVARD	Maurice	E.N.S.H.G	MOREAU	René	E.N.S.H.G
BRISSONNEAU	Pierre	E.N.S.I.E.G	MORET	Roger	E.N.S.I.E.G
BUYLE BODIN	Maurice	E.N.S.E.R.G	MOSSIERE	Jacques	E.N.S.I.M.A.G
CAVAIGNAC	Jean-François	E.N.S.I.E.G	PARIAUD	Jean-Charles	E.N.S.E.E.G
CHARTIER	Germain	E.N.S.I.E.G	PAUTHENET	René	E.N.S.I.E.G
CHENEVIER	Pierre	E.N.S.E.R.G	PERRET	René	E.N.S.I.E.G
CHERADAME	Hervé	U.E.R.M.C.P.P	PERRET	Robert	E.N.S.I.E.G
CHERUY	Arlette	E.N.S.I.E.G	PIAU	Jean-Michel	E.N.S.H.G
CHIAVERINA	Jean	U.E.R.M.C.P.P	POLOUJADOFF	Michel	E.N.S.I.E.G
COHEN	Joseph	E.N.S.E.R.G	POUPOT	Christian	E.N.S.E.R.G
COUMES	André	E.N.S.E.R.G	RAMEAU	Jean-Jacques	E.N.S.E.E.G
DURAND	Francis	E.N.S.E.E.G	RENAUD	Maurice	U.E.R.M.C.P.P
DURAND	Jean-louis	E.N.S.I.E.G	ROBERT	André	U.E.R.M.C.P.P
FELICI	Noël	E.N.S.I.E.G	ROBERT	François	E.N.S.I.M.A.G
FONLUPT	Jean	E.N.S.I.M.A.G	SABONNADIÈRE	Jean-Claude	E.N.S.I.E.G
FOULARD	Claude	E.N.S.I.E.G	SAUCIER	Gabrielle	E.N.S.I.M.A.G
GANDINI	Alessandro	U.E.R.M.C.P.P	SCHLENKER	Claire	E.N.S.I.E.G
GAUBERT	Claude	E.N.S.I.E.G	SCHLENKER	Michel	E.N.S.I.E.G
GENTIL	Pierre	E.N.S.E.R.G	SERMET	Pierre	E.N.S.E.R.G
GUERIN	Bernard	E.N.S.E.R.G	SILVY	Jacques	U.E.R.M.C.P.P
GUYOT	Pierre	E.N.S.E.E.G	SOHM	Jean-Claude	E.N.S.E.E.G
IVANES	Marcel	E.N.S.I.E.G	SOUQUET	Jean-Louis	E.N.S.E.E.G
JALINIER	Jean-Michel	E.N.S.I.E.G	VEILLON	Gérard	E.N.S.I.M.A.G
JAUSSAUD	Pierre	E.N.S.I.E.G	ZADWORNY	François	E.N.S.E.R.G

Professeurs Associés

BLACKWELDER	Ronald	E.N.S.H.G	PURDY	Gary	E.N.S.E.E.G
HAYASHI	Hirashi	E.N.S.I.E.G			

Professeurs Université des Sciences Sociales (Grenoble II)

BOLLIET	Louis		CHATELIN	Françoise	
---------	-------	--	----------	-----------	--

Chercheurs du C.N.R.S

FRUCHART	Robert	Directeur de recherche	GUELIN	Pierre	Maître de recherche
JORRAND	Philippe	Directeur de recherche	HOPFINGER	Emil	Maître de recherche
VACHAUD	Georges	Directeur de recherche	JOUD	Jean-Charles	Maître de recherche
ALLIBERT	Michel	Maître de recherche	KAMARINOS	Georges	Maître de recherche
ANSARA	Ibrahim	Maître de recherche	KLEITZ	Michel	Maître de recherche
ARMAND	Michel	Maître de recherche	LANDAU	Ioan-Dore	Maître de recherche
BINDER	Gilbert	Maître de recherche	LASJAUNIAS	Jean-Claude	Maître de recherche
BORNARD	Guy	Maître de recherche	MERMET	Jean	Maître de recherche
CARRE	René	Maître de recherche	MUNIER	Jacques	Maître de recherche
DAVID	René	Maître de recherche	PIAU	Monique	Maître de recherche
DEPORTES	Jacques	Maître de recherche	PORTESEIL	Jean-Louis	Maître de recherche
DRIOLE	Jean	Maître de recherche	THOLENCE	Jean-Louis	Maître de recherche
GIGNOUX	Damien	Maître de recherche	VERDILLON	André	Maître de recherche
GIVORD	Dominique	Maître de recherche	SUERY	Michel	Maître de recherche

Personnalités habilitées à diriger des travaux de recherche
(Decision du Conseil Scientifique)

E.N.S.E.E.G.

ALLIBERT BERNARD BONNET CAILLET CHATILLON CHATILLON COULON	Colette Claude Roland Marcel Catherine Christian Michel	DIARD EUSTATHOPOULOS FOSTER GALERIE HAMMOU MALMEJAC MARTIN GARIN	Jean Paul Nicolas Panayotis Alain Abdelkader Yves (CENG) Regina	NGUYEN TRUONG RAVAINE SAINFORT SARRAZIN SIMON TOUZAIN URBAIN	Bernadette Denis (CENG) Pierre Jean Paul Philippe Georges (Laboratoire des ultra-réfractaire ODEILLO).
--	---	--	---	--	--

E.N.S.E.R.G.

BARIBAUD BOREL CHOVET	Michel Joseph Alain	CHEHIKIAN DOLMAZON	Alain Jean Marc	HERAULT MONLLOR	Jeanny Christian
-----------------------------	---------------------------	-----------------------	--------------------	--------------------	---------------------

E.N.S.I.E.G.

BORNARD DESCHIZEAUX GLANGEAUD	Guy Pierre François	KOFMAN LEJEUNE	Walter Gérard	MAZUER PERARD REINISCH	Jean Jacques Raymond
-------------------------------------	---------------------------	-------------------	------------------	------------------------------	----------------------------

E.N.S.H.G.

ALEMANY BOIS DARVE	Antoine Daniel Félix	MICHEL OBLED	Jean Marie Charles	ROWE VAUCLIN WACK	Alain Michel Bernard
--------------------------	----------------------------	-----------------	-----------------------	-------------------------	----------------------------

E.N.S.I.M.A.G.

BERT CALMET COURTIN	Didier Jacques Jacques	COURTOIS DELLA DORA	Bernard Jean	FONLUPT SIFAKIS	Jean Joseph
---------------------------	------------------------------	------------------------	-----------------	--------------------	----------------

U.E.R.M.C.P.P.

CHARUEL Robert

C.E.N.G.

CADET COEURE DELHAYE DUPUY	Jean Philippe (LETI) Jean Marc (STT) Michel (LETI)	JOUVE NICOLAU NIFENECKER	Hubert (LETI) Yvan (LETI) Hervé	PERROUD PEUZIN TAIEB VINCENDÓN	Paul Jean Claude (LETI) Maurice Marc
-------------------------------------	---	--------------------------------	---------------------------------------	---	---

Laboratoires extérieurs :

C.N.E.T.

DEMOULIN DEVINE	Eric R.A.B.	GERBER	Roland	MERCKEL PAULEAU	Gérard Yves
--------------------	----------------	--------	--------	--------------------	----------------

I.N.S.A. Lyon

GAUBERT C.

ECOLE NATIONALE SUPERIEURE DES MINES DE SAINT-ETIENNE

Directeur : Monsieur M. MERMET
Directeur des Etudes et de la formation : Monsieur J. LEVASSEUR
Directeur des recherches : Monsieur J. LEVY
Secrétaire Général : Mademoiselle M. CLERGUE

Professeurs de 1ère Catégorie

COINDE	Alexandre	Gestion
GOUX	Claude	Métallurgie
LEVY	Jacques	Métallurgie
LOWYS	Jean-Pierre	Physique
MATHON	Albert	Gestion
RIEU	Jean	Mécanique - Résistance des matériaux
SOUSTELLE	Michel	Chimie
FORMERY	Philippe	Mathématiques Appliquées

Professeurs de 2ème catégorie

HABIB	Michel	Informatique
PERRIN	Michel	Géologie
VERCHERY	Georges	Matériaux
TOUCHARD	Bernard	Physique Industrielle

Directeur de recherche

LESBATS	Pierre	Métallurgie
---------	--------	-------------

Maîtres de recherche

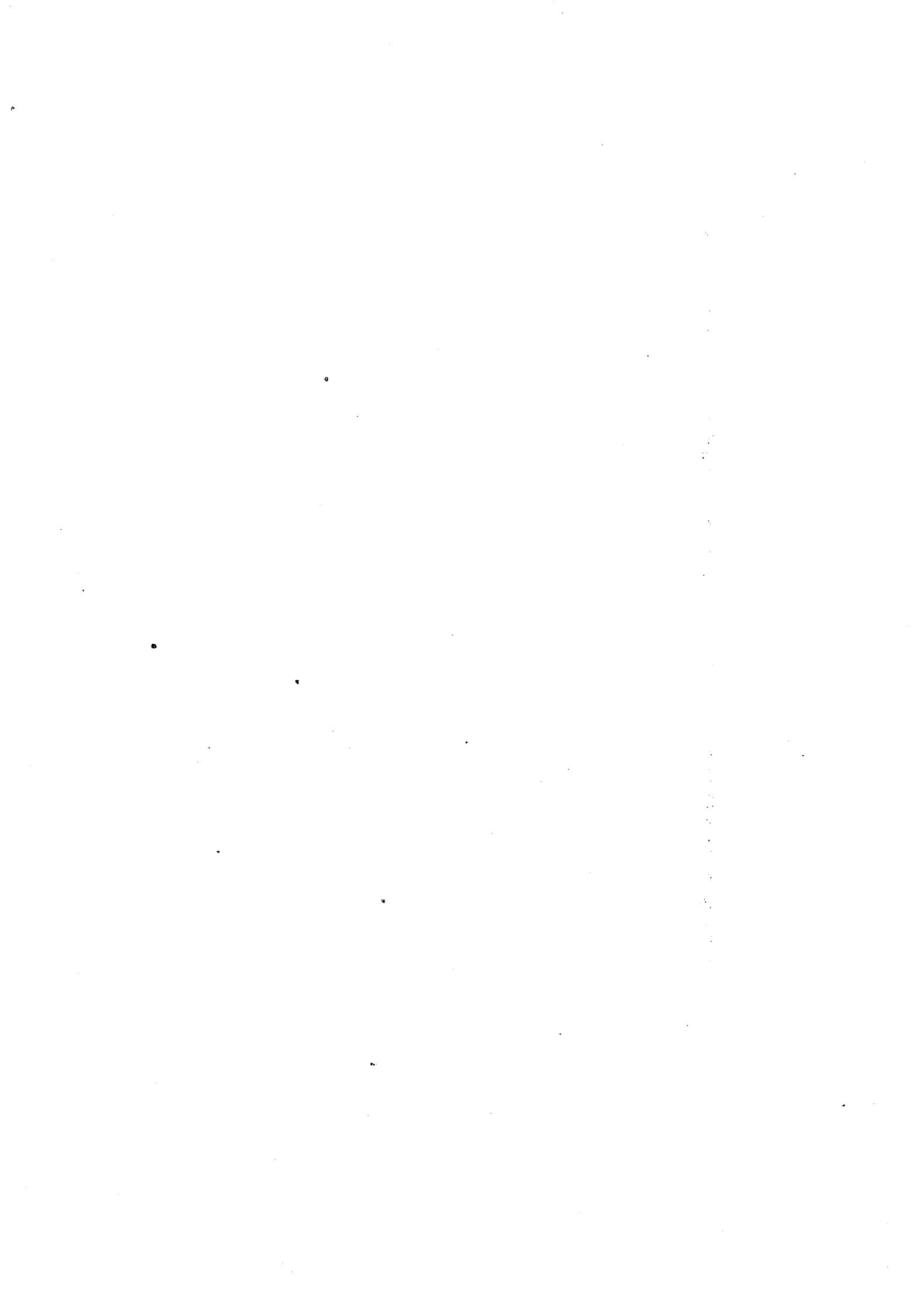
BISCONDI	Michel	Métallurgie
DAVOINE	Philippe	Géologie
FOURDEUX	Angeline	Métallurgie
KOBYLANSKI	André	Métallurgie
LALAUZE	René	Chimie
LANCELOT	Francis	Chimie
LE COZE	Jean	Métallurgie
THEVENOT	François	Chimie
TRAN MINH	Canh	Chimie

Personnalités habilitées à diriger des travaux de recherche

DRIVER	Julian	Métallurgie
GUILHOT	Bernard	Chimie
THOMAS	Gérard	Chimie

Professeur à l'UER de Sciences de Saint-Etienne

VERGNAUD	Jean-Maurice	Chimie des Matériaux & chimie industrielle
----------	--------------	--



Cette thèse a été préparée au Laboratoire d'Electronique et de Technologie de l'Informatique (LETI), en collaboration avec le Laboratoire d'Informatique et de Mathématiques Appliquées de Grenoble (IMAG). Qu'il me soit permis de remercier Monsieur GARIOD qui m'a accueilli dans le Laboratoire Mesures, Contrôle et Traitement Electronique du LETI, qu'il dirige.

Je remercie Monsieur F. ROBERT, Professeur à l'ENSIMAG d'avoir accepté de présider le jury.

Monsieur C. CARASSO, Professeur à l'Université de St-Etienne, a bien voulu venir à Grenoble pour cette occasion, je l'en remercie.

Monsieur F. NATTERER, Professeur à l'Université de Münster (RFA), a fait un large détour pour venir, je lui en suis vivement reconnaissant : ce mémoire fait largement référence à ses travaux qui sont des contributions fondamentales à de nombreux aspects des problèmes de reconstruction d'images.

Monsieur P.J. LAURENT, Professeur à l'Université Scientifique et Médicale de Grenoble a dirigé mon travail et m'a constamment soutenu, je l'en remercie.

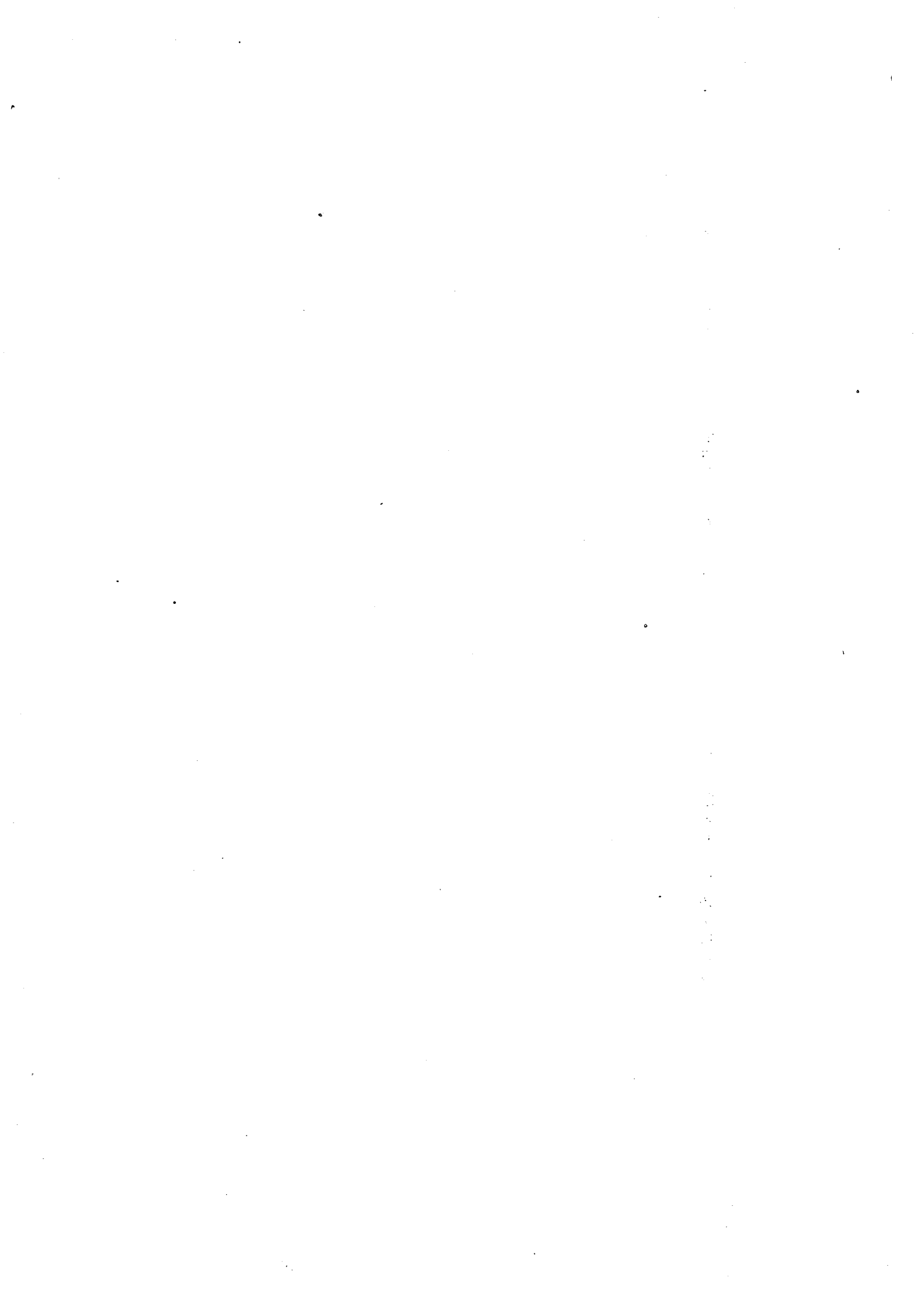
Je tiens à remercier aussi Messieurs Ph. GARDERET et R. CAMPAGNOLO, du LETI, qui m'ont initié aux aspects pratiques des problèmes étudiés ici, et pour l'intérêt qu'ils ont toujours manifesté à ce travail.

SOMMAIRE

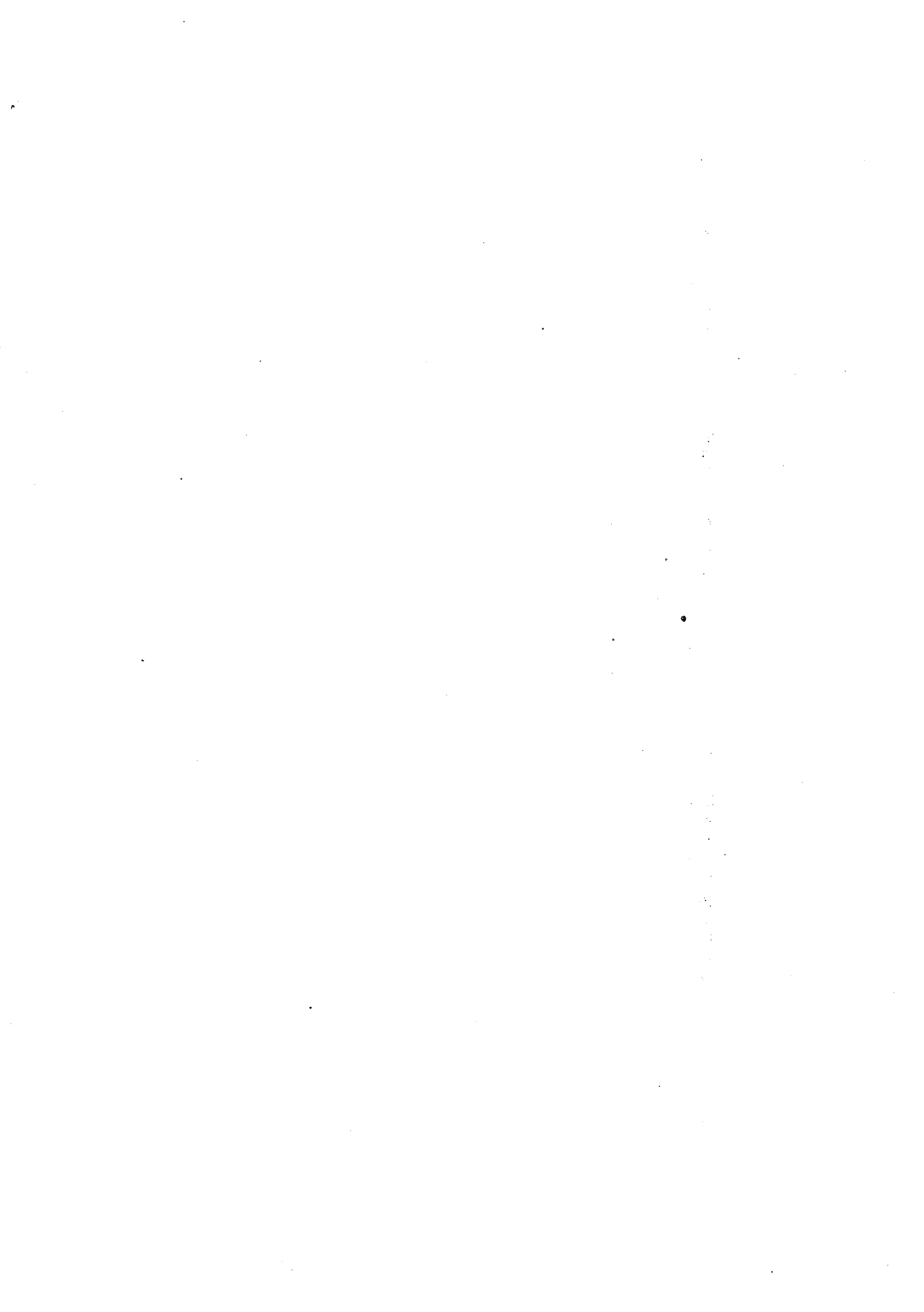
	<u>Pages</u>
INTRODUCTION	1
Chapitre I : FORMULATION DES PROBLEMES DE RECONSTRUCTION TOMOGRAPHIQUE	5
A. Formulation générale	5
B. Trois exemples de modélisation en imagerie médicale	6
1. Tomodensitomètre à rayons X	6
2. Tomographie par émission de positrons	7
3. Tomographie à positrons et à temps de vol	9
Chapitre II : SURVOL DES METHODES CONNUES DE RECONSTRUCTION D'IMAGES	13
A. L'approche analytique	13
1. Cas de la tomographie classique	13
1.1. La transformée de Radon et son inversion	13
1.2. L'épandage des projections filtrées	18
1.3. Le filtrage des projections épandues	20
2. Cas de la tomographie à temps de vol	21
B. Les méthodes algébriques	23
1. L'approche générale	23
2. Exemples standards de méthodes algébriques	24
2.1. Pixels et techniques itératives	24
2.2. D'autres décompositions	25
3. Les formulations possibles définissant l'estimation du modèle discret	28

Chapitre III : L'APPROCHE PROPOSEE PAR LES METHODES DE REGULARISATION OPTIMALE	32
A. Introduction	32
B. Quatre exemples de solution régularisée de paramètre ρ	34
1. Le critère d'écart aux mesures	34
2. Régularisation par la norme $L_2(D)$	35
3. Avec l'énergie du gradient	37
4. Avec l'énergie de flexion	38
5. En dimension finie	39
C. Estimation de la valeur optimale de ρ en fonction des mesures	40
1. La méthode du résidu	42
2. Méthode statistique	43
2.1. Avec l'erreur de prédiction	43
2.2. Avec l'erreur de reconstruction	46
3. La méthode de validation croisée	49
D. Sur le choix des critères de régularisation	52
1. Le choix de l'espace de recherche et du critère de régularité	52
2. Le choix d'une translation du problème	53
Chapitre IV : REGULARISATION PAR LA NORME L_2 AVEC PARAMETRE OPTIMAL : ALGORITHME DIRECT DE CALCUL DE LA SOLUTION D'ENERGIE MINIMALE SOUS CONTRAINTE OPTIMALE	56
A. Introduction	56
B. Mise en oeuvre numérique en tomographie classique	58
1. Les symétries	58
2. Le calcul de f_ρ	59
3. Le calcul du paramètre ρ optimal	60
4. Le cas du bruit poissonnien	63
5. Généralisation de ce type d'algorithmes à d'autres tomographes	64
C. L'efficacité du choix automatique du paramètre ρ	66

	<u>Pages</u>
Chapitre V : REGULARISATION OPTIMALE EN DIMENSION n : ALGORITHMES DIRECTS D'APPROXIMATION DANS LE CAS DE SEMI-NORMES INVARIANTES PAR ROTATION	72
A. Introduction	72
B. Une simplification du calcul du ρ optimal quand $m \gg n$	74
1. Un problème de taille n équivalent	74
2. Le calcul du ρ optimal sur ce nouveau problème	75
C. La mise en oeuvre numérique	77
1. Les symétries	77
2. Le calcul de f_{ρ}	79
3. Le calcul du paramètre ρ optimal avec Δ inversible	
D. L'efficacité du choix automatique du paramètre ρ	83
Chapitre VI : EVALUATION DES ALGORITHMES PROPOSES	88
A. En tomographie classique	88
B. En tomographie à positrons et à temps de vol	93
C. Discussion	96
CONCLUSION	99
ANNEXE 1 : Etude générale de l'existence, l'unicité et la caracté- risation de la solution régularisée de paramètre ρ	
ANNEXE 2 : Practical optimal regularization of large linear systems	
BIBLIOGRAPHIE	



INTRODUCTION



INTRODUCTION

Le problème appelé 'reconstruction d'une image à partir de ses projections' est apparu indépendamment dans de nombreux domaines scientifiques.

Sous sa forme la plus simple, ce problème est celui de la recherche de la valeur d'un paramètre scalaire en tout point d'un disque D alors que le type de mesure que l'on peut réaliser donne la valeur moyenne de ce paramètre le long d'une ligne droite connue.

La résolution numérique d'un tel problème apparaît dans de nombreux systèmes d'imagerie médicale qui diffèrent par la nature physique de la fonction f^0 recherchée, le nombre de mesures, m , leur géométrie et leur précision. Cette fonction est la densité du tissu en tomographie à rayons X, ou la concentration en émetteurs de positrons en tomographie par émission. Dans ce second contexte, on peut accéder aussi à des moyennes plus locales le long de chaque ligne. Une formulation générale de ces problèmes et quelques modèles d'acquisition sont présentés au chapitre I.

Les algorithmes (de reconstruction) couramment utilisés sont des discrétisations de formules d'inversion explicites, comme celle de la transformée de Radon. Ces méthodes analytiques sont en effet performantes dès que l'on a un nombre suffisant de bonnes mesures ($> 10^4$) uniformément réparties et que la discrétisation est faite avec précaution, en particulier avec des lissages indispensables, d'autant plus fort que les mesures sont plus bruitées.

Ces algorithmes sont décrits au chapitre II qui constitue un survol des méthodes connues.

Mais le principal défaut de ces algorithmes est leur déterminisme alors que la force de ces lissages devrait être adaptée aussi bien à la précision des mesures qu'à la forme de f^0 .

Les nouveaux algorithmes proposés dans cette thèse y remédient en réalisant une adaptation de ce type, déduite automatiquement des mesures

et éventuellement améliorée à l'aide d'une estimation de leur variance, pour un coût de calcul équivalent aux méthodes analytiques (excepté la place en mémoire). Ils ont aussi l'avantage de pouvoir prendre en compte des modèles de mesures plus complexes.

L'approche utilisée est celle des méthodes de régularisation optimale que l'on présente au chapitre III: elle consiste à définir la solution comme la moins irrégulière des fonctions qui sont ajustées aux contraintes fournies par les mesures avec un écart quadratique fixé ou optimisé. Cette dernière optimisation est réalisée à travers le calcul du paramètre de régularisation, ρ , fixant le compromis entre l'irrégularité et l'ajustement aux contraintes, par la méthode de validation croisée ou une méthode statistique. Différents choix de l'espace de recherche F et d'une semi-norme J mesurant l'irrégularité sont discutés.

Bien que cette approche théorique soit attrayante, elle conduit avec $F = L_2(\mathbb{D})$ et $J(f) = \int f^2$, à un grand système linéaire, paramétré par ρ , à m inconnues.

L'invariance par rotation de la géométrie des tomographes classiques a déjà permis la résolution directe d'un tel système avec ρ fixé a priori pour un coût compétitif.

C'est grâce à ces symétries que l'on montre au chapitre IV que, dans cette régularisation par l'énergie, le choix optimal de ρ peut être effectuée en même temps que le calcul de la solution régularisée associée, pour un supplément de coût négligeable.

On montre ensuite au chapitre V comment quand F est de dimension fini n une telle régularisation optimale peut aussi être efficacement mise en oeuvre (avec $n \approx 10^{4.5}$) indépendamment de m , dès qu'une famille engendrant F et J ont une certaine invariance par rotation.

Une technique nouvelle, non spécifique aux problèmes étudiés, permettant l'application de ces méthodes d'adaptation du paramètre à de grandes tailles de données pourvu que le nombre d'inconnues reste raisonnable, s'est avérée ici indispensable.

Des simulations numériques montrent finalement au chapitre VI l'efficacité de ces méthodes appliquées à la tomographie classique et à la tomographie temps de vol ($m \approx 10^6$).

Chapitre I

FORMULATION DES PROBLEMES DE RECONSTRUCTION TOMOGRAPHIQUE

A - Formulation générale

B - Trois exemples de modélisation en imagerie médicale

1. Tomodensitomètre à rayons X
2. Tomographie par émission de positrons
3. Tomographie à positrons et à temps de vol

A - FORMULATION GENERALE

Le but d'un système tomographique classique est de former une image de la distribution f^0 du paramètre étudié, dans un plan donné, uniquement à partir d'un nombre fini, m , de mesures y_1, y_2, \dots, y_m donnant une valeur approchée des intégrales $y_i^0 = \iint_{L_i} f^0(x,y) dx dy$ sur m bandes étroites L_i données.

Une modélisation générale des problèmes dits de reconstruction d'image conduit au problème mathématique suivant :

f^0 étant une fonction inconnue à valeur réelle, positive ou nulle, définie sur le domaine borné $D \subset \mathbb{R}^2$ et vérifiant les m contraintes :

$$y_i = (\phi_i, f^0) + \varepsilon_i \quad i = 1, \dots, m$$

où

{	a) ϕ_i $i = 1, \dots, m$ sont des distributions connues définissant l'appareil de mesure,
	b) y_i $i = 1, \dots, m$ sont les mesures réalisées sur f^0 ,
	c) les erreurs inconnues ε_i $i = 1, \dots, m$ satisfont certaines propriétés statistiques connues,

on cherche une 'bonne' fonction f proche de f^0 .

Le plus souvent ϕ_i peut être assimilé à une fonction de $L_2(D)$, l'espace des (classes de) fonctions de carré intégrable sur D avec

$$(\phi_i, f) = \iint_D \phi_i(x,y) f(x,y) dx dy.$$

On devine que le sens de 'bonne' fonction devra être choisi de manière d'une part à faciliter la construction numérique et, d'autre part, à assurer que celle-ci soit cohérente avec les informations connues a priori sur f^0 .

Nous verrons que si f est une solution construite par une méthode de type régularisation nous pouvons calculer une estimation d'une mesure de la proximité de f à f^0 .

B - TROIS EXEMPLES DE MODELISATION EN IMAGERIE MEDICALE

1. Tomodensitomètre à rayons X

Chaque mesure est ici celle du facteur d'atténuation de l'intensité d'un mince faisceau de rayons X qui traverse l'objet suivant une ligne droite donnée, et le logarithme de ce facteur est sensiblement proportionnel à l'intégrale le long de cette ligne de la densité de l'objet étudié.

Si l'on restreint l'ensemble des lignes de mesure à appartenir à un même plan que l'on munit d'un repère orthonormé $(0,x,y)$ dans lequel la densité $f(x,y)$ en tout point (x,y) de ce plan est à support contenu dans le disque unité $D = \{(x,y) / x^2 + y^2 < 1\}$, les mesures sont liées à f suivant le modèle:

Si y_i est la mesure exacte associée à la ligne L_i de largeur $d_i > 0$ et de paramètre (θ_i, u_i) dans $(0,x,y)$ c'est-à-dire :

$$L_i = \{(x,y) \in D / u_i - \frac{d_i}{2} < x \cos \theta_i + y \sin \theta_i < u_i + \frac{d_i}{2}\} \quad (1.1)$$

alors

$$y_i = \frac{1}{d_i} \int_{u_i - \frac{d_i}{2}}^{u_i + \frac{d_i}{2}} g_{\theta_i}(u) du \quad (1.2)$$

où

$g_{\theta_i} : [-1,1] \rightarrow \mathbb{R}$ est définie par :

$$g_{\theta_i}(u) = P_{\theta_i} f(u) = \int_{-\sqrt{1-u^2}}^{+\sqrt{1-u^2}} f(u \cos \theta_i - v \sin \theta_i, u \sin \theta_i + v \cos \theta_i) dv \quad (1.3)$$

$P_{\theta_i} f$ est appelée projection (ou radiographie) de f suivant la direction θ_i .

Si $d_i = 0$, on modélise généralement y_i par $g_{\theta_i}(u_i)$ mais, même pour f assez simple, g_{θ_i} n'existe pas toujours en tout point u .

Par contre dès que $d_i > 0$ il suffit de supposer f de carré intégrale sur D pour que y_i soit toujours bien défini puisque (1.2) et (1.3) s'écrivent :

$$y_i = \iint_D \frac{1}{d_i} \chi_{L_i}(x,y) f(x,y) dx dy = \left(\frac{1}{d_i} \chi_{L_i}, f \right)$$

$$\text{où } \chi_{L_i}(x,y) = \begin{cases} 1 & \text{si } (x,y) \in L_i \\ 0 & \text{sinon} \end{cases} \quad (\text{fonction caractéristique de } L_i)$$

Ce modèle linéaire est à la base de toutes les méthodes de reconstruction utilisées dans les tomodesitométriques. L'aide au diagnostic, considérable, fournie par les performances actuelles de ces scanners est une confirmation supplémentaire de la validité de ce modèle (voir par ex. A.M. Cormack (1980), G.T. Herman (1980), K.T. Smith (1977)).

2. Tomographie par émission de positrons

Des radioéléments émetteurs de positrons sont injectés dans l'organisme à étudier. Par exemple une inhalation d'oxygène marqué permet une étude du flux sanguin. Chaque émission d'un positron produit toujours, au même point, l'émission d'un couple de photons s'écartant immédiatement l'un de l'autre à la vitesse de la lumière le long d'une même droite. Si les trajets de ces 2 photons rencontrent 2 détecteurs opposés, distants d'environ 1 mètre, ces rencontres sont sensiblement simultanées. Pour un tel couple de détecteurs une horloge électronique est utilisée pour ne pas prendre en compte un photon émis suivant une ligne droite n'atteignant qu'un seul des détecteurs (car une telle détection n'est presque jamais accompagnée d'une détection simultanée sur l'autre détecteur). Le nombre de coïncidences (2 détections simultanées) comptées par ce couple est alors aussi le nombre de couples de photons émis suivant une ligne atteignant ces 2 détecteurs, si on néglige en première phase l'atténuation de ces photons.

L'activité en tout point du tissu du processus d'émission de positrons caractérisant la concentration au même point de la particule marquée, c'est la restriction $f^0(x,y)$ de cette activité à un plan d'observation que l'on souhaite estimer à partir des nombres de coïncidences y_1, \dots, y_m

comptés par m couples de détecteurs placés dans ce plan, durant un temps total t . (Pour un survol sur ces techniques d'imagerie, voir M.M. Ter - Pogossian (1980)).

Pour exprimer chacune de ces mesures en fonction de f^0 , on suppose que le processus d'émission des photons est un processus ponctuel de Poisson dans $\mathbb{R}^2[0, \pi[$ d'activité $\frac{t}{\pi} f^0(x,y)$. On peut alors montrer que y_i est la réalisation d'une variable aléatoire Y_i de Poisson de paramètre (ϕ_i, f^0) :

$$Y_i \sim P((\phi_i, f^0)) \quad (2.1)$$

avec, sachant que la largeur d_i de la ligne L_i joignant le couple i de détecteur est petite devant sa longueur ℓ_i ,

$$(\phi_i, f^0) = t \frac{d_i}{\pi} \iint_{L_i} f^0(x,y) dx dy \quad (2.2)$$

Signalons que si on ne peut raisonnablement négliger l'atténuation des photons, ce modèle est corrigé par l'introduction de la fonction $\mu(x,y)$ atténuation locale, simplement en remplaçant ϕ_i par $a_i(\mu)\phi_i$ où $a_i(\mu)$ est une constante mesurant l'atténuation des photons le long de L_i

$$a_i(\mu) = e^{-\frac{1}{d_i} \iint_{L_i} \mu(x,y) dx dy}$$

En pratique on peut accéder à une bonne valeur de ces coefficients par une simple mesure préalable.

