UNIVERSITÉ DE MONTRÉAL

APPROCHE PÉNALISÉE EN TOMOGRAPHIE HÉLICOÏDALE EN VUE DE L'APPLICATION À LA CONCEPTION D'UNE PROTHÈSE PERSONNALISÉE DU GENOU

MARC ALLAIN
INSTITUT DE GÉNIE BIOMÉDICAL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION
DU DIPLÔME DE PHILOSOPHIÆ DOCTOR (Ph.D.)
(GÉNIE BIOMÉDICAL)
DÉCEMBRE 2002

© Marc Allain, 2002.

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisisitons et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-612-81712-1
Our file Notre référence
ISBN: 0-612-81712-1

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou aturement reproduits sans son autorisation.

Canada

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée:

APPROCHE PÉNALISÉE EN TOMOGRAPHIE HÉLICOÏDALE EN VUE DE L'APPLICATION À LA CONCEPTION D'UNE PROTHÈSE PERSONNALISÉE DU GENOU

présentée par: ALLAIN Marc

en vue de l'obtention du diplôme de: Philosophiæ Doctor

a été dûment acceptée par le jury d'examen constitué de:

M. DUSSAULT Jean-Pierre, Ph.D., président
M. GOUSSARD Yves, Ph.D., membre et directeur de recherche
M. IDIER Jérôme, Ph.D., membre et codirecteur de recherche
Mme BLANC-FÉRAUD Laure, Ph.D., membre
M. SAUER Kenneth, Ph.D., membre
M. DEMOMENT Guy, Ph.D., membre

À ma plus belle rencontre,

à toi Sandrine...

REMERCIEMENTS

En premier lieu, je tiens à remercier les différents membres du jury. Je remercie en particulier Laure BLANC-FÉRAUD et Ken SAUER pour leur relecture avisée et minutieuse de mon manuscrit. Je suis également très reconnaissant à Jean-Pierre DUSSAULT d'avoir examiné plus spécifiquement les parties touchant à l'optimisation. Enfin, je remercie Bernard ROUGÉ et Jean DANSEREAU d'enrichir de leur expertise ce jury déjà fort étoffé.

Une thèse en cotutelle... Deux laboratoires, deux directeurs de recherche, deux continents : avec moi, il y eu finalement beaucoup d'individus à se lancer dans cette petite musique à « deux partitions ».

La première partie de mon doctorat (1999-2001) s'est déroulée au sein de l'*Institut de génie biomédical*, à l'École polytechnique de MONTRÉAL. C'est grâce à Yves GOUSSARD, surnommé « le crabe aux pinces d'or » par ses partenaires d'escalade (dont je fais partie), que j'ai pu faire mes premiers pas dans un laboratoire de recherche universitaire. Au sein de cette institution québecoise, Yves a été un directeur de recherche de grande qualité : je lui dois évidemment beaucoup, et pas uniquement sur les aspects scientifiques ou linguistiques. Je suis également très redevable envers Gilles SAVARD qui aura influencé durablement mes travaux de recherche en m'enseignant l'optimisation à l'École polytechnique. Merci également à Nicolas VILLAIN et à David SAVÉRY qui m'ont accompagné amicalement et scien-

tifiquement pendant mes études de doctorat. Enfin, je remercie Diane GIROUX et Louise CLÉMENT qui m'ont permis tant de fois de sortir vivant de « la grosse machine administrative ».

En FRANCE, le *Groupe problèmes inverses* (GPI) m'a accueilli pour une troisième et dernière année de thése. Orphelin d'une vie québecoise, j'ai eu malgrés tout la chance d'être adopté par de fameux personnages que je souhaite ici remercier. C'est ainsi avec beaucoup d'émotion que je tire mon chapeau à Jérôme IDIER pour le remercier de son implication et de son encadrement scientifique hors paire. Je remercie également chaleureusement Guy DEMOMENT pour ses multiples tuyaux (non percés) universitaires et pour avoir accepeté de participer à mon jury. Je remercie Charles SOUSSEN, Philippe CIUCIU et Vicent SAMSON pour leur soutient, leur joie de vivre, et finalement pour leur amitié. Merci également à Jean-Francois GIOVANELLI qui a toujours su se rendre disponible quand l'informatique me donnait des maux de tête. Merci également à Ali pour ses blagues, Ishem et Medhi pour leur bonne humeur, ainsi qu'à l'équipe administrative du L2S pour son soutien et son travail.

Enfin, j'aimerais dédier ce mémoire à certaines très belles personnes qui m'ont accompagné ces dernières années : vous avez rendu ma vie si dense... Ils se prénoment Maryse, Isabelle, Roger, Martin, David, Prométhée, Constantin, Nicolas ou Benjamin au QUÉBEC, ou encore Catherine, Jérôme, François ou Yann en FRANCE.

RÉSUMÉ

Ce travail s'inscrit dans le cadre d'un projet franco-québecois de conception et de fabrication d'un implant personnalisé du genou. On s'intéresse ici plus particulièrement à la reconstruction d'images tridimensionnelles de précision en tomographie hélicoïdale. Pour apporter une réponse efficace à ce problème, cette étude se développe suivant deux composantes complémentaires relevant d'une part des aspects méthodologiques adoptés pour améliorer la précision des reconstructions et d'autre part des aspects de mise en œuvre très délicats en imagerie 3D.

Mise en contexte

La tomographie à rayons X est un procédé rapide et peu invasif produisant des images tridimensionnelles (3D) à partir d'un ensemble de projections. En milieu hospitalier, la saisie des projections (i.e. l'échantillonnage du volume d'intérêt) est effectuée selon deux modalités distinctes :

(i) en mode *axial*, la source décrit une trajectoire circulaire dans des plans perpendiculaires à l'axe du tomographe et un jeu de projections est obtenu dans chaque plan ;

(ii) en mode *hélicoïdal*, la source décrit une hélice autour du volume d'intérêt et les projections sont saisies en continu.

Ces dernières années, le mode hélicoïdal a largement supplanté le mode axial : le volume d'intérêt est ainsi imagé plus rapidement permettant une réduction simultanée des artefacts associés aux mouvements (volontaires ou non) du patient et de la dose de rayonnement administrée.

Indépendamment du mode de saisie choisi, le volume d'intérêt est usuellement obtenu par empilement d'une série de coupes bidimensionnelles (2D) préalablement reconstruites. Si la qualité de ces reconstructions 3D suffit souvent pour le diagnostic médical, *la précision est en général largement insuffisante pour utiliser le tomographe à des fins métrologiques.* En particulier, la précision submillimétrique requise par la conception d'une prothèse personnalisée du genou est loin d'être atteinte par les tomographes médicaux disponibles actuellement sur le marché. Cette problématique est par ailleurs particulièrement sensible en tomographie hélicoïdale dans la mesure où il se produit en général une dégradation significative de la précision des reconstructions accompagnée éventuellement d'artefacts propres à ce mode de saisie.

On admet maintenant largement que les faiblesses des reconstructions standard sont en partie imputables à la *rétroprojection convoluée* (RPC), algorithme à la base des méthodes de reconstruction 2D, 3D, axiales ou hélicoïdales actuellement implantées dans les tomographes. Par ailleurs, la reconstruction hélicoïdale s'appuie sur une interpolation *ad hoc* des projections introduite à la seule fin de permettre

l'emploi de la RPC : à notre sens, cette interpolation est une entrave supplémentaire pour permettre la reconstruction d'images 3D de précision en mode hélicoïdal.

Une approche pénalisée en tomographie hélicoïdale

Cette thèse démontre que le problème de reconstruction hélicoïdal peut être reformulé de manière à s'affranchir de l'interpolation des projections et de la RPC. Notre démarche consiste à adapter au cadre hélicoïdal les *approches pénalisées* dont l'efficacité en tomographie axiale est aujourd'hui reconnue. Le problème de reconstruction sera alors résolu au travers de la *minimisation* d'un critère des *moindres carrés pénalisés* de très grande taille.

La méthode nécessite de construire un *opérateur d'observation* modélisant l'échantillonnage hélicoïdal du volume d'intérêt. Ce modèle ne suffit pas cependant pour obtenir une solution robuste puisque le problème de reconstruction d'images tomographiques est intrinsèquement instable. Pour permettre d'aboutir à une solution de qualité, nous régularisons le problème de reconstruction en introduisant un *modèle d'image* choisi *a priori*. Nous montrons en particulier que l'emploi de modèles *convexes* favorisant l'apparition de zones homogènes séparées par des interfaces franches est à ce titre intéressant. Ces modèles améliorent sensiblement la qualité des reconstructions sans alourdir inconsidérément le coût d'implantation informatique ; des tests sur données synthétiques indiquent un gain significatif de précision sur les volumes ainsi reconstruits en tomographie hélicoïdale.

Problématiques d'implantation

La mise en œuvre de ces approches pénalisées soulève des problématiques d'implantation associées à la taille très importante du problème d'optimisation. Une telle taille rend inadéquates les stratégies de minimisation standard, et seule une étude minutieuse de la structure du problème conduit à des schémas algorithmiques efficaces. Sur ces problématiques de mise en œuvre, notre apport est double :

(i) nous montrons qu'une invariance spatiale du modèle direct peut être exploitée afin de limiter l'explosion du coût mémoire. Une forme algorithmique de type *successive overrelaxation* (SOR) est implanté ce qui permet de garder la maîtrise du coût mémoire et du volume de calcul à chaque itération tout en conservant une vitesse de convergence intéressante ;

(ii) par ailleurs, nous montrons qu'une approximation du modèle d'observation conduit à une réduction sensible du coût d'implantation : moyennant une légère perte de qualité sur l'image finale, cette approche permet de considérer la reconstruction de chaque plan du volume comme un sous-problème de dimension réduite.

Apports méthodologiques en algorithmie

Cette thèse apporte finalement une contribution originale sur certains *points méthodologiques* ayant trait à la minimisation des critères pénalisés. Plus préci-

sément, les algorithmes « semi quadratiques » (SQ) employés en restauration et reconstruction d'images sont réexaminés au travers du lien fort qui les unit à des algorithmes standard *à pas fixe* — formes newtoniennes à pas fixe, relaxation scalaire ou par bloc à pas fixe. Dans un deuxième temps, nous rattachons ces algorithmes SQ aux algorithmes de WEISZFELD *généralisés* dont les représentants les plus connus sont l'algorithme *Expectation-Maximization* et ses nombreuses variantes.

Sur ces aspects méthodologiques touchant à l'algorithmie, nos travaux présentent des intérêts multiples. D'une part, nous éclairons de manière fructueuse les propriétés de l'algorithme SOR que nous avons choisi pour le problème pénalisé en tomographie hélicoïdale. D'autre part, nous obtenons certains résultats novateurs concernant les propriétés de convergence globale et de vitesse asymptotique de ces algorithmes SQ. En particulier, nous déduisons de cette étude des conditions de convergence affaiblies ainsi que des variantes algorithmiques plus rapides.

ABSTRACT

In order to design an ergonomic knee prosthesis, we present a new reconstruction method that produces significant improvement in the precision of *helical* tomographic reconstructions. Whereas the standard approach is based on interpolation and *convolution backprojection*, our technique relies on a *penalized* approach; in this framework, the 3D image is defined as the minimizer of a *penalized least-square* criterion, which leads to a very large scale optimization problem. An adequate regularization of the tomographic problem is provided by a *convex* penalization yielding a precise localization of the edges in the image at a reasonable numerical cost. Experiments carried out on synthetic data show that our method produces a significant improvement in precision over standard reconstruction techniques. However, the very large-scale nature of the numerical reconstruction problem leads to major *implementation* difficulties. In order to keep the computer cost reasonable, we used a spatial *invariance* of the observation model and minimized the penalized criterion with a *successive over relaxation* algorithm. Alternatively, an approximation in the observation model leads to a *separable* 3D reconstruction problem : as a result, the 3D image can be obtained by successive 2D problems of reduced size. Finally, our work yields an in-depth study of "Half-quadratic" (HQ) algorithms widely used in image or data processing. We were able to point out the connections between these HQ algorithms and already known algorithmic forms; weaker global

convergence conditions were provided and faster HQ variants were deduced.

TABLE DES MATIÈRES

DÉDICACE iv

REMERCIEMENTS v

RÉSUMÉ vii

ABSTRACT xii

TABLE DES MATIÈRES xiv

LISTE DES FIGURES xx

LISTE DES NOTATIONS ET DES SYMBOLES xxviii

LISTE DES ANNEXES xxxi

INTRODUCTION 1

I Tomographie en imagerie médicale : approches standard de reconstruction. 11

CHAPITRE 1 TOMOGRAPHIE À RAYONS X EN GÉOMÉTRIE AXIALE 12

1.1 Introduction 12

1.2 Tomographie axiale et imagerie médicale 12

1.3 Inversion de la TR en géométrie axiale 22

1.3.1 Notations et définitions 22

1.3.2 Formulation du problème de reconstruction 25

1.3.3 Difficultés méthodologiques d'inversion 27

1.3.4 Unicité et stabilité de l'inversion 28

CHAPITRE 2 MÉTHODES NUMÉRIQUES STANDARD DE RECONSTRUCTION AXIALE 32

2.1 Inversion basée sur l'expression analytique 32

2.2 Approches algébriques : une discrétisation de la TR 41

2.2.1 Formulation 41

2.2.2 Abordons le problème de reconstruction 45

2.3 En résumé 58

CHAPITRE 3 TOMOGRAPHE HÉLICOÏDAL 60

3.1 Tomographe hélicoïdal simple coupe 60

3.1.1 Technique de reconstruction standard en mode hélicoïdal . 64

3.1.2 Analyse des dégradations 67

3.2 Le cas du tomographe hélicoïdal multicoupes 69

3.3 En résumé... 71

II Inversion régularisée en imagerie **74**

CHAPITRE 4 APPROCHE RÉGULARISÉE EN IMAGERIE : LE CADRE DÉTERMINISTE **75**

4.1 Régularisation de l'inversion numérique 75

4.2 Contrôle de dimension 79

4.3 Approches pénalisées (régularisation de TIKHONOV généralisée) . 81

4.3.1 Modèles d'images introduit *a priori* 84

4.3.1.1 Pénalisation quadratiques 84

4.3.1.2 Pénalisation non quadratique 90

CHAPITRE 5 INTERPRÉTATION PROBABILISTE ET INFÉRENCE BAYÉSIENNE **102**

5.1 Vraisemblance et adéquation aux données 103

5.2 Inférence bayésienne 107

5.2.1 Vraisemblance *a posteriori* et estimateurs bayésiens 108

5.2.2 Modèles d'image probabilistes à base de champ de MARKOV 111

5.3 Quelques apports du cadre probabiliste 117

5.3.1 Modèle aléatoire d'observation : le cas poissonien 118

5.3.2 Estimation des hyperparamètres du modèle 122

III Approches pénalisées en tomographie : du mode axial au mode hélicoïdal **126**