Notons que les coïncidences fortuites et les effets de diffusion sont des phénomènes que l'on a négligés et qui semblent difficiles à modéliser correctement.

Si l'on suppose faibles les erreurs de mesures, y_i (ou $\frac{y_i}{a_i(\mu)}$ si l'atténuation n'est pas négligeable) est donc liée à f^0 par une équation analogue à celle de la tomographie à rayon X. Les mêmes algorithmes de reconstruction peuvent donc être utilisés pour ces 2 problèmes quand la géométrie des lignes de mesure est identique. La principale différence est que la précision relative des mesures est souvent très mauvaise en tomographie par émission (et d'autant plus que t , d_i ou $\int_{L_i} f^0$ diminuent)

et un algorithme de reconstruction performant en tomographie X risque fort d'être très instable vis à vis de ces erreurs. Une information dont il sera tiré profit pour l'élaboration de nouveaux algorithmes est le caractère poissonnien de chaque mesure acquise.

3. Tomographie à positions et à temps de vol

L'incorporation d'une information supplémentaire au système de mesure précédent est actuellement possible : c'est la mesure du temps de vol. R.E. Campagnolo (1980) a étudié l'utilisation de cette technique. Pour chaque coïncidence détectée, on mesure approximativement (erreur gaussienne) l'écart (ou temps de vol), ne dépassant pas quelques nanosecondes, entre les dates de détection des 2 photons. Cette mesure permet de classer les y_i coïncidences détectées dans L classes de temps de vol. Notons C_ℓ le segment de l'axe de L_i associé à cette classe de temps de vol, y_i^ℓ le nombre de couples détectés dans cette direction et dans cette classe. On admet que l'erreur sur la mesure de temps de vol a une distribution gaussienne, d'écart type σ connu. Le modèle est alors :

$$(\phi_i^\ell, f) = \iint_D t \cdot \frac{\binom{d_i}{\ell_i}}{\pi} \cdot \chi_{\left[u_i - \frac{d_i}{2}, u_i + \frac{d_i}{2}\right]}(u) \left[\int_{C_\ell} G_\sigma(v-s) ds \right] f_{\theta_i}(u,v) dv du$$

où

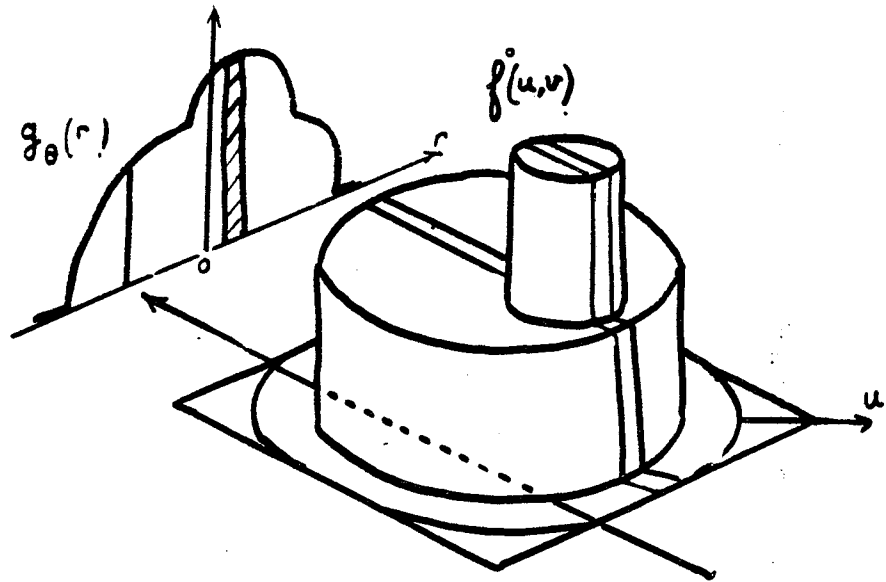
$$\left\{ \begin{array}{l} f_{\theta_i}(u,v) = f(u \cos\theta_i - v \sin\theta_i, u \sin\theta_i + v \cos\theta_i) \text{ est l'expression} \\ \text{de l'activité recherchée dans le repère attaché à la} \\ \text{direction } \theta_i \\ G_\sigma(v) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{v^2}{\sigma^2}\right) \end{array} \right.$$

qui peut aussi s'écrire comme une intégration locale de la fonction de 2 variables :

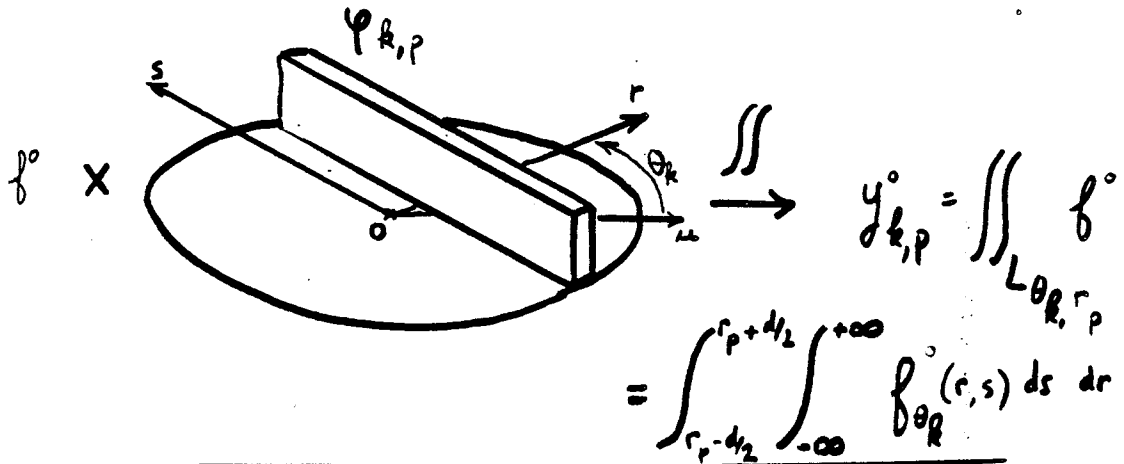
$$H_{\theta_i}^\ell(f)(u, \cdot) = G_\sigma * f_{\theta_i}(u, \cdot) \quad \text{i.e.} \quad H_{\theta_i}^\ell(f)(u, s) = \int G_\sigma(v-s) f_{\theta_i}(u, v) dv$$

appelée histoprojection de f d'angle θ_i , puisque :

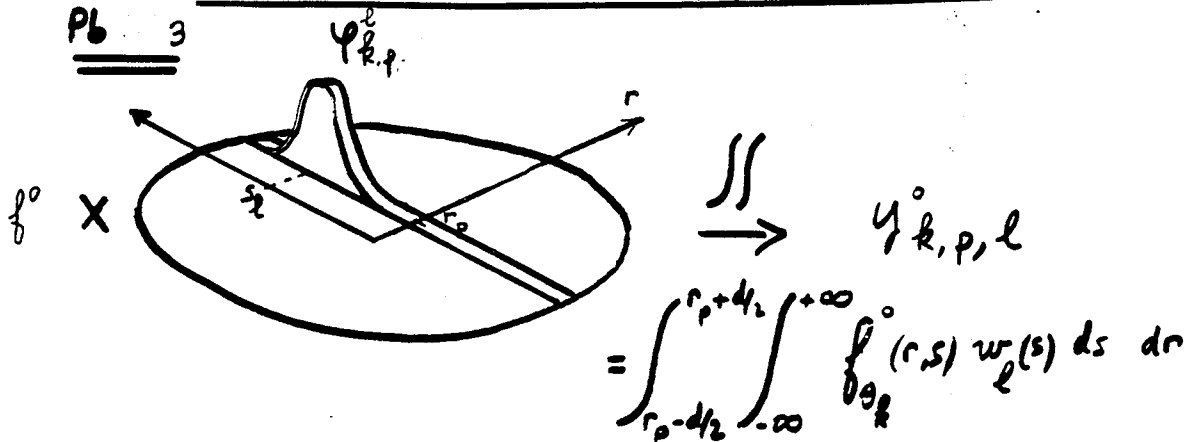
Modèle des mesures y_i en tomographie.



Pb 1 et 2



Pb 3



où $w_l = \chi_{c_l} * G_\sigma$

$$(\phi_i^\ell, f) = \int_{u_i - \frac{d_i}{2}}^{u_i + \frac{d_i}{2}} \int_{C_\ell} H_{\theta_i}(f)(u,s) ds du$$

et on montre que $y_{i\ell}$ est une réalisation de la variable aléatoire $Y_{i\ell}$ poissonnienne :

$$Y_{i\ell} \sim P((\phi_i^\ell, f))$$

On peut faire la même remarque qu'en 2 au sujet de l'atténuation des photons (ϕ_i^ℓ devient $a_i(\mu)\phi_i^\ell$).

4. Remarques sur la géométrie de ces mesures

a) ϕ_i et ϕ_i^ℓ sont des fonctions séparables dans le repère $(0,u,v)$ obtenu par rotation de θ_i de $(0,x,y)$, en effet en notant Δ_u^d la fonction

$$\chi_{\left[u - \frac{d}{2}, u + \frac{d}{2}\right]}$$

$$\phi_i(x,y) = \Delta_{u_i}^{d_i}(u) \quad \text{et} \quad \phi_i^\ell(x,y) = \Delta_{u_i}^{d_i}(u) \cdot (\chi_{C_\ell} * G_\sigma)(v)$$

(à un coefficient de proportionnalité près)

b) les coordonnées (θ_i, u_i) de chaque ligne support de ϕ_i et ϕ_i^ℓ sont généralement choisies sur une grille rectangulaire uniforme de $[0, \pi] \times [-1, 1]$ (\mathbb{D} étant le disque de rayon unité). Elles sont notées : (θ_k, u_p) $k=1, \dots, K$, $p=1, \dots, P$ où l'indice i est remplacé par le couple (k,p) . Cette invariance par rotation permettra un calcul économique des solutions que l'on va présenter.

c) le nombre total de mesures $K.P$ ($K.P.L$ avec temps de vol) est en pratique assez grand (pour fixer une idée, un exemple typique est $K = 96$, $P = 128$, $L = 64$).



Chapitre II

SURVOL DES METHODES CONNUES DE RECONSTRUCTION D'IMAGES

A - L'approche analytique

1. Cas de la tomographie classique
 - 1.1. La transformée de Radon et son inversion
 - 1.2. L'épandage des projections filtrées
 - 1.3. Le filtrage des projections épandues
2. Cas de la tomographie à temps de vol

B - Les méthodes algébriques

1. L'approche générale
2. Exemples standard de méthodes algébriques
 - 2.1. Pixels et techniques itératives
 - 2.2. D'autres décompositions
3. Les formulations possibles définissant l'estimation du modèle discret.

Dans l'abondante littérature traitant ces problèmes, de nombreuses méthodes apparaissent et sont présentées de manières diverses. Dans ce chapitre, on rappelle les 2 types d'approches qui ont conduit à ces méthodes, et on étudie en détail les méthodes admises comme les plus performantes et devenues standard.

La possibilité de modéliser les mesures comme des échantillons d'une transformation intégrale que l'on sait analytiquement inverser est à la base de la première approche. La seconde consiste, elle, à formuler les équations exprimant l'information mesurée et à rechercher une solution les satisfaisant.

A - APPROCHE ANALYTIQUE

1. Cas de la tomographie classique (I.B.1 et I.B.2)

1.1. La transformée de Radon et son inversion

Cette approche repose sur les expressions analytiques de l'opérateur inverse de l'opérateur R :

f étant une fonction continue de \mathbb{R}^2 dans \mathbb{R} supposée nulle hors du disque unité D ,

Rf est la fonction de $[0, \pi] \times \mathbb{R} \rightarrow \mathbb{R}$ définie par :

$$Rf(\theta, u) = \int_{-1}^1 f(u \cos \theta - v \sin \theta, u \sin \theta + v \cos \theta) dv$$

R est appelé opérateur de Radon, qui en 1917 a établi l'expression de R^{-1} :

Si $g = Rf$ alors

$$f(x, y) = R^{-1}g(x, y) = \frac{1}{2\pi^2} \int_0^\pi \int_{-1}^1 \frac{1}{x \cos \theta + y \sin \theta - u} \frac{\partial g}{\partial u}(\theta, u) du d\theta \quad (1.1)$$

où l'intégrale singulière est prise au sens de la valeur principale de Cauchy.

$$\int_{-1}^1 \frac{p(u)}{v-u} du = \lim_{\varepsilon \rightarrow 0} \int_{-1}^{v-\varepsilon} \frac{p(u)}{v-u} du + \int_{v+\varepsilon}^1 \frac{p(u)}{v-u} du .$$

Signalons que ces expressions ont été ensuite étendues aux distributions sur \mathbb{D} (voir Ludwig (1966)).

Maintenant on voit que dans les modélisations I.B.1 et I.B.2, en supposant la largeur des bandes d'intégration suffisamment petite, $(\phi_{k,p}, f)$ coïncide à un facteur près (de proportionnalité connue), avec $Rf(\theta_k, u_p) = P_{\theta_k} f(u_p)$. Alors la reconstruction de f^0 au point (x, y) par une méthode analytique consiste à évaluer l'intégrale (1.1) dans laquelle $g = Rf^0$ est connue aux points (θ_k, u_p) , à l'erreur $\varepsilon_{k,p}$ près, par la mesure $y_{k,p}$. En fait d'autres expressions de R^{-1} que nous allons rappeler se sont révélées plus faciles à discrétiser.

Pour les présenter, définissons l'opérateur d'épandage (ou rétro-projection) :

$$R^* : \lambda ([0, \pi] \times \mathbb{R} \rightarrow \mathbb{R}) \rightarrow R^* \lambda (\mathbb{R}^2 \rightarrow \mathbb{R})$$

$$\text{avec } R^* \lambda(x, y) = \int_0^\pi \lambda(\theta, x \cos \theta + y \sin \theta) d\theta \quad (1.2)$$

(1.1) s'écrit alors sous la forme :

$$R^{-1} g = R^* \lambda$$

$$\text{où } \lambda : (\theta, v) \in [0, \pi] \times \mathbb{R} \rightarrow \lambda(\theta, v) = \frac{1}{2\pi^2} \int_{-1}^1 \frac{1}{v-u} \frac{\partial}{\partial u} g(\theta, u) du \quad (1.3)$$

A l'aide d'une intégration par partie, cette dernière intégrale peut aussi s'écrire :

$$\lambda(\theta, v) = -\frac{1}{2\pi^2} \lim_{\varepsilon \rightarrow 0} \left[\int_{-1}^{v-\varepsilon} \frac{g(\theta, u)}{(v-u)^2} du + \int_{v+\varepsilon}^1 \frac{g(\theta, u)}{(v-u)^2} du - 2 \frac{g(\theta, v)}{\varepsilon} \right] \quad (1.4)$$

Introduisons la transformation de Fourier F_1 par rapport à u :

$$g \left([0, \pi] \times \mathbb{R} \rightarrow \mathbb{R} / \int_{-\infty}^{\infty} g(\theta, u)^2 du < \infty \right) \rightarrow F_1 g \left([0, \pi] \times \mathbb{R} \rightarrow \mathbb{C} \right)$$

$$\text{avec } F_1 g(\theta, U) = \int_{-\infty}^{\infty} g(\theta, u) e^{-2\pi i u U} du$$

Alors (1.4) qui est une convolution, équivaut formellement à :

$$F_1^{-1} \lambda(\theta, U) = |U| F_1 g(\theta, U)$$

Pour une formulation rigoureuse, voir L. Schwartz (1961).

$$\text{Ainsi } R^{-1} = R^* F_1^{-1} |U| F_1 \quad (1.5)$$

L'introduction de la transformée de Fourier F_2 à 2 dimensions donne deux autres expressions équivalentes de R^{-1} et permet de retrouver (1.5) :

Rappelons que $F_2 f$ est une fonction de \mathbb{R}^2 dans \mathbb{C} définie par :

$$F_2 f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-2\pi i (xX + yY)} dx dy$$

Notons \mathcal{C} la conversion en polaire

$$\hat{f}(\mathbb{R}^2 \rightarrow \mathbb{C}) \rightarrow \mathcal{C}\hat{f}([0, \pi] \times \mathbb{R} \rightarrow \mathbb{C})$$

$$\text{avec } \mathcal{C}\hat{f}(\theta, U) = \hat{f}(U \cos \theta, U \sin \theta)$$

On peut facilement montrer le théorème dit d'équivalence coupe-projection dans le domaine de Fourier : $\mathcal{C} F_2(f) = F_1 R(f)$ (1.6)

D'où la nouvelle décomposition :

$$R^{-1} = F_2^{-1} \mathcal{C}^{-1} F_1 \quad (1.7)$$

Remarquons que (1.6) implique que si g est la transformée de Radon d'une fonction, $F_1 g$ est assimilable à une représentation polaire, contrairement à g .

On peut maintenant établir (1.5) par le changement de variable $x = U \cos\theta$, $y = U \sin\theta$ dans l'intégrale bien connue exprimant F_2^{-1} , donnant :

$$f(x,y) = \int_0^\pi \int_{-\infty}^\infty \mathcal{E} F_2 f(U,\theta) e^{2\pi i(x U \cos\theta + y U \sin\theta)} |U| dU d\theta \quad (1.8)$$

puisque l'on reconnaît ici que $f = R^* F_1^{-1} |U| (\mathcal{E} F_2 f)$ et qu'on peut substituer ensuite $F_1 R f$ à $\mathcal{E} F_2 f$ d'après (1.6).

Enfin pour la 3ème décomposition, introduisons l'épandage de g , noté e , défini par :

$$e = R^* g = R^* R f \quad (1.9)$$

En écrivant $e = R^* F_1^{-1} F_1 R f$ et en substituant, cette fois, $\mathcal{E} F_2 f$ à $F_1 R f$, on obtient :

$$e(x,y) = \int_0^\pi \int_{-\infty}^\infty \mathcal{E} F_2 f(U,\theta) e^{2\pi i U (x \cos\theta + y \sin\theta)} dU d\theta \quad (1.10)$$

d'où si $F_2 e$ est la transformée de Fourier de e (au sens des distributions car e n'est pas globalement intégrable sur \mathbb{R}^2 dès que f est à valeur ≥ 0 , non partout nulle) la comparaison de (1.8) et (1.10) indique que :

$$\mathcal{E} F_2 f = |U| \mathcal{E} F_2 e \quad (1.11)$$

D'où d'après (1.9)

$$R^{-1} = F_2^{-1} \mathcal{E}^{-1} |U| \mathcal{E} F_2 R^* \quad (1.12)$$

Ces 3 décompositions formelles de R^{-1} ont servi de guides à

l'élaboration de 3 types de méthodes d'évaluation numérique de f à partir d'échantillons de Rf , qui sont connues sous les noms respectifs :

- épannage des projections filtrées (décomposition (1.5))
- inversion de la transformée de Fourier à 2 dimensions (1.7)
- filtrage de l'image épannée (1.12)

Il est certain que le calcul exact, ainsi décomposé de $f(x,y)$ nécessite la connaissance de Rf en tout point θ, u . Or en pratique Rf est mesurée en un nombre fini d'angles de vue θ et d'abscisses u et ces mesures sont plus ou moins bruitées (assez fortement en tomographie par émission).

Ces 3 décompositions sont donc discrétisées avec précaution afin d'assurer une certaine insensibilité aux erreurs de mesures.

L'avantage de ces décompositions est qu'on peut très facilement les modifier de manière à ce qu'elle produise une approximation à bande limitée f_b au lieu de f .

Si $\mathcal{E}F_2f_b(U, \theta) = F_b(U) \mathcal{E}F_2f(U, \theta)$, où $F_b(U)$ est une fenêtre en fréquence, nulle si $|U| > b$, la modification souhaitée consiste à remplacer $|U|$ par $|U| F_b(U)$ dans (1.5) ou (1.12) ou F_1 par $F_b F_1$ dans (1.7).

Signalons un 4ème type d'expression de Rg^{-1} basé sur les décompositions azimutales (ou en harmoniques angulaires) de f et g (Cormack (1963), Perry (1975)).

La définition de $g = Rf$ étant étendue à tous les angles θ de \mathbb{R} , $g(., u)$ et $\mathcal{E}f(., r)$ sont des fonctions périodiques de période 2π et on a les décompositions azimutales :

$$g(\theta, u) = \sum_{n=-\infty}^{\infty} g_n(u) e^{i n \theta} \quad \text{et} \quad \mathcal{E}f(\theta, r) = \sum_{n=-\infty}^{\infty} f_n(r) e^{i n \theta}$$

où chaque harmonique g_n se calcule par :

$$g_n(u) = \frac{1}{2\pi} \int_0^{2\pi} g(\theta, u) e^{-i n \theta} d\theta$$

Et la formule de Cormack est :

$$f_n(r) = -\frac{1}{\pi r} \int_r^1 \frac{\frac{\partial}{\partial u} g_n(u) T_n(u/r)}{(u^2/r^2 - 1)^{3/2}} du$$

où T_n sont les polynômes de Tchébychev.

Cette formule fait apparaître que $f(x,y)$ peut être théoriquement calculé uniquement avec les valeurs : $g(\theta,u)$ pour $u^2 > x^2 + y^2$, mais elle se révèle moins facile à discrétiser.

Les deux paragraphes suivant décrivent les deux méthodes standard, déduites des formules (1.5) et (1.12), utilisées sur les tomographes commercialisés.

On suppose, pour ne pas compliquer, une répartition uniforme des échantillons de Rf :

$$(\theta_k, u_p) = (k \frac{\pi}{M_\theta}, p \Delta u) \quad k = 1, \dots, M_\theta \quad p = -M_u, \dots, M_u$$

Δu étant tel que le support de f soit dans le disque centré en 0 de rayon $M_u \Delta u$.

1.2. L'épandage des projections filtrées

La fonction f cherchée étant généralement représentée sur un écran vidéo, on cherche seulement ses valeurs aux points d'une grille rectangulaire, qui sont ensuite présentées par des niveaux de gris différents.

Cette méthode se décompose en 2 étapes :

a) dans chaque direction θ_k , la restriction à $[-1,1]$ de $\lambda(\theta_k, \cdot)$ (appelée projection filtrée) est estimée à partir des $2 M_u + 1$ valeurs $y_{k,p}$. (λ est définie par (1.3)).

b) une approximation de $R^* \lambda$ à l'aide des M_θ coupes précédentes de λ donne une estimation de $f(x,y)$, en tout point (x,y) souhaité. ($R^* \lambda$ est définie par (1.2)).

L'étape b) a toujours été réalisée par une somme classique de Riemann, appelée épannage discret de M_θ fonctions.

$$R_{M_\theta}^* \{ \lambda(\theta_k, \cdot) / k = 1, \dots, M_\theta \} = \frac{\pi}{M_\theta} \sum_{k=1}^{M_\theta} \lambda(\theta_k, x \cos \theta_k + y \sin \theta_k)$$

Par contre les formules d'estimation de $\lambda_{\theta_k} = \lambda(\theta_k, \cdot)$ sont nombreuses. Voir S.W. Rowland (1979).

Indiquons seulement le type de formule le plus rapide et le plus souvent utilisé : c'est l'implémentation dans le domaine de Fourier.

Il est bien connu en analyse de Fourier que l'échantillonnage de g_θ aux points $p \Delta u$, ne permet de retrouver qu'une approximation de $F_1 g_\theta$ sur $[-\frac{1}{2\Delta u}, \frac{1}{2\Delta u}]$ d'autant meilleure que les amplitudes $|F_1 g_\theta(U)|$ sont faibles pour $|U| > \frac{1}{2\Delta u}$ (effet de retournement de fréquence ou "aliasing"), mais cet échantillonnage permet d'accéder à une fonction à bande limitée \hat{g}_θ , estimation lisse de g_θ , donnée par, pour $\theta = \theta_k$:

$$\hat{F}_1 g_{\theta_k}(U) = F_b(U) \cdot \mathcal{D} F_{y_k}(U)$$

$$\mathcal{D} F_{y_k}(U) = \sum_{n=-\infty}^{\infty} F_1 g_{\theta_k}(U + \frac{n}{\Delta u}) = \Delta u \sum_{p=-M_u}^{M_u} y_{k,p} e^{-2\pi i U p \Delta u}$$

où $\mathcal{D} F_{y_k}(\frac{q}{2M' \Delta u})$ $q = -M', \dots, M'$ est la transformée de Fourier discrète des $y_{k,p}$

$F_b(U)$ est une fenêtre en fréquence nulle si $|U| > b$

b choisi $< \frac{1}{2\Delta u}$

Remarquons que si $F_b = \chi_{[-1/2 \Delta u, 1/2 \Delta u]}$, alors \hat{g}_{θ_k} interpole les valeurs $y_{k,p}$ en $u = p\Delta u$. Alors $2M_u + 1$ échantillons de λ_{θ_k} sont estimés en remplaçant g_{θ_k} par \hat{g}_{θ_k} :

$$\hat{\lambda}_{k,p} = (F_1^{-1} |U| F_1 \hat{g}_{\theta_k})(p\Delta u) = \int_{-b}^b |U| F_1 \hat{g}_{\theta_k}(U) e^{2\pi i U p \Delta u} dU$$

et en calculant cette intégrale pour une somme de Riemann aux points $U = \frac{q}{2M' \Delta u}$ $q = -M', \dots, M'$ (c'est une transformée de Fourier discrète inverse).

Ensuite une interpolation de Lagrange à 1 ou 2 points donne la fonction $\hat{\lambda}_{\theta_k}$. Notons que cette méthode utilise avec profit les algorithmes de transformée de Fourier rapide, et permet de tester facilement diverses fenêtres F_b . Un exemple standard est la fenêtre de Hamming :

$$F_b(U) = \begin{cases} \left(\frac{1}{2} \left(1 + \cos \pi \frac{U}{b}\right)\right)^n & \text{si } |U| < b \\ 0 & \text{sinon} \end{cases}$$

où les paramètres b et n , respectivement appelés fréquence de coupure et exposant du filtre, sont choisis afin de prévenir l'instabilité numérique du calcul.

1.3. Le filtrage des projections épandues est l'algorithme d'approximation de R_g^{-1} sur une grille rectangulaire déduit de la 3ème décomposition (1.12).

Il consiste à estimer R_g^* puis à "filtrer" cette fonction en multipliant ses fréquences $F_2 R_g^*(X,Y)$ par $\sqrt{X^2 + Y^2}$. L'implémentation numérique nécessite de supposer borné le support de R_g^* et se décompose en 4 étapes.

a) δ étant choisi a priori (voisin de Δu), J étant choisi suffisamment grand pour que R_g^* puisse être négligé hors de $[-J\delta, J\delta]^2$, on calcule en tout point (x,y) de la grille $\{(j'\delta, j''\delta) / j', j'' = -J, \dots, J\}$ l'épandage discret déjà vu

$$\frac{\pi}{M_\theta} \sum_{k=1}^{M_\theta} \hat{g}(\theta_k, x \cos \theta_k + y \sin \theta_k)$$

où $\hat{g}(\theta_k, u)$ est une fonction interpolant les $y_{k,p}$ en $u = p\Delta u$ pour $p = -M_u, \dots, M_u$

b) on calcule la transformée de Fourier discrète de cette matrice $(2J+1) \times (2J+1)$, ce qui nous donne une approximation de $F_2 R_g^*$ sur la grille $\left\{ \left(\frac{j'}{2J\delta}, \frac{j''}{2J\delta} \right) / j', j'' = -J, \dots, J \right\}$

c) on multiplie chaque coefficient de cette nouvelle matrice par

$$U_{j',j''} = \frac{1}{2j\delta} \sqrt{j'^2 + j''^2}$$

d) on calcule enfin la transformée de Fourier inverse discrète, que nous donne \hat{f} sur la grille $(j'\delta, j''\delta)$ $j', j'' = -J, \dots, J$

Rappelons que l'approximation de $F_2 Rg^*$ calculée en b) est souvent très mauvaise au bord du pavé $[-\frac{1}{2\delta}, \frac{1}{2\delta}]^2$. Pour y remédier un lissage de \hat{f} est réalisé en multipliant en c) par $U F_b(U)$ au lieu de U où F_b est une fenêtre analogue aux précédentes.

2. Cas de la tomographie à temps de vol (I.B.3)

L'approche analytique conduit ici à une méthode du type filtrage de l'image épandue, devenue la méthode standard. En effet dans chaque direction $\theta_k = k \frac{\pi}{M_\theta}$ les mesures $y_{k,p}^\ell = (\phi_{k,p}^\ell, f)$ sont les échantillons aux points $p\Delta u, \ell\Delta v$ de l'histoprojection $H_{\theta_k} f = g_{\theta_k}$ définie au I.B.3 par :

$$g_{\theta_k}(u,v) = \int_{-1}^1 f(u \cos\theta - v' \sin\theta, u \sin\theta + v' \cos\theta) G_\sigma(v-v') dv'$$

$$\text{Et on peut montrer que l'image épandue } e(x,y) = \int_0^\pi h_\theta(x,y) d\theta$$

$$\text{où } h_\theta(x,y) = g_\theta(x \cos\theta + y \sin\theta, -x \sin\theta + y \cos\theta)$$

(h_θ est la représentation de g_θ dans le repère fixe Oxy).

est liée à f par une relation de convolution analogue à (1.12) avec

$$\frac{\pi^2 \sigma^2 |U|^2}{\pi \cdot I_0(\pi^2 \sigma^2 |U|^2)} \text{ au lieu de } |U| \quad (2.1)$$

où I_0 fonction de Bessel modifiée (cf R. Campagnolo (1980))

L'implémentation est alors faite de manière semblable au § 1.3 avec cette nouvelle fonction (de transfert) utilisée en c) et un épannage discret en a) qui consiste ici à calculer en chaque point (x,y) de la grille $(j'\delta, j''\delta)$ une estimation $\hat{e}(x,y)$ de $e(x,y)$ par la somme de Riemann :

$$\hat{e}(x,y) = \frac{\pi}{M_\theta} \sum_{k=1}^{M_\theta} \hat{h}_k(x,y) \quad (2.2)$$

où chaque $\hat{h}_k(x,y)$ est une estimation de $g_{\theta_k}(x \cos\theta + y \sin\theta, -x \sin\theta + y \cos\theta)$ calculée par une interpolation (généralement bilinéaire) des échantillons $y_{k,p}^\ell = g_{\theta_k}(p\Delta u, \ell\Delta v)$ mesurés.

Il est important de remarquer que, dans ce système de mesure, les $y_{k,p}^\ell$ sont des échantillons très bruités des histoprojections (on compte en effet en général quelques dizaines d'émissions pour chaque $y_{k,p}^\ell$, avec une variance du même ordre, en raison de la faible dose de traceur injecté et du temps limité). L'algorithme de reconstruction actuellement implémenté tente d'y remédier en lissant chaque histoprojection par la convolution suivante :

$$\tilde{g}_\theta(u,v) = \int_{-\infty}^{+\infty} g_\theta(u,v) G_{\sigma'}(v-v') dv' \quad (2.3)$$

où σ' est un paramètre de ce lissage.

Comme auparavant $\tilde{e}(x,y) = \int_0^\pi \tilde{h}_\theta(x,y) d\theta$ est lié à f par une convolution dont la fonction de transfert peut être explicitée et est évaluée sur la grille $(\frac{j'}{2J\delta}, \frac{j''}{2J\delta})$. L'estimation de \tilde{e} est obtenue dans cet algorithme en réalisant l'épandage discret (cf (2.2)) où les $y_{k,p}^\ell$ sont remplacés par $\tilde{y}_{k,p}^\ell$:

$$\tilde{y}_{k,p}^\ell = \sum_{\ell'=1}^L w_{\ell'} - \ell y_{k,p}^{\ell'} \quad (2.4)$$

où $w_\ell = G_{\sigma'}(\ell\Delta v)$

(Δv est le pas des classes de temps de vol)

Il a été observé que $\sigma' = \sigma$ convenait le mieux à cette reconstruction.

B - LES METHODES ALGEBRIQUES

1. L'approche générale

Contrairement aux méthodes analytiques où la discrétisation des formules apparaît à la fin de la procédure, l'approche consiste ici à discrétiser dès le début l'estimation cherchée de f^0 . Pour cela on choisit un nombre fini de fonctions simples $\Psi_j : \mathbb{D} \rightarrow \mathbb{R}$ ($j = 1, \dots, n$) et on cherche cette estimation sous la forme :

$$f(x,y) = \sum_{j=1}^n \lambda_j \Psi_j(x,y) \quad \text{en tout point } (x,y) \text{ de } \mathbb{D} \quad (1.1)$$

Le choix d'une telle décomposition doit évidemment d'une part assurer la possibilité d'une bonne modélisation de f^0 et d'autre part permettre pour un coût raisonnable, la recherche de n coefficients $\lambda_1, \dots, \lambda_n$ tels que $\sum_{j=1}^n \lambda_j \Psi_j$ vérifie les informations mesurées ou connues a priori sur f^0 .

Ces informations mesurées sont les valeurs y_i , plus ou moins précises liées à f^0 par :

$$y_i = (\phi_i, f^0) + \varepsilon_i \quad i = 1, \dots, m \quad (\text{cf formulation I.A})$$

Or on souhaite ici que $\sum_{j=1}^n \lambda_j \Psi_j$ soit proche de f^0

A fortiori, parce que chaque fonctionnelle (ϕ_i, \cdot) est linéaire et continue sur $L_2(\mathbb{D})$, on veut que $\sum_{j=1}^n \lambda_j (\phi_i, \Psi_j)$ soit voisin de (ϕ_i, f^0) , l'écart autorisé étant appelé erreur de discrétisation. Après substitution dans l'équation ci-dessus on obtient les m équations que doivent satisfaire les λ_j :

$$y_i = \sum_{j=1}^n \lambda_j (\phi_i, \Psi_j) + e_i \quad i = 1, \dots, m \quad (1.2)$$

où $\left\{ \begin{array}{l} (\phi_i, \Psi_j) \\ e_i \end{array} \right.$ sont des coefficients calculables
est une tolérance non exactement connue mais qui doit ressembler à la somme d'une erreur de mesure probable et d'une erreur de discrétisation

En notation matricielle, on cherche un vecteur $\lambda \in \mathbb{R}^n$ satisfaisant

$$y = H \lambda + e$$

$$\text{où } y \text{ est le vecteur } m \times 1 : (y_1, y_2, \dots, y_m)^t \quad (1.3)$$

$$\lambda \text{ est le vecteur } n \times 1 : (\lambda_1, \lambda_2, \dots, \lambda_n)^t$$

$$e \text{ est le vecteur } m \times 1 : (e_1, e_2, \dots, e_m)^t$$

$$H \text{ est la matrice } m \times n \text{ d'élément } i, j : (\phi_i, \psi_j)$$

Il est important de rappeler que H est une matrice très grande en pratique : un cas typique en tomographie par émission est $n = 10^4$ et m encore plus grand (10^6 avec le temps de vol).

2. Exemples standard de méthodes algébriques

2.1. Pixels et techniques itératives

Un exemple courant des ψ_j est celui des indicatrices des pixels d'une grille carrée recouvrant D de pas δ , c'est-à-dire :

$$n = 4 J^2 \quad , \quad \psi_{j', j''}(x, y) = \chi_{[j'\delta, (j'+1)\delta]}(x) \chi_{[j''\delta, (j''+1)\delta]}(y)$$

$$\text{pour } j' = -J, \dots, J-1$$

$$j'' = -J, \dots, J-1$$

Cette décomposition est en effet bien adaptée à la représentation de l'estimation f sur un écran vidéo. La matrice H est très creuse mais n'a pas de réelle symétrie particulière. De nombreuses méthodes itératives ont alors été proposées pour la recherche de λ . A titre d'exemples voici les 2 méthodes les plus connues (sous les noms respectifs de ART et de SIRT), où les procédures de correction à chaque itération ont une interprétation physique intuitive (pour plus de détails, voir Y. Censor (1981)).

a) la première est la procédure de Kaczmarz qui converge vers la solution de norme minimale de $H\lambda = y$, si ce système admet une solution, pour $0 < \alpha < 2$:

$\lambda^1 \in H^t (\mathbb{R}^m)$ (l'image de H^t)

$$\lambda^{r+1} = \lambda^r + \alpha \frac{y_{i_r} - h_{i_r}^t \lambda^r}{h_{i_r}^t} h_{i_r} \quad \begin{array}{l} i_r = r \pmod{m} \\ r = 1, 2, \dots, +\infty \end{array}$$

où h_i est la i ème ligne de H .

Depuis son apparition en tomographie (une des premières méthodes utilisées commercialement) des algorithmes du même type ont été élaborés pour traiter des contraintes du type : $z_i \leq h_i^t \lambda \leq w_i$ à la place de $y_i = h_i^t \lambda$, permettant d'assurer l'existence d'une solution en prenant en compte une tolérance ϵ_i supposée.

b) la deuxième méthode est un algorithme (de Richardson) voisin de celui du gradient :

$\lambda^1 \in H^t (\mathbb{R}^m)$

$$\lambda^{r+1} = \lambda^r + \alpha H^t (y - H \lambda^r) \quad r = 1, 2, \dots, +\infty$$

qui converge, avec α suffisamment petit, vers la solution pseudo-inverse, c'est-à-dire de norme minimale parmi celles qui vérifient $H^t H \lambda = H^t y$ (précisé au § 3).

L'avantage de ces 2 méthodes est évidemment la facilité de calcul de λ^{r+1} à partir de λ^r (aucune inversion matricielle mais des produits scalaires avec des vecteurs creux que l'on peut générer en cours d'algorithme). Mais leur convergence se révèle lente et dépend beaucoup de l'initialisation λ^1 .

2.2. D'autres décompositions

Les symétries d'acquisition (cf I.B.4) qui existent dans la plupart des tomographes permettent de trouver des décompositions pour lesquelles la matrice H possède une structure très intéressante.

a) Une décomposition sur les 'pixels rayons' a été récemment étudiée en tomographie X : quand les $m = K P$ mesures d'intégration sur les bandes L_i

sont positionnées aux points $(\theta_i, u_i) = (\theta_k, u_p) = (k \frac{2\pi}{K}, p\Delta u)$, $i = (k-1)P+p$,
le choix de la décomposition :

$$n = K.P \quad \psi_j = \psi_{\ell, q} = \chi_{L_{\ell, q}} = \phi_{\ell, q} \quad (\text{cf I.B.1}) \quad \ell = 1, \dots, K$$

$$q = 1, \dots, P \quad (2.2)$$

$$j = (\ell-1)P+q$$

donne à H une structure bloc-circulante. En effet son élément (i, j) vaut :

$$(\phi_i, \psi_j) = (\chi_{L_{k, p}}, \chi_{L_{\ell, q}}) = \text{surface } (L_{k, p} \cap L_{\ell, q})$$

qui est invariant par rotation identique de $L_{k, p}$ et $L_{\ell, q}$, c'est-à-dire est fonction de $(k-\ell) \pmod{K}$.

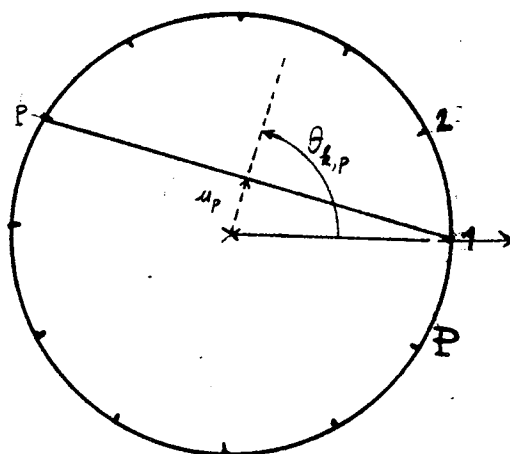
Et on montre que le calcul exact de la solution pseudo-inverse de $H\lambda = y$ nécessite alors seulement K résolutions de système d'ordre P et 2 P transformées de Fourier discrètes d'ordre K.

Après le calcul de λ , celui de $f(x, y)$ est ensuite analogue à l'épandage discret vu au A.1.2 :

$$f(x, y) = \sum_{k, p} \lambda_{k, p} \chi_{L_{k, p}}(x, y) = \sum_k \hat{\lambda}_k (x \cos \theta_k + y \sin \theta_k)$$

$$\text{où } \hat{\lambda}_k(u) = \begin{cases} \lambda_{k, p} & \text{si } |u - p\Delta u| < \Delta u/2 \\ 0 & \text{sinon} \end{cases}$$

b) Un cas de géométrie bien régulière des projections mesurées a été étudié par Robert MARR (1979): D étant le disque unité, à la circonférence duquel sont répartis uniformément P points, les $P \binom{P-1}{2}$ mesures y_i sont les intégrales de f le long des cordes joignant ces points.



Dans le repère $[-1,1] \times [0,2\pi]$, ces cordes sont décrites par les $P(P-1)$ points :

$$u_p = \cos(\pi p/P) \quad \theta_{k,p} = \frac{\pi}{P} (2k+p-2) \quad \begin{array}{l} p = 1, \dots, P \\ k = 1, \dots, P-1 \end{array}$$

et $(\phi_{k,p}, f) = (Rf)(\theta_{k,p}, u_p)$

Une base de l'espace des polynômes de degré total N ($< P-2$) sur D qui rend diagonale H_{Harr} a été mise en évidence par MARR. Ce sont les polynômes :

$$\Psi_{\ell,q}(x,y) = (x+(-1)^{\ell})^{|q|} Q_{|\ell|,q}(x^2+y^2) \quad \begin{array}{l} \ell = -\infty, \dots, 0, \dots, +\infty \\ q = 0, \dots, +\infty \end{array} \quad |q| + 2\ell \leq N \quad (2.3)$$

où les $Q_{\ell,q}$ ($q = 0, 1, \dots, +\infty$) sont les polynômes orthogonaux (de degré $q = 0, \dots, +\infty$) pour le produit scalaire ($\ell \geq 0$ fixé) sur $[0,1]$:

$$\langle f, g \rangle_{\ell} = \int_0^1 t^{\ell} f(t)g(t) dt$$

et tels que $Q_{\ell,q}(1) = 1$.

Ces polynômes peuvent être évalués numériquement par une formule de récurrence. Les 2 propriétés intéressantes des $\Psi_{\ell,q}$ sont que leurs transformées de Radon ont une expression simple :

$$R\Psi_{\ell,q}(\theta, \cos\phi) = 2(|\ell| + 2q + 1)^{-1} \sin((|\ell| + 2q - 1)\phi) e^{i\ell\theta}$$

ce qui permet de calculer les coefficients $(\phi_{k,p}, \Psi_{\ell,q})$ de H

et qu'elles sont orthogonales au sens suivant ($\delta_{\ell,\ell'} = 1$ si $\ell = \ell'$, 0 sinon) :

$$\sum_{\substack{p=1, \dots, P \\ k=1, \dots, P-1}} R\Psi_{\ell,q}(\theta_{k,p}, u_p) R\Psi_{\ell',q'}(\theta_{k,p}, u_p) = 2P^2(|\ell| + 2q + 1)^{-2} \delta_{\ell,\ell'} \delta_{q,q'}$$

ce qui signifie que $H^t H$ est diagonale.

La résolution exacte de $H^t H \lambda = H^t y$ devient alors très simple. Signalons que cette procédure ne s'applique qu'avec $N \leq P-2$. Ceci se comprend en remarquant que l'espace de ces polynômes est de dimension $n = \frac{(N+2)(N+1)}{2}$ et que $H^t H$, d'ordre n , ne pourrait être inversible avec $m = \frac{P(P-1)}{2} < n$.

Comme avec la décomposition précédente, la résolution exacte de l'équation donnant λ révèle en pratique la nécessité d'une meilleure formulation de cette équation.

3. Les formulations possibles définissant l'estimation du modèle discret

a) Si on peut raisonnablement supposer que les erreurs de mesures et de discrétisation sont négligeables, il semble légitime de chercher λ tel que le résidu $H\lambda - y$ soit le plus petit possible. Avec une norme euclidienne classique pour mesurer cette taille, il est bien connu que cela impose $H^t H \lambda = H^t y$, appelée équation normale, dont la solution d'énergie minimale (c'est-à-dire $\lambda / H^t H \lambda = H^t y$ et $\forall \mu \in \mathbb{R}^n / H^t H \mu = H^t y, \|\mu\| \geq \|\lambda\|$), appelée solution pseudo-inverse, est caractérisée par le fait qu'elle appartient à $H^t(\mathbb{R}^m)$. Or en pratique cette solution calculée exactement comme dans les 2 décompositions précédentes apparaît généralement très oscillante et numériquement instable. Ceci se retrouve dans le comportement observé des techniques itératives approchant cette solution : "faire trop d'itérations" détériore l'estimation.

Et cet 'ennui' apparaît en fait quel que soit le critère mesurant la proximité de $H\lambda$ à y . On le retrouve ainsi dans les solutions de type maximum de vraisemblance. Par exemple, en tomographie par positron (cf IB3) l'erreur de discrétisation étant généralement faible devant l'erreur poissonnienne de mesure, on peut considérer que y_i est une réalisation d'une variable aléatoire de Poisson de paramètre $h_i^t \lambda$ et la solution du maximum de vraisemblance est le vecteur λ , $\lambda_i \geq 0$, pour lequel la probabilité des réalisations conjointes y_1, \dots, y_m est maximum. (une technique itérative de calcul de cette solution est présentée par K. Lange (1984) et, pour le cas avec temps de vol, par D. Snyder (1983)).

La raison fondamentale de cet ennui est le mauvais conditionnement de la matrice H : parce que les mesures sont des moyennes globales, il existe des vecteurs μ correspondant à des fonctions très oscillantes pour lesquels toutes les mesures sont sensiblement nulles. (une étude numérique de ce mauvais conditionnement a été réalisée par J. Llacer (1982) sur des modèles réduits de tomographe standard). Une très légère modification de y dans les formulations précédentes peut alors changer complètement la solution λ en la perturbant par un tel vecteur μ .

b) Le remède bien connu à cette instabilité est l'introduction d'une information a priori sur l'estimation λ , généralement du type :

$$J(\lambda - \hat{\lambda}) \leq E \quad (1.7)$$

où $\left\{ \begin{array}{l} J(\lambda) \text{ est une forme quadratique } \lambda^t \Delta \lambda \geq 0, \text{ avec } \Delta \text{ matrice } (n,n) \\ \text{symétrique,} \\ \hat{\lambda} \text{ est une valeur 'moyenne' de } \lambda, \text{ supposée connue,} \\ \text{on connaît un ordre de grandeur de } E, \end{array} \right.$

et l'optimisation précédente du critère de proximité de $H\lambda$ à y est réalisée sous cette contrainte. Quand cette proximité est mesurée par une forme quadratique $(H\lambda - y)^t S (H\lambda - y)$, il est connu que ce nouveau problème d'optimisation conduit finalement à modifier le système linéaire donnant λ en :

$$[H^t S H + \rho \Delta] \lambda = H^t S y + \rho \Delta \hat{\lambda} \quad (1.8)$$

où le choix du paramètre réel $\rho > 0$ est équivalent à celui de E .

Cette technique est un exemple de méthodes de régularisation étudiées par Tikhonov (1977) dans un cadre abstrait.

Dans le cas de la décomposition sur les pixels carrés, des méthodes itératives ont été élaborées pour approcher une telle estimation et différents choix de ρ , S , Δ ont été expérimentalement comparés (voir E. Artzy (1979)) : ces choix se révèlent très importants pour la qualité de la reconstruction mais en même temps très dépendants aussi bien de la précision des mesures y que de la fonction f^0 cherchée.

c) Signalons que l'estimation statistique est un cadre théorique souvent utilisé pour introduire cette régularisation. Par exemple, on suppose que λ est une réalisation d'un vecteur aléatoire Gaussien de moyenne $\hat{\lambda}$, de variance la matrice V_λ et que, conditionnellement à un λ donné, y suit une loi de Gauss de moyenne $H\lambda$, de variance V_y . Alors l'estimateur Bayésien standard est donné par (1.8) où $S = V_y^{-1}$, $\rho\Delta = V_\lambda^{-1}$. Cette solution a de nombreuses propriétés statistiques (voir O.N. Strand (1968)).

Signalons que, alors que la taille du système (1.8) augmente avec n , la richesse de la décomposition, cet estimateur est aussi donné par la résolution du système $m \times m$:

$$(H V_\lambda H^t + V_y) z = y - H \hat{\lambda}$$

puisque, en multipliant les 2 termes de cette équation par $H^t V_y^{-1}$ à gauche, on obtient que $\hat{\lambda} + V_\lambda H^t z$ est solution de (1.8). Par exemple avec $V_\lambda = \tau^2 I_n$ où I_n est la matrice identité τ réel > 0 , l'estimateur Bayésien calculé sur les pixels carrés d'une grille très fine (donc n grand) a en fait une décomposition sur les 'pixels rayons' (§ 2.2) ϕ_i dont les coefficients sont calculés en inversant la matrice $HH^t + \frac{1}{\tau^2} V_y$, où $(HH^t)_{i,j} =$

$$\sum_{\ell=1}^n (\phi_i, \psi_\ell) (\psi_\ell, \phi_j) \rightarrow (\phi_i, \phi_j) \text{ quand } n \rightarrow +\infty, \text{ qui conserve donc la struc-}$$

ture bloc-circulante signalée au 2.2 si V_y la possède aussi, et peut alors être directement inversée.

Chapitre III

L'APPROCHE PROPOSEE PAR LES METHODES DE REGULARISATION OPTIMALE

A - Introduction

B - Quatre exemples de solution régularisée de paramètre ρ

1. Le critère d'écart aux mesures
2. La régularisation par la norme L_2
3. Avec l'énergie du gradient
4. Avec l'énergie de flexion
5. En dimension finie

C - Estimation de la valeur optimale de ρ en fonction des mesures

1. La méthode du résidu
2. Méthode statistique
 - 2.1. Avec l'erreur de prédiction
 - 2.2. Avec l'erreur de reconstruction
3. La méthode de validation croisée

D - Sur le choix du critère de régularisation

1. Le choix de l'espace de recherche et du critère de régularité
2. Le choix d'une translation du problème

A - INTRODUCTION

Les méthodes de reconstruction d'images actuellement utilisées dans les tomographes commercialisés sont les méthodes analytiques décrites au chapitre II.A principalement en raison de leurs performances jugées satisfaisantes dès que l'on a un assez grand nombre ($> 10^5$) de bonnes données uniformément réparties, et aussi à cause de la plus grande lenteur des méthodes itératives. Mais ces méthodes utilisent un filtre (fenêtre F_b du II.A) dont le choix est difficile à justifier alors que son adaptation à chaque application particulière est très importante pour la qualité de la reconstruction (voir par exemple T. Chang (1980)). Cet inconvénient est surtout flagrant quand les données sont bruitées comme en tomographie à positron puisque le lissage, ici indispensable, réalisé par ce filtre est choisi empiriquement alors qu'il devrait être adapté aussi bien à la précision des mesures qu'à la forme de f^0 .

L'approche des problèmes de reconstruction étudiée ici, outre le fait qu'elle peut s'appliquer à un vaste ensemble de modélisations des mesures (on pourra ainsi prendre en compte un échantillonnage moins complet, la largeur et la sensibilité de chaque détecteur), va permettre de définir cette adaptation souhaitée.

Cette approche est celle des méthodes de régularisation avec détermination du paramètre de régularisation (ou de lissage) optimal (appelées aussi méthodes d'estimation optimale par A.A. Melkman and C.A. Miccheli (1978) dans leur analyse de ses propriétés théoriques; on donne aussi le nom de fonctions splines d'ajustement aux solutions ainsi construites) :

Par exemple si on peut supposer que les erreurs de mesures sont non corrélées, de même variance σ^2 , il est naturel de dire que l'ensemble des fonctions satisfaisant les équations (formulées en I.A) données par les mesures est :

$$E_{\sigma} = \{f \in L_2(D) / \frac{1}{m} S(f) \leq \sigma^2\}$$

où

$$S(f) = \sum_{i=1}^m [(\phi_i, f) - y_i]^2$$

Or il est bien connu qu'un tel ensemble contient toujours des fonctions très oscillantes (par exemple K.T. Smith (1977)) dès qu'il est non vide (les données inexactes y_i sont presque sûrement incompatibles avec les relations de dépendance existant entre les ϕ_i et E_σ devient alors vide si σ est choisi trop petit) et cet ennui persiste même si l'on restreint E_σ à son intersection avec un sous-espace F de $L_2(\mathbb{D})$ qui doit rester raisonnablement grand pour bien modéliser f^0 (on l'a déjà signalé au II.B.3 avec F de dimension finie). On a donc besoin d'une information a priori qui consiste à préférer le moins d'oscillations possibles dans l'estimation de f^0 .

Pour cela on définit une forme quadratique J qui mesure l'irrégularité $J(f)$ de toute fonction f de F et on choisit comme estimation de f^0 la moins irrégulière des fonctions de F satisfaisant les mesures (c'est-à-dire la fonction de $E_\sigma \cap F$ pour laquelle J est minimum).

Le critère $S(f)$ mesurant l'écart de f aux mesures est discuté et l'existence, l'unicité et la caractérisation de la solution de ces problèmes d'optimisation sont étudiées avec trois types de fonctionnelles J (l'énergie (ou norme L_2), l'énergie du gradient, l'énergie de flexion) et dans le cas plus simple où F est de dimension finie (section B).

En fait excepté pour des valeurs incohérentes de σ , il existe un réel $\rho > 0$ pour lequel cet estimateur est la fonction f_ρ de F minimisant sans contrainte $J_\rho(f)$ dans F :

$$J_\rho(f) = S(f) + \rho J(f)$$

Ce paramètre de régularisation ρ est explicitement utilisé puisque le calcul de f_ρ se ramène, dans les cas étudiés, à la résolution d'un système linéaire paramétré par ρ . Et le choix d'une bonne valeur de ρ est crucial.

Dans l'exemple ci-dessus où σ est connu, ce paramètre ρ est déterminé par σ et peut être calculé en fonction des données y_i : c'est la méthode du résidu. Deux autres méthodes de choix de ρ sont aussi présentées à la section C : une méthode statistique qui nécessite de connaître (à peu près) la matrice de variance des erreurs de mesures, et qui consiste à

choisir ρ qui minimise une estimation, calculée en fonction de ρ , y et cette matrice, de la proximité de f_ρ à f^0 ; et une troisième méthode, la validation croisée qui a l'avantage de ne nécessiter aucune hypothèse statistique. (une nouvelle méthode statistique est aussi présentée).

On étudie en D une propriété, très intéressante en pratique, de ces 2 méthodes : elles permettent aussi le choix entre différentes formes quadratiques J car elles donnent accès à une mesure (plus précise pour la méthode statistique) de la qualité de la régularisation optimale faite avec J donnée.

B - QUATRE EXEMPLES DE SOLUTION REGULARISEE DE PARAMETRE ρ

1. Le critère d'écart aux mesures

Indiquons d'abord les définitions possibles de la fonctionnelle $S(f)$ mesurant l'écart (ou l'infidélité) de f aux mesures y_i , $i = 1, \dots, m$. Rappelons que nous savons que f^0 vérifie les équations :

$$y_i = (\phi_i, f^0) + \varepsilon_i \quad i = 1, \dots, m$$

où les fonctions ϕ_i et les valeurs réelles y_i sont connues, mais, suivant les cas, on ne connaît que des propriétés statistiques sur les erreurs ε_i .

Une hypothèse courante est de supposer que les ε_i sont des réalisations de variables aléatoires \mathcal{E}_i d'espérance nulle, et de matrice de variance V_ε définie positive:

$$E(\mathcal{E}_i) = 0$$

$$E(\mathcal{E}_i \mathcal{E}_j) = V_{ij} = (V_\varepsilon)_{i,j}$$

Dans ce cas il est naturel de définir $S(f)$ par :

$$S(f) = \sum_{i,j=1}^m s_{ij} ((\phi_i, f) - y_i) ((\phi_j, f) - y_j)$$

$$\text{où } s_{ij} = (S)_{ij} \text{ et } S = V_\varepsilon^{-1}$$

(avec l'hypothèse Gaussienne, $S(f)$ est la fonctionnelle minimisée dans l'estimation du maximum de vraisemblance).

Un exemple qui nous intéresse est celui du bruit Poissonien que l'on suppose en tomographie par émission (Cf I.B. 2 et 3) : alors $V_{\varepsilon} = \text{diag}(y_1^0, \dots, y_m^0)$ matrice diagonale d'éléments $v_{i,i} = y_i^0 = (\phi_i, f^0)$. y_i^0 n'est évidemment pas connu mais peut être approché par y_i et des poids s_{ij} prenant en compte ce bruit Poissonien sont :

$$s_{i,i} = \begin{cases} \frac{1}{y_i} \\ s \quad \text{arbitrairement grand si } y_i = 0 \end{cases}$$

$$s_{i,j} = 0 \quad \text{si } i \neq j$$

Dans la suite on suppose choisie la matrice S symétrique définie positive.

Notations

Nous noterons R l'opérateur de $L_2(\mathbb{D})$ dans \mathbb{R}^m qui à $f \in L_2(\mathbb{D})$ associe

$$Rf = ((\phi_1, f), (\phi_2, f), \dots, (\phi_m, f))^t$$

et y le vecteur $(y_1, y_2, \dots, y_m)^t$.

Le produit scalaire Euclidien classique dans \mathbb{R}^m de y et $z = (z_1, z_2, \dots, z_m)^t$

est noté $\langle y, z \rangle = \sum_{j=1}^m y_j z_j = y^t z$ et $\langle y, y \rangle = \|y\|^2$

et celui pondéré par S : $\langle y, z \rangle_S = y^t S z$ et $\langle y, y \rangle_S = \|y\|_S^2$.

Alors $S(f) = \left\| Rf - y \right\|_S^2$.

2. Régularisation par la norme $L_2(\mathbb{D})$

Cette première méthode consiste à chercher l'estimation de f^0 dans $F = L_2(\mathbb{D})$ et à mesurer l'irrégularité de toute solution éventuelle f par

$$J_0(f) = \|f\|_{L_2}^2 = \iint_{\mathbb{D}} f(x,y)^2 \, dx dy \quad (2.1)$$

Pour le critère d'écart aux mesures $S(f)$ défini au paragraphe précédent, le problème régularisé de paramètre $\rho > 0$ est donc par définition :

Trouver f_ρ telle que $J_\rho(f_\rho) \leq J_\rho(f)$ pour toute fonction f de $L_2(\mathbb{D})$

$$\text{où } J_\rho(f) = S(f) + \rho J(f) = \left\| Rf - y \right\|_S^2 + \rho \left\| f \right\|_{L_2}^2$$

L'existence et l'unicité du minimiseur de J_ρ sur F sont démontrées en annexe 1 où est établie aussi sa caractérisation :

$$\exists \lambda_\rho = (\lambda_1, \dots, \lambda_m)^t \quad \left\{ \begin{array}{l} \forall f \in L_2(\mathbb{D}), (f_\rho, f)_{L_2} = \langle \lambda_\rho, Rf \rangle \\ \rho \lambda_\rho = S(y - Rf_\rho) \end{array} \right. \quad (2.2)$$

En développant $\langle \lambda_\rho, Rf \rangle = \sum_{i=1}^m \lambda_i (\phi_i, f)_{L_2}$ et en le substituant dans (2.2) on en déduit que

$$(2.2) \Leftrightarrow f_\rho = \sum_{i=1}^m \lambda_i \phi_i, \text{ ce que l'on note } f_\rho = R^* \lambda_\rho \quad (2.4)$$

(R^* est en fait l'adjoint de R) et (2.3) est un système linéaire inversible donnant les λ_i :

$$\rho \lambda_\rho = S(y - R R^* \lambda_\rho) \text{ ou } (R R^* + \rho S^{-1}) \lambda_\rho = y \quad (2.5)$$

où $R R^*$ est la matrice H d'élément i, j : $H_{ij} = (\phi_i, \phi_j)_{L_2}$ $i, j = 1, 2, \dots, m$

Remarquons que (2.4) signifie que f_ρ se décompose sur les fonctions ϕ_i de mesure. On a signalé qu'une telle décomposition avait déjà été étudiée en tomographie classique (Cf II.B.2.2) où elle équivaut à un épannage classique, avec une interpolation au plus proche voisin, et que les symétries apparaissant alors dans la matrice H permettent la résolution directe du système (2.5) avec m très grand. Ce sont ces symétries qui nous permettent aussi une mise en oeuvre économique du choix du ρ optimal présentée au chapitre IV.

Cette décomposition (2.4) indique aussi que f_ρ est nulle hors de la réunion des supports des ϕ_i : comme on pouvait s'y attendre avec le critère J minimisé, l'estimation est nulle là où il n'y a pas eu de mesure. Ceci peut s'avérer indésirable et gênant quand les mesures sont incomplètes. Les trois autres régularisations présentées remédient à cet inconvénient.

On obtient une solution du même type en généralisant le choix de $L_2(\mathbb{D})$ et la norme (2.1) à tout espace de Hilbert $F, (\cdot, \cdot)_F$ sur lequel R est continu (voir F. Natterer (1980)) :

Avec $J(f) = (f, f)_F$, la seule modification dans la caractérisation (2.2) et (2.3) est $(f_\rho, f)_F$ à la place de $(f_\rho, f)_{L_2}$ (cf annexe 1) .

Par contre l'adjoint R^* n'a plus généralement une expression connue comme (2.4). Le cas $F = L_2(\mathbb{D})$ et $J(f) = \int_{\mathbb{D}} w(x,y) f(x,y)^2 dx dy$

où w est une fonction poids ($\forall (x,y) \in \mathbb{D} \quad w(x,y) > 0$) est lui facile à résoudre :

$$f_\rho = \frac{1}{w} \sum_{i=1}^m \lambda_i \phi_i$$

$$(H + \rho S^{-1}) \lambda_\rho = y \quad \text{où } h_{ij} = (\phi_i, \frac{1}{w} \phi_j)_{L_2} \quad i, j = 1, \dots, m$$

Notons que si $\{f \in L_2(\mathbb{D}) / Rf = y\}$ est non vide, son élément minimisant J est unique et a la caractérisation ci-dessus avec $\rho = 0$ (Cf Théorème 2 de l'annexe 1).

3. Avec l'énergie du gradient

Ici l'espace dans lequel on cherche l'estimation est $F = H^1(\mathbb{D})$ l'espace des fonctions de $L_2(\mathbb{D})$ dont les dérivées partielles premières (au sens des distributions) sont aussi dans $L_2(\mathbb{D})$, et l'irrégularité est mesurée par l'énergie du gradient de f :

$$J^1(f) = \iint_{\mathbb{D}} \left(\frac{\partial f}{\partial x}(x,y) \right)^2 + \left(\frac{\partial f}{\partial y}(x,y) \right)^2 dx dy \quad (3.1)$$

Rappelons que $J_0(f) + J^1(f)$ est une norme munissant $H^1(\mathbb{D})$ d'une structure d'espace de Hilbert. $H^1(\mathbb{D})$ est aussi le complété de $C^1(\mathbb{D})$, ensemble des fonctions continues sur $\bar{\mathbb{D}}$, ayant des dérivées partielles premières continues sur \mathbb{D} , pour cette norme. On peut supposer qu'il n'y a pas de fonctions constantes non nulles dans le noyau de R (c'est toujours vrai en pratique). Alors on démontre (voir annexe 1) que le problème :

trouver $f_\rho \in H^1(\mathbb{D})$ / $J_\rho^1(f_\rho) \leq J_\rho^1(f) \quad \forall f \in H^1(\mathbb{D})$

$$\text{où } J_\rho^1 = S(f) + \rho J^1(f) = \|Rf - y\|_S^2 + \rho \int_{\mathbb{D}} \left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2$$

a une solution unique caractérisée par :

$$\exists \lambda = (\lambda_1, \dots, \lambda_m) \quad \left\{ \begin{array}{l} \forall f \in H^1(\mathbb{D}) \quad \int_{\mathbb{D}} \frac{\partial f_\rho}{\partial x} \frac{\partial f}{\partial x} + \frac{\partial f_\rho}{\partial y} \frac{\partial f}{\partial y} = \langle \lambda, Rf \rangle \quad (3.2) \\ \rho \lambda = S(y - Rf_\rho) \end{array} \right.$$

En utilisant la formule de Green, on obtient (Cf N. Dyn and G. Wahba (1979)) une caractérisation de f_ρ comme solution d'équations aux dérivées partielles (où Δf est le Laplacien $\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$) avec conditions aux limites :

$$\left\{ \begin{array}{l} R^* S R f_\rho - \rho \Delta f_\rho = R^* S y \\ \text{la dérivée normale sur le bord de } \mathbb{D}, \quad \frac{\partial f_\rho}{\partial \bar{n}}, \text{ est nulle} \end{array} \right.$$

4. Avec l'énergie de flexion

Ici F est l'espace des distributions de Schwartz dont les dérivées secondes partielles sont de carré sommable dans \mathbb{R}^2 , et la mesure d'irrégularité est :

$$J_2(f) = \int_{\mathbb{R}^2} \left(\frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right)^2$$

la minimisation de $S(f) + \rho J_2(f)$ dans F a une solution unique f_ρ qui a une représentation explicite (établie par J. Duchon (1976) avec une formulation plus générale des contraintes de mesures) :

$$f_\rho(x, y) = \sum_{i=1}^m \lambda_i \xi_i(x, y) + a_0 + a_1 x + a_2 y$$

$$\text{où } \xi_i(x, y) = \iint_{\mathbb{D}} \phi_i(u, v) E(x, y, u, v) \, du dv$$

$$\text{avec } E(x, y, u, v) = d^2 \text{Log } d \quad \text{où } d = \sqrt{(x-u)^2 + (y-v)^2}$$

et les coefficients $\lambda = (\lambda_1, \dots, \lambda_m)^t$, $a = (a_0, a_1, a_2)^t$ satisfont le système linéaire :

$$\begin{cases} (K + \rho S^{-1}) \lambda + T a = y \\ T^t \lambda = 0 \end{cases}$$

où K est la matrice m, m de coefficient $i, j = (\phi_i, \xi_j)$

$$T \text{ matrice } m, 3 / T_{i,1} = \int_{\mathbb{D}} \phi_i, \quad T_{i,2} = \iint_{\mathbb{D}} x \phi_i(x, y) dx dy,$$

$$T_{i,3} = \iint_{\mathbb{D}} y \phi_i(x, y) dx dy$$

De plus si $Rf = y$ a des solutions dans F , celle qui minimise J_2 est unique et a la même caractérisation avec $\rho = 0$.

Signalons que $J_2(f)$ est une approximation de l'énergie de flexion de la plaque mince constituée des points de coordonnées $(x, y, f(x, y))$.

5. En dimension finie

Nous avons vu que les régularisations précédentes conduisent à la recherche de m coefficients λ_i . Or quand m est grand (10^6 en tomographie avec temps de vol) et que les données sont bruitées, on devine qu'un degré de liberté plus petit que m dans l'estimation cherchée serait largement suffisant. (pour une justification théorique voir F. Natterer (1977)).

On choisit donc un espace F de dimension finie n , capable de bien modéliser les éventuelles solutions et facile à utiliser, pour lequel on connaît une base de n fonctions $\psi_1, \psi_2, \dots, \psi_n$ de $L_2(\mathbb{D})$

Le critère de régularité de f dans F est alors défini par une forme quadratique, $J(f)$ donnée par sa matrice (n, n) symétrique, Δ dans cette base :

$$f = \sum_{j=1}^n \lambda_j \psi_j \Rightarrow J(f) = I(\lambda) = \lambda^t \Delta \lambda \text{ où } \lambda = (\lambda_1, \dots, \lambda_n)^t \quad (5.1)$$

avec Δ choisie semi-définie positive, i.e $\forall \lambda, I(\lambda) \geq 0$

Le choix de Δ est donc très vaste et nous verrons des exemples au chapitre V provenant de l'approximation, par une méthode d'éléments finis, des solutions régularisées définies aux paragraphes 3 et 4.

Dans ce contexte la solution régularisée f_ρ de paramètre $\rho > 0$ doit vérifier :

$$J_\rho(f_\rho) \leq J_\rho(f) \text{ pour toute } f \text{ de } F$$

$$\text{où } J_\rho(f) = \|Rf - y\|_S^2 + \rho J(f)$$

ce qui équivaut à, en écrivant $f_\rho = \sum_{j=1}^n \lambda_{\rho,j} \psi_j$ et $\lambda_\rho = (\lambda_{\rho,1}, \dots, \lambda_{\rho,n})^t$,

$$\forall \lambda \in \mathbb{R}^n, I_\rho(\lambda_\rho) \leq I_\rho(\lambda)$$

$$\text{où } \begin{cases} I_\rho(\lambda) = \|H\lambda - y\|_S^2 + \rho \lambda^t \Delta \lambda \\ H \text{ est la matrice } m, n \text{ de coefficients } H_{ij} = (\phi_i, \psi_j) \end{cases}$$

Ce problème matriciel a la solution bien connue (c'est aussi une conséquence du théorème 1 de l'annexe 1):

S'il n'y a pas de vecteur non nul commun au noyau de H et à celui de Δ alors f_ρ est unique et

$$\begin{cases} f_\rho = \sum_{j=1}^n \lambda_{\rho,j} \psi_j \end{cases} \quad (5.2)$$

$$\begin{cases} (H^t S H + \rho \Delta) \lambda_\rho = H^t S y \end{cases} \quad (5.3)$$

où $H^t S H + \rho \Delta$ est une matrice inversible.

Signalons que quand $\rho \rightarrow 0$, la solution λ_ρ du système (5.3) converge vers la solution pseudo-inverse de $H\lambda = y$ c'est-à-dire la solution qui minimise $\lambda^t \Delta \lambda$ parmi celles qui minimisent $\|H\lambda - y\|_S^2$ dans \mathbb{R}^n .

C - L'ESTIMATION DE LA VALEUR OPTIMALE DE ρ EN FONCTION DES MESURES

L'espace de recherche F et le critère de régularité J sont supposés choisis parmi les quatre types d'exemples précédents. Ce choix sera toujours fait de manière à assurer l'unicité de f_ρ (condition sur le noyau de J).

On peut vérifier que dans toutes les régularisations envisagées, f_ρ est une fonction linéaire du vecteur des données $y = (y_1, \dots, y_m)^t$ (cette propriété

découle, dans le cas général de l'annexe 1, de la linéarité de la projection (auto-adjointe) sur un sous-espace fermé dans un Hilbert). Il existe donc une matrice (m,m) $A(\rho)$ symétrique telle que :

$$Rf_{\rho} = ((\phi_1, f_{\rho}), \dots, (\phi_m, f_{\rho}))^t = A(\rho) y$$

Et il est théoriquement possible de calculer $A(\rho)$ pour un ρ fixé dans le cas de la régularisation par l'énergie (B.2), l'énergie de flexion (B.4) et en dimension finie (B.5). Il est facile de vérifier sur la définition de f_{ρ} que f_{ρ} est a fortiori la fonction minimisant $J(f)$ parmi les fonctions interpolant $A(\rho) y : \{f \in F / Rf = A(\rho)y\}$ et nous verrons que $J(f_{\rho})$ diminue quand ρ augmente.