CHAPITRE 6 TOMOGRAPHIE EN GÉOMÉTRIE AXIALE . . 127

6.1 Choix d'une approche pénalisée en tomographie 128

6.1.1 Motivation 128

6.1.2 Reconstruction pénalisée en tomographie 131

6.2 Une première mise en œuvre 137

6.3 Limitations de la méthode pénalisée 145

6.4 En résumé 157

CHAPITRE 7 TOMOGRAPHIE EN GÉOMÉTRIE HÉLICOÏDALE160

7.1 Introduction 160

7.2 Méthodologie 162

7.2.1 Régularisation $\ell_2\ell_1$ 3D 164

7.2.2 Modèle d'observation en géométrie hélicoïdale 166

7.2.3 Critère pénalisé pour la tomographie hélicoïdale 171

7.3 Résultats de simulations 175

7.3.1 Fantômes synthétiques de petite taille 175

7.3.2 Fantôme synthétique de grande taille 185

7.4 Choix algorithmiques en tomographie hélicoïdale 188

7.4.1 Stratégies de minimisation 192

7.4.2 Relaxation semi quadratique et tomographie hélicoïdale . . 199

7.5 Conclusion 206

IV Minimisation des critères pénalisés 209

CHAPITRE 8 MINIMISATION DES CRITÈRES PÉNALISÉS . 210

8.1 Formulation primale du problème 211

8.2 Algorithmes de minimisation des critères pénalisés 217

8.2.1 Algorithmes de relaxation 218

8.2.2 Algorithmes de type POCS 221

8.2.3 Algorithmes à directions de descente 222

8.2.4 Algorithmes semi quadratiques 231

8.2.5 Autres algorithmes de minimisation 231

8.3 Formulations primales/duales « semi quadratiques » 232

8.3.1 Constructions « semi quadratiques » 232

8.3.2 Algorithmes de relaxation semi quadratiques 240

8.4 Formes primales des algorithmes semi quadratiques 253

8.5 En résumé 264

CHAPITRE 9 AUTOUR DE LA CONVERGENCE DES ALGORITHMES SQ 267

9.1 Définition du schéma itératif considéré 268

9.2 Convergence globale en optimisation non contrainte 271

9.3 Convergence globale à pas fixe *via* la règle d'ARMIJO 281

9.4 Admissibilité du pas fixe pour les algorithmes SQ 289

9.4.1 Convergence globale des algorithmes SQ de GY 291

9.4.2 Convergence globale de l'algorithme SQ de GR 302

9.5 Convergence locale et vitesse asymptotique 308

9.6 En résumé 322

CONCLUSION **324**

RÉFÉRENCES **335**

LISTE DES FIGURES

1 Exemple de prothèse du genou, partie fémorale : (←) demi prothèse standard, (→) prothèse de recouvrement. 2

2 ← : coupe axiale au dessus du genou (haut) et coupe tomographique à rayons X correspondante (bas). → : table des atténuations massiques exprimée en cm^2/g à 40 keV ; tirée de [Macovski, 1983]. . . . 5

3 Procédé d'estimation de la géométrie des condyles fémoraux développé par l'équipe au court des années 1990 ; tiré de [Villain, 1997]. 6

1.1 Transmission du faisceau X infiniment mince au travers d'une couche de matériaux 14

1.2 Géométrie d'acquisition en mode axial : (a) rayons parallèles ou (b) rayons divergents. 16

1.3 Trajectoire de la source pour le mode axial. 17

1.4 Illustration de la TR en dimension 2. 24

2.1 Échantillonnage radiale de la TF de x. 37

2.2 Calcul des éléments de $\boldsymbol{R}$ pour un modèle à base d'indicatrices sur les pixels avec indexation lexicographique ; (a) représente le cas d'un faisceau sans épaisseur avec des « pixels carrés », (b) le cas d'un faisceau de type « bande », et (c) le cas d'un faisceau de type « bande » avec des « pixels cylindriques ». 44

2.3 Mise en œuvre extraite de [Censor *et al.*, 2001] des algorithmes ART, BICAV et CAV sur données bruitées ; (haut) : évolution de l'erreur relative mesurée sur la reconstruction du fantome ; bas : fantôme synthétique utilisé, et meilleurs reconstructions obtenues pour les trois algorithmes. 56

3.1 Trajectoire de la source en mode hélicoïdal. 60

3.2 Trajectoire de la source en mode hélicoïdal suivant l'axe du tomographe (←), et dans le plan de l'image $\xi_1 O \xi_2$ associé à l'angle $\bar{\varphi}_0$ sur l'hélice(→). Pour une position angulaire $\bar{\varphi} = \bar{\varphi}_0$ selon l'axe du tomographe, l'angle de projection dans le plan de l'image est $\varphi_0 = \mathrm{mod}_{2\pi}(\bar{\varphi}_0)$. 62

3.3 Procédés usuels d'interpolation adopté pour recréer un jeu « axial consistant » dans le plan défini par $\bar{\varphi} = \bar{\varphi}_0$. Chaque *pseudo* projection axiale d'angle φ découle de deux projections obtenues dans les plans $\bar{\varphi}_0 + \Delta\varphi$ et $\bar{\varphi}_0 - 2\pi + \Delta\varphi$ pour le mode « full-scan » (←), ou $\bar{\varphi}_0 + \Delta\varphi$ et $\bar{\varphi}_0 - \pi + \Delta_\varphi$ pour le mode « half-scan » (→) ; $\Delta_\varphi = \varphi - \mathrm{mod}_{2\pi}(\bar{\varphi}_0)$. 66

3.4 Échantillonnage hélicoïdal multicoupe : (←) ensemble de quatres rangées de capteurs montées en parallèle ; (→) illustration du mode de saisie *simple coupe* vs. *multicoupes.* 71

4.1 Exemples de fonctions coût proposées dans la littérature : (a) quadratique, (b) compromis convexe $\ell_2\ell_1$, (c) non convexe $\ell_2\ell_0$. 94

5.1 Voisinages d'ordre zéro, un et deux dans le plan avec les cliques associées (←) ; voisinage du premier ordre en 3D (→) ; cette figure est reproduite avec l'aimable autorisation de Nicolas Villain. 114

6.1 Effets de bords exclus, chaque élément dans l'image 2D est mutuellement voisin de quatre (←) à huit pixels (→) ; les différences premières entre éléments voisins correspondent donc aux différences verticales et horizontales auxquelles il faut éventuellement ajouter les différences entre éléments diagonaux. 134

6.2 Fantôme de SHEPP et LOGAN utilisé pour les mises en œuvre (←) ; tracé de la 63-ème colonne de ce fantôme (↑) ; Sinogrammes composé de 151 projections échantillonnées sur 175 rayons obtenu pour un faisceau d'épaisseur nulle et sans bruit d'observation (→). 138

6.3 Cas ① et ② : reconstruction par RPC du fantôme à partir de projections obtenues pour : une épaisseur de faisceau nulle et sans perturbation (①) ; une épaisseur de faisceau égale à la largeur d'un pixel et un bruit additif ou multiplicatif (②) sur les projections — cf. texte pour les caractéristiques du bruit. La fréquence de coupure du filtre a été ajusté manuellement de manière à obtenir le meilleur compromis visuel entre la résolution et le niveau de bruit. 140

6.4 Cas ③ : reconstructions par RPC du fantôme à un niveau de perturbation « limite » pour la robustesse de la méthode (bruit additif gaussien décorrélé, centré et d'écart type $\sigma = 0,008$). 141

6.5 Cas ① et ② : reconstruction par approche pénalisée du fantôme à partir de projections obtenues pour une épaisseur de faisceau nulle et sans bruit (①) ; une épaisseur de faisceau égale à la largeur d'un pixel et un bruit additif (②) ou multiplicatif (②). Les paramètres α et s ont été ajustées manuellement de manière à obtenir le meilleur compromis visuel entre la résolution et le niveau de bruit. 142

6.6 Cas ③ : limite de robustesse au bruit de la reconstruction pénalisée (bruit additif décorrélé, gaussien centré d'écart type $\sigma = 0,01$) ; cas ④ test de robustesse à une excursion angulaire incomplète : 151 projections produites uniformément dans $[0; 5\pi/6[$ perturbées par un bruit additif décorrélé, gaussien centré d'écart type $\sigma = 0,001$. La colonne du centre illustre les performances de la RPC (filtre de Hamming, $\nu_c = 0,8$), celle de droite illustre celles de l'approche pénalisée (modèle $\ell_2\ell 1$, $\alpha = 2,5.10^{-3}$; $s = 8.10^{-5}$). 144

6.7 Fantôme de HERMAN utilisé pour les mises en œuvre (←) et sa version seuillée de manière à laisser apparaître les détails intracraniens (↑) ; tracé de la soixante troisième colonne de ce fantôme (→). . . . 146

6.8 Cas ① : reconstruction du fantôme de la figure 6.7 par RPC (←), par approche pénalisée à partir des même projections (↑) et pour des projections produites pour le fantôme pixelisé. 148

7.1 Trajectoires de la source en mode axiale (←) et en mode hélicoïdal (→). 161

7.2 Système de voisinage 3D retenu : le voxel courant (croix) interagit avec 8 voisins dans le plan de l'image et 2 voisins dans l'axe du tomographe. 165

7.3 Projection en géométrie hélicoïdale dans le plan $\xi_2 O \xi_3$ (←) et dans le plan $\xi_1 O \xi_2$ (→). 167

7.4 Fantôme utilisé pour créer le jeux de projections hélicoïdales (haut) ; plans du fantôme choisis pour la reconstruction (←) et emplacement par rapport au fantôme (→). 177

7.5 Fantôme synthétique utilisé pour l'estimation de la SSP (←) ; plans de reconstruction du fantôme choisis (↑) et énergie normalisée contenue dans chacun des plans du fantôme destinés à être reconstruit. . 178

7.6 Reconstruction du fantôme de la figure 7.4 pour un faisceau d'épaisseur nulle (←) ou égale au coté d'un voxel (→) : par RPC+LI-360° **(a)**, RPC+LI-180° **(b)**, approche pénalisée $\ell_2\ell_1$ **(c)**. RPC mise en œuvre avec un filtre de HAMMING et une fréquence de coupure réduite $\nu_c = 0,2$; les hyperparamètres de l'approche pénalisée ont été ajustés manuellement. 179

7.7 Reconstruction du fantôme 7.4 pour des données bruitées ($\sigma = 1$) et une épaisseur de faisceau nulle (←) ou égale à la largeur d'un plan de voxels (→) : RPC+LI-180° (haut) ou approche pénalisée $\ell_2\ell_1$ (bas). La RPC a été mise en œuvre avec un filtre de reconstruction de type HAMMING et une fréquence de coupure réduite $\nu_c = 0,2$. 182

7.8 Tracées des SSP : RPC+LI-180° **(a)**, RPC+LI-360° **(b)**, app. pénalisée $\ell_2\ell_1$ **(c)**. 183

7.9 Fantôme 3D de grande taille : (→) vue en perspective, (← haut) variation du profile suivant l'axe du tomographe, (← bas) coupes 10, 19 et 36 extraite du fantôme. 185

7.10 Reconstruction des coupes 36, 19 et 10 par LI-180° + RPC (haut) ou par approche pénalisée $\ell_2\ell_1$ (bas) et tracé des courbes segmentées à 0,7 pour le volume reconstruit par approche pénalisée. La RPC a été mise en œuvre avec un filtre de reconstruction de type cosinus et une fréquence de coupure réduite $\nu_c = 0,3$; les hyperparamètres de l'approche pénalisée ont été ajustés manuellement. 187

7.11 Tracé de l'EQM dans chaque tranche pour la reconstruction exacte. 188

7.12 Reconstruction de la coupe 19 par l'approche pénalisée $\ell_2\ell_1$ exacte (←) et inexacte (↑) ; contours segmentés pour les plan 10, 19 et 36 pour la reconstruction du volume de manière inexacte (→) 188

7.13 Opérateur d'observation et matrice $\boldsymbol{Q}$ obtenue en géométrie axiale (haut) et en géométrie hélicoïdale (bas) sur un problème de petite taille (15×15 pixels en axial et 5 plans de 15×15 voxels en hélicoïdal) ; dans $\boldsymbol{H}$, l'opérateur $\boldsymbol{H}_r$ a été délimité par des pointillés afin de simplifier l'interprétation de sa structure. 191

7.14 Reconstruction de la coupe 19 par approche pénalisée $\ell_2\ell_1$ après 70 itération de l'algorithme (←) pour une initialisation par RPC (haut) et par une image uniformément nulle (bas) ; tracés segmentées à 0,7 extraits des coupes 10, 19 et 36 obtenues après 70 itération pour l'initialisation par un vecteur nul. 206

9.1 Illustration de la règle d'ARMIJO. 275

9.2 Domaines de convergence découlant : de l'étude de convexité du critère primal/dual (domaine standard) ; de l'admissibilité du pas fixe pour la règle d'ARMIJO (←) ou de la condition d'approximation supérieure (→). 299

9.3 Tracé de $\phi''(u)$ et de $\phi'(u)/u$ pour $\phi(u) = \sqrt{u^2 + s^2}$; on remarquera que le maximum de $\phi'(u)/u$ est atteinte en zéro et qu'elle correspond à la courbure maximale de ϕ. 315

9.4 Évolution du rayon spectral $\sigma(\theta, \boldsymbol{u})$ en fonction du pas fixe θ. 318

9.5 Courbes de niveaux pour $\sigma_{\text{GY}}(\theta, a)$ à l'intérieur du nouveau domaine de convergence défini par $\theta \in]0; 2[$ et $a \in]2\widehat{a}/\theta[$ pour $\widehat{a} = 0, 1$; le meilleur taux de convergence noté $\sigma^\star$ est atteint sur la frontière. . . 321

LISTE DES NOTATIONS ET DES SYMBOLES

2D bidimensionnel ;

3D tridimensionnel ;

ART technique de reconstruction algébrique (Algebraic Reconstruction Technic) ;

BF basses fréquences ;

BFGS BROYDEN-FLETCHER-GOLDFARB-SHANNO ;

BICAV moyennage sur les composante par bloc (Block Iterative Component Averaging) ;

CAV moyennage sur les composante (Component Averaging) ;

EM espérance puis maximisation (Expectation-Maximization) ;

EQM erreur quadratique moyenne ;

GC gradient conjugué ;

GCP gradient conjugué préconditionné ;

GR GEMAN et REYNOLDS ;

GRF champ aléatoire de GIBBS (GIBBS Random Field) ;

GS GAUSS-SEIDEL ;

GY GEMAN et YANG ;

HF hautes fréquences ;

l-BFGS algorithme BFGS à mémoire limitée ;

LI-180 interpolation linéaire sur 180 degrès ;

LI-360 interpolation linéaire sur 360 degrès ;

MAP maximum *a posteriori* ;

MRF champ aléatoire de MARKOV (MARKOV Random Field) ;

MV maximum de vraisemblance ;

Nw NEWTON ;

POCS projection sur des ensembles convexes (Projection Onton Convex Sets) ;

qNw quasi NEWTON ;

RPC retroprojection convoluée ;

RSB rapport signal à bruit ;

SART technique de reconstruction algébrique simultanée (Simultaneous Algebraic Reconstruction Technic) ;

SD algorithme du gradient (Steepest Descent) ;

SOR sur-relaxations successices (Successive Over relaxation) ;

SQ semi quadratrique ;

SSP profile de sensibilité ùne tranche (Slice Sensibility Profil) ;

SSU mise à jour pour un seul site (Single Site Update) ;

SVD décomposition en valeurs singulières (Singular Value Decomposition) ;

TF transformée de FOURIER ;

TFR transformée de FOURIER rapide ;

TR transformée de RADON ;

TSVD décomposition en valeurs singulières tronquée (Truncated Singular Value Decomposition).