Une régularisation de paramètre ρ peut donc être interprétée comme une transformation linéaire de y qui remplace ces données inexactes par des données $A(\rho)y$ compatibles (appartenant à l'image de R) et plus lisse (avec un lissage exprimé sur f_{ρ}) suivie d'une interpolation à J minimum. La force de ce lissage étant contrôlée par ρ , nous sommes dans le cadre d'application des méthodes de lissage optimal, précédemment étudiées pour le lissage de données par fonctions splines, principalement par C.H. Reinsch (1971), G. Wahba (1975), P. Craven et G. Wahba (1979) et généralisées par G. Wahba (1977). Avant de présenter les principes de ces méthodes, notons le résultat suivant d'algèbre statistique, nécessaire dans la suite :

Lemme Si Y est un vecteur aléatoire à valeurs dans \mathbb{R}^m , de moyenne y^0 , de variance $\text{Var}(Y) = V$,

Si A est une matrice (n,m) , z un vecteur de \mathbb{R}^n
et S est une matrice symétrique (n,n)

$$\text{alors } E (\| z + A(Y-y^0) \|_S^2) = \| z \|_S^2 + \text{Trace} (S A V A^t)$$

Démonstration : on développe, avec $S^{\frac{1}{2}}$ matrice symétrique racine carrée de S :

$$\| z + A(Y-y^0) \|_S^2 = \| z \|_S^2 + 2 z^t S A(Y-y^0) + \| S^{\frac{1}{2}} A(Y-y^0) \|^2$$

or il est connu que pour toute matrice $B (n,m)$

$$E(B(Y-y^0)) = 0 \text{ donc l'espérance du 2ème terme est nulle.}$$

et $\text{Var}(BY) = B \text{Var}(Y) B^t$,

$$\begin{aligned} \text{donc } E(\|B(Y-y^0)\|^2) &= E(\text{Trace } B(Y-y^0) [B(Y-y^0)]^t) \\ &= \text{Trace } \text{Var}(BY) \\ &= \text{Trace } B \text{Var}(Y) B^t \end{aligned}$$

avec $B = S^{\frac{1}{2}} A$ et la propriété $\text{Trace } BC = \text{Trace } CB$ facile à vérifier, on en déduit $E(\|S^{\frac{1}{2}} A(Y-y^0)\|^2) = \text{Trace}(S A V A^t)$, d'où le résultat. ■

1. La méthode du résidu

Nous avons déjà vu dans l'introduction A, que si on fait l'hypothèse d'un bruit blanc de variance σ^2 connue, une idée intuitive est de dire alors que f est suffisamment ajustée aux données si $\|Rf - y\|^2 \leq m\sigma^2$, puisque f^0 vérifie cette contrainte en moyenne ($E(\|Rf^0 - y\|^2) = m\sigma^2$).

Une généralisation de cette majoration du résidu est naturellement, si on suppose Y de moyenne Rf^0 , de variance V connue :

$$S(f) \leq \varepsilon^2 \quad \text{avec } \varepsilon^2 = \text{Trace}(S V)$$

où $S(f) = \|Rf - y\|_S^2$, puisque le lemme ci-dessus entraîne $E(S(f^0)) = \varepsilon^2$

Nous avons signalé au A que si ε est choisi trop petit ($< \varepsilon_0$) cette contrainte pouvait être irréalisable. D'autre part si ε est choisi trop grand, ($> \varepsilon_\infty$) celle-ci est vérifiée par un grand nombre de fonctions annulant J . En pratique l'hypothèse d'une valeur de ε cohérente (éventuellement déduite de l'hypothèse de V) est toujours telle que $\varepsilon_0 < \varepsilon < \varepsilon_\infty$

Notons $s(\rho) = S(f_\rho)$. On a alors :

Théorème 1 f_ρ est l'unique solution du problème : $\text{Min}_{f \in F/S(f) \leq s(\rho)} J(f)$.

la démonstration est immédiate en supposant, par l'absurde, qu'il y a une fonction $g \neq f_\rho$ telle que $S(g) \leq s(\rho)$ et $J(g) < J(f_\rho)$ ce qui entraînerait que f_ρ n'est pas minimiseur de J_ρ , et pour l'unicité, on suppose cette fois que $J(g) = J(f_\rho)$ d'où $J(g) + \rho S(g) \leq J(f_\rho) + \rho s(\rho)$ qui est en contradiction avec l'unicité du minimiseur de J_ρ . ■

Maintenant, comme nous vérifierons, dans les régularisations mises en oeuvre, que s est une fonction croissante de ρ , décrivant $]\varepsilon_0, \varepsilon_\infty[$ quand ρ décrit l'axe des réels > 0 , (et c'est généralement vrai), le problème

$$\text{Min}_{f \in F/S(f) \leq \varepsilon^2} J(f) \quad \text{où } \varepsilon \text{ est cohérent}$$

a pour solution unique :

$$f_{\rho_\varepsilon} / \rho_\varepsilon \text{ vérifie } s(\rho_\varepsilon) = \varepsilon^2$$

Cette méthode dite de résidu se formule alors par :

$$\rho_\varepsilon \text{ solution de } \|A(\rho)y - y\|_S^2 = \varepsilon^2$$

2. Méthode statistique

2.1. Avec l'erreur de prédiction

a) Définissons maintenant le paramètre optimal ρ^0 selon Wahba (1977) : ρ^0 est le minimiseur de $r(\rho)$, erreur quadratique de prédiction des données exactes :

$$r(\rho) = \frac{1}{m} \sum_{i=1}^m [(\phi_i, f_\rho) - (\phi_i, f^0)]^2 = \frac{1}{m} \|A(\rho)y - y^0\|^2$$

où $y^0 = Rf^0$ est le vecteur des données exactes (inconnues).

Comme au § 1 les modèles statistiques supposés dans la suite concernent uniquement les erreurs de mesures : y est la réalisation d'un vecteur aléatoire Y de moyenne y^0 . $\frac{1}{m} \|A(\rho)Y - y^0\|^2$ est alors une variable aléatoire dont la moyenne est fonction de ρ , y^0 et de la variance de Y .

Le paramètre optimal en moyenne, par rapport à la répartition des erreurs de mesure est :

$$\overline{\rho^0} \quad \text{qui minimise } E(r(\rho))$$

où, par abus de notation, $r(\rho)$ représente aussi la variable aléatoire fonction de Y .

Ni $r(\rho)$, ni son espérance ne peuvent être connus en pratique puisque y^0 ne l'est pas. Mais P. Craven et G. Wahba (1979) ont mis en évidence un estimateur de $E(r(\rho))$, dans le cas d'un bruit blanc de variance σ^2 connue :

$$\hat{r}(\rho) = \frac{1}{m} \| A(\rho)y - y \|^2 - \frac{\sigma^2}{m} \text{Trace } (A(\rho) - I)^2 + \frac{\sigma^2}{m} \text{Trace } (A(\rho))^2$$

qui, lui, est calculable et est un estimateur non biaisé de $E(r(\rho))$, c'est à dire :

$$E(\hat{r}(\rho)) = E(r(\rho))$$

La recherche de ρ_r minimiseur de $\hat{r}(\rho)$ constitue alors une seconde méthode de choix du ρ .

b) Présentons maintenant une généralisation immédiate de cette estimation de $E(r(\rho))$.

Théorème 2 : Si le vecteur de mesure y est de moyenne $y^0 = Rf^0$, de variance V

Si on connaît un estimateur sans biais C_y de V (i.e $E(C_y) = V$) et si l'erreur de prédiction est mesurée par

$$r(\rho) = \| A(\rho)y - y^0 \|^2_{S'}, \quad \text{où } S' \text{ est une matrice symétrique définie positive,}$$

alors

$$\hat{r}(\rho) = \| A(\rho)y - y \|^2_{S'} - \text{Trace } S' [A(\rho) - I] C_y [A(\rho) - I] + \text{Trace } S' A(\rho) C_y A(\rho)$$

est un estimateur sans biais de $E(r(\rho))$.

Démonstration : En écrivant $r(\rho) = \| A(\rho)y^0 - y^0 + A(\rho)(Y - y^0) \|^2_{S'}$, et en appliquant le lemme du début

$$E(r(\rho)) = \| A(\rho)y^0 - y^0 \|^2_{S'} + \text{Trace } S' A(\rho) V A(\rho)$$

D'autre part en écrivant :

$$s'(\rho) = \| A(\rho)Y - Y \|^2_{S'} = \| A(\rho)y^0 - y^0 + [A(\rho) - I](Y - y^0) \|^2_{S'},$$

on obtient aussi :

$$E(s'(\rho)) = \| A(\rho)y^0 - y^0 \|^2_{S'} + \text{Trace } S' [A(\rho) - I] V [A(\rho) - I]$$

De ces 2 expressions, on déduit alors, puisque $E(C_y) = V$,

$$E(s'(\rho) - \text{Trace } S' [A(\rho) - I] C_y [A(\rho) - I] + \text{Trace } S' A(\rho) C_y A(\rho)) = E(r(\rho))$$

Ce théorème démontre en particulier le cas "bruit blanc" où $C_y = V = \sigma^2 I_m$ ci-dessus. Une autre application importante est le cas "bruit poissonnien" :

où $C_y = \text{diag}(y_1, \dots, y_m)$ est une approximation connue de $V = \text{diag}(y_1^0, \dots, y_m^0)$

on a bien $E(C_y) = V$, alors le théorème nous dit, par exemple avec $S' = I_m$, que

$$\hat{r}(\rho) = \|A(\rho)y - y\|^2 + \sum_{i=1}^m y_i - 2 \text{Trace } \text{diag}(y_1, \dots, y_m) [I - A(\rho)]$$

est une estimation sans biais de $E(\|A(\rho)y - y^0\|^2)$.

Nous avons signalé au B.1 quelle matrice de pondération S on pouvait choisir dans ce contexte poissonnien : les mêmes idées sont valables pour S' dans la définition de $r(\rho)$. Dans le cas où $S' = \text{diag}(s'_1, \dots, s'_n)$ d'élément > 0 , on déduit du théorème :

$$\hat{r}(\rho) = \|A(\rho)y - y\|_{S'}^2 + \sum_{i=1}^m s'_i y_i - 2 \text{Trace } \text{diag}(s'_1 y_1, \dots, s'_m y_m) [I - A(\rho)]$$

est un estimateur sans biais de $E(\|A(\rho)y - y^0\|_{S'}^2)$

c) Signalons que dans ces estimations de $r(\rho)$, on n'a pas utilisé le fait que ρ est un paramètre contrôlant la force du lissage : elles sont généralisables à toute transformation linéaire symétrique $y \rightarrow Ay$ et permettent alors d'estimer l'aptitude de cette transformation à approcher y^0 . On peut donc imaginer que l'on peut ajuster des paramètres de cette transformation de manière à améliorer cette aptitude.

Remarquons enfin que le choix de ρ par cette méthode statistique est aussi équivalent à minimiser $t(\rho)$:

$$t(\rho) = \|A(\rho)y\|_{S'}^2 - 2 \langle A(\rho)y, y \rangle_{S'} + 2 \text{trace } S' A(\rho) C_y \quad (2.7)$$

(il est en effet facile de vérifier que $\hat{r}(\rho) - t(\rho)$ est une constante par rapport à ρ). $t(\rho)$ est en fait une estimation sans biais de $E(r(\rho)) - \|y^0\|_{S'}^2$ (facile à montrer avec le lemme).

2.2. Avec l'erreur de reconstruction

On peut préférer mesurer la proximité de f_ρ à f^0 par une distance exprimée directement sur ces 2 fonctions, plutôt que, comme précédemment, sur Rf_ρ et y^0 . J.P. Butler, J.A. Reeds et S.V. Dawson (1981) ont étudié, dans le cas particulier de la régularisation par l'énergie, l'erreur L_2 d'estimation de f^0 :

$$e(\rho) = \int_{L_2(\mathbf{D})} (f_\rho - f^0)^2 \quad (2.8)$$

En écrivant $e(\rho) = \|f_\rho\|_{L_2}^2 - 2(f_\rho, f^0)_{L_2} + \|f^0\|_{L_2}^2$ et en substituant l'expression de $f_\rho = R^*(H + \rho S^{-1})^{-1}y$ où $H = RR^*$ (cf B.1. équations (2.4) et (2.5)), ils ont observé que $e^*(\rho) = e(\rho) - \|f^0\|_{L_2}^2$ est calculable en fonction du vecteur $\varepsilon = y - y^0$:

$$e^*(\rho) = y^t \Lambda_\rho H \Lambda_\rho y - 2 y^t \Lambda_\rho y + 2 y^t \Lambda_\rho \varepsilon \quad (2.9)$$

où $\Lambda_\rho = (H + \rho S^{-1})^{-1}$

Avec l'hypothèse du bruit blanc de variance σ^2 connue, il est alors facile d'expliciter la valeur maximum $h(\rho)$ de cette erreur par rapport aux erreurs possibles ε telles que $\|\varepsilon\|^2 \leq m \sigma^2$:

$$h(\rho) = y^t \Lambda_\rho H \Lambda_\rho y - 2 y^t \Lambda_\rho y + 2 \sigma \sqrt{m} \|\Lambda_\rho y\|$$

Les auteurs cités ont proposé de choisir comme paramètre ρ la valeur qui minimise cette majoration $h(\rho)$ (leur analyse des variations de cette fonction leur a permis de montrer l'unicité de ce minimum et son expression explicite). En fait, en pratique, cette méthode est équivalente à celle du résidu.

Ici nous généralisons cette définition de $e(\rho)$ et nous présentons un estimateur sans biais de $e(\rho)$, à une constante près, déduit d'une expression généralisant (2.9).

Définissons l'erreur vraie d'estimation de f^0 par f_ρ (ou erreur de reconstruction) par :

$$e(\rho) = J(f_\rho - f^0) \quad (2.10)$$

A l'aide du théorème de caractérisation de f_ρ , il est facile de montrer (cf annexe 1, § 4) que pour toute fonction f de F :

$$J(f_\rho - f) - J(f) = \left\langle \frac{1}{\rho} S(I-A(\rho))y, A(\rho)y \right\rangle - 2 \left\langle \frac{1}{\rho} S(I-A(\rho))y, Rf \right\rangle \quad (2.11)$$

On en déduit avec $f = f^0$ que $e^*(\rho)$ défini par $e^*(\rho) = e(\rho) - J(f^0)$, ne dépend de f^0 qu'à travers ε comme (2.9).

On a alors l'estimateur sans biais $\hat{e}^*(\rho)$ de $e^*(\rho)$ suivant :

Théorème 3 : Avec les mêmes hypothèses que le théorème 2,

$\hat{e}^*(\rho)$ défini par :

$$\hat{e}^*(\rho) = \frac{1}{\rho} \left[3 \langle A(\rho)y, y \rangle_S - \|A(\rho)y\|_S^2 - 2\|y\|_S^2 + 2 \text{trace } S(I-A(\rho))C_y \right] \quad (2.12)$$

est un estimateur sans biais de $E(e^*(\rho))$

Démonstration

L'espérance du dernier terme de (2.11) où $f = f^0$ est $2 \left\langle \frac{1}{\rho} S(I-A(\rho))y^0, y^0 \right\rangle$

D'autre part le lemme du début nous permet d'écrire :

$$E\left(\left\langle \frac{1}{\rho} S(I-A(\rho))y, y \right\rangle\right) = \left\langle \frac{1}{\rho} S(I-A(\rho))y^0, y^0 \right\rangle + \text{Trace } \frac{1}{\rho} S(I-A(\rho))V$$

Par combinaison de ces 2 relations, on déduit :

$$E\left(\left\langle \frac{1}{\rho} S(I-A(\rho))y, A(\rho)y \right\rangle - 2 \left\langle \frac{1}{\rho} S(I-A(\rho))y, y \right\rangle + 2 \text{Trace } \frac{1}{\rho} S(I-A(\rho))C_y\right) = E(e^*(\rho)) \quad \blacksquare$$

Le choix de ρ_e minimiseur de $\hat{e}^*(\rho)$ constitue alors sans doute une bonne méthode d'approximation de minimiseur de $e(\rho)$ défini par (2.10).

Pour pouvoir comparer cette nouvelle variante de la méthode statistique à

la précédente (minimiseur de $\hat{r}(\rho)$), il serait intéressant d'analyser les variances respectives des deux estimateurs utilisés.

Signalons enfin une autre approche théorique développée par J.W. Hilgers (1982) en dimension finie, qui consiste à évaluer la qualité de f_ρ par une distance (norme quadratique) entre f_ρ (que l'on note $f_{\rho,y}$) et f_{o,y^0} solution pseudo-inverse avec les données exactes y^0 à la place de y . On devine que cette distance s'exprime alors en fonction de ε . Une étude analytique de cette fonction pour $\|\varepsilon\|$ petit lui permet d'obtenir des estimateurs de celle-ci mais qui semblent difficiles à utiliser en pratique puisqu'ils s'expriment en fonction de $Rf_{o,y}$ d'évaluation numérique instable.

Concluons ces estimations d'erreurs avec une remarque concernant la convergence avec ε et m de ces solutions régularisées : dans le cas où les mesures sont des échantillons de la transformée de Radon (cf II.A.1) alors, par exemple :

si $f^0 \in H^1(D)$, $f_{c,\varepsilon,y}$ où $F = H^1(D)$, $J = J_0 + J^1$ (cf III.B.3) converge vers f^0 (au sens de l'erreur (2.8)) quand $\varepsilon = \|y - y^0\| \rightarrow 0$ et $m \rightarrow \infty$ avec un échantillonnage uniforme..

La convergence est assurée quel que soit la constante c mais le choix de celle-ci (ou ρ) reste crucial en pratique où ε est > 0 et m fini.

C'est un des résultats de majoration d'erreur présentés par F. Natterer (1980.a).

3. La méthode de validation croisée

Cette troisième méthode a l'avantage de ne nécessiter aucune connaissance statistique sur les erreurs. Elle consiste à choisir ρ_v qui minimise :

$$v(\rho) = \frac{\frac{1}{m} \| A(\rho)y - y \|^2}{\left| \frac{1}{m} \text{Trace} (I - A(\rho)) \right|^2} \quad (3.1)$$

Ce critère, dit de validation croisée généralisée, est dû à Wahba (1977) qui a aussi montré qu'avec principalement l'hypothèse d'un bruit blanc et $J(f^0)$ raisonnable, ρ_v est un estimateur de $\overline{\rho_0}$, minimiseur de $E(r(\rho)) = E(\| A(\rho)y - y^0 \|^2)$, d'autant meilleur que m est grand, au sens où :

$$\lim_{m \rightarrow +\infty} \frac{E(r(\rho_v))}{E(r(\overline{\rho_0}))} = 1$$

Cette méthode est d'usage courant pour le lissage automatique (car non paramétré) de données par fonctions spline (elle est implémentée dans la bibliothèque scientifique IMSL (1980)). Et elle a en fait pour origine une idée simple due à M. Stone (1974) : elle consiste à laisser de côté une donnée et à étudier la validité de différents ρ par l'aptitude des estimateurs f_ρ associés à prédire la donnée non utilisée. On choisit alors le ρ qui assure la meilleure prédiction en moyenne sur les m séparations d'une donnée possibles.

La dérivation de (3.1) à partir de cette idée est intéressante à construire : soit f_ρ^k l'estimation obtenue quand on laisse de côté la donnée y_k , i.e.

$$f_\rho^k \text{ minimiseur de } \sum_{\substack{i=1 \\ i \neq k}}^m ((\phi_i, f) - y_i)^2 + \rho J(f) \quad (3.2)$$

La quantité $((\phi_k, f_\rho^k) - y_k)^2$ mesure l'aptitude de f_ρ^k à prédire y_k

Quand toutes les données interviennent de manière symétrique, c'est-à-dire plus précisément qu'une permutation circulaire des coefficients de y ne modifie le vecteur lissé $A(\rho)y$ que par cette même permutation (i.e. $A(\rho)$

est circulant), le ρ optimum au sens de la validation est celui, suivant l'idée de M. Stone, qui minimise :

$$V(\rho) = \frac{1}{m} \sum_{k=1}^m ((\phi_k, f_{\rho}^k) - y_k)^2 \quad (3.3)$$

Heureusement l'évaluation de $V(\rho)$ ne nécessite pas la résolution des m sous-problèmes (3.2) mais seulement la connaissance de $A(\rho)$. Pour le montrer, il faut d'abord observer que f_{ρ}^k est aussi la solution du problème initial avec (ϕ_k, f_{ρ}^k) à la place de y_k . On écrit alors que la $k^{\text{ième}}$ valeur prédite par f_{ρ}^k est la $k^{\text{ième}}$ ligne du produit de $A(\rho) = (a_{k,\ell})_{\substack{k=1,\dots,m \\ \ell=1,\dots,m}}$ par ce nouveau vecteur de données :

$$(\phi_k, f_{\rho}^k) = \sum_{\substack{\ell=1 \\ \ell \neq k}}^m a_{k,\ell} y_{\ell} + a_{kk} (\phi_k, f_{\rho}^k)$$

ce qui s'écrit aussi après une simple manipulation algébrique :

$$(1 - a_{kk}) [(\phi_k, f_{\rho}^k) - y_k] = \left[\sum_{\ell=1}^m a_{k\ell} y_{\ell} \right] - y_k$$

Or nous sommes dans le cas où $A(\rho)$ est circulante, donc $1 - a_{kk} = \frac{1}{m} \text{Trace}(I - A(\rho))$ et en sommant ces égalités élevées au carré pour $k = 1, \dots, m$, on obtient $V(\rho)$:

$$V(\rho) = \frac{\frac{1}{m} \|[I - A(\rho)]y\|^2}{\left[\frac{1}{m} \text{Trace } I - A(\rho)\right]^2} = v(\rho)$$

Dans le cas où il n'y a pas cette symétrie dans les données, on peut théoriquement circulariser le problème, c'est-à-dire qu'il existe une matrice $Q = (q_{i,j})_{\substack{i=1,\dots,m \\ j=1,\dots,m}}$ unitaire telle que :

$$\tilde{\phi}_i = \sum_{j=1}^m q_{ij} \phi_j$$

$$\tilde{y} = Q y$$

donnent mis à la place des ϕ_i et y dans J_{ρ} un nouveau problème équivalent (car Q conserve la norme) pour lequel $\tilde{A}(\rho)$ est circulante ($\tilde{A}(\rho)$ est par

définition la matrice / $((\overset{\sim}{\phi}_1, \tilde{f}_\rho), \dots, (\overset{\sim}{\phi}_m, \tilde{f}_\rho))^t = \tilde{A}(\rho) \tilde{y}$).

L'idée de Stone (3.3) s'y applique alors et consiste ici à minimiser :

$$\tilde{V}(\rho) = \frac{\frac{1}{m} \| (I - \tilde{A}(\rho)) \tilde{y} \|^2}{\left(\frac{1}{m} \text{Trace } I - \tilde{A}(\rho) \right)^2} .$$

Or un simple calcul algébrique permet de vérifier que $\tilde{V}(\rho)$ est en fait indépendant de la matrice Q puisque $\tilde{V}(\rho) = v(\rho)$.

D - LE CHOIX DES CRITERES DE REGULARISATION

Maintenant que nous avons vu comment, pour F et J donnés, réaliser une régularisation optimale, on peut se demander comment choisir F et surtout J. Un choix supplémentaire offert est celui de la translation du problème, c'est-à-dire de remplacer, dans $J_\rho(f)$ minimisé, $J(f)$ par $J(f - \bar{f})$ où \bar{f} est une fonction connue que l'on suppose a priori proche de f^0 .

1. Le choix de l'espace de recherche et du critère de régularité

Dans le cas où la variance du bruit est approximativement connue (par la matrice C_y du C.2) l'estimateur $\hat{r}(\rho)$ de l'erreur de prédiction des données exactes, donné au théorème 2 du C.2, est évidemment fonction de F et J à travers $A(\rho)$, notons le $\hat{r}_{F,J}(\rho)$ pour indiquer cette dépendance et notons $\hat{r}_{F,J}$ sa valeur minimale associée à la régularisation optimale suivant ce critère :

$$\hat{r}_{F,J} = \inf_{\rho > 0} \hat{r}_{F,J}(\rho) \quad (1.1)$$

Sans chercher évidemment à optimiser F, J, on peut choisir entre quelques possibilités celle pour laquelle $\hat{r}_{F,J}$ est le plus petit.

Par contre on ne peut utiliser l'estimation $\hat{e}^*(\rho) + J(f^0)$ de l'erreur de reconstruction définie au théorème 3 que pour le choix de F en prenant pour critère que

$$\hat{e}_{F,J}^* = \inf_{\rho > 0} \hat{e}_{F,J}^*(\rho) \quad (1.2)$$

soit le plus petit.

Si la variance du bruit est inconnue, la méthode de validation croisée permet encore de faire un tel choix : nous avons vu que $v(\rho)$ mesure une moyenne des aptitudes de f_ρ^k à prévoir la $k^{\text{ième}}$ donnée non utilisée (cf C(3.3)) ; or cette aptitude est aussi fonction évidemment de F et de J ; notons-la $v_{F,J}(\rho)$. Entre plusieurs couples F, J en compétition, il semble alors naturel de choisir celui qui est associé à la plus

petite valeur de $v_{F,J}$ où

$$v_{F,J} = \inf_{\rho > 0} v_{F,J}(\rho) \quad (1.3)$$

Cette idée a été proposée par G. Wahba (1977) et on peut en outre la justifier par la propriété suivante démontrée par P. Craven et G. Wahba (1979) sur un cas simple et avec l'hypothèse d'un bruit blanc de variance σ^2 :

$$E(v_{F,J}(\rho)) \approx E(r_{F,J}(\rho)) + \sigma^2$$

2. Le choix d'une translation du problème

Ce nouveau paramètre peut être pris en compte facilement dans les estimations (1.1) et (1.3). En effet le problème translaté, par \bar{f} , c'est-à-dire

$$\text{Min}_{f \in F} \left\| Rf - y \right\|_S^2 + \rho J(f - \bar{f})$$

est équivalent au problème que l'on sait résoudre, où $\bar{y} = R\bar{f}$:

$$\text{Min}_{f \in F} \left\| Rf - (y - \bar{y}) \right\|_S^2 + \rho J(f)$$

au sens où la solution \bar{f}_ρ du premier se déduit de celle $f_{\rho, y-\bar{y}}$ du second par :

$$\bar{f}_\rho = \bar{f} + f_{\rho, y-\bar{y}}$$

Alors l'erreur de prédiction des données exactes par \bar{f}_ρ est :

$$r(\rho) = \left\| R \bar{f}_\rho - y^0 \right\|^2 = \left\| A(\rho)(y-\bar{y}) - (y^0 - \bar{y}) \right\|^2$$

donc $\hat{r}(\rho)$ et $v(\rho)$ se déduisent des expressions précédentes (cf C(2.4) et (3.1)) en remplaçant y par $y - \bar{y}$. Notons $\hat{r}_{F,J,\bar{y}}(\rho)$, $v_{F,J,\bar{y}}(\rho)$ ces nouvelles fonctions. Leur valeur minimale par rapport à ρ est une estimation calculable de la qualité d'un triplet F, J, \bar{y} et est donc un critère de choix.

Mais ceci n'est pas généralisable à l'estimateur (1.2) puisque l'erreur de reconstruction est maintenant:

$$e(\rho) = J(\bar{f}_\rho - f^0) = J(f_{\rho, y-\bar{y}} - (f^0 - \bar{f}))$$

et c'est la quantité $e^*(\rho) = e(\rho) - J(f^0 - \bar{f})$ que l'on sait estimer : son estimateur s'obtient en remplaçant y par $y-\bar{y}$ dans son expression $e^*(\rho)$ du théorème 3. Sa valeur minimum $\hat{e}_{F, J, \bar{y}}^*$ est donc calculable, mais c'est la valeur de

$$e_{F, J, R\bar{f}}^* + J(f^0 - \bar{f})$$

qui définit maintenant la qualité du triplet F, J, \bar{f} et son second terme n'est généralement pas connu.



Chapitre IV

REGULARISATION PAR LA NORME L_2 AVEC PARAMETRE ρ OPTIMAL : ALGORITHME DIRECT DE CALCUL DE LA SOLUTION D'ENERGIE MINIMALE SOUS CONTRAINTE OPTIMALE

A - Introduction

B - Mise en oeuvre numérique en tomographie classique

1. Les symétries

2. Le calcul de f_ρ

3. Le calcul du paramètre ρ optimal

4. Le cas du bruit poissonnien

5. Généralisation de ce type d'algorithmes à d'autres tomographes

C - L'efficacité du choix automatique du paramètre ρ

A. INTRODUCTION

Dans ce chapitre on étudie la mise en oeuvre numérique du calcul pour ρ fixé de la solution d'énergie minimale f_ρ définie au III.B.2, et de l'estimation du paramètre ρ optimal par l'une des trois méthodes décrites au III.C.

Rappelons que f_ρ est la fonction de $L_2(\mathbb{D})$ qui minimise

$$J_\rho(f) = \sum_{i=1}^m \left[\int_{\mathbb{D}} \phi_i f - y_i \right]^2 + \rho \int_{\mathbb{D}} f^2 \quad (1)$$

et nous avons montré au III.B.2. que f_ρ est combinaison linéaire des ϕ_i $i = 1, \dots, m$:

$$f_\rho = \sum_{i=1}^m \lambda_{\rho,i} \phi_i \quad (2)$$

avec un vecteur de coefficients, $\lambda_\rho = (\lambda_{\rho,1}, \dots, \lambda_{\rho,m})^t$, donné par le système linéaire :

$$(H + \rho \cdot I) \lambda_\rho = y \quad (3)$$

$$H \text{ matrice } m \times m \text{ de coefficient } i, j \quad H_{i,j} = \int_{\mathbb{D}} \phi_i \phi_j \quad (4)$$

(dans le cas où la somme dans $J_\rho(f)$ est pondérée par une matrice S , I est remplacée par S^{-1}).

Nous voyons donc que la reconstruction d'une image par cette méthode se ramène à 3 étapes (après avoir choisi ρ) :

- . calcul de H pour construire le système linéaire (3)
- . résolution de ce système donnant λ_ρ
- . évaluation de f_ρ à partir de (2) sur une grille uniforme

Nous avons signalé au chapitre I que typiquement m est de l'ordre de 10^4 en tomographie classique à positrons et 10^6 avec le temps de vol : les 2 premières étapes ne sont donc pas directement envisageables sur un ordinateur actuel.

Heureusement la plupart des tomographes ont une géométrie possédant une certaine invariance par rotation qui permet une décomposition de H simplifiant énormément ces 2 étapes (on est typiquement ramené à des matrices d'ordre $m/100$).

Cette symétrie a déjà permis la mise en oeuvre du calcul de f_{ρ} dans le cas de la tomographie classique où la 3ème étape est équivalente à un épandage déjà vu au II.A.2. F. Natterer (1980) a montré qu'une telle reconstruction est toujours au moins aussi performante que la méthode analytique standard (cf II.A.2) quand chacune utilise de 'bons' paramètres, et qu'elle a un coût de calcul compétitif quand ces paramètres sont préalablement fixés.

On a signalé la nécessité d'une adaptation de paramètre ρ dépendante non seulement du niveau de bruit des données mais aussi des données elles-mêmes. Et nous avons vu au III.C qu'une bonne valeur de ρ peut théoriquement être calculée en fonction de y et éventuellement de sa variance supposée.

Un des principaux résultats obtenus ici est la démonstration que la mise en oeuvre de ces méthodes de calcul de ρ est en fait très peu coûteuse.

On présente ici un algorithme qui détermine la valeur ρ^* optimale suivant l'un des 3 critères du III.C et calcule f_{ρ^*} pour un coût total (temps de calcul et place en mémoire) presque identique au coût minimum avec ρ fixé a priori.

Le cas du bruit poissonnien (cf III.C.2) est aussi pris en compte. Nous regardons ensuite comment cet algorithme est généralisable à la tomographie avec temps de vol.

Des simulations numériques sont présentées en C sur un exemple type de tomographie classique (96 x 128 données). Elles révèlent l'efficacité, assez impressionnante, de la méthode de validation croisée même avec un bruit non blanc, par exemple poissonnien, où la méthode statistique améliore seulement légèrement l'estimation du paramètre optimal ρ^0 (défini au III.C.2).

B. MISE EN OEUVRE NUMERIQUE EN TOMOGRAPHIE CLASSIQUE

1. Les symétries

Pour fixer les idées, les tomographes envisagés sont d'abord ceux décrits au I.B.1 et 2 avec $\phi_i = \chi_{L_i}$, $i = 1, \dots, m$, où

$$L_i = \left\{ (x, y) \in \mathbb{D} / r_i - \frac{d}{2} < x \cos \theta_i + y \sin \theta_i < r_i + \frac{d}{2} \right\}$$

et la répartition uniforme des bandes étroites L_i suivante :

$$m = K \cdot P \quad d = \frac{1}{P} \quad \text{et } (\theta_i, r_i) = \left(\frac{(k-1)2\pi}{K}, -\frac{d}{2} + p \cdot d \right) \quad \begin{array}{l} i=(k-1)P+p \\ k=1, \dots, K \text{ et } p=1, \dots, P \\ i=1, \dots, m \end{array} \quad (1.1)$$

Ainsi L_i est noté $L_{k,p}$, le premier indice caractérisant sa direction, le second sa distance au centre o , et les P premières bandes $L_{1,p}$ forment une partition de la moitié du disque unité \mathbb{D} et sont parallèles à l'axe Oy et toutes les autres $L_{k,p}$ sont obtenues par rotation successive de $L_{1,p}$ autour de o d'angle $\frac{(k-1)2\pi}{K}$.

Avec un arrangement lexicographique des $L_{k,p}$, la matrice H (définie par A(4)) a alors la structure bloc-circulante suivante :

$$H = \begin{bmatrix} H_{1,1} & - & - & H_{1,K} \\ H_{2,1} & & & H_{2,K} \\ \vdots & & & \vdots \\ H_{K,1} & - & - & H_{K,K} \end{bmatrix} = \begin{bmatrix} H_1 & & & H_K \\ H_K H_1 & & & H_{K-1} \\ & \ddots & & \\ H_2 & \dots & H_K H_1 & \end{bmatrix} \quad \text{où } H_k \text{ matrice } P \times P \quad (1.2)$$

puisque $(H_{k,l})_{p,q} = \int_{\mathbb{D}} \phi_{k,p} \phi_{l,q}$ est simplement la surface de l'inter-

section $L_{k,p} \cap L_{l,q} \cap \mathbb{D}$ et que cette surface est invariante si on ajoute modulo K un même entier à k et l . De plus, on vérifie facilement que :

$$H_k = H_{K+2-k} \quad \text{et } H_k \text{ symétrique} \quad (1.3)$$

Or il est connu qu'une telle matrice a la bloc-diagonalisation suivante : notons W la matrice de transformée de Fourier discrète (T.F.D) d'ordre K

$$(W \text{ a pour coefficients } W_{k,l} = e^{-\frac{2\pi(k-1)(l-1)i}{K}})$$

Notons E la matrice KP x KP de permutation définie par :

$$E y = (y_{1,1}, y_{2,1}, \dots, y_{K,1}, \dots, y_{1,P}, \dots, y_{K,P})$$

alors

$$W_v = E^t \begin{bmatrix} W & & o \\ & W & \\ o & & \ddots \\ & & & W \end{bmatrix} E \quad \text{est la matrice telle que } W_v y \text{ est obtenu}$$

par T.F.D des P_i colonnes du tableau

$$(y_{k,p}) \quad \begin{matrix} k = 1, \dots, K \\ p = 1, \dots, P \end{matrix}$$

et on a : $H = \frac{1}{K} \bar{W}_v \begin{bmatrix} M_1 & & o \\ & M_2 & \\ o & & \ddots \\ & & & M_K \end{bmatrix} W_v$ (1.4) (où $\bar{W}_v = W_v^{-1}$ est la matrice

conjuguée de W_v)

où les M_k sont des matrices $P \times P$ définies par

$$\begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_K \end{bmatrix} = W_v \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_K \end{bmatrix} \quad (1.5)$$

2. Le calcul de f_ρ

Alors de la décomposition (1.4) on déduit l'expression de λ_ρ solution de A(3).

$$\lambda_\rho = \frac{1}{K} \bar{W}_v \begin{bmatrix} (M_1 + \rho I)^{-1} \\ & (M_2 + \rho I)^{-1} \\ & & \ddots \\ & & & (M_K + \rho I)^{-1} \end{bmatrix} W_v y$$

Cette expression est à la base de l'algorithme, déjà connu, de calcul exact de λ_ρ , qui se décompose en :

- (a) calcul des blocs M_k par (1.5)
- (b) calcul de $W_v y$
- (c) résolution des K systèmes linéaires d'ordre P (inversion de $M_k + \rho I$)
- (d) multiplication par \bar{W}_v

En fait parce que la taille totale des blocs M_k n'est pas énorme, on a intérêt à effectuer l'étape (a), indépendante des données, au préalable et à mémoriser ces M_k . Ce précalcul nécessite celui des P premières lignes de H (calcul de surfaces élémentaires) puis P^2 T.F.D d'ordre K suivant (1.5). La mémorisation est celle de $\frac{K}{2}$ matrices réelles symétriques puisque les M_k sont réelles, symétriques et $M_k = M_{K+2-k}$ d'après (1.3).