LISTE DES ANNEXES

ANNEXE I Annexe au chapitre 8 356

ANNEXE II Annexe au chapitre 9 361

II.1 Preuve du théorème 1 361

II.2 Preuve des lemmes 362

II.3 Preuve des propositions 368

INTRODUCTION

Cette thèse s'est déroulée dans le cadre d'une convention de cotutelle FRANCE-QUÉBEC *établie entre l'École polytechnique de* MONTRÉAL *et l'Université* PARIS-*Sud.*

L'implant de recouvrement pour l'articulation du genou

L'arthrite (arthrose, arthrite rhumatoïde) est un problème de santé publique qui ne cesse de croître dans la plupart des pays occidentaux[1]. Les conséquences de cette pathologie vont d'une simple inconfort à la perte complète de la mobilité, et bien souvent, la pose d'une prothèse articulaire est le seul moyen pour retrouver une certaine qualité de vie.

Vers un implant ergonomique...

La pose des prothèses standard (dites « prothèses totales ») nécessite l'ablation d'une quantité importante de tissus osseux pour pallier à la forme simpliste de l'interface os/implant (voir figure 1-gauche). Cette manière de procéder cumule plusieurs inconvénients dont le principal est de rendre difficile toute révision ultérieure de l'implant une fois sa durée de vie atteinte (usuellement de 8 à 15 ans).

[1]Avec un total de plus de 22000 implants sur l'année 1999-2000, la prothèse du genou est actuellement la prothèse la plus implantée au CANADA ; par ailleurs le nombre d'implant a subit une progression de plus de 45 % depuis 1995 [David, 2002].

Cette dernière raison explique pourquoi la pose de tels implants est refusée quasi systématiquement aux patients de moins de 55 ans.

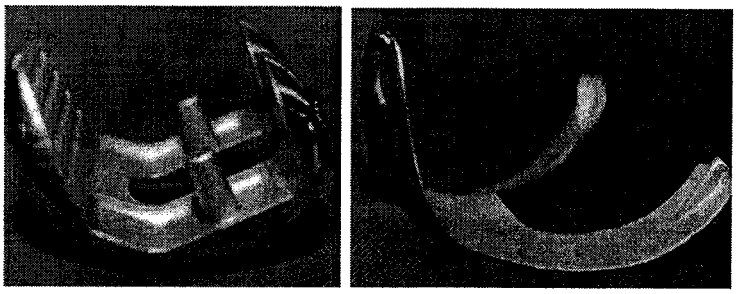


Figure 1: Exemple de prothèse du genou, partie fémorale : (←) demi prothèse standard, (→) prothèse de recouvrement.

A contrario une prothèse dite de « recouvrement » peut être mise en place par un acte chirurgical moins important et peut subir un nombre de révision élevé. Le principe d'un tel implant est de remplacer le cartilage endommagé par une coquille d'alliage biocompatible adaptée à la forme du fémur (voir figure 1-b). Les premières utilisations cliniques sont dues au DR. GARIÉPY, chirurgien orthopédiste à l'HÔTEL-DIEU à MONTRÉAL, qui a développé un procédé artisanal pour la fabrication de telles prothèses entre 1960 et 1980. Au terme de ces vingt années, il est ressorti que ces implants présentaient d'indéniables qualités mais que leur tenue dans le temps nécessitait qu'ils soient ajustés très précisément sur les condyles fémoraux pour réduire les risques de rupture par fatigue mécanique. Il est donc nécessaire d'effectuer une estimation suffisamment précise de la géométrie articulaire propre à chaque patient si on souhaite assurer la pérennité de l'implant.

La conception de la prothèse personnalisée : un projet commun

La conception de l'implant personnalisé du genou repose sur un projet commun impliquant le *Groupe de recherche en biomécanique et biomatériaux* de l'*École polytechnique de* MONTRÉAL, le *Département de génie mécanique* de l'*École de technologie supérieure de* MONTRÉAL, et le *Groupe de problème inverse* dépendant du *Laboratoire des signaux et systèmes* en FRANCE. Les objectifs de ce projet initié en 1994 sont de permettre la conception et la fabrication de cette prothèse personnalisée en tirant parti des progrès récents survenus dans les domaines de l'imagerie médicale, de la conception/fabrication assistée par ordinateur, et de l'ingénierie des matériaux. Soulignons que l'estimation de la géométrie fémorale est le premier maillon d'une chaîne conséquente de traitements informatiques et mécaniques [Allain *et al.*, 2000]. Dès lors, estimer cette géométrie avec précision permet de se mettre dans des conditions favorables pour produire un implant viable.

Le rôle de l'imagerie médicale

L'estimation de la géométrie articulaire s'appuiera sur l'imagerie par *tomographie à rayons X*. Cette modalité d'imagerie, largement disponible dans le milieu hospitalier, permet d'obtenir de très forts contrastes entre les tissus osseux et les autres tissus du corps humain tout en restant peu invasive ; cf. figure 2. Un point doit cependant retenir notre attention : conçu principalement pour fournir une information qualitative aux cliniciens, le tomographe à rayon X peut manquer cruelle-

ment de précision pour certaines applications métrologiques. Et c'est effectivement le cas dans notre situation. Une étude préalable a montré que la précision requise pour assurer une durée de vie satisfaisante (10 ans en moyenne) à l'implant était inférieure au millimètre. Or, les tomographes disponibles en milieu hospitalier — ceux auxquels nous aurons accès — sont loin d'atteindre une telle précision. Dans la suite de ce document, nous verrons que ce manque de précision n'est pas intrinsèque à l'appareil de mesure, mais qu'il découle plutôt des techniques de reconstruction implantées dans les appareils commerciaux. L'objectif de notre étude est donc de développer des approches alternatives de reconstruction de manière à améliorer la précision des images tomographiques.

Travaux préalables en imagerie

En mai 1998, date de mon intégration au projet, un certain nombre de travaux significatifs avaient déjà été menés afin d'estimer la surface des condyles fémoraux avec la précision requise. En particulier, les travaux menés par SYLVIE DORÉ ont permis de caractériser expérimentalement le flou (c.à.d. la réponse impulsionnelle) introduit dans les images obtenues en sortie d'un tomographe axial typique (largeur à mi-hauteur, anisotropie, variance spatiale). *Cette étude à permis en outre de quantifier expérimentalement leur résolution spatiale et de souligner leur incapacité à fournir une précision submilimétrique* [Doré et Goussard, 1997]. Par la suite, NICOLAS VILLAIN a tiré parti de cette étude pour mettre en œuvre une *res-*

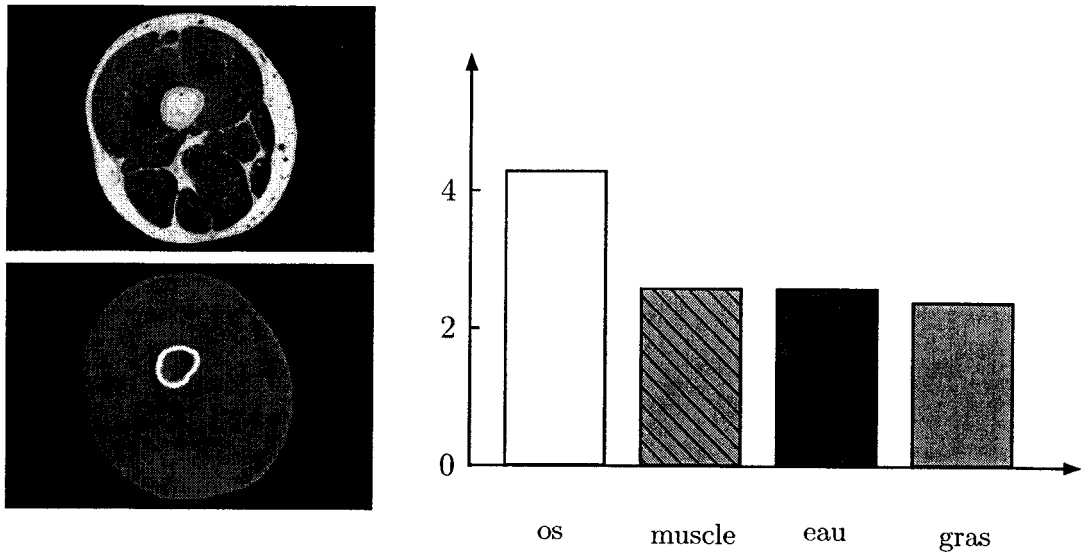


Figure 2: ← : coupe axiale au dessus du genou (haut) et coupe tomographique à rayons X correspondante (bas) ; extrait de http ://www.meddean.luc.edu/lumen. → : table des atténuations massiques exprimée en cm^2/g à 40 keV ; tirée de [Macovski, 1983].

tauration tridimensionnelle des images fournies par le tomographe ; une validation expérimentale de la méthode a alors permis de montrer que cette démarche pouvais fournir une précision meilleure que le millimètre [Villain *et al.*, 2001]. Le procédé qui est actuellement mis en œuvre pour estimer la géométrie des condyles fémoraux est illustré par la figure 3. On notera qu'une étape d'interpolation par *krigeage* permet d'intercaler une série de plans entre chaque image restaurée : cette étape est destinée à « combler » l'écart minimal séparant deux plans axiaux produits par le tomographe et dont l'ordre de grandeur est au mieux d'un millimètre.

Un avantage majeur de la méthode développées par NICOLAS VILLAIN est sa

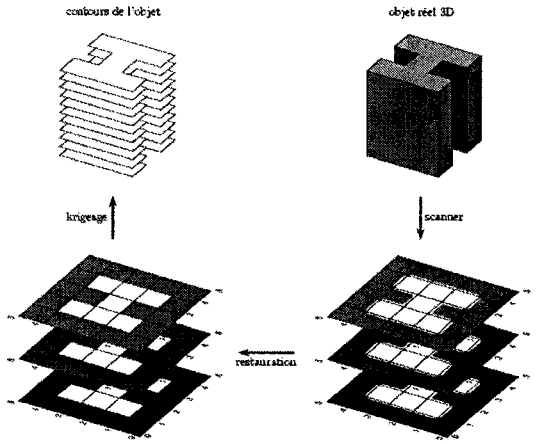


Figure 3: Procédé d'estimation de la géométrie des condyles fémoraux développé par l'équipe au court des années 1990 ; tiré de [Villain, 1997].

souplesse de mise en œuvre en contexte hospitalier : d'une part elle permet de traiter directement les images rendues disponibles par le tomographe, d'autre part elle s'adapte aisément à n'importe quel tomographe médical. En terme de précision finale, néanmoins, on peut légitimement supposer qu'un traitement direct sur les mesures de projections pourrait produire de meilleures reconstructions si on le compare à ce *post traitement* sur les images[2].

Vers une tomographie hélicoïdale de haute précision

L'imagerie en tomographie à rayons X est marquée depuis une dizaine d'années par l'arrivée massive des tomographes hélicoïdaux. L'avantage de ces tomographes est de permettre une réduction simultanée du temps d'examen et de la dose de

[2]Cette conjecture a déjà été vérifiée sur des données synthétiques [Villain *et al.*, 2001], et nous espérons pouvoir la valider sur données expérimentales dans un futur proche. Cette validation expérimentale reste pour le moment difficile puisqu'il nous est impossible d'accéder aux mesures de projection sur les tomographes hospitalier mis à notre disposition.

radiation administrée au patient. Ces avancées se sont faites néanmoins au prix d'une dégradation de la qualité des reconstructions qui peut être significative en pratique. Le remplacement du classique tomographe axial par un tomographe hélicoïdal a évidemment un impact sur le projet d'implant personnalisé : les méthodes précédentes ayant été développées et validées en tomographie axiale, il s'agit alors de produire un travail similaire pour une approche adaptée au cadre hélicoïdal. De manière plus générale, on trouvera un intérêt évident à développer une méthode alternative permettant de pallier, au moins en partie, aux dégradations constatées en tomographie hélicoïdale. Ceci constitue l'objectif principal de ma thèse de doctorat, et nous montrons clairement dans ce document qu'une approche « pénalisée » de la tomographie hélicoïdale constitue une alternative prometteuse pour reconstruire des images précises. Un certain nombre de travaux de validation sur données réelles doivent néanmoins être menés pour valider cette méthode, notamment dans le cadre du projet d'implant personnalisé.

Aspects algorithmiques : mise en œuvre et travaux méthodologiques

La mise en œuvre d'une inversion pénalisée en imagerie 3D présente habituellement de sévères difficultés d'implantation. La résolution du problème d'optimisation associé à la méthodologie d'inversion est à ce titre particulièrement délicate. En effet, la taille du problème numérique est telle qu'elle empêche l'implantation de la plupart des algorithmes de minimisation standard. La reconstruction d'une image

3D passe donc par la recherche d'une solution algorithmique viables adaptée au problème en tomographie hélicoïdale. Dans ce contexte, nous avons proposé deux approches qui permettent la résolution de problèmes de reconstruction de grande taille ; des tests sur données synthétiques ont permis de démontrer leur efficacité ; là encore des tests sur données réelles sont nécessaires de manière à valider les résultats encourageants de ces algorithmes.

Les travaux algorithmiques effectués dans le cadre de la reconstruction d'image ont également motivé une analyse approfondie des algorithmes « semi quadratiques ». Bien que répandus dans la communauté, cette famille d'algorithmes reste encore peu étudiée. Dans ce document, nous établissons un lien clair entre ces algorithmes et des algorithmes newtoniens à pas fixe, et certains résultats originaux sur leurs propriétés de convergence sont établis. Ces travaux ont conduit en particulier à des mises en œuvre rapides de ces algorithmes.

Organisation du document

Le présent manuscrit a été organisé en suivant un découpage en 4 parties : les *trois premières parties* (parties A, B et C) constituent un exposé des questions méthodologiques et pratiques associées à la reconstruction d'images en tomographie axiale et hélicoïdale. La quatrième partie D traite de la minimisation des critères pénalisés et rassemble nos contributions sur le plan algorithmique.

L'exposé débute par une présentation du cadre méthodologique associé au problème de reconstruction en tomographie — *chapitre* 1. Les algorithmes de reconstruction standard employés dans le cadre de la tomographie axiale sont présentés dans le *chapitre* 2, et le *chapitre* 3 introduit le problème de reconstruction en géométrie hélicoïdale et décrit le principe standard de reconstruction dans ce mode. Ce groupe de trois chapitres forme la première partie du manuscrit.

La seconde partie du manuscrit — *chapitres* 4 et 5 — traite de la régularisation en traitement d'images et présente les approches pénalisées qui seront à la base de notre méthode de reconstruction. Cette partie est découpée en deux chapitres traitant respectivement de la régularisation dans un cadre *déterministe* (régularisation au sens de TIKHONOV généralisée) et *probabilistes* (approche bayésienne en traitement d'image).

La troisième partie traite de l'inversion pénalisée en géométrie axiale et de son extension à la géométrie hélicoïdale. Dans le cadre simplifié que constitue la géométrie axiale, le *chapitre* 6 permet de motiver certains choix et de souligner certaines problématiques qui resterons valides en géométrie hélicoïdale. L'approche pénalisée conçue pour la tomographie hélicoïdale et la description de l'algorithme de reconstruction sont décris au *chapitre* 7; ce chapitre rassemble une part de nos contributions originales.

La quatrième partie du manuscrit traite de la minimisation des critères pénalisés

et étudie largement des algorithmes *semi quadratiques.* Le *chapitre* 8 introduit les algorithmes semi quadratiques et établit des liens clairs entre cette famille et des algorithmes *newtoniens à pas fixe* scalaires ou multivariés. Nous établissons également clairement dans ce chapitre le lien unissant ces algorithmes à l'algorithme *Expectation-Maximization* et ses nombreuses variations. Le *chapitre* 9 tire parti du lien établi au chapitre précédent pour étudier les propriétés de convergence globale et de vitesse asymptotique de ces algorithmes. Ces deux chapitres rassemblent les éléments algorithmiques originaux de cette thèse.

Enfin, le un chapitre de conclusion résume les principaux éléments de ce travail de recherche et présente un certain nombre de perspectives.

PARTIE I

Tomographie en imagerie médicale : approches standard de reconstruction.

CHAPITRE 1

TOMOGRAPHIE À RAYONS X EN GÉOMÉTRIE AXIALE

1.1 Introduction

Au sens strict, le problème de reconstruction tomographique consiste à *inverser* la transformation de Radon (abrég. TR) que nous définissons ci-dessous. Il est cependant d'usage d'élargir cette désignation à *tout* problème d'inversion mettant en jeu une TR. Ces problèmes de reconstruction sont omniprésents non seulement dans la majorité des problèmes d'imagerie médicale moderne[1] mais aussi dans un nombre conséquent de problèmes de mesure physique (géophysique, astronomie, contrôle non destructif, radar...).

Dans cette étude, on s'intéresse « simplement » à la reconstruction d'image médicale (éventuellement 3D) par tomographie à rayons X : problème qu'on ramènera formellement à l'inversion de la transformée de Radon 2D.

1.2 Tomographie axiale et imagerie médicale

Cette section décrit brièvement le principe de mesure en tomographie axiale et la méthodologie standard de reconstruction. La physique des phénomènes observés est

[1]En tomographie de transmission et d'émission PET-SPECT, mais également en imagerie de réflection ultra sonore et en imagerie par résonance magnétique (IRM).

décrite *via* le modèle d'observation de BEER LAMBERT. Nous donnons également les principales limites de ce modèle et nous fixons les hypothèses valides pour la suite dans ce document.

Physique du phénomène

Considérons un faisceau de photons X infiniment fin se propageant suivant un angle d'incidence donné dans un matériau (figure 1.1). Les rayons X interagissent dans la matière de différentes manières. Dans la gamme d'énergie adoptée en tomographie clinique ($\approx$ 120 keV), l'effet combiné de la diffraction (effet Compton) et de l'absorption (effet photoélectrique) produit une atténuation exponentionelle du faisceau X au travers du matériau. Dans la mesure où la source est *monochromatique* (*i.e.* le faisceau X est mono énergétique), la loi de BEER LAMBERT décrit alors correctement le phénomène physique :

$$N = N_0 \exp\left(-\int_{\boldsymbol{\xi}\in L} x(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi}\right) \qquad \text{avec} \qquad \boldsymbol{\xi} = (\xi_1, \xi_2) \in \mathbb{R}^2, \tag{1.1}$$

avec $x(\boldsymbol{\xi})$ la distribution *d'atténuation* des photons X, et N_0 et N, respectivement, le nombre de photons X émis par seconde par la source et reçus par le détecteur. Cette atténuation étant directement liée à la densité du matériau considéré, la tomographie à rayon X est également appelée *tomodensitométrie.*

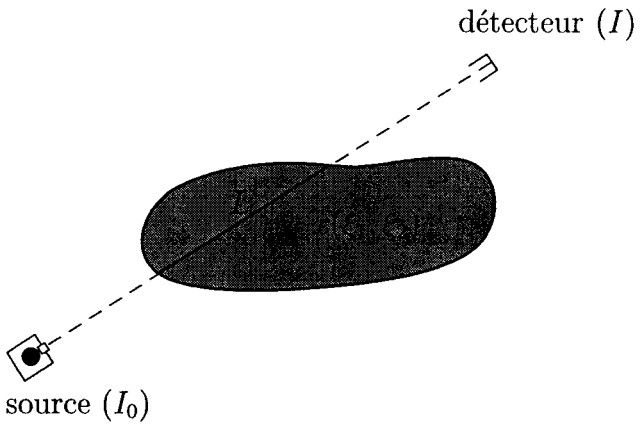


Figure 1.1: Transmission du faisceau X infiniment mince au travers d'une couche de matériaux

Procédé de mesure

Le processus de mesure en tomographie *axiale* à rayons X consiste à « illuminer » une section infiniment mince de l'objet à imager sous un certain angle d'incidence $\varphi = \bar{\varphi}$. D'après (1.1), l'énergie d'un faisceau infiniment fin est reliée à l'intégrale linéique de $x(\boldsymbol{\xi})$ par

$$-\log\left(\frac{N}{N_0}\right) = \int_{\boldsymbol{\xi}\in L} x(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi}. \tag{1.2}$$

L'ensemble des intégrales linéiques de $x(\xi_1, \xi_2)$ selon des droites *parallèles* est appelée « projection » et s'identifie à la *transformée de Radon* $g(s, \bar{\varphi})$ prise en $\bar{\varphi}$ avec s une distance algébrique séparant le rayon d'une direction parallèle de référence. En faisant varier l'angle d'incidence φ dans le plan, l'ensemble des projections se

relie formellement à la TR de x qui s'écrit

$$g(s,\varphi) = \int_{\mathbb{R}^2} x(\boldsymbol{\xi})\, \delta(s - \langle \boldsymbol{\xi}, \boldsymbol{\theta} \rangle)\, \mathrm{d}\boldsymbol{\xi}$$

$$\text{où} \quad \boldsymbol{\theta} = \begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix}, \quad \boldsymbol{\xi} = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}, \quad \text{et} \quad s \in \mathbb{R},\ \varphi \in [0;\pi[. \tag{1.3}$$

En milieu hospitalier, l'arrangement *parallèle* des rayons dans le plan a plutôt laissé le pas à un arrangement en *éventails*. Notons que ces deux arrangements de faisceaux sont reliées par :

$$s = r \sin\alpha \qquad \text{et} \qquad \varphi = \beta + \alpha - \frac{\pi}{2}, \tag{1.4}$$

où les angles β et α repèrent le rayon dans un arrangement éventail tel qu'illustré sur la figure 1.2.(b). On souligne que le passage d'un type d'arrangement à un autre ne présente pas de difficulté méthodologique particulière, et dans cet exposé, on se concentrera sur la géométrie à rayons parallèles qui permet des écritures formelles plus simples.

Reconstructions axiales 2D

On cherche maintenant à estimer la fonction bidimensionnelle (abrég. 2D) $x(\xi_1, \xi_2)$ sous-jacente ayant conduit aux projections axiales selon un des procédés de mesure

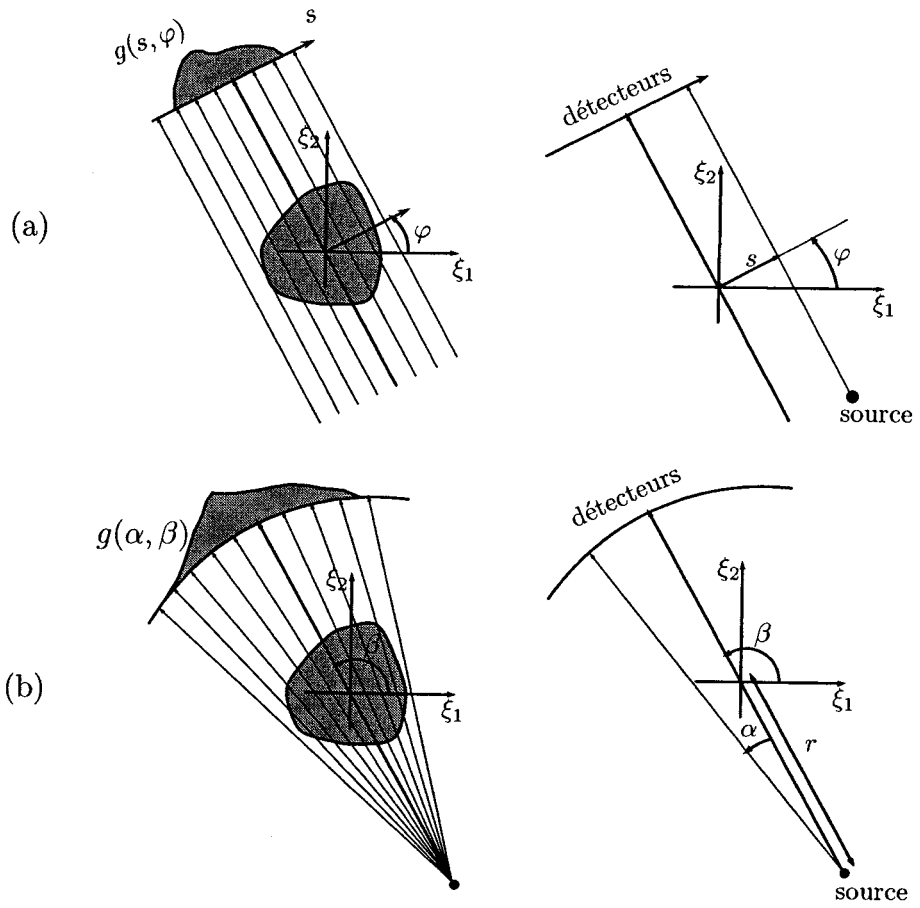


Figure 1.2: Géométrie d'acquisition en mode axial : (a) rayons parallèles ou (b) rayons divergents.

illustrés figure 1.2. C'est l'étape de *reconstruction* qui consiste en une *inversion numérique* de la transformation (1.3) modélisant le système d'imagerie. En pratique, ce problème de reconstruction en géométrie axiale est principalement abordé selon deux approches « concurrentes » :

(A) la première s'appuie sur la formulation *analytique* de l'inverse de la TR de x pour en déduire un algorithme de reconstruction. Cette démarche, présentée en section 2.1, a conduit aux algorithmes implantés dans les tomographes

médicaux : les algorithmes de type *Rétroprojection convoluée* (abrég. RPC) ;

(B) la seconde approche s'appuie sur une décomposition préalable de x sur une base finie de fonctions appropriées (typiquement une base d'indicatrices de *pixels*) ; l'inversion numérique a ensuite lieu pour reconstruire « au mieux » cette approximation. Cette démarche conduit aux algorithmes dits « algébriques » qui seront présentés en section 2.2.

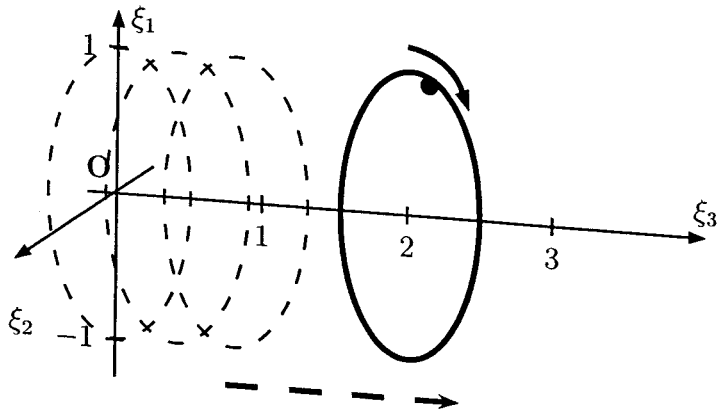


Figure 1.3: Trajectoire de la source pour le mode axial.

Reconstructions en 3D coupe par coupe

La reconstruction d'images tomographiques tridimensionnelles (abrég. 3D) est d'un intérêt majeur pour le diagnostic médical. En pratique, la reconstruction d'un volume en tomographie *axiale* n'est pas posé dans son cadre intrinsèque (c'est-à-dire l'inversion d'une transformation de Radon en dimension 3). Ce problème est plutôt appréhendé comme une succession de problèmes de reconstruction *indépendants* en 2D. La tomographie axiale 3D classique fonctionne donc suivant le principe

« stop and go » [2] (voir figure 1.3). Ce choix a évidemment le mérite de minimiser les coûts économiques et informatiques puisqu'il permet d'exploiter directement les algorithmes et le matériel conçus dans le cadre de la reconstruction axiale 2D. Nous verrons que la même démarche a été adoptée pour la tomographie hélicoïdale avec des conséquences parfois problématiques.

Principales limites du modèle physique

Assimiler la mesure tomographique en rayons X à une transformation de Radon (*via* la loi de BEER LAMBERT) est évidemment une approximation plus ou moins légitime, dépendamment du contexte, c.à.d. des caractéristiques physiques de la source d'émission, des capteurs de mesure et de l'objet d'intérêt. Sur les images reconstruites, ces erreurs de modèle conduisent à des *artefacts* dûment référencés dans la littérature. Les plus sévères sont décris ci-dessous :

1. Si la source est *polychromatique*, on constate un « durcissement » [3] de faisceau qui conduit à surestimer les atténuations. L'atténuation x est donc également

[2] Les données sont obtenues dans un plan de mesure et une image tomographique 2D est reconstruite (de 0,7 à 3 secondes, typiquement 1 seconde), la table est déplacée (de 3 à 10 secondes), et le processus est répété. Ce procédé permet typiquement l'acquisition d'une douzaine d'images 2D par minute.

[3] Pour une source polychromatique, l'atténuation des photons de basse énergie est plus probable : le faisceau a tendance à se « durcir » puisque son spectre d'énergie a tendance à éliminer les composantes de faible énergie.

une fonction de l'énergie et le modèle (1.1) doit plutôt s'écrire

$$N = \int S_0(E) \exp\left(- \int_{\boldsymbol{\xi} \in L} x(\boldsymbol{\xi}, E)\, \mathrm{d}\boldsymbol{\xi} \right) \mathrm{d}E, \tag{1.5}$$

où $S_0(E)$ est la densité spectrale d'énergie. Outre le fait que les constructeurs aient amélioré la monochromatie des sources, ces effets sont généralement bien compensés, soit par des pré et/ou post traitements, soit en modélisant la dépendance en énergie de $x(\boldsymbol{\xi}, E)$.