Si en outre ρ est fixé a priori, le calcul le plus rapide de λ_ρ est obtenu en mémorisant $(M_k + \rho I)^{-1}$ au lieu de M_k et se réduit à 2 P T.F.D d'ordre K (étapes b et d) et K multiplications par une matrice d'ordre P. C'est l'algorithme proposé par F. Natterer (1980) qui signale que le coût de ce calcul est alors du même ordre que celui de l'épandage de λ_ρ , étape finale donnant f_ρ . Un tel épandage apparaissant aussi dans la méthode analytique standard (cf II.A.1.2), le coût de cette régularisation à ρ fixé s'avère compétitif.

Remarquons que pour conserver une telle réduction avec une matrice de pondération S différente de I, il est nécessaire que S soit aussi bloc-circulante comme H, fixé a priori (les blocs mémorisés sont dans ce cas $(M_k + \rho S_k^{-1})^{-1}$, S_k étant défini par (1.4) avec S, S_k à la place de H, M_k). On peut ainsi prendre en compte une sensibilité des détecteurs supposée connue et invariante par rotation.

3. Le calcul du paramètre ρ optimal

La méthode du résidu, les deux méthodes statistiques et la méthode de validation croisée définies au III.C nécessitent soit, pour la première, le calcul du zéro soit la minimisation des fonctions de ρ respectives $s(\rho) - \varepsilon^2$, $\hat{r}(\rho)$, $\hat{e}^*(\rho)$ et $v(\rho)$ dépendantes de y et suivant les cas de S, S', C_y .

Pour ne pas trop compliquer, nous regardons d'abord le cas $S = S' = I$ et $C_y = \sigma^2 I$. L'évaluation de ces fonctions se ramène alors à celles de $\|y - A(\rho)y\|$ et Trace $(I - A(\rho))$. Nous allons voir comment l'évaluation de ces expressions peut être rapide, permettant alors d'approcher ce zéro ou ce minimum par une méthode itérative.

La substitution de l'expression de f_ρ (A (2) et (3)) dans la définition de $A(\rho)$ ($A(\rho)y = R f_\rho$) donne immédiatement :

$$A(\rho) = H (H + \rho I)^{-1} \quad (3.1)$$

On a donc :

$$y - A(\rho)y = \rho (H + \rho I)^{-1} y = \rho \lambda_\rho$$

Mais répéter, avec différents ρ , le calcul de λ_ρ est beaucoup plus

coûteux que le calcul de sa seule norme suivant l'algorithme suivant. Celui-ci repose sur la possibilité, grâce aux symétries, de calculer une diagonalisation de H donc aussi de A(ρ), à partir de (1.4).

En effet chaque bloc M_k , réel et symétrique, est diagonalisable par transformation orthogonale (c'est-à-dire $V_k V_k^t = V_k^t V_k = I$):

Soit $M_k = V_k D_k V_k^t$ où $D_k = \text{diag}(d_{k,1}, \dots, d_{k,p})$ cette diagonalisation.

En substituant ces K expressions dans (1.4) on obtient une diagonalisation de H :

$$H = \bar{U}^t D U \quad (3.2) \quad \text{où } U = \frac{1}{\sqrt{K}} \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_K \end{bmatrix} \quad W_V \quad (3.3) \quad \text{et } D = \begin{bmatrix} D_1 \\ & D_2 \\ & & \ddots \\ & & & D_K \end{bmatrix} \quad (3.4)$$

et donc d'après (3.1) et $U^{-1} = \bar{U}^t$

$$I - A(\rho) = \bar{U}^t (I - D(D + \rho I)^{-1}) U$$

d'où la trace de I-A(ρ) : $\text{Trace } I - A(\rho) = \sum_{\substack{k=1, K \\ p=1, P}} \left(1 - \frac{d_{k,p}}{d_{k,p} + \rho}\right) \quad (3.5)$

et en notant :

$$\tilde{y} = U y \quad ,$$

on a :

$$\| [I - A(\rho)] y \|^2 = \sum_{\substack{k=1, K \\ p=1, P}} \left(1 - \frac{d_{k,p}}{d_{k,p} + \rho}\right)^2 |\tilde{y}_{k,p}|^2 \quad (3.6)$$

Comme d'autre part, on a aussi :

$$\lambda_\rho = \bar{U}^t (D + \rho I)^{-1} \tilde{y} \quad , \quad (3.7)$$

on en déduit que le calcul du paramètre ρ^* optimal et de la solution associée λ_{ρ^*} se décompose en 3 étapes, après avoir préalablement calculé et mémorisé les matrices V_k et les valeurs propres $d_{k,p}$ (au lieu, comme en 2, de mémoriser $(M_k + \rho I)^{-1}$) :

- (a) calcul de $\tilde{y} = U y$ (U défini par (3.3)) (P T.F.D d'ordre $K + KP^2$ opérations)
- (b) calcul de ρ^* : par exemple par la méthode itérative de la section dorée pour minimiser $v(\rho)$ qui s'évalue suivant (3.5) et (3.6) en $3 K P$ opérations

- c) calcul de λ_{ρ} suivant (3.7) (coût identique à (a))
 (où 1 opération = 1 addition + 1 multiplication)

Comme en pratique une vingtaine d'itérations suffisent largement en b), le temps de calcul est globalement inférieur à deux fois le temps minimum (cf 2) nécessaire au calcul d'un seul λ_{ρ} .

Remarques

- a) On pourrait montrer que si $S = S'$ et S fixé a priori bloc-circulante, on a encore un algorithme économique du même type où seules changent les matrices V_k et D_k .

Sans rentrer dans les détails, signalons seulement que si S est associée aux blocs S_1, \dots, S_k (par une relation analogue à (1.4)) les K diagonalisations orthogonales à calculer sont alors celles de :

$$S_k \frac{1}{2} M_k S_k \frac{1}{2} = U_k D'_k U_k^T$$

et ce sont les K produits $S_k \frac{1}{2} U_k$ et les diagonales D'_k qui sont mémorisées et utilisées à la place des V_k et D_k dans l'algorithme précédent ((3.6) et (3.5) calculent ici $\| [I-A(\rho)]y \|^2$ et $\text{Trace}[S(I-A(\rho))]$).

- b) L'expression (3.6) indique immédiatement le comportement de $s(\rho) = \| (I-A(\rho))y \|^2$ en fonction de ρ :

$s(\rho)$ est une fonction croissante sur $]0, +\infty[$

$$s(\rho) \xrightarrow{\rho \rightarrow 0} \varepsilon_0^2 = \sum_{\substack{k=1, K \\ p=1, P}} |\tilde{y}_{kp}|^2 \quad \text{et} \quad s(\rho) \xrightarrow{\rho \rightarrow +\infty} \varepsilon_\infty^2 = \sum_{\substack{k=1, K \\ p=1, P}} |\tilde{y}_{kp}|^2$$

$/d_{kp} = 0$

On pourrait facilement montrer que $\varepsilon_0^2 = \text{Min}_{f \in L_2} \|y - Rf\|^2$ et $\varepsilon_\infty^2 = \|y\|^2$

4. Le cas du bruit poissonnien

Nous regardons ici comment $r(\rho)$ et $e^{\mathbf{x}}(\rho)$ peuvent être calculés en fonction de \mathbf{y} et C_y , l'approximation de la variance de \mathbf{y} dans le cas où C_y est diagonale. L'exemple type est le bruit poissonnien :

$$C_y = \text{diag}(y_{1,1}, \dots, y_{1,P}, \dots, y_{K,1}, \dots, y_{K,P}) \quad (4.1)$$

Supposons $S = S' = I$; alors la nouvelle quantité à évaluer pour différents ρ est :

$$\text{Trace } C_y (I - A(\rho))$$

Or, avec U et D définies par (3.3) et (3.4), on a :

$$\begin{aligned} \text{Trace } C_y (I - A(\rho)) &= \text{Trace } (U C_y U^{-1} U (I - A(\rho)) U^{-1}) \\ &= \text{Trace } L (I - D(D + \rho I)^{-1}) \end{aligned}$$

$$\text{où } L = U C_y U^{-1}$$

Cette Trace s'exprime donc en fonction des coefficients diagonaux de L :

$$L_{k,p} \quad \begin{array}{l} k = 1, \dots, K \\ p = 1, \dots, P \end{array}$$

puisque

$$\text{Trace } C_y [I - A(\rho)] = \sum_{\substack{k=1, K \\ p=1, P}} L_{k,p} \left(1 - \frac{d_{k,p}}{d_{k,p} + \rho}\right) \quad (4.2)$$

Or, parce que C_y est diagonale, on peut observer que :

$$\frac{1}{K} \bar{W}_v C_y W_v = \begin{bmatrix} C_1 & C_2 & & C_K \\ C_K & C_1 & & \\ \vdots & \vdots & \ddots & \vdots \\ C_2 & & C_K & C_1 \end{bmatrix} \quad \text{avec } \begin{cases} C_k \text{ diagonale} \\ C_{1,p} = \frac{1}{K} \sum_{k=1} y_{k,p} \end{cases} \quad (4.3)$$

d'où (cf (3.3))

$$\bar{L} = \bar{U} C_y U^t = \frac{1}{K} \begin{bmatrix} v_1^t & & & \\ & v_2^t & & \\ & & \ddots & \\ & & & v_K^t \end{bmatrix} \bar{W}_v C_y W_v \begin{bmatrix} v_1 & & & \\ & v_2 & & \\ & & \ddots & \\ & & & v_K \end{bmatrix} = \begin{bmatrix} v_1^t C_1 v_1 & v_1^t C_2 v_2 & \dots & v_1^t C_K v_K \\ v_2^t C_K v_1 & v_2^t C_1 v_2 & & \\ \vdots & & & \\ & & & v_K^t C_1 v_K \end{bmatrix}$$

et chaque coefficient L_{kp} se calcule donc par le produit scalaire $\langle V_k^p, V_k^p \rangle_{C_1}$ où V_k^p est le p -ième vecteur colonne de V_k et (cf 4.3) $C_1 = \text{diag}(C_{1,1}, \dots, C_{1,p})$. L'étape b) du calcul du minimum $\hat{f}(\rho)$ (ou de $\hat{e}^*(\rho)$) est donc augmentée du calcul des L_{kp} dépendants de y qui coûte $2 K P^2$ opérations.

5. Généralisation de ce type d'algorithmes à d'autres tomographes

C'est la structure bloc-circulante de H (cf (2)) qui permet le calcul de λ_ρ avec ρ optimal par le type d'algorithme présenté.

La condition nécessaire à cette structure est que l'ensemble des m fonctions ϕ_i caractéristiques de l'appareil soit formé de la réunion de K ensembles de m_r fonctions ($\{\phi_{k,1}, \phi_{k,2}, \dots, \phi_{k,m_r}\}$ est le $k^{\text{ième}}$ ensemble) telles que l'on puisse écrire :

$$(\phi_{k,p}, \phi_{l,q}) = h_{[(k-l) \bmod K] + 1, p, q}$$

Ceci est vérifié dès que le $k^{\text{ième}}$ ensemble est obtenu par rotation du 1er autour de 0 d'angle $\theta_k = (k-1) \frac{2\pi}{K}$, c'est-à-dire :

$$\phi_{k,p}(x,y) = \phi_{1,p}(u,v) \quad (5.1) \quad \text{où} \quad \begin{cases} u = x \cos \theta_k - y \sin \theta_k \\ v = x \sin \theta_k + y \cos \theta_k \end{cases}$$

En tomographie classique il suffit donc que les domaines $L_{k,p}$ d'intégration associés à $\phi_{k,p} = \chi_{L_{kp}}$ aient cette invariance par rotation.

Dans les exemples B.1 et B.2 du chapitre I où les L_i sont des bandes de largeur d_i de position θ_i, u_i (cf I.B(1.1)), la répartition sur une grille rectangulaire de $[0, 2\pi] \times [-1, 1]$ uniforme en θ de ces positions :

$$\theta_i, u_i = \left(k \frac{2\pi}{K}, u_p \right)$$

et l'indépendance de d_i par rapport à θ_i : $d_i = d_p$, assurent cette invariance. On peut donc prendre en compte les cas de données incomplètes, appelés 'projections creuses' ou 'projections tronquées' s'il n'y a pas de mesures près du centre ($\forall p, |u_p| > c > 0$) ou près du bord ($\forall p, |u_p| < b < 1$), sans

traitement préalable de ces données alors que les méthodes analytiques nécessitent de les compléter, généralement par une simple interpolation qui n'assure pas la compatibilité de ces nouvelles données.

En tomographie à temps de vol cette invariance par rotation des bandes $L_{k,p}$ support de $\phi_{k,p}^\ell$ (cf I.B.4) est aussi suffisante puisque :

$$\phi_{k,p}^\ell(x,y) = \Delta_{u_p}^d(u) (\chi_{C_\ell} * G_\sigma)(v) = \phi_{1,p}^\ell(u,v)$$

Mais ici les blocs V_1, \dots, V_K sont d'ordre $m_r = PL$ et, même si leur calcul n'est effectué qu'une seule fois, leur mémorisation est en pratique prohibitive (typiquement $P = \frac{128}{2}$ et $L = 64$).

Remarquons que l'étape finale de la reconstruction

$$f_\rho(x,y) = \sum_{k,p,\ell} \lambda_{\rho,k,p,\ell} \phi_{kp}^\ell$$

est ici identique à l'épandage des histoprojections convoluées par G_σ défini au II.A.2 (équations II.A(2.2) et (2.4) avec f_ρ, λ_ρ à la place de \hat{e}, y).

C. L'EFFICACITE DU CHOIX AUTOMATIQUE DU ρ

Cet algorithme a été implémenté sur un mini-ordinateur VAX 780 pour le type de géométrie décrite au B.1 : 2 P bandes juxtaposées de largeur $d = 1$ forment une partition du disque de rayon P (les intégrales sur ces bandes constituent une projection de direction 0) et $\frac{K}{2}$ projections sont réparties uniformément sur $[0, \pi[$ c'est-à-dire associées aux directions $(k-1) \frac{2\pi}{K}$ $k = 1, \dots, \frac{K}{2}$.

C'est la géométrie parallèle standard avec un nombre pair de mesures par projection. Remarquons que dans le cas d'un nombre impair ($= 2P+1$) on se ramène aussi immédiatement à la géométrie de B.1 en remplaçant $-\frac{1}{2} + p$ par $-1+p$ dans B(1.1) et, puisque chaque bande centrale est alors décrite 2 fois, en pondérant les $L_{k,1}$ et $y_{k,1}$ par $\frac{1}{\sqrt{2}}$ dans S.

Ainsi avec 96 projections de 128 échantillons, ($P=64$ et $K=192$) la mémorisation nécessaire est celle de 97 matrices 64×64 (V_k) et d'une diagonale 97×64 (leur calcul préalable demande environ 6 minutes, la diagonalisation étant effectuée par une procédure d'IMSL).

L'observation (fig.IV.1) des valeurs propres D de RR^* ainsi obtenues montre bien la redondance dans les données exactes et permet d'appréhender l'instabilité numérique de l'équation (3.7) quand $\rho \rightarrow 0$, vis-à-vis des erreurs dans y ou même seulement des erreurs d'arrondis.

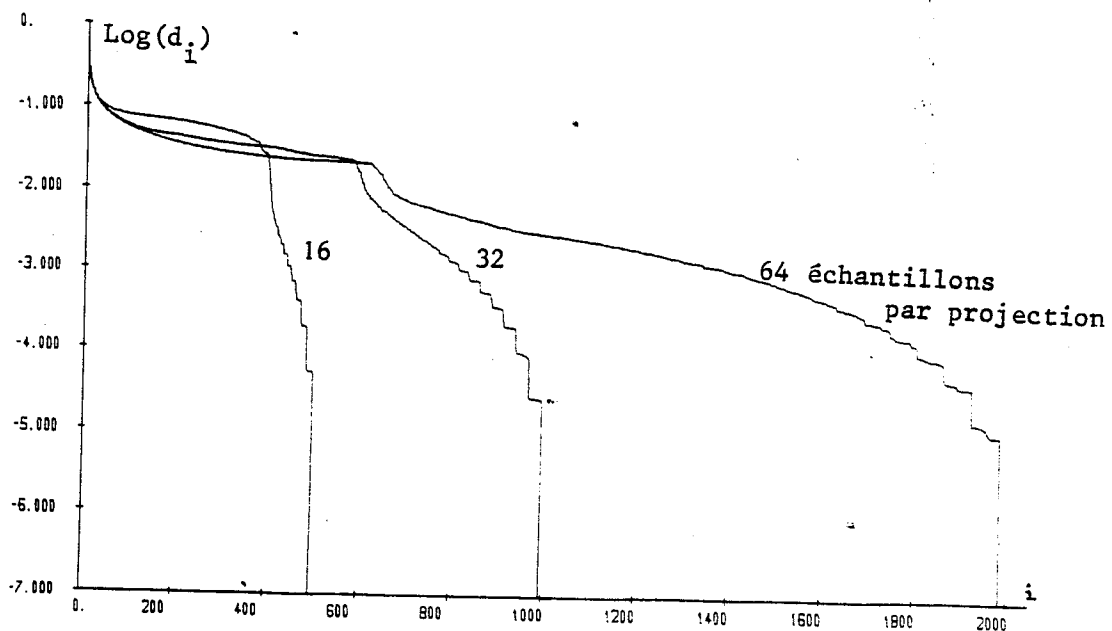


Fig. IV.1 : les valeurs propres (ordonnées) de RR^* avec 32 projections

Pour évaluer nos méthodes de choix de ρ , nous avons créé des objets test f^0 de valeur constante donnée sur des disques disjoints ($D(x,y,r)$ désigne dans la suite le disque de centre (x,y) de rayon r). On a alors des formules analytiques simples permettant de calculer précisément $y^0 = Rf^0$ le vecteur des données exactes.

La génération d'un bruit aléatoire (IMSL) permet de simuler un vecteur de données y à partir duquel on calcule f_ρ et ρ^* donné par l'une des méthodes de choix : pour comparer ρ^* avec le paramètre optimal ρ^0 nous avons évalué :

$$R(\rho) = \frac{1}{KP} \|A(\rho) y - y^0\|^2$$

grâce à la formule (se déduit comme (3.6)) :

$$R(\rho) = \frac{1}{KP} \|D(D+I)^{-1} Uy - Uy^0\|^2$$

qui nécessite le calcul de Uy^0 parallèlement à celui de Uy (étape (a) du 3).

$$1. \quad f^0 = \chi_{D(0,0,30)} + \chi_{D(0,15,15)} + \chi_{D(0,10,2)} + \chi_{D(0,20,4)} \quad (1)$$

y est obtenu en ajoutant à y^0 un bruit blanc de variance $\sigma^2 = 9$

L'évaluation numérique rapide (K.P opérations)

$$\text{de } S(\rho) = \frac{1}{KP} \|A(\rho) y - y\|^2 \quad \text{et} \quad V(\rho) = \frac{S(\rho)}{\left| \frac{1}{KP} \text{Trace } I-A(\rho) \right|^2}$$

a permis de tracer leurs courbes (figure IV.2) : on peut observer la très bonne efficacité de la validation croisée, meilleure que la méthode du résidu ($S(\rho) = 9$), pour estimer ρ^0 .

L'utilité de ce critère de choix est confirmée par le comportement en fonction de ρ de l'erreur de reconstruction, moyenne quadratique sur $D' = D(0,0,30)$ de $(f_\rho - f^0)$:

$$E(\rho) = \left[100 \cdot 1/\text{Surf}(D') \right] \sqrt{\int_{D'} (f_\rho - f^0)^2} \quad (2)$$

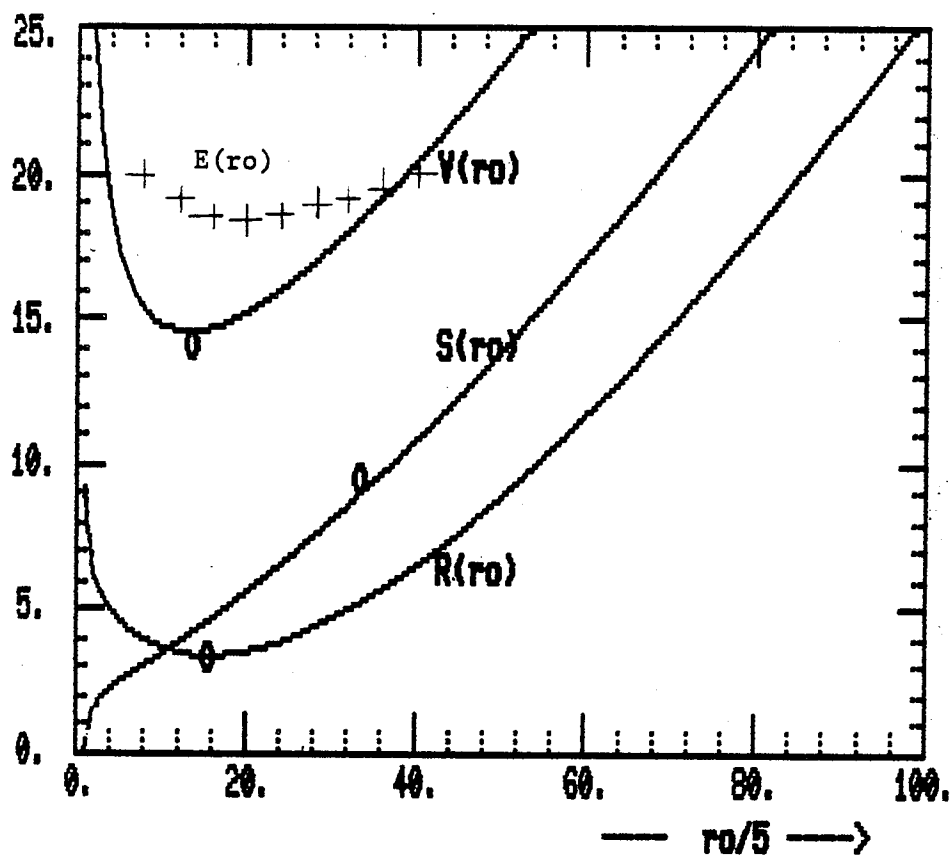


Fig.IV.2 : 96 projections de 128 données.

Critère d'énergie minimale. Bruit blanc

2. Identique à 1 avec un bruit poissonnien

$$\tilde{R}(\rho) = S(\rho) + \frac{1}{KP} \left[\sum_{k,p} y_{k,p} - 2 \text{Trace Diag}(y_{k,p}, k=1, \dots, K, p=1, \dots, P) (I-A(\rho)) \right]$$

C'est un excellent estimateur de $R(\rho)$; on ne les distingue pas sur la figure IV.3.

3. Identique à 2 avec 20 projections de 64 données, et

$$f^0 = 3 \left[X_{D(0,0,28)} + X_{D(0,14,12)} + X_{D(0,15,2)} \right] \quad (3)$$

\tilde{R} reste encore un bon estimateur, mais le critère de validation croisée est aussi satisfaisant (cf figure IV.4).

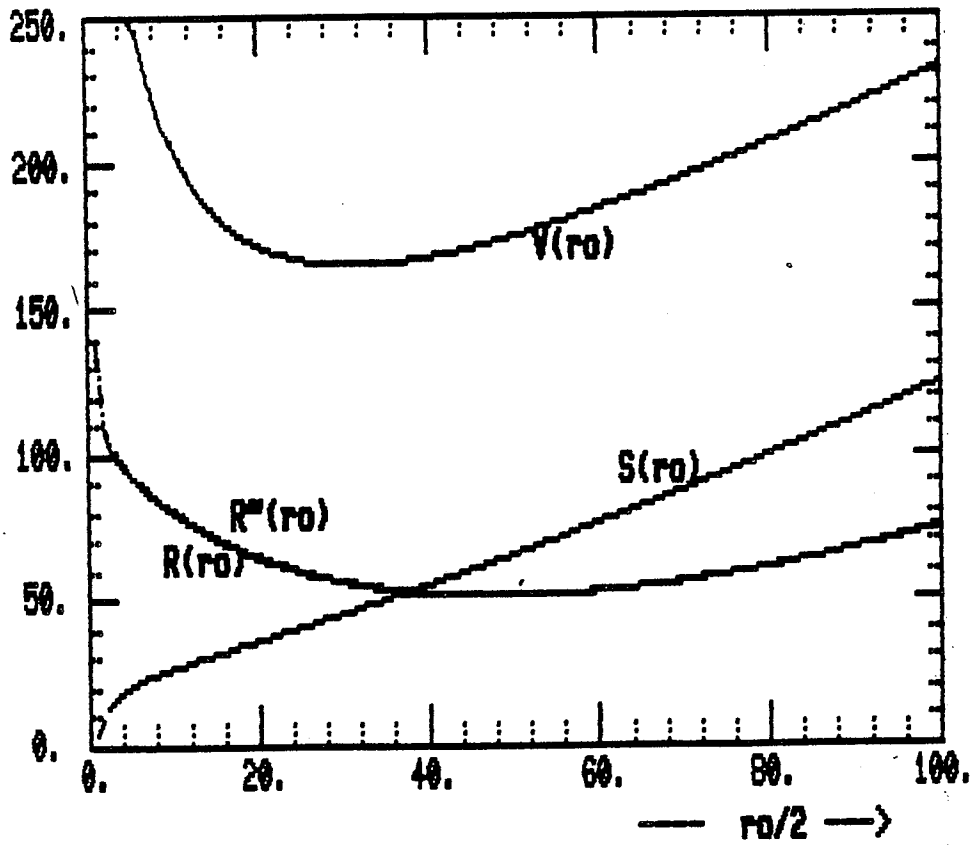


Fig. IV.3 : idem Fig IV.2 bruit poissonnien

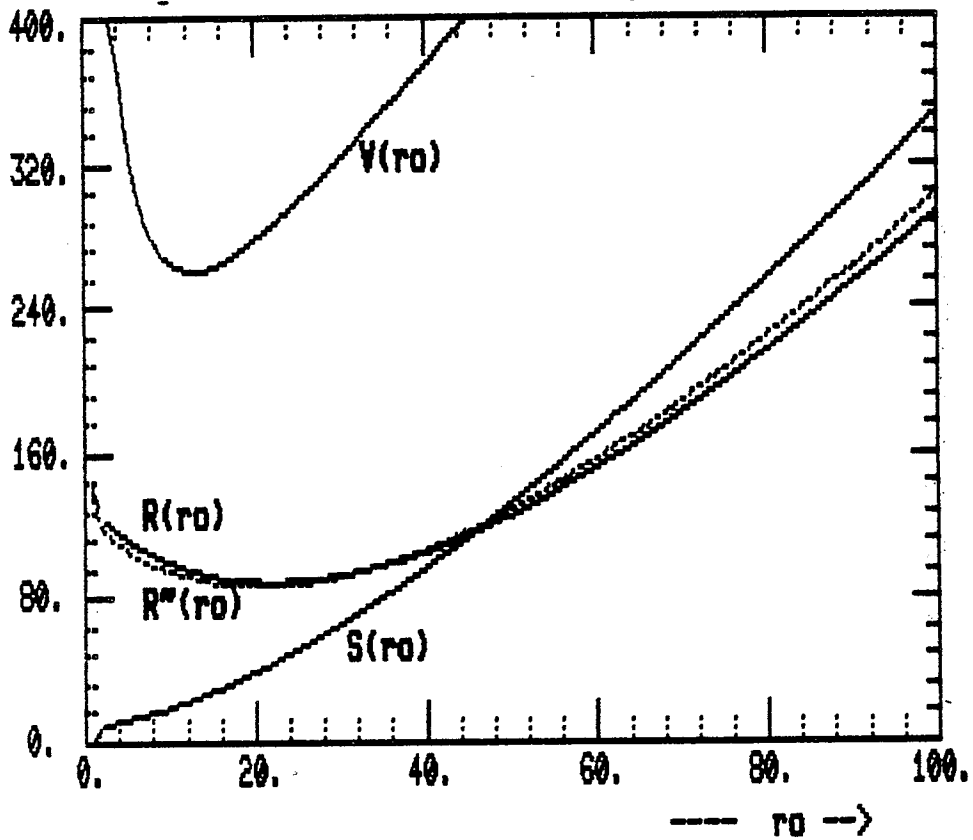


Fig. IV.4. : idem Fig. IV.3

20 projections de 64 données

f^0 donnée par (3)

De nombreux autres tests ont été effectués. Dans le cas du bruit blanc, ils ont toujours révélé une très bonne efficacité de la méthode de la validation croisée pour estimer ρ^0 alors que la méthode du résidu bien qu'elle utilise la variance du bruit donne généralement une estimation trop grande (et donc un lissage plus fort). La dépendance de la tolérance idéale $S(\rho^0)$ en fonction de f^0 a été confirmée. Un exemple extrême est le cas $f^0 = 0$: les mêmes autres conditions qu'en 1 ont conduit à ρ^0 très grand avec cette fois une efficacité identique des 2 méthodes d'estimations.

Les simulations d'autres types de bruit ont mis en évidence l'assez bonne robustesse de la validation croisée, l'excellent comportement de la méthode statistique quand on connaît le modèle probabiliste du bruit.

Au sujet du coût de ces reconstructions indiquons qu'avec 96 x 128 données, le calcul du ρ^* minimisant $v(\rho)$ et de f_{ρ^*} (étape (a), (b), (c) et épannage des coefficients) prend moins de 2 mn dont 1 pour l'épannage, sans tirer profit des importants parallélismes existants.

Chapitre V

REGULARISATION OPTIMALE EN DIMENSION n : ALGORITHMES DIRECTS D'APPROXIMATION DANS LE CAS DE SEMI-NORMES INVARIANTES PAR ROTATION

A - Introduction

B - Une simplification du calcul du ρ optimal quand $m \gg n$

1. Un problème de taille n équivalent
2. Le calcul du ρ optimal sur ce nouveau problème

C - La mise en oeuvre numérique

1. Les symétries
2. Le calcul de f_{ρ}
3. Le calcul du paramètre ρ optimal avec Δ inversible

D - L'efficacité du choix automatique du paramètre ρ

A. INTRODUCTION

Nous avons montré au chapitre précédent comment mettre en oeuvre efficacement (en tomographie classique) le calcul de la solution d'énergie minimale sous une contrainte choisie (méthode du résidu) ou optimisée (si le calcul du paramètre ρ est effectué suivant l'une des autres méthodes du III.C). Mais ce critère d'énergie ne se révèle pas toujours bien choisi : on lui préfère souvent un critère contrôlant les oscillations locales de la solution tel que celui de l'énergie J^1 du gradient (ou d'un autre opérateur différentiel) (cf III.B.3). Parce qu'on n'a pas de formules explicites d'une telle solution (ou parce que cette formule a trop de paramètres), on étudie ici son approximation par une méthode de type éléments finis. C'est-à-dire qu'on remplace l'espace F par un sous-espace de dimension n . Le choix d'une base $\psi_1, \psi_2, \dots, \psi_n$ de F conduit alors à un problème du type III.B.5 où Δ est la matrice sur cette base de la forme quadratique J envisagée et où H est la matrice $m \times n$: $H_{i,j} = (\phi_i, \psi_j)$.

Dès que la condition sur le noyau de Δ est vérifiée, on rappelle que la solution régularisée de paramètre ρ est la fonction de F minimisant

$$\sum_{i=1}^m [(\phi_i, f) - y_i]^2 + \rho J(f) \text{ dans } F \text{ ou } \|H\lambda - y\|^2 + \rho \lambda^t \Delta \lambda \text{ dans } \mathbb{R}^n \quad (1)$$

et est caractérisée par :

$$f_\rho = \sum_{j=1}^n \lambda_{\rho,j} \psi_j$$

avec $\lambda_\rho = (\lambda_{\rho,1}, \dots, \lambda_{\rho,n})^t$ solution de

$$(H^t H + \rho \Delta) \lambda_\rho = H^t y \quad (2)$$

Si on choisit les ψ_i avec des supports restreints, H est une matrice creuse et nous verrons que le calcul de $H^t y$ est possible à un coût raisonnable même pour m de l'ordre de 10^6 (complexité analogue au calcul de l'image épanouie vu au II.A.2). Il reste ensuite à résoudre un système d'ordre n .

De telles reconstructions avec une technique itérative pour approcher la solution de ce système ne sont pas nouvelles en tomographie classique (comme on l'a signalé au II.B.3.b). Mais le principal résultat que nous montrons (section B) est que l'adaptation de ρ aux données (par une des méthodes de III.C) peut être réalisé avec une faible augmentation du coût minimum du calcul direct de λ_{ρ} , tant que la taille des inconnues (n) ne rend pas prohibitif les précalculs et leur mémorisation.

Et nous montrons ensuite (section C) que, comme au chapitre précédent, l'invariance par rotation de la géométrie de mesures et du critère J choisi permet, non seulement une résolution directe de (2) mais aussi l'application du résultat précédent, avec un n assez grand (10^4) pourvu que les fonctions ψ_i possèdent aussi une telle symétrie. Le calcul de f_{ρ} avec ρ optimal a alors un coût compétitif avec celui des méthodes analytiques.

On présente ensuite (section D) l'application numérique de ces algorithmes en tomographie classique et avec temps de vol (jusqu'à $m = 96 \times 128 \times 64$) avec une base correspondant à une discrétisation sur une grille polaire ($n = 96 \times 128$) et $J = J_0 + J^1$ (cf III.B.3 et 4) : les expérimentations montrent encore une grande efficacité de la méthode de validation croisée.

B. UNE SIMPLIFICATION DU CALCUL DU ρ OPTIMAL QUAND $m \gg n$

Nous avons vu au chapitre IV que la diagonalisation par transformation unitaire utilisée pour le calcul de λ_ρ donne immédiatement celle de $A(\rho)$. Ici la matrice $H^t H + \rho \Delta$ à inverser pour calculer λ_ρ est beaucoup plus petite que $A(\rho)$ et il est alors d'un grand intérêt pratique d'observer que sa diagonalisation unitaire est encore suffisante pour mettre en oeuvre une méthode telle que la validation croisée. Pour simplifier nous présentons ici le cas $\Delta = I$. Les cas où Δ est inversible s'en déduisent immédiatement, et pour une généralisation nous renvoyons à l'article en annexe 2

1. Un problème de taille n équivalent

Notons $H^t H = V D^2 V^t$ (1.1) la diagonalisation unitaire de $H^t H$ (où D est une matrice diagonale, d'éléments > 0 et $V^t V = I$).

Il est connu qu'on peut écrire aussi : $H = U D V^t$ (1.2)

où les colonnes de U (resp. V) forment une base orthogonale de l'image de H (resp. du complémentaire du noyau de H).

Alors pour tout λ de \mathbb{R}^n , puisque $U U^t y - y$ est perpendiculaire à l'image de H , on a :

$$\|H\lambda - y\|^2 = \|H\lambda - U U^t y\|^2 + \|U U^t y - y\|^2 \quad (1.3)$$

Le deuxième terme étant constant par rapport à λ , on est ramené au problème équivalent :

$$\text{Min}_{\lambda \in \mathbb{R}^n} (\|\tilde{H}\lambda - \tilde{y}\|^2 + \rho \lambda^t \lambda) \quad (1.4)$$

$$\text{où} \quad \tilde{H} = U^t H = D V^t$$

$$\tilde{y} = U^t y$$

(obtenu en remplaçant $\|H\lambda - y\|^2$ par $\|H\lambda - UU^t y\|^2$ qui vaut aussi $\|DV^t \lambda - U^t y\|^2$ puisque $U^t U = I$).

C'est sur ce nouveau problème à moins de n équations (autant d'équations que de colonnes dans U) que nous allons choisir le paramètre ρ . Le ρ optimum selon Wahba (cf III.C.2) est bien identique pour les 2 problèmes, puisque si on note $\tilde{y}^0 = U^t y^0$, $\|H\lambda_\rho - y^0\|^2$ est égal à une constante près (d'après (1.3)) à $\|\tilde{H}\lambda_\rho - \tilde{y}^0\|^2$.