2. L'effet de « volume partiel » intervient lorsque l'*épaisseur* du faisceau n'est plus négligeable vis-à-vis des variations géométriques de l'objet. Cet effet se caractérise par un flou dans l'image reconstruite qui peut rendre incertaine la localisation précise des interfaces. Il est possible de tenir compte de cet effet en introduisant une *fonction de faisceau* $\psi : \mathbb{R} \to \mathbb{R}$ dans le modèle de mesure. Dans ce cas, la relation liant l'image aux projection n'est plus (1.3) mais,

$$g_\psi(s, \varphi) = [\mathcal{R}_\psi x](s, \theta) \int \psi(s - \langle \boldsymbol{\xi}, \boldsymbol{\theta} \rangle)\, x(\boldsymbol{\xi})\, \mathrm{d}\boldsymbol{\xi}, \tag{1.6}$$

voir par exemple [Hanson et Wechsung, 1983, (1)]. En pratique, la compensation de ce type d'effet passe plus par un investissement sur l'instrumentation que sur la méthodologie d'inversion.

3. La *fonction d'ouverture* de la source peut introduire une distorsion si ses

dimensions ne sont plus négligeables par rapport à l'objet. Le phénomène peut être modélisé par un flou de convolution ; cependant, pour un arrangement éventail, ce flou n'intervient pas de manière identique sur toute l'image (le noyau de convolution n'est pas spatialement invariant). L'emploi d'une source fortement colimatée semble réduire ce problème.

4. La taille des capteurs est loin d'être *ponctuelle* et des distorsions sont introduites dans l'image suite à une convolution avec la fonction d'ouverture du capteur.
5. Pour des capteurs insuffisamment colimatés, la détection des photons *diffractés* lors de leur propagation peut conduire à un artefact de reconstruction. Ce phénomène se compense correctement par un post traitement sur l'image.

On souligne néanmoins que la transformation (1.3) est habituellement un modèle *fiable* du procédé de mesure axial en imagerie médicale. Cet état de fait résulte en bonne partie du travail des constructeurs qui soignent particulièrement l'instrumentation de manière à limiter les dégradations. En pratique, il est alors souvent suffisant de travailler sur un tomographe idéal défini par (1.3) pour développer des méthodes de reconstruction d'image.

Les hypothèses de travail pour ce document

Pour rendre l'exposé qui suit plus clair et la résolution du problème de reconstruction plus aisée, nous ferons un certain nombre d'hypothèses simplificatrices.

En particulier, on ne tiendra pas compte des artefacts de reconstruction dûs au durcissement et la diffraction du faisceau, et nous supposerons que les capteurs n'introduisent pas de distorsion (voir section précédente). D'autre part, même si la totalité des tomographes actuellement disponibles utilisent une géométrie en éventail, nous baserons notre exposé majoritairement sur un arrangement à rayons parallèles, la géométrie en éventail étant néanmoins adoptée afin de traiter des données réelles en géométrie hélicoïdale.

On souligne que ces choix ne portent pas atteinte à la généralité de nos travaux et permettent de se concentrer sur les performances intrinsèques des méthodes présentées sans en alourdir la présentation et la mise en œuvre.

Bibliographie annotée

La physique des phénomènes ainsi que les différentes modalités de mesure sont décrites de manière complète dans [Kak et Slaney, 1987, chap.4]. Pour une présentation complémentaire, nous renvoyons également à [Macovski, 1983, chap.3] et [Herman, 1980, chap.3]. Pour les artefacts de reconstruction, si [Kak et Slaney, 1987] semblent produire la synthèse la plus récente sur le sujet, la lecture de [Herman, 1980, chap.5] est instructive et finalement complémentaire.

1.3 Inversion de la TR en géométrie axiale

La transformée de Radon revêt un intérêt tout particulier en imagerie médicale à rayons X pour laquelle, sous certaines hypothèses réalistes, l'inversion de cette transformation peut être formellement associée à la résolution du problème d'imagerie. Conformément aux développements de la section 1.2, on suppose donc que la tomographie axiale se caractérise par un faisceau de rayons parallèles infiniment fins se propageant sans réfraction, ni diffraction. Il faut néanmoins souligner qu'en dépit des simplifications, ce problème d'inversion n'a rien de trivial. En outre, la compréhension des difficultés méthodologiques et pratiques qu'il révèle sont incontournables pour qui veut s'attacher à résoudre des problèmes plus complexes faisant intervenir une transformation de Radon.

N.B. Dans la mesure où notre objectif principal est d'étudier des méthodes numériques d'inversion, nous supposerons que l'existence des relations de ce document (majoritairement des intégrations, différentiations...) sont garanties de fait.

1.3.1 Notations et définitions

Dans ce qui suit, les lettres italiques grasses sont des vecteurs dont les composantes seront indicées par la lettre en italique, ex. $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_n)^T$. Les lettres capitales en italique sont des ensembles ou des espaces en dimension finie ou in-

finie ; ex. X. Les éléments d'espaces fonctionnels hilbertiens seront notés par des minuscules italiques (ex. x) et les capitales calligraphiées désigneront des opérateurs linéaires bornés sur ces espaces (ex. $\mathcal{R}$). On note $\mathrm{Dom}(x)$ le support de x, $\mathrm{Ker}(\mathcal{R})$ et $\mathrm{Im}(\mathcal{R})$ désignent respectivement le noyau et l'image de l'opérateur $\mathcal{R}$, $\overline{\mathrm{Im}(\mathcal{R})}$ la fermeture topologique de $\mathrm{Im}(\mathcal{R})$. Enfin, $\langle \cdot, \cdot \rangle$ et $|| \cdot ||$ désignent respectivement le produit scalaire et la norme euclidienne dans $\mathbb{R}^N$: $\forall \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^N$, $\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \sum_i a_i b_i$ et $||\boldsymbol{a}|| = \langle \boldsymbol{a}, \boldsymbol{a} \rangle^{1/2}$.

Définitions

Soit $x(\xi_1, \xi_2)$ la *distribution d'atténuation* que l'on cherche à reconstruire ; par la suite, on supposera que le support de x s'inscrit dans une *région de reconstruction* de rayon $\rho > 0$; *i.e.* $\mathrm{Dom}(x) \subset \{\boldsymbol{\xi} \in \mathbb{R}^N : ||\boldsymbol{\xi}|| \leq \rho\}$ — hypothèse légitime puisque le patient ou l'objet doit s'insérer à l'intérieur du tomographe.

L'interprétation visuelle de la transformation de Radon est celle des intégrales de ligne de $x(\xi_1, \xi_2)$ illustrée par la figure 1.4 : $\boldsymbol{\theta} = (\cos\varphi, \sin\varphi)$ décrit le cercle unité S^1 et $\Theta^\perp$ est la droite définie par la direction du vecteur $\boldsymbol{\theta}^\perp = (-\sin\varphi, \cos\varphi)$. C'est ce que formalise la définition suivante.

Définition 1 *Soit S^1 la sphère unité dans $\mathbb{R}^2$. Pour $\boldsymbol{\theta} \in S^1$, on introduit $\Theta^\perp$ la direction perpendiculaire à $\boldsymbol{\theta}$. Pour $x : \mathbb{R}^2 \to \mathbb{R}$ une fonction intégrable, la*

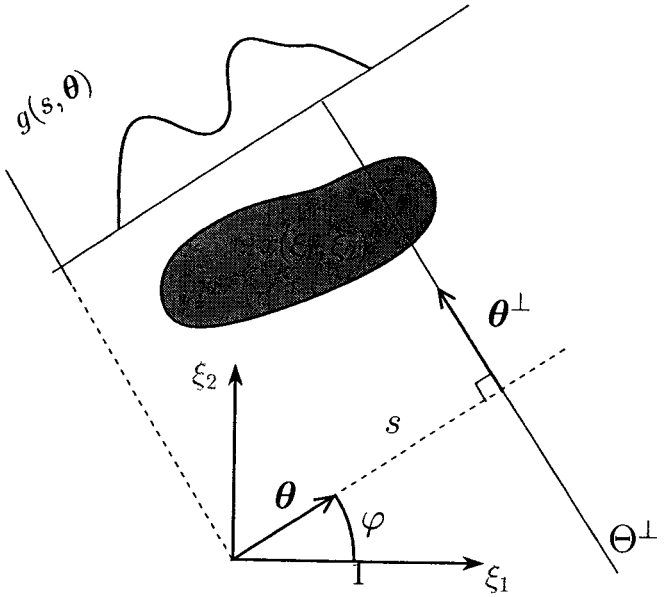


Figure 1.4: Illustration de la TR en dimension 2.

transformée de Radon $g : \mathbb{R} \times S^1 \to \mathbb{R}$ *s'écrit de manière équivalente*

$$g(s,\boldsymbol{\theta}) = \int_{\Theta^\perp} x(\boldsymbol{\xi} + s\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\xi}, \qquad ou \qquad g(s,\boldsymbol{\theta}) = \int_{\mathbb{R}^2} \delta(s - \langle \boldsymbol{\xi}, \boldsymbol{\theta} \rangle)\, x(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi} \quad (1.7)$$

où on a posé $\boldsymbol{\xi} = (\xi_1, \xi_2)$, $\mathrm{d}\boldsymbol{\xi} = \mathrm{d}\xi_1\,\mathrm{d}\xi_2$, *et où* $\delta(s)$ *est la distribution de Dirac monodimensionnelle. On écrira cette transformation sous forme fonctionnelle*

$$g = \mathcal{R}x \quad (1.8)$$

avec $\mathcal{R} : X \to G$ *l'opérateur de Radon reliant entre eux les espaces de Hilbert de l'objet et des projections.*

Cette définition s'étend directement pour les fonctions $x : \mathbb{R}^N \to \mathbb{R}$. Dans ce

cas, la TR se définit en intégrant suivant tous les hyperplans de dimension $N-1$, [Deans, 1983, Chap. 2]. Bien que notre propos est de traiter de reconstructions tomographiques *tridimensionnel* (3D), la transformée de Radon *bidimensionnel* (2D) sera suffisante dans la plupart de nos développements.

La *transformation de Fourier* (abrégé par TF) d'une fonction sera notée par un « $\hat{}$ » *e.g.,* ; ainsi pour une fonction réelle $x(\xi_1, \xi_2)$ intégrable, la TF $\hat{x}(\nu_1, \nu_2)$ est en général à valeur complexe et définie par

$$\hat{x}(\boldsymbol{\nu}) = (2\pi)^{-1} \int_{\mathbb{R}^2} x(\boldsymbol{\xi})\, e^{-i\langle \boldsymbol{\xi}, \boldsymbol{\nu} \rangle}\, \mathrm{d}\boldsymbol{\xi}, \quad \text{où} \quad \boldsymbol{\nu} = \begin{pmatrix} \nu_1 \\ \\ \nu_2 \end{pmatrix} \in \mathbb{R}^2, \tag{1.9}$$

ce qu'on écrira sous forme fonctionnelle $\hat{x} = \mathcal{F}x$, avec $\mathcal{F} : X \to \hat{F}$ l'opérateur de Fourier 2D. Pour les fonctions définies sur $\mathbb{R} \times S^1$, on introduit également la transformation de Fourier *par rapport à la première variable* définie par

$$\hat{v}(\omega; \boldsymbol{\theta}) = (2\pi)^{-1/2} \int_{\mathbb{R}} v(s, \boldsymbol{\theta})\, e^{-is\omega}\, \mathrm{d}s, \quad \text{avec} \quad \omega, s \in \mathbb{R},\ \boldsymbol{\theta} \in S^1. \tag{1.10}$$

1.3.2 Formulation du problème de reconstruction

En pratique, on ne dispose évidemment que d'un nombre fini d'angles de projection, chaque projection étant elle-même échantillonnée : les données sont sonc issues d'un *échantillonnage* de g suivant P angles de projection distincts et M

échantillons par projection :

$$g(s_l, \varphi_p) = [\mathcal{R}x](s_l, \varphi_p) \tag{1.11}$$

$$l = -L+1, \cdots, L; \qquad p = 1, \cdots, P$$

où $\{\varphi_p = p\Delta\varphi\}$ avec $\Delta\varphi = \pi/P$, et $\{s_l = l\Delta s\}$ avec $\Delta s = \rho/L$. Ceci nous conduit à formuler le problème de reconstruction de la manière suivante :

1ère question :

Comment reconstruire $x(\xi_1, \xi_2)$ à partir des données $g(s_l, \varphi_p)$?

On doit prendre conscience qu'une telle formulation du problème est naïve : pour des raisons qui vont devenir bientôt claire, il est *vain* d'espérer retrouver x à partir des mesures disponibles en nombre fini. Il est par contre possible de donner une réponse assez complète à une question moins ambitieuse mais finalement d'un intérêt majeur :

2ème question :

Comment construire une approximation « pertinente » de $x(\xi_1, \xi_2)$ à partir de $g(s_l, \varphi_p)$?

Pour tenter de répondre à cette question, une étude d'unicité et de stabilité du problème de reconstruction est utile et finalement instructive pour une compréhension des lacunes des méthodes de reconstruction standard.

1.3.3 Difficultés méthodologiques d'inversion

Si on dispose de g, on sait depuis les travaux effectué par JOAN RADON en 1917 que la TR ci-dessus a une inverse explicite[4] :

$$x(\boldsymbol{\xi}) = \frac{1}{4\pi^2} \int_0^{2\pi} \int_{\mathbb{R}} \frac{g'(s, \boldsymbol{\theta})}{\langle \boldsymbol{\xi}, \boldsymbol{\theta} \rangle - s} \, \mathrm{d}s \, \mathrm{d}\varphi \qquad \text{avec} \qquad \boldsymbol{\theta} = \begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix} \tag{1.12}$$

avec g' la dérivée de g par rapport à la variable s. Bien que cette dernière égalité résolve *en principe* notre problème de reconstruction tomographique, le résultat formulé en ces termes masque des difficultés méthodologiques et pratiques qui ne manqueront pas de se révéler puisqu'on sait depuis la fin des années 70 que l'inversion de la TR est un problème *mal posé*. On rappelle d'abord que, pour ce problème, les trois conditions de HADAMARD qui définissent un problème *bien posé* sont [Tikhonov et Arsénine, 1976, p. 14] :

$$\begin{array}{lll} \text{(A)} & \forall g \in G,\ g \in \mathrm{Im}(\mathcal{R}), & \text{(existence)} \\ \text{(B)} & \mathrm{Ker}(\mathcal{R}) = \{0\}, & \text{(unicité)} \\ \text{(C)} & \mathrm{Im}(\mathcal{R}) = \overline{\mathrm{Im}(\mathcal{R})} & \text{(continuité).} \end{array} \tag{1.13}$$

[4] L'intégrale (1.12) est interprétée au sens de *la valeur principale de Cauchy* au point de discontinuité $\langle \boldsymbol{\xi}, \boldsymbol{\theta} \rangle = s$, c.à.d.

$$\int \frac{f(u)}{x-u} \mathrm{d}u = \lim_{\epsilon \to 0} \left[\int_{-\infty}^{\epsilon} \frac{f(u)}{x-u} \, \mathrm{d}u + \int_{\epsilon}^{-\infty} \frac{f(u)}{x-u} \, \mathrm{d}u \right]$$

quand cette limite existe.

Un problème ne vérifiant pas ces trois conditions est appelé *mal posé*, et il est bien connu que la résolution numérique de tels problèmes est sinon impossible, tout du moins particulièrement délicate, [Demoment, 1989; Nashed, 1981]. Le problème d'inversion de la TR définie par (1.12) est mal posé puisqu'on ne peut garantir [Natterer, 1986] :

ni l'*unicité* pour tout nombre *fini* d'angles de projection (B),

ni la *continuité* de la solution face aux perturbations (C),

1.3.4 Unicité et stabilité de l'inversion

Si le problème initial est formellement insoluble, notre première idée consiste à le modifier suffisamment pour retrouver l'unicité et la continuité de la solution.

Unicité de la solution

Comme indiqué plus haut, l'unicité de l'inversion n'est pas garantie. Soyons maintenant plus précis et introduisons

$$G_{\text{donnes}} = \{[\mathcal{R}x](s, \theta_p) \mid p = 1, \cdots, P\}$$

l'ensemble des P projections prisent dans un cône quelconque de $\mathbb{R}^2$. D'après [Smith *et al.*, 1977, Sec. 4], l'objet peut être déterminé de manière unique par *tout ensemble infini* d'angles de projection ($P = \infty$), mais par *aucun ensemble fini* ($P < \infty$) ;

ce dernier cas correspondant bien évidemment au cadre pratique. Cependant, il est possible de montrer que les éléments *non triviaux* de $\mathrm{Ker}(\mathcal{R})$ ont pour support fréquentiel

$$\Omega^c = \{\boldsymbol{\nu} \in \mathbb{R}^2 : |\boldsymbol{\nu}| > P/\rho\}, \tag{1.14}$$

Pour $P < \infty$, l'ensemble G_{donnes} détermine de manière *unique* $\hat{x}$ sur Ω, le complémentaire de Ω^c ; on peut donc assurer l'unicité de l'*approximation* $\bar{x}$ de x défini par

$$\hat{\bar{x}}(\boldsymbol{\nu}) = [\mathcal{F}\bar{x}](\boldsymbol{\nu}) = \begin{cases} \hat{x}(\boldsymbol{\nu}) \text{ si } \quad \boldsymbol{\nu} \in \Omega \\ 0 \qquad \text{sinon.} \end{cases}$$

Continuité $\not\Rightarrow$ stabilité de la solution

Comme nous le verrons au paragraphe suivant, le problème *numérique* de reconstruction est entièrement discrétisé. Dans ce contexte, $\mathcal{R}$ se décompose sur une base de dimension *finie* et devient un opérateur matriciel $\boldsymbol{R} : \mathbb{R}^N \to \mathbb{R}^M$. Dans ce cadre, la continuité de l'inversion numérique est restaurée puisqu'en dimension fini, on a toujours $\mathrm{Im}(\boldsymbol{R})$ fermé. Ainsi, la reconstruction d'une approximation $\bar{x}$ (cf. ci-dessus) de x sur N points est un problème formellement bien posé permettant d'accéder aux composantes « basse fréquence » de x.

On déduit directement de (1.14) que la *résolution spatiale* de cette approximation est ρ/P. Cependant, la qualité de l'approximation dépend également (et

souvent de manière dramatique) de la *stabilité* du problème numérique puisqu'il devient évident que le problème fait intervenir une version « perturbée » de (1.11)

$$y(s_l, \varphi_p) = g(s_l, \varphi_p) + \varepsilon(s_l, \varphi_p) \tag{1.15}$$

où $\varepsilon(s_l, \varphi_p)$ regroupe les bruits d'instrumentation et les erreurs de modèle. Le *mauvais conditionnement* [Demoment et Idier, 2001*a*] compromet alors sévèrement la robustesse de la reconstruction. L'instabilité numérique est particulièrement prononcée si l'excursion angulaire est restreinte (cf. section 6.2). Toutefois, si l'excursion angulaire est suffisante, le problème de reconstruction peut être suffisamment stabilisé par un simple filtrage passe bas : c'est la technique retenue en pratique.

Épilogue

On notera que le problème d'inversion associé à la tomographie axiale à rayons X est *structurellement simple* (opérateur linéaire et forme explicite de l'inverse). D'autre part, on montre que son caractère mal posé n'est pas très sévère puisque l'inversion de la TR est seulement *modérément mal posée* [Faridani, 1999, Th. 4.1]. On doit également ajouter que le contexte de mesure contribue également à rendre l'inversion plus commode puisque (1) le rapport signal à bruit (abrég. RSB) demeure assez important (ce qui n'est d'ailleurs pas sans contrepartie puisque le RSB est une image de la dose de rayonnement administrée au patient), et (2)

la source effectue une rotation complète autour du patient. Ce deuxième aspect est crucial pour la stabilité de l'inversion puisqu'une excursion angulaire restreinte conduit à une instabilité numérique accrue lors de la reconstruction. On verra justement que ce phénomène contribue à diminuer la précision des reconstruction en géométrie hélicoïdale.

Bibliographie annotée

Pour une introduction complète et lisible sur la TR, nous conseillons la remarquable monographie de [Deans, 1983] ; on trouvera aussi dans cet ouvrage une revue de littérature impressionnante sur les applications de la TR. La monographie de [Herman, 1980, Chap. 16] constitue une lecture complémentaire accessible. Les difficultés méthodologiques d'inversion ainsi que les principaux résultats mathématiques dans le domaine continu sont étudiés de manière complète dans la monographie de [Natterer, 1986] ainsi que dans le remarquable article de synthèse de [Faridani, 1999]. Ces deux dernières références sont d'une lecture ardue mais instructive. Pour une présentation des difficultés d'inversion des problèmes mal posés et mal conditionnés, nous renvoyons le lecteur aux monographies de [Tikhonov, 1963] et à [Demoment et Idier, 2001*a*] ainsi qu'aux articles de synthèse de[Demoment, 1989] et de [Nashed, 1981].

CHAPITRE 2

MÉTHODES NUMÉRIQUES STANDARD DE RECONSTRUCTION AXIALE

Après une rapide présentation des difficultés méthodologiques d'inversion de la TR 2D, on s'intéresse dans ce chapitre plus particulièrement aux méthodes *numériques* permettant la reconstruction. Soulignons à ce propos que ce chapitre dédié aux algorithmes de reconstruction en tomographique *axiale* est important pour au moins deux raisons : il met en évidence les écueils rencontrés par les approches standard (ex. RPC) — cf. section 2.1, et il ouvre la voie d'une analyse des limites des approches adoptées en reconstruction hélicoïdales standard.

N.B. Notre propos n'est pas ici de faire un inventaire de toutes les approches de reconstruction possibles et de tous leurs modes d'implantation. On se contentera plutôt d'un inventaire que nous pensons pertinent et qui reflète les possibilités des tomographes standard.

2.1 Inversion basée sur l'expression analytique

La démarche la plus répandue traite du problème d'inversion dans sa formulation analytique et nous présentons deux de ses représentants les plus populaires : la

rétroprojection convoluée (abrég. RPC) et la *synthèse de Fourier*. Au cours de cet exposé, on tentera de mettre en relief les raisons qui les rendent inadaptées pour notre application cherchant à produire des images de précision.

L'algorithme de rétroprojection convoluée

Cette méthode de reconstruction, de loin la plus répandue dans les tomographes, s'appuie principalement sur une implantation numérique de (1.12). On privilégie néanmoins ici une approche tirée de [Natterer, 1999, Sec. 2] ne faisant pas intervenir une intégrale singulière. On introduit pour une fonction $g : \mathbb{R} \times S^1 \to \mathbb{R}$, l'opérateur de *rétroprojection*[1] $\mathcal{B}$ tel que :

$$[\mathcal{B}g](\boldsymbol{\xi}) = \int_{S^1} g(\langle \boldsymbol{\theta}, \boldsymbol{\xi} \rangle, \boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta}; \tag{2.1}$$

on montre alors la propriété suivante

$$V \star x = \mathcal{B}(v * g) \tag{2.2}$$

[1] $\mathcal{B}$ n'est autre que l'opérateur *adjoint* de $\mathcal{R}$, c.à.d. le seul opérateur tel que $\forall x \in X$, $g \in G$, $\langle \mathcal{R}x, g \rangle_X = \langle x, \mathcal{B}g \rangle_Y$, avec $\langle \cdot, \cdot \rangle_X$ et $\langle \cdot, \cdot \rangle_Y$, respectivement, le produits scalaire associé à X et à Y.

où $\star$ et $*$ désignent la convolution bidimensionnelle et mono dimensionnelle, respectivement, *i.e.*

$$\int_{\mathbb{R}^2} V(\boldsymbol{\xi}-\boldsymbol{y})x(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} = \int_{S^1}\int_{\mathbb{R}} v(\langle\boldsymbol{\theta},\boldsymbol{\xi}\rangle - s,\boldsymbol{\theta})g(s,\boldsymbol{\theta})\,\mathrm{d}s\,\mathrm{d}\boldsymbol{\theta}. \tag{2.3}$$

où $V : \mathbb{R}^2 \to \mathbb{R}$ et $v : \mathbb{R}\times S^1 \to \mathbb{R}$ sont reliés par leurs transformées de Fourier respectives,

$$\hat{V}(\boldsymbol{\nu}) = 2(2\pi)^{1/2}||\boldsymbol{\nu}||^{-1}\hat{v}(||\boldsymbol{\nu}||;\boldsymbol{\theta}), \quad \boldsymbol{\nu}\not\equiv \mathbf{0};$$

$\hat{v}(\omega,\boldsymbol{\theta})$ étant défini par (1.10). Dans (2.2), le membre de gauche « contrôle l'approximation » faite sur la reconstruction, l'idée étant que V soit une approximation de la distribution de Dirac δ puisqu'alors $V\star x$ est proche de x. Pour $V\equiv\delta$, on a $\hat{V}(\nu) = (2\pi)^{-1}$ et le *filtre de reconstruction* v s'écrit,

$$\forall\boldsymbol{\theta}\in S^1, \qquad \hat{v}(||\boldsymbol{\nu}||;\boldsymbol{\theta}) = \hat{v}(||\boldsymbol{\nu}||) = 2(2\pi)^{-3/2}||\boldsymbol{\nu}||$$

qui est indépendant de l'angle de projection. On déduit une forme analytique équivalente à (1.12) pour inverser la TR

$$x(\boldsymbol{\xi}) = \frac{1}{2}(2\pi)^{-3/2}\int_{S^1}\int_{\mathbb{R}} |\omega|\hat{g}(\omega;\boldsymbol{\theta})\,e^{i2\pi\omega s}\,\mathrm{d}\omega\,\mathrm{d}\boldsymbol{\theta} \qquad \omega\in\mathbb{R}.$$

où $\hat{g}(\omega;\boldsymbol{\theta})$ est la TF par rapport à la première variable de $g(s,\boldsymbol{\theta})$. Le « filtre » de reconstruction $v(||\boldsymbol{\nu}||) = ||\boldsymbol{\nu}||$ reflète le caractère instable de l'inversion et n'est d'ailleurs pas physique. En pratique, on fixe une bande passante Ω qui définira la résolution de l'approximation, puis on impose en général la forme paramétrée suivante à V :

$$\hat{V}_\Omega(\boldsymbol{\nu}) = \frac{1}{2\pi}\hat{\phi}\left(\frac{||\boldsymbol{\nu}||}{\Omega}\right) \qquad \text{avec,} \qquad \hat{\phi}(\omega) = \begin{cases} \approx 1, & |\omega| \leq 1 \\ = 0, & |\omega| > 1 \end{cases}$$

la famille de filtres de reconstruction présente alors une symétrie de révolution et s'écrit

$$\hat{v}_\Omega(||\boldsymbol{\nu}||) = 2(2\pi)^{3/2}\hat{\phi}\left(\frac{||\boldsymbol{\nu}||}{\Omega}\right);$$

les exemples les plus répandus étant la fonction porte, cosinusoïdale, sinus cardinal. L'algorithme de *rétroprojection convoluée* (abrég. RPC) implanté dans les tomographes médicaux s'appuie sur la « méthode des trapèzes » pour estimer le membre de droite de (2.3). Ainsi partant de $g(s_l,\boldsymbol{\theta}_p)$ défini par (1.11), on écrit

$$[V_\Omega \star x](\boldsymbol{\xi}) = \int_S\int_{\mathbb{R}} v_\Omega(\langle\boldsymbol{\theta},\boldsymbol{\xi}\rangle - s)g(s,\boldsymbol{\theta})\,\mathrm{d}s\,\mathrm{d}\boldsymbol{\theta} \tag{2.4}$$

$$\approx \frac{2\pi\rho}{LP}\sum_{p=1}^{P}\sum_{l=-L+1}^{L} v_\Omega(\langle\boldsymbol{\theta}_p,\boldsymbol{\xi}\rangle - s_l)g(s_l,\boldsymbol{\theta}_p). \tag{2.5}$$

Afin d'éviter le calcul d'une somme double pour chaque point de reconstruction, on introduit l'intermédiaire

$$h(s, \boldsymbol{\theta}_p) = \frac{\rho}{L} \sum_{l=-L+1}^{L} v_\Omega(s - s_l) g(s_l, \boldsymbol{\theta}_p)$$

ce qui permet de ré-écrire (2.4)

$$[V_\Omega \star x](\boldsymbol{\xi}) \approx \frac{2\pi}{P} \sum_{p=1}^{P} h(\langle \boldsymbol{\theta}_p, \boldsymbol{\xi} \rangle, \boldsymbol{\theta}_p).$$

Moyennant une interpolation, la mise en œuvre ne requiert plus qu'une seule somme par point de reconstruction, cf. Natterer [1999] pour les détails. Cet algorithme requière en principe $O(L^2 P)$ pour reconstruire une grille de $L \times L$ éléments. La résolution spatiale de $1/\Omega$ mètres de cet algorithme impose des conditions sur l'échantillonnage :

$$L \geq \frac{1}{\pi} \rho \Omega \qquad\qquad P \geq \Omega \rho; \tag{2.6}$$

la première condition se déduit du théorème de SHANNON *via*[2] le théorème de la tranche de FOURIER présenté ci-dessous, la seconde est la condition donnée en section 1.3.4. Ces conditions ne garantissent néanmoins pas la stabilité de la solution numérique si l'excursion angulaire est insuffisante ou si le rapport signal à

[2]Ce théorème permet d'affirmer que $x(\xi)$ et ses projections ont essentiellement le même support fréquenciel. Ainsi, si le pas d'échantillonnage des projections est $\Delta_s = \rho/L$, alors le théorème de SHANNON indique que le support fréquentiel $\Omega \leq L\pi/\rho$ peut être reconstruit sans repliement spectral.

bruit est défavorable.