Remarquons aussi que l'hypothèse de bruit blanc de variance σ^2 faite éventuellement sur y est conservée sur \tilde{y} (puisque $U^t y$ a pour matrice de variance $U^t \sigma^2 I U = \sigma^2 I$)

2. Le calcul du ρ optimal sur ce nouveau problème

Notons $\tilde{A}(\rho)$ la matrice de lissage des données pour ce problème. L'intérêt de cette transformation est d'avoir une expression de $\tilde{A}(\rho)$ très simple et celle de \tilde{y} calculable facilement à partir de $H^t y$. En effet :

$$(1.4) \text{ a pour solution : } \lambda_\rho = (H^t H + \rho I)^{-1} H^t y = H^t (H H^t + \rho I)^{-1} y$$

$$= V D (D^2 + \rho I)^{-1} \tilde{y} \quad (2.1)$$

donc (2.1) porté dans la définition $\tilde{A}(\rho)\tilde{y} = \tilde{H}\lambda_\rho$ de $\tilde{A}(\rho)$ donne :

$$\tilde{A}(\rho) = D^2 (D^2 + \rho I)^{-1} \quad (2.2)$$

$$\text{et de } H^t y = V D U^t y \text{ on déduit : } \tilde{y} = D^{-1} V^t (H^t y) \quad (2.3)$$

On observe ainsi que pour le calcul de λ_ρ , $\tilde{s}(\rho) = \|(\tilde{A}(\rho) - I)\tilde{y}\|^2$ et trace $\tilde{A}(\rho)$, la seule 'grosse' matrice utilisée est H dans le calcul unique de $H^t y$.

On comprend que cette technique est très bien adaptée au cas où H est creuse de sorte que $H^t y$ puisse se calculer simplement et où l'on peut précalculer et mémoriser la décomposition (1.1) de $H^t H$ puisqu'alors le calcul du paramètre optimal et de la solution associée est réduit à :

- calcul de $H^t y$
- calcul de $\tilde{y} = D^{-1} V^t (H^t y)$ (n^2 opérations)
- recherche de ρ^* par minimisation d'une fonction (par ex :

$$\tilde{v}(\rho) = \frac{\|(\tilde{A}(\rho) - I)\tilde{y}\|^2}{[\text{Trace } \tilde{A}(\rho) - I]^2}$$
) qui s'évalue en n opérations.
- calcul de $\lambda_{\rho}^* = V D (D^2 + \rho^* I)^{-1} \tilde{y}$ (n^2 opérations)

Le coût global de ce calcul est ainsi comparable au coût minimum du calcul de λ_{ρ} pour un ρ choisi a priori (avec mémorisation préalable de $(H^t H + \rho I)^{-1}$).

Remarque

Le cas où Δ est inversible se traite de la même manière : la seule différence concerne les matrices D et V mémorisées qui sont remplacées ici par D' et B :

$$\text{où } \left\{ \begin{array}{l} D' \text{ est donnée par la décomposition en valeur propre de} \\ L^{-1} H^t H L^{-1} = V' D'^2 V'^t \\ \text{où } L \text{ est une matrice symétrique, racine carrée de} \\ \Delta : \Delta = L^2, \\ B = L^{-1} V' \end{array} \right.$$

(ceci s'obtient immédiatement par le changement de variable $\mu = L\lambda$, $\lambda = L^{-1}\mu$ dans le problème initial A(1)).

C. LA MISE EN OEUVRE NUMERIQUE

1. Les symétries

On a vu au IV.B.5 que les ϕ_i $i=1, \dots, m$ sont naturellement groupées en K angles de vue (cf IV.B.(5.1)), le passage d'un angle de vue au suivant équivalent à répéter les mêmes m_r mesures après avoir appliqué à l'objet une rotation d'angle $\frac{2\pi}{K}$ autour de 0.

Alors si $n = K \cdot Q$ et si les $\Psi_{k,q}$ $k = 1, \dots, K$ $q = 1, \dots, Q$ sont générées par rotations successives d'angle $\theta_k = \frac{(k-1)2\pi}{K}$, $k = 1, \dots, K$ des Q premières, c'est-à-dire :

$$\Psi_{k,q}(x,y) = \Psi_{1,q}(x \cos\theta_k - y \sin\theta_k, x \sin\theta_k + y \cos\theta_k) \text{ pour}$$

$$\begin{aligned} k &= 1, \dots, K \\ q &= 1, \dots, Q \end{aligned} \quad (1.1)$$

on vérifie facilement que :

$\sum_{i=1}^m (\phi_i, \Psi_{k,q})(\phi_i, \Psi_{\ell,p})$ est invariant si on ajoute modulo K un même entier à k et ℓ , ce qui signifie que $H^t H$ a une structure bloc-circulante (identique à RR^* du IV.B(1.2)) :

$$H^t H = \begin{bmatrix} T_1 & T_2 & & & T_K \\ & T_K & T_1 & T_2 & \\ & & & \backslash & \\ & & & & T_2 \\ T_2 & & & & T_1 \end{bmatrix} \quad \text{où } T_k \text{ matrice d'ordre } Q. \quad (1.2)$$

$$H^t H \text{ symétrique entraîne que } T_k = T_{K+2-k}^t \quad (1.3)$$

En outre parce qu'on a toujours pour chaque angle θ_k une invariance de $\{\phi_{k,1}, \dots, \phi_{k,m_r}\}$ par la symétrie autour de l'axe perpendiculaire à leur direction (par exemple l'axe Ox pour les mesures de direction Oy), on peut vérifier qu'avec une même invariance pour $\{\Psi_{1,1}, \dots, \Psi_{1,Q}\}$ on a T_k symétrique pour tout k (1.4).

L'exemple d'une telle base que nous proposons ici correspond à une discrétisation sur une grille polaire obtenue par partition de \mathbb{D} en K secteurs

angulaires uniformes (d'angle $\frac{2\pi}{K}$), chaque secteur étant ensuite partitionné de manière identique en Q pavés par exemple par les cercles de rayons $q \cdot \frac{1}{Q}$ pour $q = 1, \dots, Q-1$. $\Psi_{k,q}$ est alors la fonction indicatrice d'un tel pavé.

Avec une telle base H est une matrice creuse et on a :

$$(H^t y)_j = \sum_{i=1}^m (\phi_i, \Psi_j) y_i = (\Psi_j, \sum_{i=1}^m y_i \phi_i)$$

qui est alors la valeur moyenne sur le support de Ψ_j de la fonction $\sum y_i \phi_i$ dont l'évaluation sur une grille est connue sous le nom d'épandage des données (déjà signalé au chapitre IV).

Seules les Q premières colonnes de $H^t H$ sont à évaluer (le calcul de chaque colonne équivaut à un épandage) pour permettre le précalcul et la mémorisation des K matrices M_k d'ordre Q de la bloc-diagonalisation de $H^t H$ (analogue à la décomposition IV.B(1.4) de RR^*) :

$$\text{si } \begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_K \end{bmatrix} = W_V \begin{bmatrix} T_1 \\ T_2 \\ \vdots \\ T_K \end{bmatrix} \quad (1.5) \quad \text{alors } H^t H = \frac{1}{K} \bar{W}_V \begin{bmatrix} M_1 & & \\ & M_2 & \\ & & \ddots \\ & & & M_K \end{bmatrix} W_V \quad (1.6)$$

Pour tirer profit de ces symétries, il suffit de choisir un critère de régularité associé à une matrice Δ ayant la même structure bloc-circulante de $H^t H$. Une telle matrice est obtenue par la discrétisation, par exemple sur la grille polaire, de toute norme invariante par rotation autour de l'origine, par exemple :

$$J(f) = \int_D f^2 + \int_D \left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2$$

Le choix de Δ se ramène en fait au choix de K matrices, d'ordre Q , $\Delta_1, \dots, \Delta_K$ puisque l'on a aussi :

$$\Delta = \frac{1}{K} \bar{W}_V \begin{bmatrix} \Delta_1 & & \\ & \Delta_2 & \\ & & \ddots \\ & & & \Delta_K \end{bmatrix} W_V \quad (1.7)$$

En pratique (comme dans l'exemple $J(f)$ ci-dessus) ces matrices Δ_k sont calculables explicitement et faciles à générer (tridiagonales).

2. Le calcul de f_{ρ}

En portant les décompositions (1.6) et (1.7) dans le système A(2) donnant λ_{ρ} , on obtient :

$$\lambda_{\rho} = \frac{1}{K} \bar{W}_v \begin{bmatrix} (M_1 + \rho \Delta_1)^{-1} \\ (M_2 + \rho \Delta_2)^{-1} \\ \vdots \\ (M_K + \rho \Delta_K)^{-1} \end{bmatrix} W_v^t H^t y$$

On en déduit un algorithme du même type qu'au chapitre précédent (IV.B.2) excepté le calcul de $H^t y$ qui apparaît maintenant :

En effet si les $2K$ matrices M_k et Δ_k $k = 1, \dots, K$ sont précalculées et mémorisées pour un tomographe donné, le calcul des valeurs de f_{ρ} sur la grille polaire de taille $K \times Q$ se décompose en :

- (a) calcul de $\mu = H^t y$ (un épandage)
- (b) calcul de $W_v \mu$ (Q T.F.D. d'ordre K)
- (c) K résolution de système linéaire d'ordre Q
- (d) multiplication par \bar{W}_v (coût identique à (b))

Comme au IV.B.2 le coût le plus économique du calcul de λ_{ρ} avec ρ fixé une fois pour toute est obtenu en mémorisant $(M_k + \rho \Delta_k)^{-1}$ $k = 1, \dots, K$. Mais ici la taille de ces blocs est indépendante de m_r : elle est fonction de la richesse de la base choisie.

Remarquons que cette mémorisation est simplifiée par les propriétés qu'ont ces matrices d'être réelles, symétriques et deux par deux identiques (pour k et $K+2-k$).

3. Le calcul du paramètre ρ optimal avec Δ inversible

On voit dans l'algorithme précédent que le calcul de λ_{ρ} ne fait intervenir y qu'à travers $\mu = H^t y$. Et en fait même si $S(f_{\rho}) = \|H \lambda - y\|^2$ ne peut pas être exprimé en fonction de μ indépendamment de y , on a montré en B que la valeur optimale de ρ est bien estimée par le minimiseur de la fonc-

tion $\tilde{v}(\rho)$ suivante, exprimée à travers $H^t y$:

$$\tilde{v}(\rho) = \frac{\tilde{s}(\rho)}{\left(\sum_{\substack{j \\ /d_j \neq 0}} \frac{\rho}{d_j^2 + \rho}\right)^2} \quad (3.1) \quad \text{où } \tilde{s}(\rho) = \sum_{j/d_j \neq 0} \left(\frac{d_j}{d_j^2 + \rho} - \frac{1}{d_j}\right)^2 |z_j|^2 \quad (3.2)$$

$$= \sum_{j/d_j \neq 0} \left| \frac{\rho z_j}{d_j(d_j^2 + \rho)} \right|^2 \quad (3.3)$$

$$\text{avec } \begin{cases} z = (z_1, \dots, z_n)^t = \bar{V}^t L^{-1} H^t y & (3.4) \\ L \text{ matrice d'ordre } n \text{ racine carrée symétrique de } \Delta & (3.5) \\ V \text{ matrice d'ordre } n \text{ unitaire / } V \text{ diag}(d_1, \dots, d_n) \bar{V}^t = L^{-1} H^t H L^{-1} & (3.6) \end{cases}$$

$\tilde{v}(\rho)$ est en fait le critère de validation croisée appliqué au problème équivalent défini au B.1. La décomposition de $L^{-1} H^t H L^{-1}$ envisagée ici autorise que V soit à coefficient dans \mathbb{C} , contexte dans lequel les résultats du B restent valables avec \bar{V}^t (resp. \bar{U}^t) à la place de V^t (resp. U^t) et des colonnes ont été éventuellement rajoutées à la matrice V de B(2.3) sans que cela ne modifie cette expression puisque la matrice D^{-1} est ici prolongée par des zéros et de même pour D .

Les diagonales étant dans B.2 de taille la dimension de l'image de $H L^{-1}$, c'est-à-dire aussi le nombre ℓ de d_j non nulles, la méthode du résidu avec l'hypothèse d'un bruit blanc de variance σ^2 connu donne l'équation :

$$\tilde{s}(\rho) = \ell \cdot \sigma^2$$

Montrons maintenant que $L^{-1} V$ et les valeurs d_j peuvent être calculées et mémorisées pour un coût raisonnable.

En effet le calcul de L^{-1} se ramène à celui de L_k^{-1} avec L_k racine carrée symétrique de chaque blocs Δ_k puisque :

$$L^{-1} = \frac{1}{K} \bar{W}_V \begin{bmatrix} L_1^{-1} & & & \\ & L_2^{-1} & & \\ & & \ddots & \\ & & & L_K^{-1} \end{bmatrix} W_V$$

Le calcul de V se ramène au calcul des blocs $L_k^{-1} M_k L_k^{-1}$ puis à leur dia-

gonalisation par transformation orthogonales :

$$L_k^{-1} M_k L_k^{-1} = V_k \text{diag}(d_{k,1}^2, \dots, d_{k,Q}^2) V_k^t \quad (3.7)$$

puisque $V = \frac{1}{\sqrt{K}} \bar{W}_V \begin{bmatrix} V_1 & & \\ & V_2 & \\ & & \ddots \\ & & & V_K \end{bmatrix}$ est alors bien unitaire et diagonalise $L^{-1} H^t H L^{-1}$

pour les valeurs $(d_1, \dots, d_n) = (d_{1,1}, \dots, d_{1,Q}, \dots, d_{K,1}, \dots, d_{K,Q})$.

Et finalement le calcul des K blocs $B_k = L_k^{-1} V_k$ donne $L^{-1} V$ sous la forme simple :

$$L^{-1} V = \frac{1}{\sqrt{K}} \bar{W}_V \begin{bmatrix} B_1 & & \\ & B_2 & \\ & & \ddots \\ & & & B_K \end{bmatrix}$$

Ainsi le z (défini par (3.4)) nécessaire au calcul de $\tilde{s}(\rho)$ ou $\tilde{v}(\rho)$ est donné par :

$$z = \frac{1}{\sqrt{K}} \begin{bmatrix} B_1^t & & \\ & B_2^t & \\ & & \ddots \\ & & & B_K^t \end{bmatrix} W_V H^t y \quad (3.8)$$

D'autre part en portant (3.6) dans la caractérisation de λ_ρ on a :

$$\lambda_\rho = (H^t H + \rho \Delta)^{-1} H^t y = L^{-1} V \text{diag}\left(\frac{1}{d_j^2 + \rho}, j = 1, \dots, n\right) \bar{V}^t L^{-1} H^t y$$

d'où avec (3.4) et (3.8) :

$$\lambda_\rho = \frac{1}{\sqrt{K}} \bar{W}_V \begin{bmatrix} B_1 & & \\ & B_2 & \\ & & \ddots \\ & & & B_K \end{bmatrix} \text{diag}\left(\frac{1}{d_{k,p}^2 + \rho}, k = 1, \dots, K, p = 1, \dots, Q\right) z \quad (3.9)$$

On voit ainsi apparaître les 4 étapes d'un algorithme de régularisation optimale assez voisin de celui obtenu en IV.B.3.

Si au lieu de mémoriser les blocs M_k et Δ_k comme en 2, on mémorise les blocs B_k d'ordre Q et les n valeurs propres $d_{k,q}$, alors cette reconstruction avec calcul du paramètre optimal est réalisée par :

- (a) calcul de $H^t y$
- (b) calcul de z suivant (3.8) (Q T.F.D. d'ordre $K + K Q^2$ opérations)
- (c) calcul de ρ^* par exemple par la minimisation de $\tilde{v}(\rho)$ qui s'évalue en $3 K.Q$ opérations
- (d) calcul de λ_{ρ}^* suivant (3.9) (coût identique à (b)).

D. L'EFFICACITE DU CHOIX AUTOMATIQUE DU PARAMETRE ρ

L'implémentation a été réalisée pour la tomographie classique avec la géométrie de projections à rayons parallèles, uniformément réparties sur $[0, \pi[$ identique à celle étudiée au IV.C et pour la tomographie temps de vol (I.B.3) avec chaque mesure précédente décomposée en L mesures plus locales uniformément réparties.

La taille des expérimentations présentées est 96 projections de 128 données pour le premier tomographe, 96 histoprojections de 128 x 64 données pour le second (donc $K = 192$, $P = L = 64$) avec une incertitude sur l'information temps de vol $\sigma = 8$ (l'unité est le pas d'échantillonnage linéaire) et un pas des classes de temps de vol $\Delta v = 3$.

La base envisagée alors est celle proposée au C.1, associée à la grille polaire de même géométrie : $K = 192$, $Q = 64$. Nous verrons qu'une telle discrétisation se révèle suffisamment fine pour bien modéliser toutes les solutions accessibles avec cette configuration des mesures et un bruit réaliste dans celles-ci.

L'introduction de l'information temps de vol n'intervient en fait qu'au moment du calcul des blocs T_1, \dots, T_{97} et à l'étape (a), calcul de $H^t y$ (épandage de données). Les autres étapes de l'algorithme sont identiques.

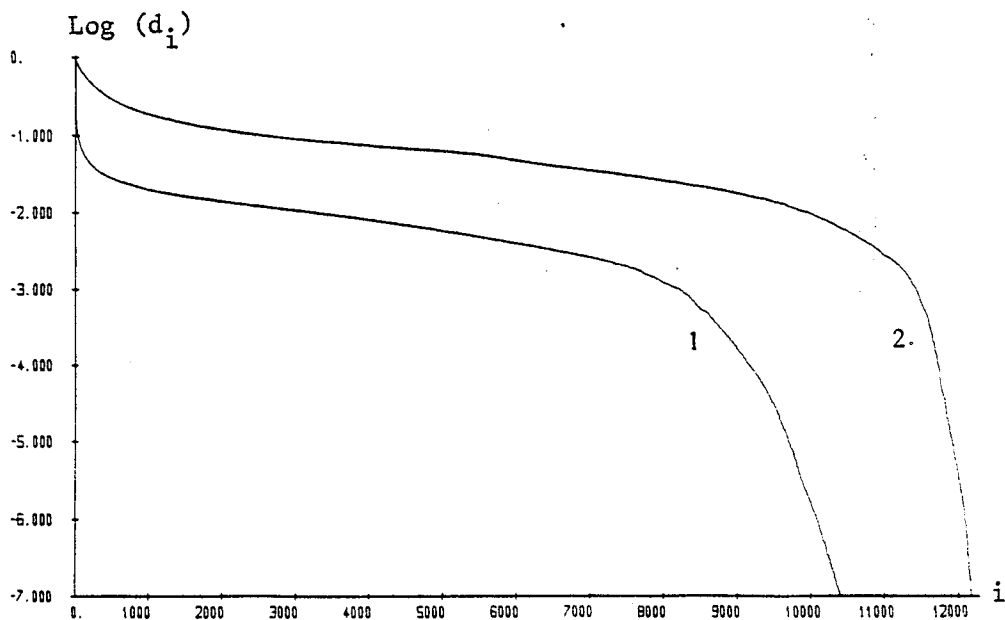


Fig. V.1 : les valeurs propres de $H^t H$ sans (1) et avec (2) temps de vol

On a principalement expérimenté le critère d'irrégularité suivant :

$$\lambda^t \Delta \lambda = \lambda^t C \lambda + \lambda^t G_1^t C G_1 \lambda + \lambda^t G_2^t C G_2 \lambda$$

discrétisant la somme de l'énergie de $f = \sum \lambda_i \Psi_i$ et de l'énergie de son gradient

avec

$$\begin{cases} C & \text{matrice diag}(C_1, \dots, C_n) / \sum C_i \lambda_i^2 \text{ approche } \int f^2 \\ G_1 \text{ (resp. } G_2) & \text{matrice de discrétisation de } \frac{\partial f}{\partial r} \text{ (resp } \frac{1}{r} \frac{\partial f}{\partial \theta}) \text{ sur} \\ & \text{la grille polaire} \end{cases}$$

Les matrices $\Delta_1, \dots, \Delta_K$ de la bloc-diagonalisation de $\Delta(B(1.7))$ sont faciles à calculer analytiquement (tridiagonales quand G_1 (et G_2) sont déterminées par des différences finies à 2 points).

Une fois calculé les blocs T_k (ce qui demande 64 épandages et est l'étape la plus coûteuse), les calculs préalables (des blocs B_k et de la diagonale D) prennent quelques minutes.

On a procédé ensuite comme IV.C avec (cf B.1 et C.3) l'erreur exacte :

$$R(\rho) = \frac{\tilde{r}(\rho)}{\ell} = \frac{1}{\ell} \|\tilde{A}(\rho) \tilde{y} - \tilde{y}^0\|^2 = \frac{1}{\ell} \sum_{j/d_j \neq 0} \left| \frac{d_j z_j}{d_j^2 + \rho} - \frac{z_j^0}{d_j} \right|^2$$

ou $z^0 = V^t H^t y^0$ est calculé en répétant les étapes (a) et (b) de l'algorithme du C.3 avec y^0 au lieu de y ; et avec :

$$S(\rho) = \frac{1}{\ell} \tilde{s}(\rho) \quad (\text{cf C.3.2}) \quad \text{et} \quad V(\rho) = \frac{S(\rho)}{\left| \frac{1}{\ell} \text{Trace} (\tilde{A}(\rho) - I) \right|^2} = \ell \cdot \tilde{v}(\rho)$$

(cf C.(3.1))

1. Le problème et les données simulées sont ici identiques au IV.C.1 : on observe sur la figure V.2 que la méthode de validation croisée produit encore une très bonne estimation du paramètre optimal. On remarque que la meilleure erreur moyenne de prédiction des données exactes $\text{Min}_\rho R(\rho)$ est ici environ deux fois plus petite qu'avec la régularisation par l'énergie du IV.C.1. En fait la seule comparaison des $\text{Min}_\rho V(\rho)$ associés à chacun de ces 2 critères de régularité laisse deviner cette supériorité, ici, du critère 'énergie + gradient', puisque l'écart entre V et R reste dans les 2 cas voisin de σ^2 près du maximum. Cette méthode de choix du critère, proposée au III.D.1, est ici efficace.

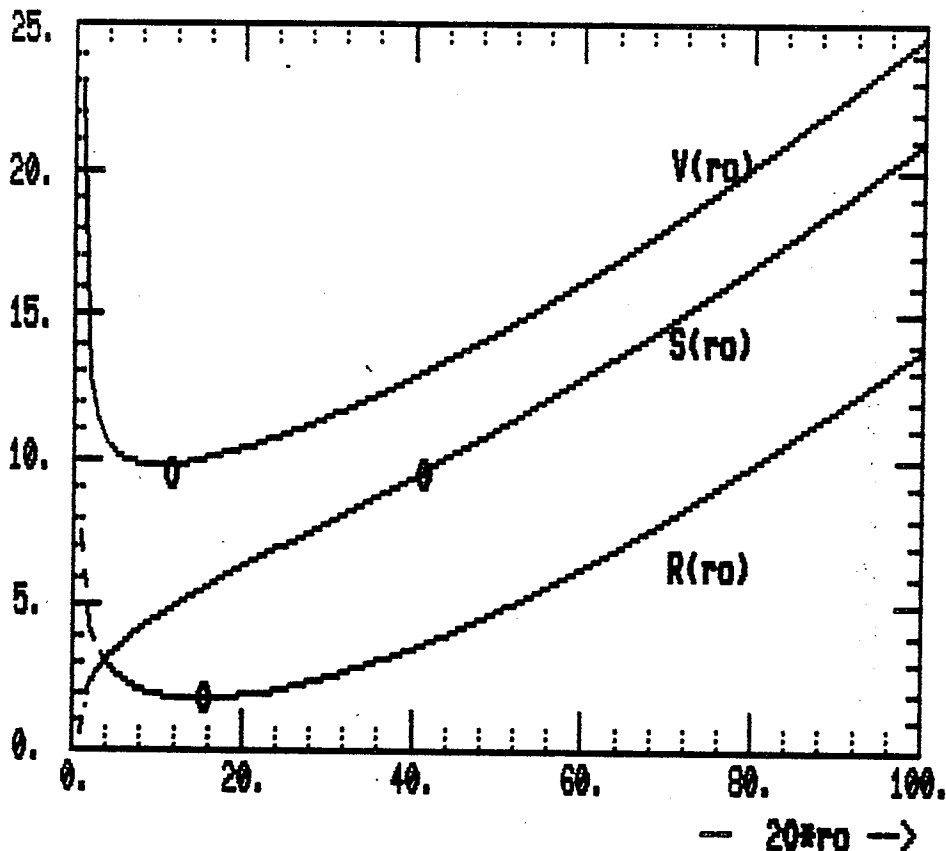


Fig.V.2 : Idem Fig.IV.1, critère énergie + gradient, discrétisation grille polaire 192×64 .

2. Ici c'est la tomographie temps de vol où la largeur des bandes est supposée nulle. On peut ainsi calculer analytiquement les valeurs exactes $y_{k,p,\ell}^0$ (cf. I.B.3) :

$$y_{k,p,\ell}^0 = \int_{C_\ell} H_{\theta_k} f^0(u_p, v) dv$$

$$\text{où } \begin{cases} \theta_k = (k-1) \frac{2\pi}{K} \\ u_p = -\frac{1}{2} + p \\ C_\ell = \left[\left(-\frac{L-1}{2} + \ell - 1\right) \Delta v, \left(-\frac{L-1}{2} + \ell\right) \Delta v \right] \end{cases}$$

avec f^0 constante sur des disques données.

f^0 est choisie proportionnelle à la fonction du 1) de manière à ce qu'une perturbation poissonnien de y^0 donne une acquisition réaliste : la valeur moyenne de ces 96 x 128 x 64 données est de l'ordre de 4.

On observe toujours la très bonne efficacité de la validation croisée sur la figure V.3.

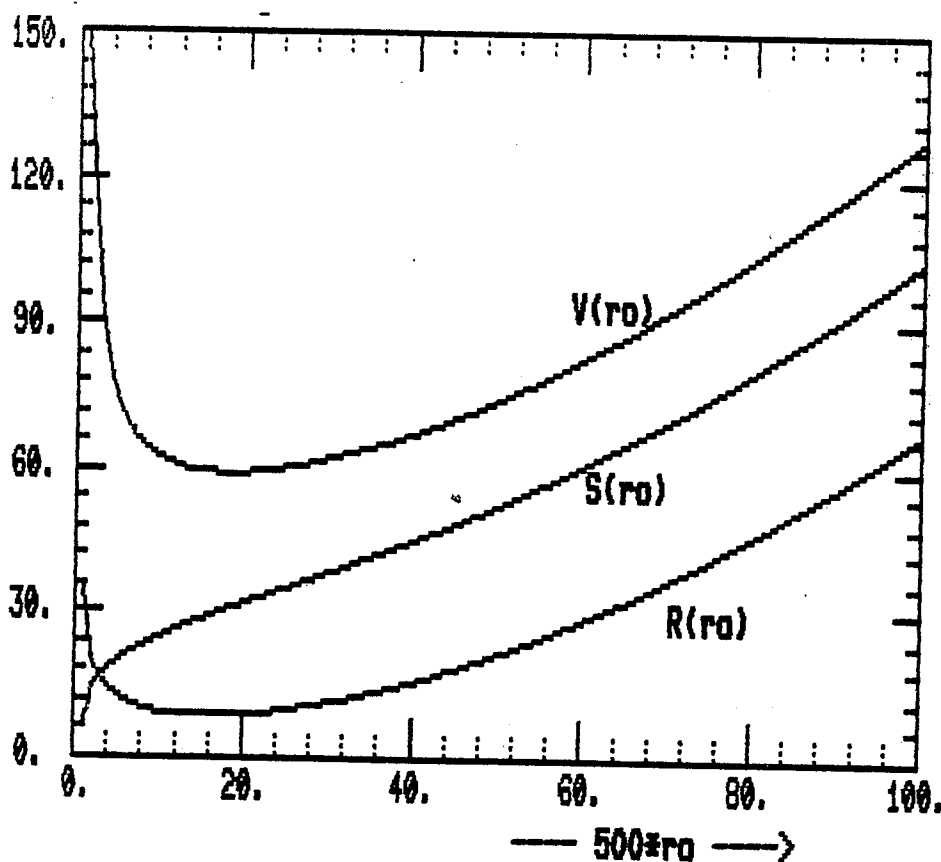
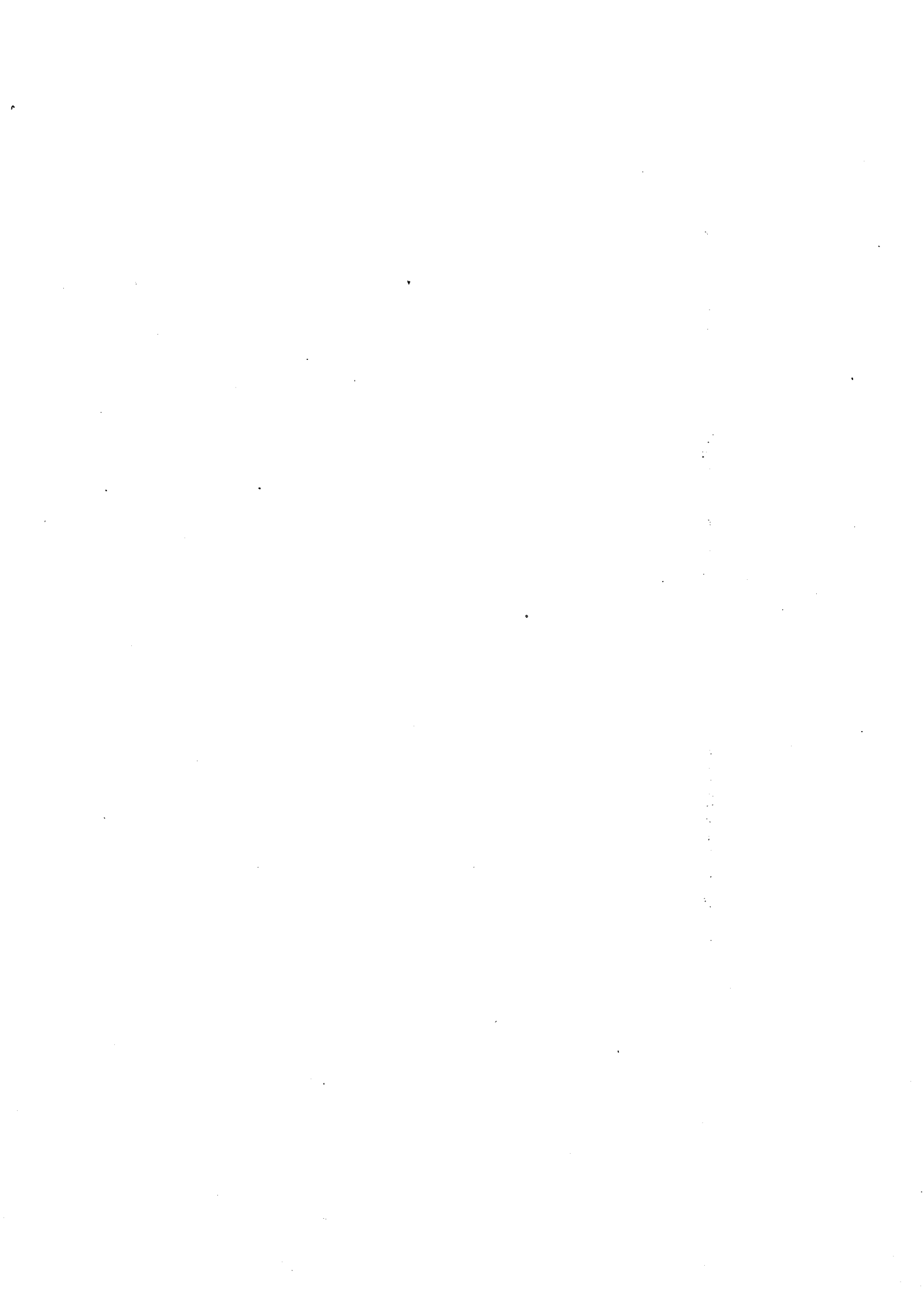


Fig.V.3 : 96 x 128 x 64 données temps de vol



Chapitre VI

EVALUATION DES ALGORITHMES PROPOSES

- A. En tomographie classique
- B. En tomographie à positrons et à temps de vol
- C. Discussion

A. EN TOMOGRAPHIE CLASSIQUE

Nous allons voir ici quel type de reconstructions produisent nos algorithmes suivant le critère de régularité choisi. On présente d'abord le cas d'un nombre réduit de données puis celui d'une taille plus grande.

- 20 x 64 données

La géométrie est celle du IV.C : 20 projections réparties sur $[0, \pi[$ chacune étant constituée des intégrales sur 64 bandes juxtaposées formant une partition du disque de rayon 32.

L'objet initial était une fonction constante sur 4 disques dont la représentation en perspective sur une grille 64 x 64 de pas 1 est montrée figure VI.1. Les données sont prises exactes (calcul analytique des intégrales).

La méthode analytique standard (épannage des projections filtrées) a produit (avec un filtre de Hamming (Cf . II.A.1.2)) la reconstruction figure VI.2. Le choix d'un filtre lissant plus la reconstruction aurait en même temps fait disparaître le plus petit détail.

On a calculé par l'algorithme du IV la solution d'énergie minimale interpolant les données ($\rho=0$ dans IV.B.2 et inversion au sens du pseudo-inverse puisque d'après III.B.2 tout vecteur λ solution de $RR^*\lambda=y$ donne f_0) : même si cette reconstruction, représentée figure VI.3, semble oscillante, sa reprojction (calcul des intégrales) coïncide (à la précision machine près) avec les données exactes. La méthode de validation croisée produit un résultat pratiquement identique : les données n'étant pas bruitées, le lissage optimal (au sens de Wahba (III.C.2.1)) est bien celui qui ne modifie pas ces données ($\rho \approx 0$). On voit ainsi que le critère de l'énergie minimale ne produit pas une reconstruction visuellement très satisfaisante.

On a ensuite calculé la solution d'énergie et gradient minimum pour un paramètre de régularisation calculé par la méthode de la validation croisée, avec une discrétisation sur la grille polaire de pas angulaire $\frac{\pi}{20}$,

linéaire 1 ($n = 40 \times 32$) (cf algorithme du chapitre V). Le choix automatique du paramètre a déterminé un écart aux données de 1 % (erreur quadratique relative) dû à l'erreur de discrétisation. On observe sur la représentation figure VI.4 de cette reconstruction, la disparition de la majorité des 'artefacts' en étoile présents dans les reconstructions VI.2 et VI.3 et dû principalement au fait que ces dernières reconstructions sont du type épannage le long des bandes des coefficients calculés.

L'aspect oscillant des artefacts de la figure VI.3, que quelques auteurs ont établi analytiquement (voir A.K. Louis (1981)), confirme la nécessité d'un critère de régularité contrôlant les variations locales comme l'énergie du gradient.