La synthèse de Fourier

Le théorème de la *tranche centrale de Fourier* met à jour la relation forte existant entre la TF par rapport à la première variable de $g(s, \boldsymbol{\theta})$ et la TF de $x(\boldsymbol{\xi})$ — cf. Natterer [1986] :

$$\hat{g}(\omega; \boldsymbol{\theta}) = (2\pi)^{1/2} \hat{x}(\omega \boldsymbol{\theta}). \tag{2.7}$$

On peut alors aborder la reconstruction directement dans le domaine de FOURIER en s'appuyant sur ce lien bijectif : l'objectif est alors d'inverser une approximation de $\hat{x}$ obtenue *via* les TF monodimensionnelles des projections. La mise en œuvre numérique fait alors appel à la *transformée de Fourier rapide* (abrég. TFR) qui est une version optimisée de la *transformée de Fourier discrète* (abrég. TFD) : [Natterer, 1999]

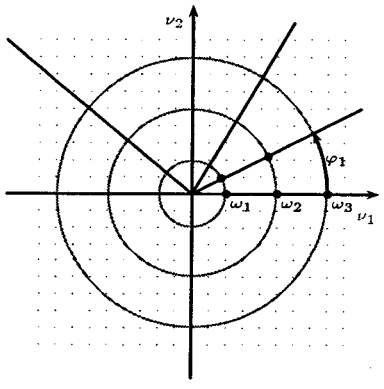


Figure 2.1: Échantillonnage radiale de la TF de x.

1. On se donne Ω, une bande passante définissant la résolution de la reconstruction et une fréquence d'échantillonnage $\omega_e = 2\pi L/\rho$ vérifiant la première condition de (2.6), c.à.d. $\omega_e \geq 2\Omega$.

2. La TFD fournit L points fréquentiels distincts[3] par projection ; ainsi en posant la fréquence discrète $\omega_r = r\omega_e/2L$, on calcule

$$\hat{g}(\omega_r; \boldsymbol{\theta}_p) = (2\pi)^{-1/2} \frac{\rho}{L} \sum_{l=-L+1}^{L} g(s_l, \boldsymbol{\theta}_p) e^{-i2\pi l \omega_r/\omega_e} \qquad r = 0, \cdots, L-1. \quad (2.8)$$

3. Par le théorème de la tranche centrale, le calcul de (2.8) pour $p = 1, \cdots, P$ fournit un ensemble de $2L \times P$ valeurs $\{\hat{x}(\omega_r \boldsymbol{\theta}_p)\}_{r,p}$ qui échantillonnent *radialement* le domaine de Fourier (cf. figure 2.1).

4. Une *interpolation* ramène l'ensemble des points fréquentiels $\hat{x}(\omega_r \boldsymbol{\theta}_p)$ répartis radialement à un ensemble de valeurs $\tilde{x}(j,k)$ sur une grille *cartésienne, i.e.*

$$\{\hat{x}(\omega_r \boldsymbol{\theta}_p)\}_{r,p} \quad \rightsquigarrow \quad \{\tilde{x}(j,k)\}_{(j,k)\in I}$$

où I est l'ensemble des indices définit par $I = \{(u,v) \in \mathbb{Z}^2 : \max(|u|, |v|) \leq L\}$ avec $\mathbb{Z}$ l'ensemble des entiers relatifs.

[3] g étant réelle, sa TF est hermitienne et les L points correspondant à $r = -1, \cdots, -L+1$, n'apportent pas d'information supplémentaire.

5. En utilisant la TFD 2D inverse, l'ensemble des $\tilde{x}(j,k)$ permet le calcul d'une approximation de x sur une grille cartésienne :

$$x(u,v) \approx \frac{\pi}{2\rho} \sum_{(j,k)\in I} \tilde{x}(j,k) e^{i\pi(ju+kv)/L}, \qquad \forall (u,v) \in I.$$

On s'en doute, l'intérêt majeur de cet algorithme est son coût d'implantation[4] puisqu'au total, la complexité algorithmique de $O(L(P+L)\log L)$ est nettement plus intéressante que celle de la RPC qui nécessite $O(L^2P)$ pour reconstruire une grille de $L \times L$ éléments. *L'algorithme tel que nous venons de le décrire n'est pourtant pas utilisé en pratique car, même dans un contexte normalement favorable, il se révèle particulièrement instable.* En fait, la découverte d'algorithmes stables basés sur ce principe est plus récente et nécessite la suppression de l'étape d'interpolation qui introduisait cette instabilité ; cf. [Cheung et Lewitt, 1991; Edholm et Herman, 1988].

Épilogue

À notre connaissance, les tomographes à rayons X implantés en milieu hospitalier utilisent tous exclusivement la RPC qui produit des images acceptables pour

[4]Une mise en œuvre à base de TFR conduit respectivement à $O(PL\log L)$ et $O(L^2\log L)$ opérations pour l'étape 2 et 5 ; si on considère que l'interpolation requière $O(1)$ par élément de la grille cartésienne, alors celle-ci peut être fait en $O(L^2)$ opérations. On en déduit finalement que la reconstruction demande de l'ordre de $L(P+L)\log L$ opérations.

le diagnostique médical. Cette technique s'avère néanmoins incapable de produire des images d'une précision inférieure au millimètre, ceci même pour un niveau de stabilité du problème satisfaisant (fort RSB et nombre important de projections) ; on pourra en particulier se reporter à [Doré et Goussard, 1997] qui estiment une précision de l'ordre de 1,4 mm sur un tomographe hospitalier classique.

Outre les instabilités introduites par l'implantation numérique de la méthode[5], *le problème provient principalement du filtre de reconstruction et de la limitation de bande passante qu'il introduit.* Ainsi, en coupant les hautes fréquences (abrég. HF) du bruit qui seraient exagérément amplifiées, ce filtre régularise l'inversion mais il coupe également les composantes HF de l'image. Les bords francs dans les images sont alors lissés, ce qui hypothèque toute extraction précise des frontières « os/tissus mous » et par là même, empêche une reconstruction nette des contours.

Bibliographie annotée

Les résultats présentés dans cette section s'étendent au cadre 3D ainsi que pour des arrangements des faisceaux divergents. Le lecteur intéressé lira avec intérêt l'article de synthèse récent [Natterer, 1999] qui présente de manière rigoureuse les diverses méthodes numériques d'inversion de la TR. La monographies de [Kak et

[5]Indiquons entre autres : (1) le passage à des sommes discrètes pour évaluer des intégrales, (2) l'interpolation nécessaire pour reconstruire l'image sur une grille cartésienne, (3) la discrétisation de la réponse impulsionnelle choisie pour la convolution ; voir [Kak et Slaney, 1987, Sec. 3.3.3] et [Natterer, 1999, Sec. 2] pour plus de détails.

Slaney, 1987] ainsi que l'article de [Natterer, 1997] couvrent une matière plus large : tomographie d'émission, de transmission, à source diffractante ou non ; la première référence adopte une présentation orientée « traitement du signal ». On pourra également lire [Herman *et al.*, 1979, Chap. 2] et [Herman, 1980, Chap.7-9] pour compléter la présentation des méthodes des reconstructions par RPC et de leurs modalités d'implantation. Enfin, on pourra noter que certains auteurs cherchent à améliorer les performances de la RPC par l'emploi d'ondelettes [Delaney et Bresler, 1995] ou par l'emploi de filtres non linéaires [Andia *et al.*, 2001].

2.2 Approches algébriques : une discrétisation de la TR

On se propose maintenant d'aborder le problème de reconstruction à partir d'une forme *discrétisée* de la scène continue $x(\boldsymbol{\xi})$. Dans ce qui suit, le domaine de reconstruction D est un sous ensemble rectangulaire de $\mathbb{R}^2$ partitionné en $N = N_{\xi_1} \times N_{\xi_2}$ *pixels*, *i.e.* en sous-ensembles disjoints formant un pavage de cette partie du plan.

2.2.1 Formulation

La « méthode d'expansion en série », introduite par HERMAN [Herman, 1980, Chap. 6], adopte une modélisation *paramétrique* de la scène continue en la décom-

posant sur une base de fonctions appropriées :

$$x(\boldsymbol{\xi}) \approx x_N(\boldsymbol{\xi}) = \sum_{n=1}^{N} x_n b_n(\boldsymbol{\xi}); \tag{2.9}$$

$\forall n \in \{1, \cdots, N\}$, $b_n : \mathbb{R}^2 \to \mathbb{R}$ est une fonction de base donnée initialement (généralement des indicatrices du support des pixels de reconstruction pris dans un ordre *lexicographique*).

Approximation numérique de l'opérateur de TR

En adoptant ce formalisme, le problème de reconstruction tomographique se « réduit » à la résolution d'un problème numérique d'inversion découlant d'une discrétisation de la TR : à partir du modèle (2.9) et en s'appuyant sur l'expression de la TR donnée par (1.7), on écrit

$$\begin{aligned} \forall (l,p) \in \{-L+1, \cdots, L\} \times \{1, \cdots, P\}, & \\ g(s_l, \boldsymbol{\theta}_p) &= \int_{\mathbb{R}^2} \delta\big(s_l - \langle \boldsymbol{\xi}, \boldsymbol{\theta}_p \rangle\big) x_N(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} \\ &= \sum_{n=1}^{N} x_n \int_{\mathbb{R}^2} \delta\big(s_l - \langle \boldsymbol{\xi}, \boldsymbol{\theta}_p \rangle\big) \, b_n(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi}, \end{aligned}$$

et en concaténant les N paramètres x_n et les $M = 2L \times P$ échantillons $g(s_l, \boldsymbol{\theta}_p)$ dans les vecteurs $\boldsymbol{x} \in \mathbb{R}^N$ et $\boldsymbol{g} \in \mathbb{R}^M$, respectivement, la dernière relation peut être

ré-écrite

$$\boldsymbol{g} = \boldsymbol{R}\boldsymbol{x} \tag{2.10}$$

avec $\boldsymbol{R} \in \mathbb{R}^{M \times N}$ telle que $(\boldsymbol{R})_{m,n}$ (c.à.d. l'élément en ligne m-colonne n) représente la contribution de b_n dans le rayon l de la projection p; où $m = 2L(p-1) + l + L$ on a :

$$(\boldsymbol{R})_{m,n} = \int_{\mathbb{R}^2} \delta\big(s_l - \langle \boldsymbol{\xi}, \boldsymbol{\theta}_p \rangle\big)\, b_n(\boldsymbol{\xi})\, \mathrm{d}\boldsymbol{\xi}.$$

Notons pour la suite que l'opérateur $\boldsymbol{R}$ est souvent de *très grande taille* mais que son taux de remplissage est habituellement très faible — $\boldsymbol{R}$ est une *matrice creuse* ; ces points critiques pour la mise en œuvre seront de nouveau abordés au chapitre 7.

Finalement on souligne que ce formalisme est particulièrement flexible et permet, en particulier, de traiter sans difficulté des arrangements de faisceau *non parallèle* et/ou à épaisseur de rayon *non nulle*. Dans ces deux cas, on aboutit à une relation linéaire similaire à (2.10) où $(\boldsymbol{R})_{m,n}$ représentent toujours la contribution de b_n dans l'échantillon m. Comme exemple, indiquons que le recours à (1.6) plutôt qu'à la TR permet d'intégrer l'épaisseur du faisceau dans la modélisation et conduit à

$$(\boldsymbol{R})_{m,n} = \int_{\mathbb{R}^2} \psi(s_l - \langle \boldsymbol{\xi}, \boldsymbol{\theta}_p \rangle)\, b_n(\boldsymbol{\xi})\, \mathrm{d}\boldsymbol{\xi}$$

où ψ est la fonction de faisceau.

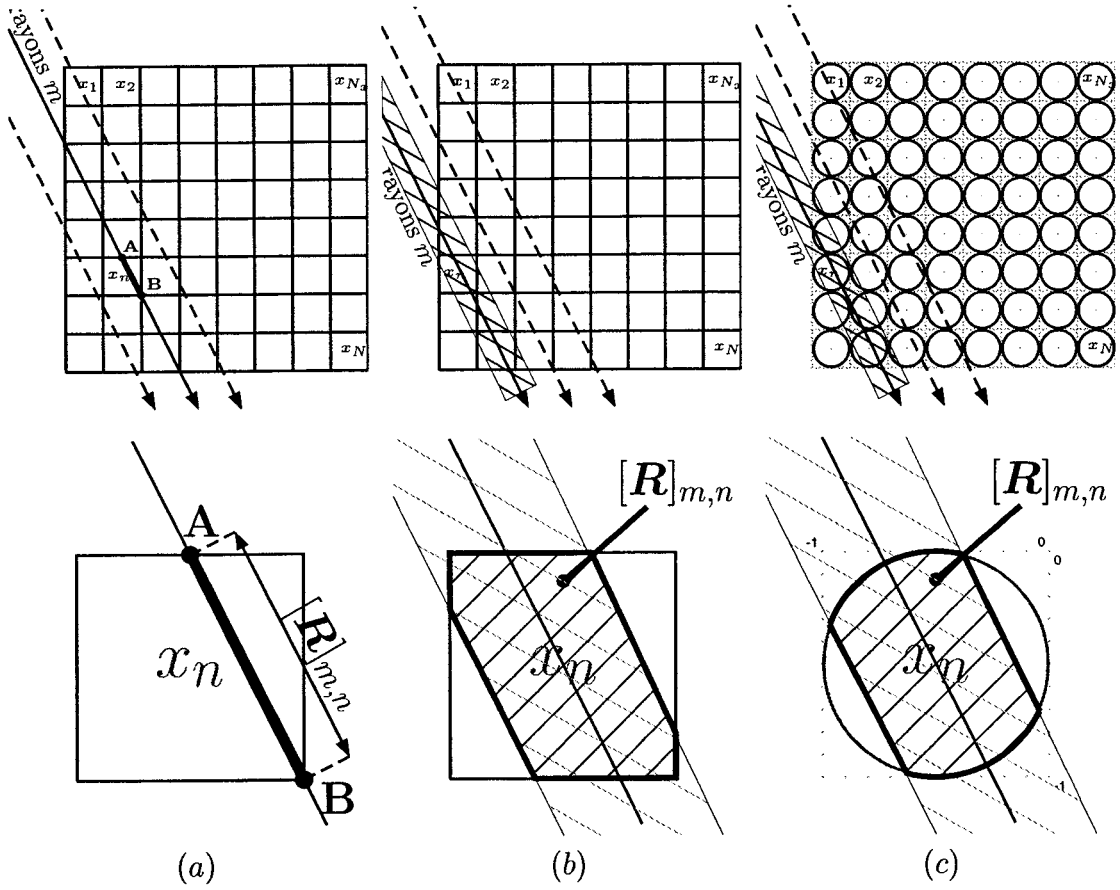


Figure 2.2: Calcul des éléments de $\boldsymbol{R}$ pour un modèle à base d'indicatrices sur les pixels avec indexation lexicographique ; (a) représente le cas d'un faisceau sans épaisseur avec des « pixels carrés », (b) le cas d'un faisceau de type « bande », et (c) le cas d'un faisceau de type « bande » avec des « pixels cylindriques ».