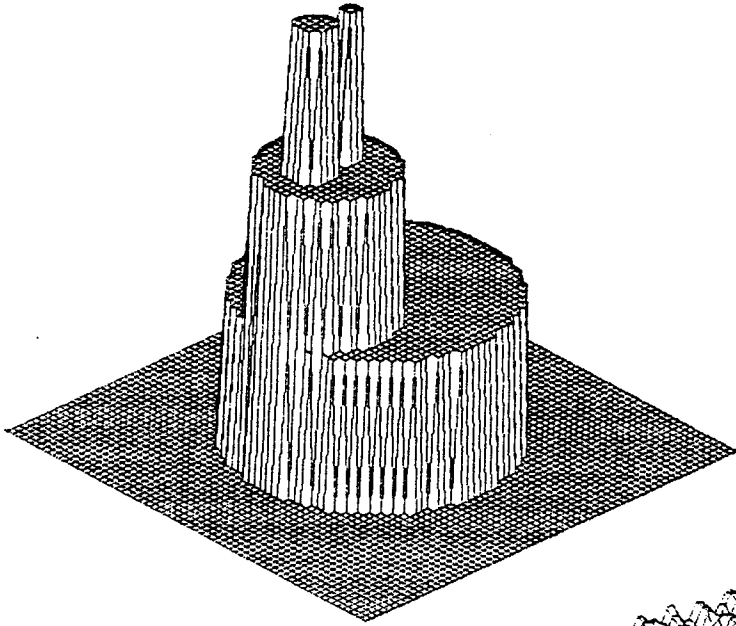


Fig.VI.1 : Objet initial

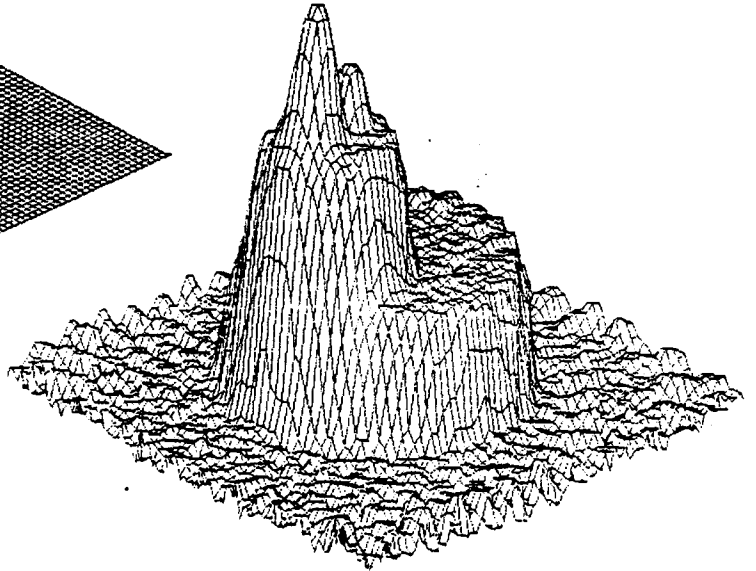


Fig.VI.2 : Epannage des projections filtrées

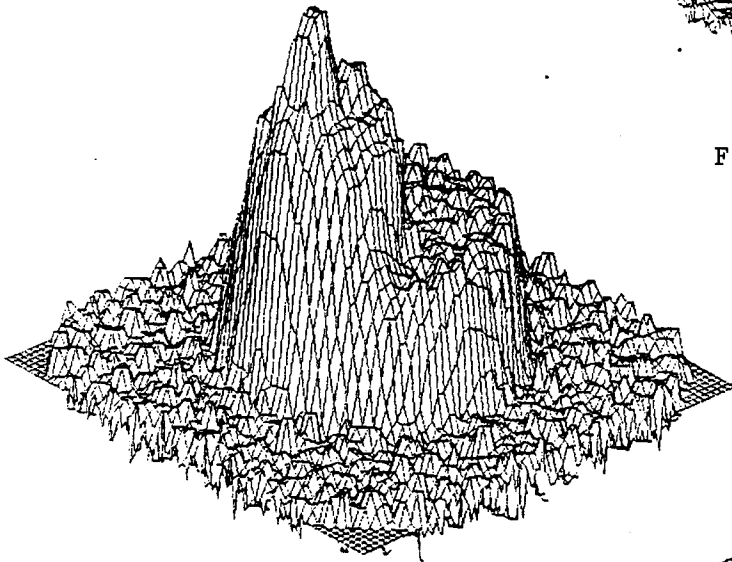


Fig.VI.3 : Interpolation d'énergie minimale

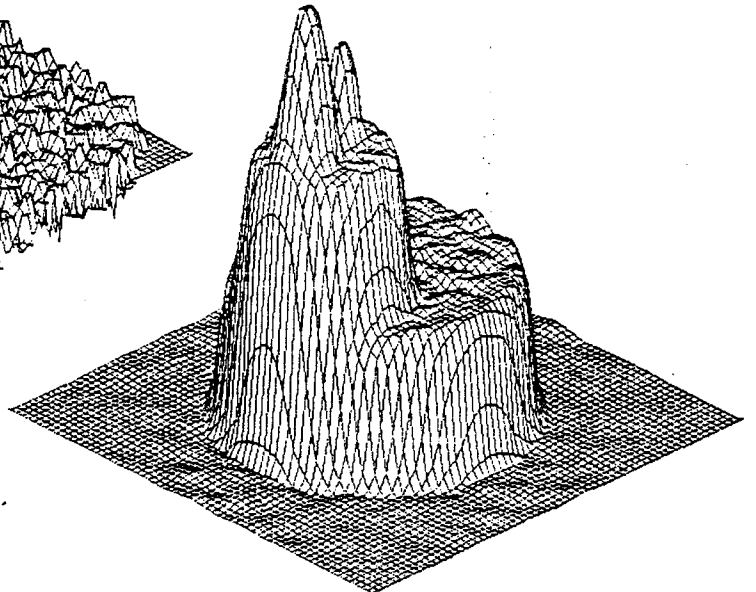


Fig.VI.4 : Régularisation optimale avec l'énergie du gradient (40 x 32 pavés polaires)

3 reconstructions avec 20 projections
de 64 échantillons

- 96 x 128 données

Ici on a reconstruit à partir de 96 projections réparties sur $[0, \pi[$, chacune étant constituée des intégrales sur 128 lignes parallèles uniformément espacées sur le disque de rayon 64. Les bandes sont prises infiniment minces pour être précisément dans le cadre d'application des méthodes analytiques.

Les données ont été calculées sur un objet initial aussi simple que le précédent (4 disques). On a d'abord expérimenté avec les données exactes, puis avec une perturbation de ces données, trois algorithmes de reconstructions :

1. La méthode analytique standard (cf ci-dessus) avec un filtre de Hamming dit à "haute résolution", c'est-à-dire peu lissant.
2. idem 1 avec un filtre de Hamming plus lissant
3. la régularisation, sur une grille polaire 192 x 64, par l'énergie du gradient, avec la validation croisée (cf ci-dessus).

Ces 3 algorithmes ont produit respectivement les solutions représentées sur les figures 5, 6, 7 avec des données exactes et 8, 9, 10 avec des données générées en ajoutant un bruit blanc aux précédentes.

On observe dans le cas sans bruit que la forme des 4 cylindres est presque aussi bien rendue par l'algorithme 1 que par le 3, le second atténuant trop le cylindre le plus fin. On remarque que la discrétisation sur la grille polaire dans l'algorithme 3 n'est pas visible : même le plus petit détail ici n'est pas déformé.

Dans le cas avec bruit on observe (fig.8) l'instabilité numérique de l'algorithme 1 : les oscillations sont beaucoup plus fortes que dans la reconstruction produite par l'algorithme 2 (fig. 9). L'algorithme 3 produit lui une reconstruction voisine de cette dernière et que l'on peut juger visuellement (fig.10) comme un bon compromis entre la résolution et l'insensibilité au bruit auxquelles on doit en même temps accéder. L'algorithme 3 est ainsi le seul qui produit toujours une bonne solution qu'il y ait ou non du bruit dans les données.

D'autres expérimentations numériques ont montré le bon comportement de cet algorithme.

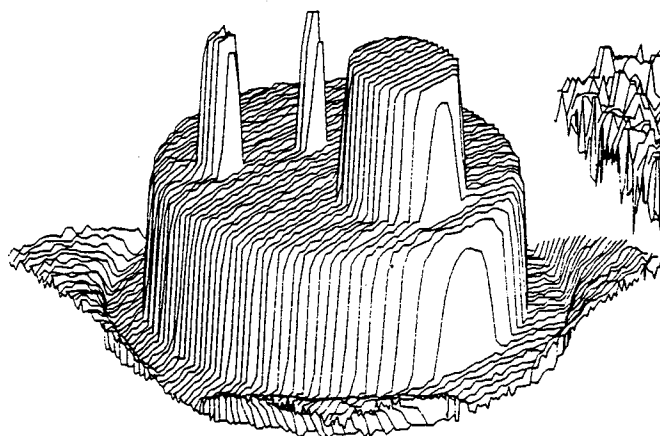


Fig.VI.5

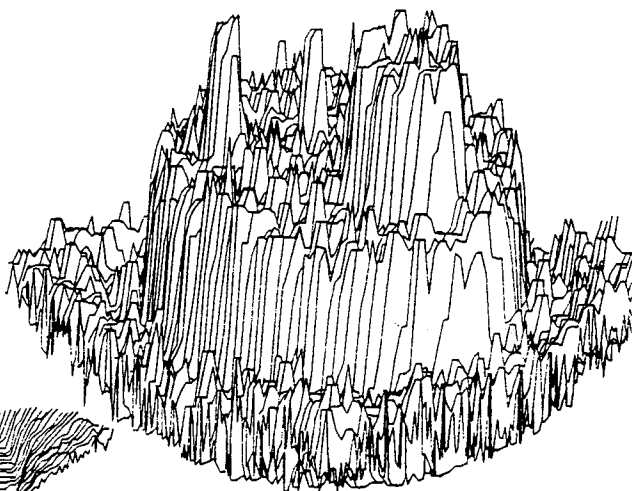


Fig.VI.8

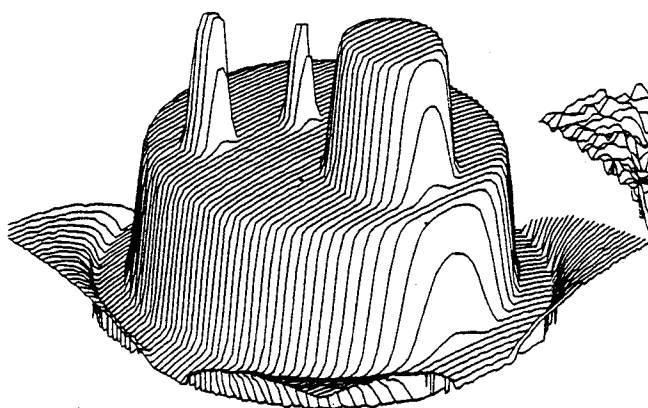


Fig.VI.6

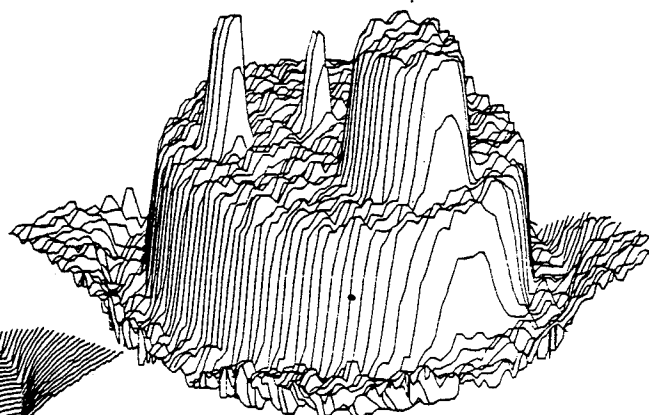


Fig.VI.9

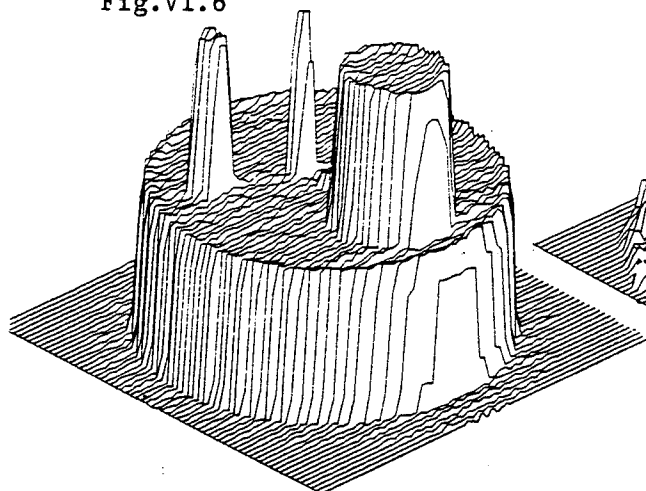


Fig.VI.7

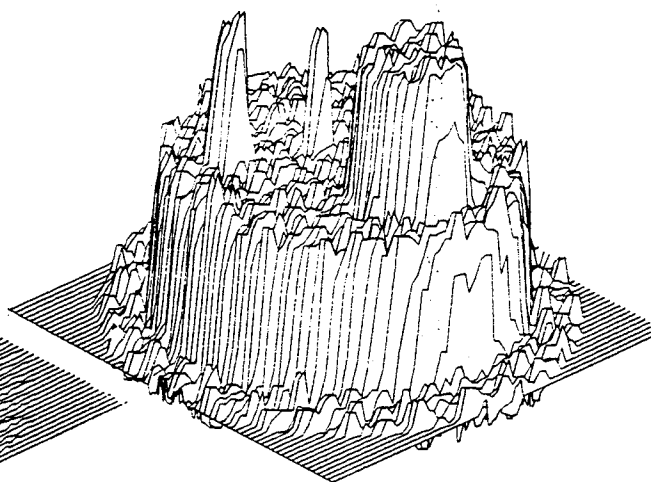


Fig.VI.10

Reconstruction avec 96 projections de 128 échantillons

sans bruit avec bruit

méthode analytique 1
 méthode analytique 2
 régularisation optimale

5	8
6	9
7	10

B. EN TOMOGRAPHIE A POSITRONS ET A TEMPS DE VOL

La géométrie utilisée pour les expérimentations est celle décrite au V.D. : 96 histoprojections de 128 x 64 données.

La largeur des bandes supports des $\phi_{k,p}^{\ell}$ (définis au I.B.4) est supposée nulle pour être dans le cadre d'application de la méthode analytique décrite au II.A.2.

L'unité étant encore le pas entre les 128 lignes parallèles, l'écart-type de l'erreur de positionnement le long de chaque ligne est pris $\sigma = 8$ et la longueur des classes de positions $\Delta v = 3$.

L'expérimentation a été réalisée avec l'objet test

$$f^0 = C [X_{D(0,0,30)} + X_{D(0,15,15)} + X_{D(0,20,4)} + X_{D(0,10,1)}]$$

C étant choisi de manière à ce que le nombre total de coïncidences détectées ($\sum y_{k,p,\ell}$) soit environ 2 millions, où chaque $y_{k,p,\ell}$ a été généré comme la réalisation d'une variable aléatoire poissonnienne de paramètre :

$$y_{k,p,\ell}^0 = \int_{C_{\ell}} H_{\theta_k} f^0(u_p, v) dv \quad (\text{cf V.C et I.B.3})$$

Comme précédemment on présente les reconstructions produites par deux algorithmes analytiques (cf II.A.2) différents par un filtre atténuant peu les hautes fréquences pour la méthode 1, et un peu plus pour la méthode 2. Et on les compare à la reconstruction produite sur la grille polaire 192 x 64 par la méthode de régularisation optimale par l'énergie et l'énergie du gradient, avec validation croisée.

Pour cela il suffit d'observer la coupe (128 échantillons) de ces 3 reconstructions le long de l'axe $y = 10$ (passant par le plus petit détail de rayon 1) : figure VI.11 et VI.12 pour les algorithmes analytiques et figure VI.13 pour la régularisation optimale. On observe le bon compromis réalisé par la validation croisée.

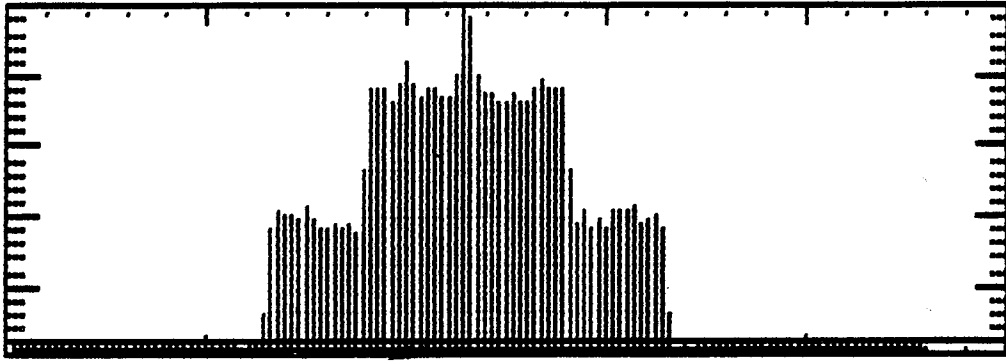


Fig.VI.11 : méthode analytique 1

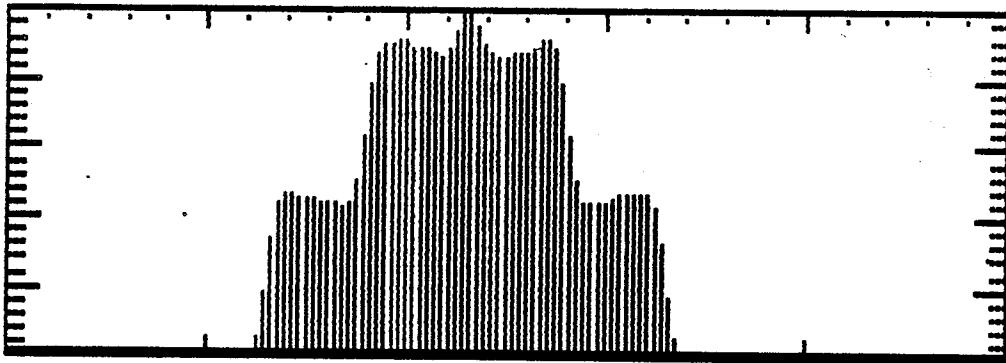


Fig.VI.12 : méthode analytique 2

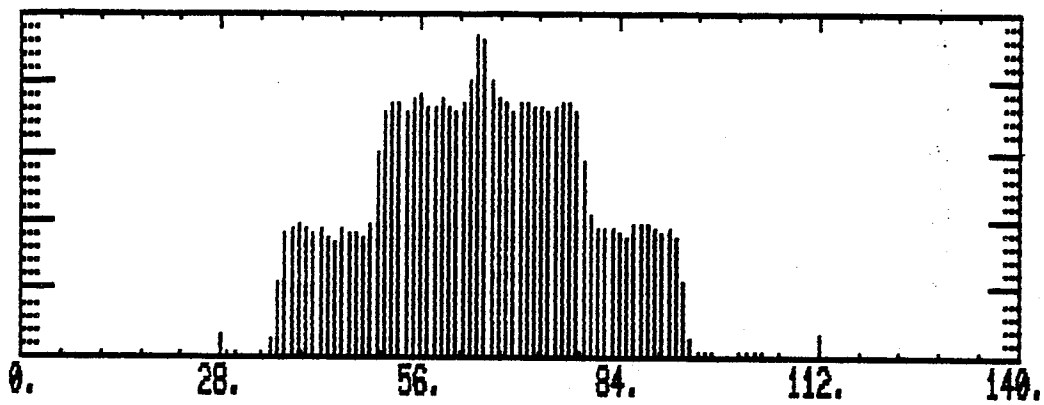


Fig.VI.13 : régularisation optimale

coupe y = 10 des 3 reconstructions avec
96x128x64 données temps de vol

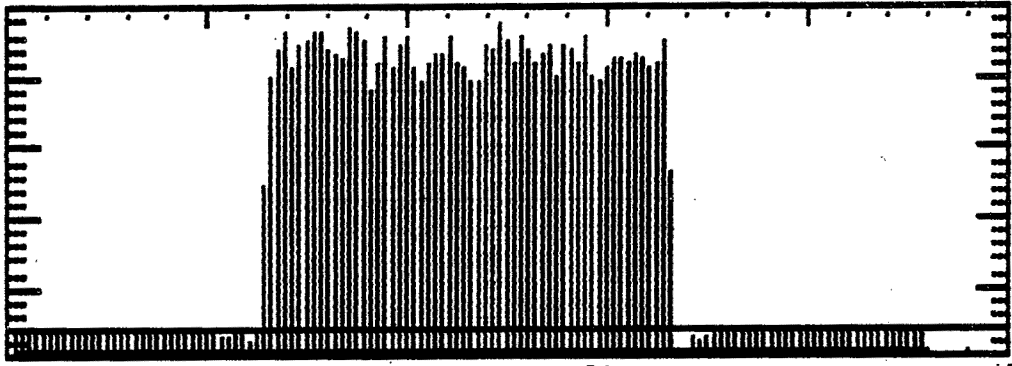


Fig.VI.14 : idem fig.VI.11

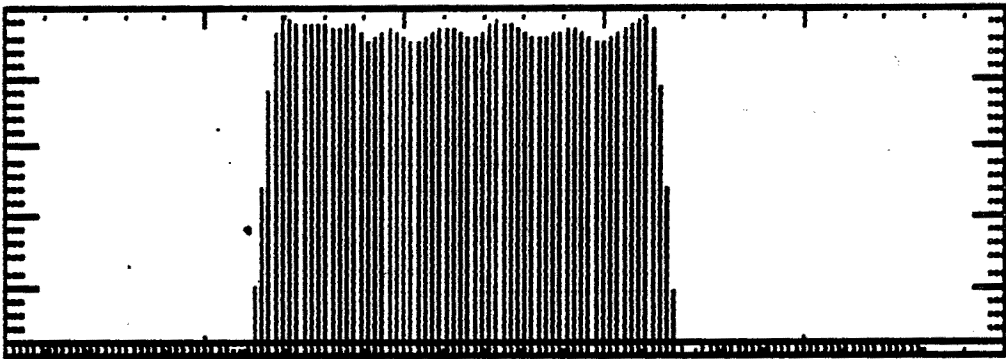


Fig.VI.15 : idem fig.VI.12

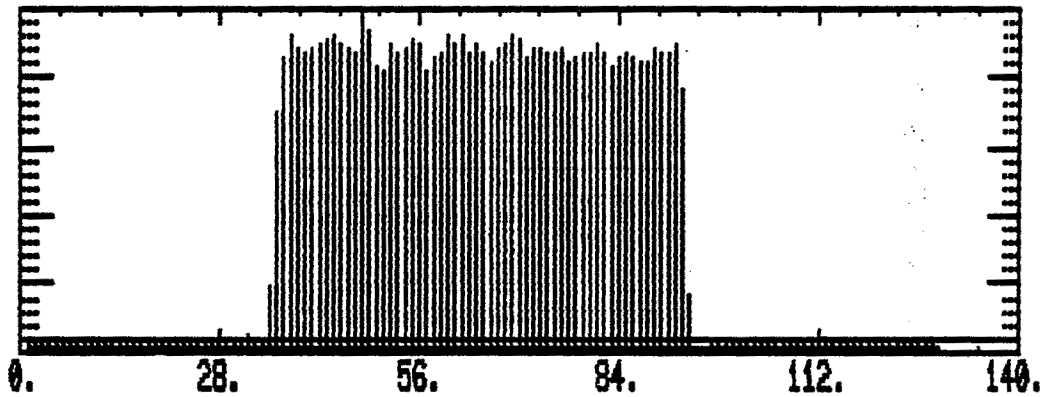


Fig.VI.16 : idem fig.VI.13

coupe $y = -10$ des 3 reconstructions avec
96x128x64 données temps de vol

Les coupes le long de l'axe $y = -10$ (resp. fig. VI.14, 15 et 16) montrent le lissage effectué par la régularisation optimale là où f^0 est constante.

On remarque que la reconstruction produite par la régularisation optimale est celle qui a le moins de valeurs non nulles hors du disque support de f^0 .

C - DISCUSSION

D'autres expérimentations numériques, dont quelques unes ont été réalisées avec des données réellement acquises, ont confirmé les performances de ces méthodes de régularisation optimale et on peut énoncer les conclusions suivantes :

- avec un nombre réduit de données ou non uniformément réparties, la supériorité de ces méthodes est flagrante. Ceci se comprend puisque l'utilisation au mieux des informations mesurées et l'incorporation d'une information a priori sont toutes deux indispensables dans ce contexte de données incomplètes et seules ces méthodes les réalisent conjointement.
- avec un grand nombre de données uniformément réparties (en tomographie classique ou avec temps de vol) le type de solution produit par les méthodes analytiques, quand de bons paramètres sont choisis, diffèrent peu visuellement des solutions de régularisations optimales. Ceci confirme les performances habituellement admises des méthodes analytiques malgré leur utilisation un peu brutale (ou optimiste si l'on préfère) des informations mesurées puisqu'on ne contrôle jamais l'écart à ces mesures. Mais l'avantage indéniable des méthodes de régularisation optimale est leur adaptabilité à travers le calcul du paramètre déterminant le meilleur compromis 'résolution insensibilité au bruit'.

Que ce calcul soit toujours efficace même avec la méthode de validation croisée qui ne nécessite aucune information a priori supplémentaire peut paraître assez impressionnant, mais s'explique sans doute par les redondances existant toujours dans des données sans bruit.

Les estimations d'erreur rendant ce calcul possible ont aussi un attrait en elles-mêmes puisqu'elles fournissent une évaluation de la qualité de chaque reconstruction.

CONCLUSION



CONCLUSION

Cette thèse propose un nouveau type de méthodes de reconstruction d'images, après la formulation de ces problèmes (chapitre I) et un survol indispensable des algorithmes connus (chapitre II).

L'approche théorique (chapitre III) conduisant à ces méthodes montre l'attrait des solutions que l'on y définit : parmi toutes les fonctions de deux variables satisfaisant convenablement les données (relativement à leur précision) on choisit la moins irrégulière.

Après l'étude de différents critères d'irrégularité, on expose les méthodes envisageables (du moins théoriquement) pour calculer la valeur optimale du paramètre de régularisation déterminant l'ajustement aux données (ou, si l'on préfère, le compromis 'résolution - insensibilité au bruit'). Parmi celles-ci (méthode du résidu, statistique, validation croisée), une nouvelle méthode statistique est proposée.

Les principaux résultats pratiques obtenus concernent l'élaboration d'algorithmes de reconstruction réalisant ces régularisations optimales pour un coût compétitif avec celui des algorithmes standard (chapitres IV et V). On a eu besoin ici d'un résultat nouveau de réduction de la taille des données (m) à celle des inconnues (n) pour appliquer ces méthodes d'ajustement automatique (annexe 2) avec $m \gg n$ comme en tomographie à temps de vol et on a tiré profit des symétries de révolution propres à ces tomographes pour traiter de grande valeur de n (10^5).

L'implantation de ces algorithmes sur l'ordinateur VAX 11/780 du LETI/MCTE a permis de réaliser diverses expérimentations numériques en tomographie classique et avec temps de vol.

On a noté l'efficacité assez impressionnante de la méthode de validation croisée qui devine, en quelque sorte, le niveau de bruit des données.

Au sujet de la performance de ces reconstructions, on peut conclure qu'en présence d'un nombre limité de données une méthode de régularisation optimale produit toujours un meilleur résultat, et qu'avec suffisamment de données une méthode analytique produit d'aussi bons résultats, mais à

condition de bien choisir le filtre utilisé, ce qu'on ne sait faire qu'empiriquement.

Il faut noter que ces comparaisons ont été faites dans les contextes où les méthodes analytiques standards sont applicables.

L'avantage qu'ont les méthodes de régularisation optimale de pouvoir prendre en compte des modèles d'acquisition plus complexes peut s'avérer très intéressant dans de nombreuses situations : projections tronquées (cf IV.B.5), largeur des bandes d'intégration non négligeable et sensibilité variable d'un couple de détecteurs à un autre comme en tomographie par émission. Tous ces aspects n'apparaissent dans les algorithmes proposés qu'à travers les coefficients de la matrice H et ne modifient pas le coût de ces reconstructions s'ils conservent l'invariance par rotation initiale.

Des expériences numériques supplémentaires s'avèreraient très utiles pour évaluer l'importance de l'amélioration des reconstructions obtenues avec cette prise en compte.

Le rôle, déjà considérable, de ces systèmes d'imagerie, dans l'élaboration du diagnostic médical, justifie ces efforts algorithmiques visant à utiliser au mieux les données mesurées, sans en rendre le traitement plus complexe.

Annexe 1

ETUDE GENERALE DE
L'EXISTENCE, L'UNICITE ET LA CARACTERISATION
DE LA SOLUTION REGULARISEE DE PARAMETRE ρ

Rappelons que R est l'opérateur de $L_2(\mathbb{D})$ dans \mathbb{R}^m , $Rf = ((\phi_1, f), \dots, (\phi_m, f))^t$, et $y \in \mathbb{R}^m$ est le vecteur des données, $\|y\|_S^2 = y^t S y$ où S matrice symétrique définie positive.

§1. Définition 1

Soit $F, (\cdot, \cdot)_F$ un sous-espace de Hilbert de $L_2(\mathbb{D})$ tel que R est continue de F dans \mathbb{R}^m

Soit $G, (\cdot, \cdot)_G$ un espace de Hilbert et T un opérateur linéaire continu surjectif de F dans G , de noyau $\mathcal{N}(T)$ de dimension fini .

Le problème $P_\rho(F, G, T)$ où ρ est un paramètre réel $\rightarrow 0$ est par définition :

Trouver $f_\rho \in F$ tel que

$$\|Rf_\rho - y\|_S^2 + \rho \|Tf_\rho\|_G^2 = \min_{f \in F} (\|Rf - y\|_S^2 + \rho \|Tf\|_G^2)$$

On a besoin du théorème suivant :

Théorème 1 : Si $\mathcal{N}(T) \cap \mathcal{N}(R) = \{0\}$, alors :

$P_\rho(F, G, T)$ a une solution unique et on a la caractérisation suivante :

$$f_\rho \text{ est solution} \Leftrightarrow \forall f \in F \quad \langle Rf_\rho, Rf \rangle_S + \rho \langle Tf_\rho, Tf \rangle_G = \langle y, Rf \rangle_S$$

Démonstration

Ceci est un cas particulier des résultats plus généraux que l'on peut trouver au chapitre IV du livre de P.J. Laurent (1972). Nous décrivons ici la démarche et les théorèmes fondamentaux utilisés. On introduit l'espace produit cartésien $P = \mathbb{R}^m \times G$ muni du produit scalaire $((p_1, p_2)) = \langle y_1, y_2 \rangle_S + \rho \langle g_1, g_2 \rangle_G$ où $p_1 = [y_1, g_1]$ et $p_2 = [y_2, g_2]$ appartiennent à P et de la norme induite :

$$\|p_1\|^2 = \|y_1\|_S^2 + \rho \|g_1\|_G^2$$

Il est facile de montrer que $(P, \|\cdot\|)$ est un espace de Hilbert.

Maintenant notons L l'application qui à $f \in F$ associe $Lf = [Rf, Tf]$, élément de P .

Notons $p = [y, 0] \in P$

Le problème $P_\rho(F, G, T)$ s'écrit alors :

$$\text{trouver } f_\rho \in F / \left\| \int f_\rho - p \right\|^2 = \text{Min}_{f \in F} \left\| \int f - p \right\|^2$$

Or parce que $\mathcal{N}(T) \cap \mathcal{N}(R) = \{0\}$, L est une bijection de F dans $R(L) = \{v \in P / \exists f \in F / v = Lf\}$.

On est donc ramené au problème équivalent :

$$\text{trouver } v_\rho \in R(L) / \left\| v_\rho - p \right\|^2 = \text{Min}_{v \in R(L)} \left\| v - p \right\|^2$$

Maintenant en utilisant les hypothèses R et T continus, T surjectif, on pourrait montrer que $R(L)$ est un sous-espace fermé de P (c'est l'étape la plus délicate, analysée au théorème (4.4.3) de P.J. Laurent).

Alors le problème est celui de la recherche de la meilleure approximation de p dans $R(L)$, posée dans P . C'est le théorème fondamental de projection orthogonale sur un sous-espace fermé d'un espace de Hilbert qui nous assure alors l'existence et l'unicité de $v_\rho = Lf_\rho$ et sa caractérisation :

$v_\rho - p$ est orthogonal à $R(L)$, c'est-à-dire $\forall v \in R(L), ((v_\rho - p, v)) = 0$

Substituer $[\int f_\rho, T f_\rho]$ à v_ρ et $\{[\int f, T f] / f \in F\}$ à $\{v / v \in R(L)\}$ donne la caractérisation de f_ρ énoncée dans le théorème.

Corrolaire 1 :

Autre caractérisation équivalente de f_ρ .

On a immédiatement (avec toujours $\mathcal{N}(T) \cap \mathcal{N}(R) = \{0\}$) :

$$f_\rho \text{ est la solution de } P_\rho(F, G, T) \Leftrightarrow \exists \lambda \in \mathbb{R}^m / \begin{cases} \forall f \in F (T f_\rho, T f)_G = \langle \lambda, R f \rangle \\ \rho \lambda = S(y - R f_\rho) \end{cases}$$

où $\langle \lambda, R f \rangle$ s'écrit aussi $(\sum_{i=1}^m \lambda_i \phi_i, f)_{L_2}$ avec $\lambda = (\lambda_1, \dots, \lambda_m)^t$

§2. Définition 2 :

Si y appartient à l'image de R , sous les mêmes hypothèses définissant F, G, T , on peut définir le problème $P_0(F, G, T)$ d'interpolation des mesures avec $\| \int f \|$ minimum :

$$\text{trouver } f_0 \in F / R f_0 = y \text{ et } \forall f \in F, R f = y \Rightarrow \| \int f \| \geq \| \int f_0 \|$$

On pourrait montrer le théorème suivant (arguments analogues) qui nous dit qu'il suffit de poser $\rho = 0$ dans la caractérisation précédente :

Théorème 2 : Si $\mathcal{N}(T) \cap \mathcal{N}(R) = \{0\}$

$P_0(F, G, T)$ a une solution unique et on a la caractérisation

$$f_0 \text{ solution} \Leftrightarrow \begin{cases} \exists \lambda \in \mathbb{R}^m / \forall f \in F (Tf_0, Tf)_G = \langle \lambda, Rf \rangle \\ Rf_0 = y \end{cases}$$

§ 3. Applications

a) Cas où $F, (,)_F = G, (,)_G, T = I$

Alors il suffit de vérifier que R est continu de F dans \mathbb{R}^m pour pouvoir appliquer ce théorème, qui donne :

la minimisation de $\|Rf - y\|_S^2 + \rho \|f\|_F^2$ dans F a une solution unique f_ρ caractérisée par:

$$\exists \lambda \in \mathbb{R}^m / \begin{cases} \forall f \in F (f_\rho, f)_F = \langle \lambda, Rf \rangle \\ \rho \lambda = S(y - Rf_\rho) \end{cases}$$

Un tel exemple est $F = L_2(\mathbb{D})$, $(f, g)_F = \int_{\mathbb{D}} w(x, y) f(x, y) g(x, y) dx dy$, w fonction poids continue > 0

R est continu car :

$$(\phi_i, f)_{L_2} = \left(\frac{\phi_i}{w}, f\right)_F \leq \left\| \frac{\phi_i}{w} \right\|_F \|f\|_F \text{ d'après l'inégalité de Schwarz}$$

et $\forall f \in F (f_\rho, f)_F = \langle \lambda, Rf \rangle$ s'écrit $\forall f \in L_2(\mathbb{D}) (wf_\rho, f)_{L_2} = \left(\sum_{i=1}^m \lambda_i \phi_i, f\right)_{L_2}$

$$\text{ce qui entraîne } f_\rho = \frac{1}{w} \sum_{i=1}^m \lambda_i \phi_i$$

b) Cas où $F = H^1(\mathbb{D})$ (défini au III.B.3),

$$\|f\|_{H^1}^2 = \|f\|_{L_2}^2 + \left\| \frac{\partial f}{\partial x} \right\|_{L_2}^2 + \left\| \frac{\partial f}{\partial y} \right\|_{L_2}^2$$

La continuité de R sur $L_2(\mathbb{D})$ entraîne celle sur F .

Posons $G = L_2(\mathbb{D}) \times L_2(\mathbb{D})$ si $g = [g_1, g_2]$, $g' = [g'_1, g'_2] \in G$

$$\text{alors } (g, g')_G = (g_1, g'_1)_{L_2} + (g_2, g'_2)_{L_2}$$

$$T : f \in H^1(\mathbb{D}) \rightarrow \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right] \in G$$

Il est facile de vérifier que T satisfait les hypothèses de la définition.
Le théorème devient :

si $\mathcal{N}(T) \cap \mathcal{N}(R) = \{0\}$, la minimisation de $\|Rf - y\|_S^2 + \rho \left(\left\| \frac{\partial f}{\partial x} \right\|_{L_2}^2 + \left\| \frac{\partial f}{\partial y} \right\|_{L_2}^2 \right)$

a une solution unique f_ρ caractérisée par :

$$\begin{cases} \exists \lambda \in \mathbb{R}^m / \forall f \in H^1(\mathbb{D}), \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial x} \right)_{L_2} + \left(\frac{\partial f}{\partial y}, \frac{\partial f}{\partial y} \right)_{L_2} = \langle \lambda, Rf \rangle \\ \rho \lambda = S(y - Rf_\rho) \end{cases}$$

§ 4. Remarque

Du théorème 1, on déduit facilement que :

$$\forall f \in F, \quad \|T(f_\rho - f)\|^2 - \|Tf\|^2 = \frac{1}{\rho} \langle y - Rf_\rho, Rf_\rho - 2Rf \rangle_S$$

En effet la caractérisation de f_ρ appliquée à $f = f_\rho$ donne

$$\langle Rf_\rho, Rf_\rho \rangle_S + \rho (Tf_\rho, Tf_\rho)_G = \langle y, Rf_\rho \rangle_S$$

D'où en la combinant avec la formule générale :

$$\rho (Tf_\rho, Tf_\rho)_G - 2\rho (Tf_\rho, Tf)_G = \langle y - Rf_\rho, Rf_\rho \rangle - 2 \langle y - Rf_\rho, Rf \rangle_S$$

Annexe 2

PRACTICAL OPTIMAL REGULARIZATION
OF LARGE LINEAR SYSTEMS



PRACTICAL OPTIMAL REGULARIZATION
OF LARGE LINEAR SYSTEMS

ABSTRACT

We seek an estimate of $x^0 \in \mathbb{R}^n$, given a noisy value ($y \in \mathbb{R}^m$) of $y^0 = Kx^0$ where K is a known $m \times n$ matrix. The regularization method produces a set of estimates x_ρ obtained by minimizing $J_\rho(x)$ in \mathbb{R}^n :

$$J_\rho(x) = \|Kx - y\|_m^2 + \rho \|Lx\|_{n-p}^2$$

where $\|\cdot\|_m$ is the Euclidean norm in \mathbb{R}^m , L is an a priori chosen $(n-p) \times n$ matrix and ρ is the regularization parameter.

Three methods which give an automatic choice (ρ^*) of ρ are recalled. We study here the numerical implementation of these practical approximations and present, for each of the three methods, an algorithm which estimates ρ^* and computes x_{ρ^*} (the minimizer of J_{ρ^*}) with a cost (number of operations and memory place) nearly identical to the minimum cost of the computation of one estimate $x_\rho = (K^t K + \rho L^t L)^{-1} K^t y$ (the preliminary computations, independent of y , are not taken in account).

This algorithm is applied to a tomographical picture reconstruction problem where K is a large sparse matrix ($m \gg n$) and $K^t K$ and L are block-circulant. Simulations show that the cross-validation method is impressively efficient at choosing ρ^* .

REGULARISATION OPTIMALE PRATIQUE DE GRANDS SYSTEMES LINEAIRES

RESUME

On recherche une approximation de $x^0 \in \mathbb{R}^n$ à partir de la connaissance d'une valeur approchée y de $y^0 = Kx^0$ où K est une matrice (m,n) connue. La méthode de régularisation étudiée consiste à choisir une matrice L $(n-p,n)$ telle que la norme euclidienne $\|Lx\|_{n-p}$ mesure l'irrégularité de x , et à minimiser dans \mathbb{R}^n ,

$$J_{\rho}(x) = \|Kx-y\|_m^2 + \rho \|Lx\|_{n-p}^2$$

où ρ est le paramètre de régularisation >0 . 3 critères de choix d'un paramètre ρ^* optimal sont rappelés. On étudie ici la mise en oeuvre numérique d'une telle approximation quand $m \gg n$. Après s'être ramené à un problème équivalent à n équations, on propose une méthode de choix automatique de ρ^* pour chacun des 3 critères, et de calcul de x_{ρ^*} minimisant $J_{\rho^*}(x)$, pour un coût total (place en mémoire et temps de calcul) comparable au coût minimum de la résolution directe de l'équation normale $(K^tK + \rho L^tL)x_{\rho} = K^ty$, sans prendre en compte les pré-calculs indépendants de K^ty .

Cette méthode est appliquée à un problème de reconstruction d'image en tomographie où K est une grande matrice creuse, K^tK et L sont bloc-circulantes. On vérifie l'efficacité de la validation croisée comme critère de choix de ρ^* .

§ 1 - Introduction

Let K be a real $m \times n$ matrix. We consider the system of linear equations :

$$Kx = y^0 = y + \epsilon \quad (1.1)$$

where we know only a noisy value ($y \in \mathbb{R}^m$) of the exact data vector y^0 and ϵ is a random noise vector. If we assume a "white noise" :

$$\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_m)^t \text{ with } E(\epsilon_i) = 0, E(\epsilon_i \epsilon_j) = \sigma^2 \delta_{ij} \quad i, j = 1, \dots, m \quad (1.2)$$

then an intuitive idea is to say that given y and σ^2 , there is only one constraint on x :

$$S(x) = \| Kx - y \|_m^2 \leq m \cdot \sigma^2 \quad (1.3)$$

where $\| \cdot \|_m$ is the usual Euclidean norm in \mathbb{R}^m .

We are interested in the situations where either because σ is too large or because K is too ill-conditioned, a wide range of values of x may satisfy constraint (1.3), and therefore we need prior information concerning the solution. Defining the irregularity of x by $\|Lx\|_{n-p}$ where L is a chosen $(n-p) \times n$ matrix (when the origin of system (1.1) is the discretization of an integral equation, L is normally a discrete approximation of a derivative operator), we take as the solution the least irregular x satisfying (1.3). Except for some incoherent value of σ , such a solution minimizes $J_\rho(x)$ in \mathbb{R}^n :

$$J_\rho(x) = \|Kx - y\|_m^2 + \rho \|Lx\|_{n-p}^2 \quad (1.4)$$

where $\rho > 0$ is the regularization parameter that we must choose.

It is well known that when $N(K) \cap N(L) = \{0\}$ (where $N(K)$ is the null space of K), for every $\rho > 0$, minimization of J_ρ has only one solution x_ρ given by :

$$(K^t K + \rho L^t L) x_\rho = K^t y \quad (1.5)$$

The value of ρ is critical to the quality of x_ρ : if ρ is too small the data error induces a solution which is too irregular and if ρ is too large x_ρ is less sensitive to noise but Kx_ρ may be too far from the data.

In this paper we study methods of automatic choice of ρ and we show some important simplifications for the numerical implementation with $m \gg n$.

In § 2 we rapidly review three different methods. We shall see that the optimal ρ is the root of or minimizer of a real function of the variable ρ . As we have to approximate this optimal value by an iterative procedure, we need a fast algorithm for the evaluation of this function. For this evaluation, we have to compute $S(x_\rho) = \|Kx_\rho - y\|_m^2$ and except for the first method, the trace of the (m,m) matrix $A(\rho)$ satisfying $Kx_\rho = A(\rho)y$.

We will first transform our initial problem (1.4) into an equivalent problem (§3) with less than n equations :

$$\text{Min}_{x \in \mathbb{R}^n} (\tilde{S}(x) + \rho \|Lx\|_{n-p}^2)$$

where $\tilde{S}(x) = \|\tilde{K}x - \tilde{y}\|_\ell^2$, $\tilde{K} = F^t K$, $\tilde{y} = F^t y$, $\ell = \text{rank}(K)$

and F is any $m \times \ell$ matrix with columns f_i , $i = 1, \dots, \ell$, forming an orthogonal basis of the range of K .

Then we will show in § 4 that one can choose F in order to obtain the following result : if L is invertible, we can compute a $n \times \ell$ matrix B , and ℓ real values $d_i > 0$, $i = 1, \dots, \ell$, using the singular value decomposition of $L^{-1t} K^t K L^{-1}$, such that :

$$\begin{aligned} \tilde{A}(\rho) &= D^2(D^2 + \rho I)^{-1} & \text{where } \begin{cases} D \text{ is the diagonal matrix} \\ I \text{ is the identity in } \mathbb{R}^\ell \end{cases} & \begin{bmatrix} d_1 \\ d_2 \\ \dots \\ d_\ell \end{bmatrix} \\ \tilde{y} &= D^{-1} B^t K^t y \\ x_\rho &= B(D^2 + \rho I)^{-1} D \tilde{y} \end{aligned}$$

If L is not invertible, we obtain (§ 5) similar expressions, with or without boundary values on x_ρ .

In § 6 we numerically apply the above results to a tomographical picture reconstruction problem. Here $m = 96 \times 128 \times 64$ and $n = 96 \times 128$ but K is very sparse and as $K^t K$ and L are block-circulant the preliminary computation of B and D is much simplified and their memorisation takes up only $97 \times 64 \times 64$ matrices. Generalized cross-validation is implemented and simulations with generated noisy data show that it is very efficient.

§ 2 - Methods for choosing the parameter

These methods have been previously studied for the smoothing of data by Spline functions (see Reinsch [1], Wahba [4]) and they can be generalized ([3]) to any linear transformation of noisy data when this transformation is governed by a parameter which defines the degree of smoothing of these data. Here, noting $y_\rho = Kx_\rho$, the data vector predicted by x_ρ ,

$$A(\rho) = K(K^t K + \rho L^t L)^{-1} K^t \text{ is the } m \times m \text{ matrix satisfying } y_\rho = A(\rho)y$$

In the case of white noise with a known variance σ^2 (1.2) the generalisation of Reinsch's suggestion is to choose ρ_s so that :

$$S(x_{\rho_s}) = \|A(\rho_s)y - y\|_m^2 = m\sigma^2 \quad (2.1)$$

If there is not such a ρ_s , the constraint $S(x) \leq m\sigma^2$ either is never satisfied or is always satisfied with $<$ instead of \leq . Generally one such ρ_s exists and it is then easy to show that x_{ρ_s} is the solution of the problem :

$$\begin{aligned} & \text{Min} \quad \|Lx\|_{n-p}^2 \\ & x \in \mathbb{R}^n / S(x) \leq m\sigma^2 \end{aligned}$$

Using the exact data vector y^o , Wahba has defined in [2] the optimum ρ_0 as the ρ which minimizes $R(\rho)$, the mean square error of prediction of this vector :

$$R(\rho) = \frac{1}{m} \|A(\rho)y - y^o\|_m^2 \quad (2.2)$$

$R(\rho)$ cannot be known in practice, but we can easily demonstrate ([4]) that, with white noise, $\hat{R}(\rho)$, which is computable, is an unbiased estimate of $E(R(\rho))$:

$$\hat{R}(\rho) = \frac{1}{m} \|A(\rho)y - y\|_m^2 - \frac{\sigma^2}{m} \text{Trace } (A(\rho) - I)^2 + \frac{\sigma^2}{m} \text{Trace } A(\rho)^2 \quad (2.3)$$

The search of ρ which minimizes \hat{R} constitutes, then, a second method. The third one needs no statistical knowledge concerning the errors : it consists of choosing the ρ_v which minimizes $V(\rho)$:

$$V(\rho) = \frac{\frac{1}{m} \|A(\rho)y-y\|^2}{\left|\frac{1}{m} \text{Trace} (A(\rho) - I)\right|^2} \quad (2.4.)$$

This is Wahba's method of generalized cross-validation ([3],[4]). It originates from a simple idea (Stone [5]) : one omits a datum and studies the validity of different ρ by measuring how the corresponding estimates x_ρ , computed without this datum, can predict it. We choose the ρ which gives the best prediction, on the average, for all the possible omissions. Wahba has also shown that when white noise is assumed, ρ_v is a good estimate of the minimizer of $E(R(\rho))$. (and is even better when m is large).

§ 3 - Equivalent problem

We remark that for every ρ the regularized solution x_ρ does not depend directly on y but on $K^t y \in \mathbb{R}^n$. In fact we can transform our problem into an equivalent problem made of ℓ linear equations with $\ell = \text{rank} (K)$. Let P be the projection in \mathbb{R}^m on the range of K , $R(K)$. Then,

$$\|Kx-y\|_m^2 = \|Kx-Py\|_m^2 + \|Py-y\|_m^2$$

And, as $Kx-Py \in R(K)$, if F is any $m \times \ell$ matrix, with columns $f_1, f_2, \dots, f_\ell \in \mathbb{R}^m$ forming an orthogonal basis of $R(K)$, then we have :

$$\|Kx-Py\|_m^2 = \|F^t(Kx-Py)\|_\ell^2 \quad \text{and} \quad F^t(Py-y) = 0$$

Then, letting $\tilde{K} = F^t K$, $\tilde{y} = F^t y$, the minimization (1.4) is equivalent to

$$\text{Min}_{x \in \mathbb{R}^n} (\|\tilde{K}x-\tilde{y}\|_\ell^2 + \rho \|Lx\|_{n-p}^2)$$

On the basis of this formulation we will now choose the parameter ρ .

When we assume white noise with variance σ^2 then a white noise is still satisfied by \tilde{y} (because the covariance matrix of $F^t \varepsilon$ is $F^t (\sigma^2 I_m) F = \sigma^2 I_\ell$), and the methods 1 and 2 are applicable.

If we let $\tilde{y}^{\circ} = F^t y^{\circ}$ then $\| Kx_{\rho} - y^{\circ} \|_m^2$ is equal, if $y^{\circ} \in R(K)$, to $\| \tilde{K}x_{\rho} - \tilde{y}^{\circ} \|_{\ell}^2$ or else differs by a constant which is independent of ρ , so the optimum ρ (following Wahba(2.2)) is the same in the two problems. The generalized cross-validation applied to $\tilde{A}(\rho) = \tilde{K}(\tilde{K}^t \tilde{K} + \rho L^t L)\tilde{K}^t$ and \tilde{y} can give an estimate of this optimum.

In the numerical implementation of these methods, we have to evaluate $\tilde{S}(x_{\rho}) = \| \tilde{A}(\rho)\tilde{y} - \tilde{y} \|_{\ell}^2$ and $\text{Trace}(\tilde{A}(\rho))$ for several values of ρ . We see below how to choose F in order to simplify these evaluations.

§ 4 - When L is invertible, we can write :

$$\begin{aligned} x_{\rho} &= (\tilde{K}^t \tilde{K} + \rho L^t L)^{-1} \tilde{K}^t \tilde{y} \\ &= L^{-1} (L^{-1t} \tilde{K}^t \tilde{K} L^{-1} + \rho I_n)^{-1} L^{-1t} \tilde{K}^t \tilde{y} \end{aligned} \quad (4.1.)$$

Consider now the singular value decomposition of KL^{-1} (see Golub [6]) :

$$KL^{-1} = UD V^t$$

Where U and V are $m \times l$ and $n \times l$ orthogonal matrices and D is the diagonal matrix $\text{diag}(d_1, d_2, \dots, d_l)$, with $d_1 \geq d_2 \geq \dots \geq d_l > 0$.

Then we can take $F = U$, so that :

$$\tilde{K}L^{-1} = DV^t = M \quad (4.2.)$$

and

$$\tilde{y} = U^t y = D^{-1} V^t L^{-1t} K^t y \quad (4.3.)$$

Substituting (4.2) and (4.3) in (4.1) gives :

$$\begin{aligned} Lx_{\rho} &= (M^t M + \rho I_n)^{-1} M^t \tilde{y} = M^t (MM^t + \rho I_l)^{-1} \tilde{y} \\ &= VD(D^2 + \rho I_l)^{-1} D^{-1} V^t L^{-1t} K^t y \end{aligned}$$

Now let $B = L^{-1} V$. Then one obtains :

$$x_{\rho} = B \text{diag}\left(\frac{1}{d_i^2 + \rho}, i = 1, \dots, l\right) B^t K^t y \quad (4.4.)$$

$$\tilde{A}(\rho) = MM^t (MM^t + \rho I_l)^{-1} = \text{diag}\left(\frac{d_i^2}{d_i^2 + \rho}, i = 1, \dots, l\right) \quad (4.5.)$$

$$\tilde{A}(\rho) \tilde{y} - \tilde{y} = \text{diag}\left(\frac{d_i}{d_i^2 + \rho} - \frac{1}{d_i}, i = 1, \dots, l\right) B^t K^t y \quad (4.6.)$$

Note that we don't need to compute U , and D^2 and V are given by the singular value decomposition of $L^{-1t} K^t KL^{-1}$. We observe that this method is well adapted to the case where K is a large sparse matrix (so that we can easily compute $K^t y$) and when we need to solve (1.1) a great number of times with different data vectors, since we can then precompute (the most expensive part) and store the $n \times l$ matrix B and the l values d_i . Then the calculation of an optimal parameter and of the corresponding regularized solution is reduced to the four following steps :

- 1) compute $K^t y$
- 2) compute $z = B^t (K^t y)$ ($O(n^2)$ operations)
- 3) search by an iterative procedure for ρ^* , the zero or the minimizer of a function (e.g. $\tilde{V}(\rho)$), computable with $O(n)$ operations
- 4) compute $x_{\rho^*} = B \text{diag}(\frac{1}{d_i^2 + \rho^*}, i = 1, \dots, \ell) z$ ($O(n^2)$ operations)

So the whole algorithm requires the same memory space and about twice the number of operations as the calculation of x_ρ (with a priori chosen ρ and memo-risation of $(K^t K + \rho L^t L)^{-1}$).

§ 5.1 - If L is not invertible, again we can easily obtain an expression of x_ρ similar to (4.1) as soon as we assume some auxiliary boundary value on the solution.

Since we choose L , we can assume its rank is $n-p$. A frequently occurring example in the case of an integral equation of one va-riable is the discrete approximation of the derivative :

$$L = \begin{bmatrix} 1-1 & & & (0) \\ & 1-1 & & \\ & & \ddots & \\ (0) & & & \ddots \\ & & & & 1-1 \end{bmatrix}, p = 1$$

Let N be the $n \times p$ matrix with columns forming a basis of $N(L)$.

Every $w \in \mathbb{R}^n$ has an unique expression $x = L^\dagger w + Nz$ where L^\dagger is the pseudo-inverse of L , $w = Lx \in \mathbb{R}^{n-p}$, $z \in \mathbb{R}^p$.

In order to be able to express x as a function of Lx , we must add p boundary values :

$$Z^t x = c \quad \text{where} \quad \left\{ \begin{array}{l} Z \text{ is a } n \times p \text{ matrix} \\ Z^t N \text{ invertible} \\ c \in \mathbb{R}^p \end{array} \right. \quad (5.1)$$

With (5.1), we get :

$$\begin{aligned} w = Lx &\iff x = L^{\dagger}w + N(Z^{\dagger}N)^{-1} (c - Z^{\dagger}L^{\dagger}w) \\ x &= (I - N(Z^{\dagger}N)^{-1} Z^{\dagger})L^{\dagger}w + N(Z^{\dagger}N)^{-1}c \end{aligned} \quad (5.2.)$$

Let $G = (I - N(Z^{\dagger}N)^{-1} Z^{\dagger})L^{\dagger}$. Thus the problem

$$\begin{aligned} \text{Min} \quad & (S(x) + \rho \|Lx\|_{n-p}^2) \\ \text{with } x & \in \mathbb{R}^n / Z^{\dagger}x = c \end{aligned}$$

becomes :

$$\begin{aligned} \text{Min} \quad & (\|KGw - (y - KN(Z^{\dagger}N)^{-1}c)\|_m^2 + \rho \|w\|_{n-p}^2) \\ \text{with } w & \in \mathbb{R}^{n-p} \end{aligned}$$

We then use our procedure of §3 and §4 with KG , I in place of K, L . Without repeating this procedure, we note only that the singular value decomposition introduced here is :

$$KG = UD V^{\dagger}$$

Then (4.4) gives :

$$w_{\rho} = V \text{diag}\left(\frac{1}{d_i^2 + \rho}, i = 1, \dots, \ell\right) V^{\dagger}G^{\dagger}K^{\dagger}(y - KN(Z^{\dagger}N)^{-1}c)$$

Substituting this in (5.2) :

$$x_{\rho} = GV \text{diag}\left(\frac{1}{d_i^2 + \rho}, i = 1, \dots, \ell\right) V^{\dagger}G^{\dagger}K^{\dagger}(y - KN(Z^{\dagger}N)^{-1}c) + N(Z^{\dagger}N)^{-1}c$$

and (4.5), (4.6) are now :

$$\tilde{A}(\rho) = \text{diag}\left(\frac{d_i^2}{d_i^2 + \rho}, i = 1, \dots, \ell\right)$$

$$\tilde{A}(\rho) \tilde{y} - \tilde{y} = \text{diag}\left(\frac{d_i}{d_i^2 + \rho} - \frac{1}{d_i}, i = 1, \dots, \ell\right) V^{\dagger}G^{\dagger}K^{\dagger}(y - KN(Z^{\dagger}N)^{-1}c)$$

In comparison with § 4, we have to compute G in place of L^{-1} and subtract from the data vector a linear combination of the p columns of $KN(Z^{\dagger}N)^{-1}$.

§ 5.2 - Furthermore, if we don't want to add some auxiliary constraint (5.1), we can use the following relation which is always satisfied by x_ρ :

$$N^t K^t K x_\rho = N^t K^t y$$

This relation is obtained by multiplying $(K^t K + \rho L^t L) x_\rho = K^t y$ by N^t .

We verify that $Z^t = N^t K^t K$ is such that $Z^t N$ is invertible. Indeed, $Z^t N = N^t K^t K N$ is symmetric and $z^t Z^t N z = \|K N z\|^2$ is > 0 , except for $N z = 0$ hence $z = 0$, as long as $N(K) \cap N(L) = \{0\}$.

In the previous results, the constant term in (5.2) is :

$$N(Z^t N)^{-1} c = N(N^t K^t K N)^{-1} N^t K^t y$$

Let x_∞ be this term, since it's the limit of x_ρ when $\rho \rightarrow +\infty$

We note that x_∞ is the element of $N(L)$ which minimizes $\|Kx - y\|$ for $x \in N(L)$ and G is now :

$$G = |I - N(N^t K^t K N)^{-1} N^t K^t K| L^t$$

§ 6 - A numerical application : positron emission tomography with time of flight measurements.

This tomograph has detectors placed in a circle around a positron-emitting object and we want to estimate the activity of the emitter at every point in the plane of the circle. The emission of one positron produces, at the same point, two gamma rays in opposite directions. Hence, if two detectors detect gamma rays at about the same time then we know that a positron emission has occurred on the line between these detectors. The time-of-flight measurement gives an approximate localisation of this emission along this line. In the tomograph of LETI [7] the acquisition system groups all detections in 96 directions uniformly distributed throughout 180°. For each direction there are 128 parallel lines and, on each line, 64 time-of-flight groupings. Using the rotational invariance of this geometry and reorganizing with 192 directions covering 360°, these $m = 96 \times 128 \times 64$ counts can be taken as approximations of :

$$y_{k,i,j}^{\circ} = \int g_j(v) f_k(u_i, v) dv \quad \begin{array}{l} k = 1, \dots, 192 \\ i = 1, \dots, 64 \\ j = 1, \dots, 64 \end{array} \quad (6.1)$$

where $\left\{ \begin{array}{l} f_k(u, v) \text{ is the expression of the search activity in the Cartesian frame} \\ \text{where the 64 lines in the direction } k \text{ satisfy :} \\ \\ u = u_i \text{ (= distance from the center) } \quad i = 1, \dots, 64 \\ \\ g_j \text{ is a known function given by : } g_j(v) = \int_{C_j} g(v-s) ds \text{ where } C_j \text{ is} \\ \text{the interval on } 0v \text{ of the positions corresponding}^j \text{ to the time-of-flight} \\ \text{class } j \text{ and } g \text{ is the probability density of localisation error (taken} \\ \text{as a given normal density)} \end{array} \right.$

Even if we refine this model, the measurement $y_{k,i,j}$ is only a Poisson random variable of unknown rate $y_{k,i,j}^{\circ}$ and has a small signal to noise ratio since in practice, $y_{k,i,j}^{\circ}$ is approximately between 0 and 100. Hence it is not worth searching for an estimate of f with m degrees of freedom.

A discretization able to represent possible solutions well, and considerably simplifying the computation is the discretization on the polar grid which has the same geometry as that of the measurements.

where $\left\{ \begin{array}{l} \text{a) } B_k \text{ are } 64 \times \ell_k \text{ matrices } (\ell_k \leq 64 \text{ and } \ell_1 + 2(\ell_2 + \dots + \ell_{9,6}) + \ell_{9,7} = \ell) \\ \text{b) multiplying } z \in \mathbb{R}^n \text{ by } W \text{ involves } 64 \text{ parallel discrete} \\ \text{Fourier transforms of order } 192. \end{array} \right.$

This reconstruction algorithm has been implemented on a VAX 11/780. The regularization parameter is chosen by the method of generalized cross-validation. The minimization of $\tilde{V}(\rho)$ (3.1, 4.5, 4.6) uses the golden section search method. In practice, about twenty iterations are sufficient.

In the numerical simulation presented, the measurements are generated with an analytic computation of exact data $y^\circ_{k,i,j}$ (6.1) where f is a given constant on four non intersecting discs (so that the average value of $y^\circ_{k,i,j}$ is about 4), and are perturbed with Poisson random noise.

We have represented in figure 1 the Euclidean distance $\tilde{S}(\rho)$ between the predicted data and the noisy data, the cross-validation function $\tilde{V}(\rho)$ that we minimize, and the distance $\tilde{R}(\rho)$ between the predicted data and the exact data. Letting $z = B^t K^t y$, $z^\circ = B^t K^t y^\circ$, these functions are computed by :

$$\tilde{R}(\rho) = \frac{1}{\ell} \left\| \tilde{A}(\rho) \tilde{y} - \tilde{y}^\circ \right\|_\ell^2 = \frac{1}{\ell} \sum_{i=1}^{\ell} \frac{1}{d_i^2} \left| \frac{d_i^2}{d_i^2 + \rho} z_i - z_i^\circ \right|^2$$

$$\tilde{S}(\rho) = \left\| \tilde{A}(\rho) \tilde{y} - \tilde{y} \right\|_\ell^2 = \rho \sum_{i=1}^{\ell} \frac{1}{d_i^2} \left| \frac{z_i}{d_i^2 + \rho} \right|^2$$

$$\tilde{V}(\rho) = \frac{1}{\ell} \frac{\tilde{S}(\rho)}{\left(\frac{\rho}{\ell} \sum_{i=1}^{\ell} \frac{1}{d_i^2 + \rho} \right)^2}$$

Figure 1 shows that the cross-validation method produces a very good estimate of the optimal regularization parameter.

We have performed this experiment using several different f of the same type that used in the simulation described above. Each simulation performed demonstrated a similar, high efficiency of the cross-validation method and we verified that the optimal parameter was dependent on the signal-to-noise ratio of the measurements and on the shape of f .

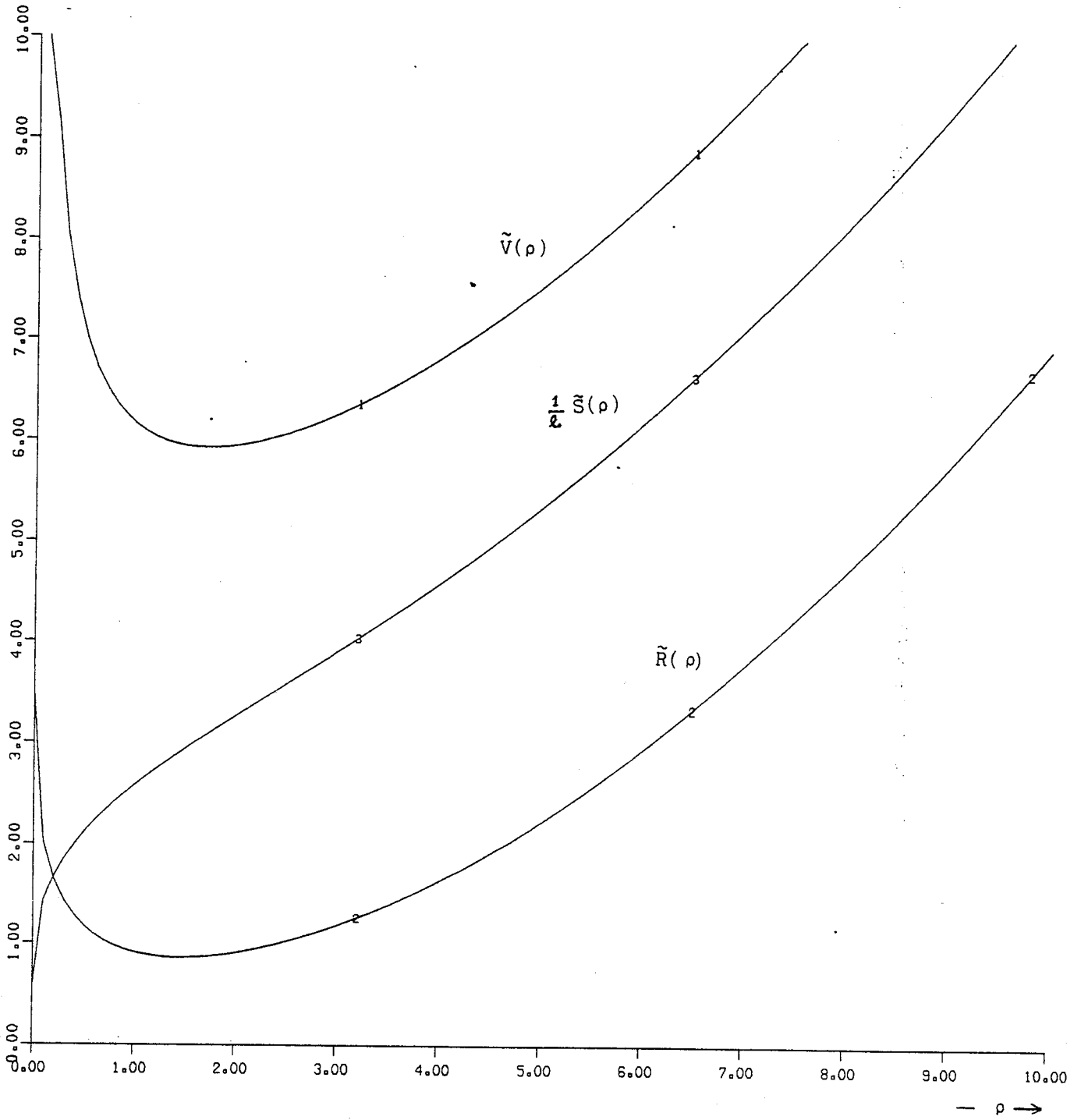


Figure 1

References

- [1] C.H. REINSCH "Smoothing by Spline functions II" Numer.Math.,
16 (1971), pp. 451-454

- [2] G. WAHBA "Smoothing noisy data with spline functions". Numer.Math.,
24 (1975), pp. 383-393

- [3] G. WAHBA "Practical approximate solutions to linear operator equations
when the data are noisy" SIAM, J. Num. Anal. 14 (1977), 651-667

- [4] P. CRAVEN, G. WAHBA "Smoothing noisy data with spline functions"
Numer.Math. 31 (1979) 377-403

- [5] M. STONE "Cross-validators choice and assessment of statistical predic-
tion (with discussion)"
J. Roy. Statist. Soc., Ser. B, 36 (1974) pp. 111-147

- [6] C. GOLUB, C. REINSCH "Singular value decomposition and least square
solution" Numer. Math. 14, 403-420 (1970)

- [7] R. ALLEMAND and al. "A new time-of-flight method for positron computed
tomography" in D.A.B. Lindberg and P.L. Reichertz, Eds., Biomedical
Images and Computers. Berlin : Springer, 1980.

* * * * *

BIBLIOGRAPHIE



- E. ARTZY, T. ELFVING, G.T. HERMAN (1979) 'Quadratic Optimization for Image reconstruction, II'. Computer Graphics and Image Processing 11, 242-261.
- J.P. BUTLER, J.A. REEDS, S.V. DAWSON (1981) 'Estimating solutions of first kind integral equations with nonnegative constraints and optimal smoothing' SIAM J. Num. Anal., 18 : 381-397
- R.E. CAMPAGNOLO (1980) "Utilisation de la technique de temps de vol pour l'amélioration de l'imagerie en tomographie par positons"
Thèse, Université de Grenoble
- T. CHANG and G.T. HERMAN (1980) "A scientific study of filter selection for a fan-beam convolution reconstruction algorithm", SIAM J. Appl. Math., 39, pp 83-105.
- A.M. CORMACK (1963) 'Representation of a function by its line integrals, with some radiological applications', J. Appl. Phys., 9 : 2722-2727 and (1964) 10 : 2908-2913
- A.M. CORMACK (1980) 'Early Tomography and related topics' in G.T. Herman and F. Natterer, Eds., 'Mathematical Aspects of Computerized Tomography' Berlin : Springer, 1981.
- J. DUCHON (1976) 'Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces' RAIRO, Analyse Num. vol 10, n° 12 : 5-12
- N. DYN and G. WAHBA (1979) 'On the estimation of functions of several variables from aggregated data' University of Wisconsin-Madison MRC. Tech. Sum. Rep. n° 1974
- G.T. HERMAN (1980) 'Image Reconstruction from Projections : The Fundamentals of Computerized Tomography. New York : Academic Press.
- J.W. HILGERS (1982) 'A theory for Optimal regularization in the finite dimensional case' Linear Alg. and Its Applications 48 : 359-379.
- IMSL (1980) International Mathematical and Statistical Libraries Inc. Edition 8. Subroutine ICSSCV, Houston, TX.
- K. LANGE, R. CARSON (1984) "EM reconstruction algorithms for emission and transmission tomography" J. Comput. Assist. Tomogr. 8(2) : 306-316

- P.J. LAURENT (1972) 'Approximation et optimisation' Hermann
- J. LLACER (1982) "Tomographic Image Reconstruction by eigenvector decomposition : its limitations and areas of applicability"
IEEE Trans. On Medical Imaging, MI 1 : 34-42.
- A.K. LOUIS (1981) ' Ghosts in tomography - the null space of the Radon transform' Math. Methods Appl. Sci. 3, pp 1-10.
- D. LUDWIG (1966) 'The Radon Transform on Euclidean space'
Comm. on Pure and Appl. M th. 19 : 49-81.
- R.B. MARR (1974) 'On the reconstruction of a function on a circular domain from a sampling of its line integrals'
J. Math. Anal. Appl., vol. 19, 357-374.
- F. NATTERER (1977) 'The finite element method for ill-posed problems'
R.A.I.R.O. Anal. Num. 11 : 271-278.
- F. NATTERER (1980) 'Efficient Implementation of 'optimal' algorithms in computerized tomography' Math. Meth. In the Appl. Sci. 2 : 545-555.
- F. NATTERER (1980.a) 'A Sobolev Space Analysis of picture reconstruction'
SIAM J. Appl. Math., 39 : 402-411.
- R.M. PERRY (1979) "Reconstructing a function by circular harmonic analysis of its Radon transform' in Dig. Topical Meet. On Image processing for 2-D and 3-D Reconstruction from Projections : Theory and Practice in Medicine and the Physical Sciences.
Washington, DC : Optical Society of America, 1975, paper Th.A.6. pp 1-4.
- S.W. ROWLAND (1979) "Computer implementation of image reconstruction formulas" in G.T. Herman, Ed., Image Reconstruction from projection. Implementation and Applications. Berlin Springer, pp 9-79.
- L. SCHWARTZ (1961) "Méthodes mathématiques pour les sciences physiques"
HERMANN
- K.T. SMITH, D.C. SOLMON, and S.L. WAGNER (1977) "Practical and mathematical aspects of the problem of reconstructing objects from radiographs"
Bul. AMS 83 : pp 1227-1270.
- D. SNYDER, D. POLITTE (1983) "Image reconstruction from list-mode data in an emission tomography system having time-of-flight measurements"
IEEE Trans. Nucl. Sci., NS 30 : 1843-9

- M. STONE (1974) "Cross-validators choice and assessment of statistical prediction" (with discussion) J. Roy. Statist. Soc., Ser. B, 36 : 111-147.
- O.N. STRAND, E.R. WESTWATER (1968) 'Minimum-RMS estimation of the numerical solution of a Fredholm integral equation of the first kind' SIAM J. Num. Anal., 5 : 287-295.
- M.M. TER-POGOSSIAN, M.E. RAICHLE, B.E. SOBEL (1980) 'Positron-Emission Tomography' Scientific American, Oct., pp. 170-181.
- A. TIKHONOV, V. ARSENIN (1977) 'Solution of ill-posed problems' (Winston Willey)
- G. WAHBA (1975) "Smoothing noisy data with spline functions". Numer. Math. 24, pp. 383-393.
- G. WAHBA (1977) " Practical approximate solutions to linear operator equations when the data are noisy" SIAM J. Num. Anal. 14 pp 651-667.

AUTORISATION de SOUTENANCE

VU les dispositions de l'article 3 de l'arrêté du 16 avril 1974,

VU les rapports de présentation de Messieurs

- . P.J LAURENT, Professeur
- . Ph. GARDERET, Ingénieur

Monsieur Didier GIRARD

est autorisé à présenter une thèse en soutenance en vue de l'obtention du diplôme de DOCTEUR-INGENIEUR, spécialité "Mathématiques Appliquées".

Fait à Grenoble, le 12 octobre 1984

Le Président de l'I.N.P.-G

D. BLOCH
Président
de l'Institut National Polytechnique
de Grenoble

P.O. le Vice-Président,