Modèles paramétriques standard

Le choix de la base $\{b_n(\boldsymbol{\xi})\}_{n=1}^N$ a bien sûr une influence directe sur la structure de $\boldsymbol{R}$ et sur son taux de remplissage. Un choix très répandu consiste à prendre pour b_n la fonction indicatrice sur le pixel n : dans ce cas, les éléments $(\boldsymbol{R})_{m,n}$ représentent la longueur de la fraction de rayons traversant le pixel. Si on souhaite tenir compte de l'*épaisseur du faisceau*, il est alors usuel de choisir pour ψ une

fonction « bande » et, dans ce cas, $(\boldsymbol{R})_{m,n}$ représente la surface de bande dans le pixel n. On trouvera une illustration de ces deux modèles sur la figure 2.2.

L'emploi d'un modèle à rayons infiniment fins rend le calcul des éléments de $\boldsymbol{R}$ aisé au prix d'un effet de « pixélisation » dans les reconstructions (*checkerboard effect*). Bien que le modèle de projection « à bande » soit susceptible d'éliminer cet effet, le calcul des aires d'intersection est suffisamment complexe pour rendre la construction de $\boldsymbol{R}$ peu attractive pour les grandes images. En pratique, le choix de fonctions b_n qui présentent une *symétrie radiale* présente un intérêt certain car le calcul des aires et simplifié. On pourra par exemple choisir comme base des *cylindres* de hauteur unité inscrits dans le support des pixels partitionnant le domaine de reconstruction ; des fonctions « douces » à faible support ont également été introduites. Nous renvoyons le lecteur à [Hanson et Wechsung, 1985] et à [Lewitt, 1992] pour une étude des différentes fonctions de base et de leur influence sur la reconstruction.

2.2.2 Abordons le problème de reconstruction

En partant de l'approximation donnée par (2.10) des intégrales de ligne $\boldsymbol{g}$, on déduit un modèle d'observation liant les paramètres $\boldsymbol{x}$ aux données observées $\boldsymbol{y}$

$$\boldsymbol{y} = \boldsymbol{R}\boldsymbol{x} + \boldsymbol{\varepsilon} \tag{2.11}$$

où $\boldsymbol{\varepsilon} \in \mathbb{R}^M$ est introduit pour tenir compte des erreurs de modélisation (tomographie $\leftrightarrow$ TR), de discrétisation de la scène, et des bruits d'instrumentation. Par la suite, cette formulation permet de ramener le problème de reconstruction tomographique à celui de l'estimation de $\boldsymbol{x}$ à partir des mesures $\boldsymbol{y}$; en ce sens, $\boldsymbol{x}$ sera appelé l' « image » tomographique. L'équation d'observation (2.11) conduit principalement à formuler le problème de reconstruction suivant deux approches : l'inversion au *sens des moindres carrés* (abrég. MC), et la recherche d'une *solution réalisable* pour le système $\boldsymbol{y} = \boldsymbol{R}\boldsymbol{x}$. Indépendamment de la méthodologie, tous les algorithmes de reconstruction basés sur l'inversion de (2.11) ont souvent été rassemblés sous le terme générique d'*approche algébrique.*

Solutions réalisables du système $\boldsymbol{y} = \boldsymbol{R}\boldsymbol{x}$

Une première manière d'aborder la reconstruction consiste à s'intéresser à l'équation d'observation (2.11) en négligeant le bruit et à considérer le système linéaire

$$\boldsymbol{y} - \boldsymbol{R}\boldsymbol{x} = \boldsymbol{0}, \tag{2.12}$$

qui peut être sur-déterminé ($M > N$) ou sous-déterminé ($M < N$), et pour lequel on définit l'ensemble convexe des *solutions réalisables*

$$S = \{\boldsymbol{x} \mid \boldsymbol{y} - \boldsymbol{R}\boldsymbol{x} = \boldsymbol{0}\}.$$

Cette démarche a conduit un nombre significatif d'auteurs à étudier et à mettre en œuvre des méthodes numériques *itératives* de type POCS (*Projection onto convexe sets*) pour obtenir une solution numérique de (2.12). On souligne que cette démarche n'est pas exempte d'obstacles méthodologiques qui poussent à certaines « acrobaties » algorithmiques non négligeables. En effet, les conditions expérimentales conduisent invariablement[6] à $\boldsymbol{y} \notin \mathrm{Im}(\boldsymbol{R})$, et le système (2.12) est toujours *inconsistant* en pratique (*i.e.* $S = \emptyset$). On notera en revanche que l'emploi de ces algorithmes POCS sur une l'équation normale découlant d'une formulation au sens des moindres carrés lève ces difficultés.

Inversion au sens des moindres carrés et inverse généralisée

Une approche naturelle est de résoudre le problème au sens des moindres carrés, c.à.d. en cherchant un des éléments de l'ensemble S_{MC} défini par

$$S_{\mathrm{MC}} = \left\{ \boldsymbol{x} \in \mathbb{R}^N : \min ||\boldsymbol{y} - \boldsymbol{R}\boldsymbol{x}||^2 \right\}.$$

Pour les problèmes en dimensions finies, S_{MC} n'est jamais vide, il est fermé et convexe [Rockafellar, 1970, p. 263] ; de plus on montre facilement que les éléments

[6] En particulier, pour un modèle d'image utilisant une base $\{b_n(\boldsymbol{\xi})\}_{n=1}^N$ d'indicatrices sur le support des pixels, $\mathrm{Im}(\boldsymbol{R})$ est uniquement constituée des fonctions constantes par morceaux.

de S_{MC} correspondent aux solutions de l'équation normale

$$\boldsymbol{R}^T\boldsymbol{y} - \boldsymbol{R}^T\boldsymbol{R}\boldsymbol{x} = \boldsymbol{0}. \tag{2.13}$$

Si $\text{Ker}(\boldsymbol{R})$ n'est pas trivial, $\boldsymbol{R}^T\boldsymbol{R} \in \mathbb{R}^{N\times N}$ n'est pas de rang plein et on n'a plus unicité de la solution de (2.13) ; le problème est de nouveau mal posé. Néanmoins, S_{MC} étant fermé et convexe, on montre[7] l'existence et l'unicité d'une solution dans S_{MC} de *norme minimale* $\boldsymbol{x}^\dagger$ définit par

$$\boldsymbol{x}^\dagger = \arg\min_{\boldsymbol{x}} ||\boldsymbol{x}||^2 \qquad \text{s.c.} \qquad \boldsymbol{x} \in S_{\text{MC}}.$$

Cette solution étant de norme minimale, elle appartient à $\text{Ker}(\boldsymbol{R})^\perp$, et s'écrit simplement en utilisant la *décomposition en valeurs singulières* (abrég. SVD) de $\boldsymbol{R}$

$$\boldsymbol{x}^\dagger = \left(\sum_{j=1}^{r} \frac{\boldsymbol{v}_j\boldsymbol{u}_j^T}{\sigma_j}\right)\boldsymbol{y} = \boldsymbol{R}^\dagger\boldsymbol{y}, \tag{2.14}$$

avec $\boldsymbol{R}^\dagger$ l'*inverse généralisée* de $\boldsymbol{R}$. Dans (2.14), r est le *rang* de $\boldsymbol{R}$, et $\{\sigma_j, \boldsymbol{u}_j, \boldsymbol{v}_j\}_{j=1}^r$ est le système singulier de l'opérateur $\boldsymbol{R}$; voir [Golub et Van Loan, 1996, Sec 5.5] pour plus de détails. En pratique, il semble qu'aucune approche cherchant explicitement à calculer $\boldsymbol{x}^\dagger$ n'ait été retenue pour être implantée dans les tomographes médicaux. Cette présentation de l'inverse généralisée $\boldsymbol{x}^\dagger$ reste pourtant fondamen-

[7] Ce résultat découle de la stricte convexité de la norme euclidienne et de la convexité de S_{MC}.

tale pour analyser le comportement des méthodes algébriques standard que nous décrivons dans le paragraphe suivant. Enfin, on notera que les résultats exposé ci-dessus s'étendent sans aucune difficultée aux solutions de l'ensemble des *moindres carrés pondérés* défini par

$$S_{\mathrm{w}} = \left\{ \boldsymbol{x} \in \mathbb{R}^N : \min ||\boldsymbol{W}^{1/2}(\boldsymbol{y} - \boldsymbol{R}\boldsymbol{x})||^2 \right\} \tag{2.15}$$

où $\boldsymbol{W}^{1/2}$ est la *racine carrée* d'une matrice $\boldsymbol{W}$ définie non négative — c.a.d. la seule matrice de $\mathbb{R}^{M\times M}$ telle que $\boldsymbol{W} = \boldsymbol{W}^{1/2}\boldsymbol{W}^{1/2}$; cf. [Golub et Van Loan, 1996, Sec. 4.2.10].

Schéma itératif de type POCS en tomographie médicale

Parmi les deux formulations que nous venons de présenter, c'est l'approche de type POCS sur le système inconsistant (2.12) qui a été retenue[8] pour reconstruire des images médicales dans ce cadre « algébrique ». Les algorithmes employés sont en majorité des dérivés de l'algorithme de KACZMARZ ; on citera par exemple [Censor *et al.*, 1983], ou [Censor *et al.*, 2001; Herman et Meyer, 1993] pour les publications les plus récentes. On présente maintenant les plus connus de ces algorithmes

[8]Bien que ce choix semble avant tout historique, il faut remarquer que de travailler sur l'équation normale nécessite de former l'opérateur $\boldsymbol{R}^T\boldsymbol{R}$ qui n'est généralement plus creux et de grande taille.

appliqués au système d'équations linéaires, éventuellement inconsistantes :

$$\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} = 0, \qquad \boldsymbol{A} \in \mathbb{R}^{M \times N}.$$

Approches itératives

(i) BICAV : Récemment apparu dans la littérature [Censor *et al.*, 2001], l'algorithme BICAV (*Block-Iterative Component Averaging*) est un des nombreux algorithmes de type POCS développé pour résoudre numériquement le système inconsistant (2.12) en tomographie. Nous présentons d'abord cet algorithme puisque de nombreux algorithmes « algébriques » largement reconnus se déduisent de sa formulation : on se donne J ensembles ordonnés d'indices

$$1 \leq j \leq J, \qquad B_j \subseteq \{1, \cdots, M\}$$

de cardinal $\#(B_j) = M_j$, ces ensembles étant choisis de manière à ce que chaque indice $\{1, \cdots, M\}$ apparaisse au moins une fois dans $B = \bigcup_j B_j$. On pose

$$\boldsymbol{A}_j^T = \left(\boldsymbol{a}_j^1 | \cdots | \boldsymbol{a}_j^{M_j}\right)$$

la matrice $N \times M_j$ constituée par les colonnes de $\boldsymbol{A}^T$ dont les indices sont B_j ; si $\boldsymbol{x}^{(k)}$ désigne la k-ième itération d'un algorithme initialisé par $\boldsymbol{x}^{(0)}$, l'algorithme procède

alors à la mise à jour *complète* des inconnues en utilisant de manière cyclique les J « blocs » de données [Censor *et al.*, 2001, Sec. III] :

$$\begin{aligned} \forall k \in \mathbb{N}, \qquad \boldsymbol{x}^{(k;0)} &= \boldsymbol{x}^{(k-1)}, \\ 1 \leq j \leq J, \quad \boldsymbol{x}^{(k;j)} &= \boldsymbol{x}^{(k;j-1)} + \alpha^{(k)} \boldsymbol{A}_j^T \boldsymbol{W}_j^{-1} \left(\boldsymbol{y}_j - \boldsymbol{A}_j \boldsymbol{x}^{(k;j-1)} \right); \end{aligned} \tag{2.16}$$

l'*itération complète* $k \longrightarrow k+1$ étant obtenue après les J dernières mises à jour :

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k;M)}.$$

Dans (2.16), $\alpha^{(k)} > 0$ est le *paramètre de relaxation*, $\boldsymbol{y}_j \in \mathbb{R}^{M_j}$ est le vecteur colonne composé des éléments de $\boldsymbol{y}$ indicés par B_j, et $\boldsymbol{W}_j \in \mathbb{R}^{M_j \times M_j}$ est une matrice *diagonale* positive telle que

$$\boldsymbol{W}_j = \operatorname{diag}\left(||\boldsymbol{S}^{1/2} \boldsymbol{a}_m^1||^2, \cdots, ||\boldsymbol{S}^{1/2} \boldsymbol{a}_m^{M_j}||^2 \right);$$

$\boldsymbol{S}^{1/2}$ est la racine carrée de matrice *diagonale* $\boldsymbol{S}$ dont l'entrée s_n, $1 \leq n \leq N$, représente *le nombre d'éléments non nuls* dans la ligne n de $\boldsymbol{A}_j^T$.

(ii) ART : L'algorithme ART (*Algebraic Reconstruction Technic*) a été le premier algorithme de reconstruction tomographique mis en œuvre. Dans sa formulation initiale introduite par Gordon *et al.* [1970], cet algorithme se déduit de la formulation BICAV en posant $\boldsymbol{S} = \boldsymbol{I}$ et en considérant autant de blocs que de mesures

(*i.e.* $J = M$). Dans ce cas, $B_j = \{j\}$ et on écrit pour $1 \leq m \leq M$ [Herman, 1980, Chap. 11] :

$$\boldsymbol{x}^{(k;m)} = \boldsymbol{x}^{(k;m-1)} + \alpha^{(k)} \frac{\boldsymbol{a}_m}{||\boldsymbol{a}_m||^2} \big(y_m - \langle \boldsymbol{a}_m, \boldsymbol{x}^{(k;m)} \rangle\big); \tag{2.17}$$

où $\boldsymbol{a}_m$ est la m-ième colonne de $\boldsymbol{A}^T$.

(iii) CAV : Quand BICAV n'utilise qu'un seul bloc ($J = 1$), on obtient l'algorithme CAV (*Component Averaging*) tel que défini par [Censor *et al.*, 2001, rel. (2.15)]. Dans ce cas, la remise à jour des inconnues devient *simultanée* et (2.16) s'écrit

$$\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(k-1)} + \alpha^{(k)} \boldsymbol{A}^T \boldsymbol{W}^{-1} \big(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{(k-1)}\big). \tag{2.18}$$

Il est aisé de vérifier que cette itération correspond à l'*algorithme du gradient* à pas $\alpha^{(k)}$ appliqué au critère $J(\boldsymbol{x}) = \frac{1}{2}||\boldsymbol{W}^{-1/2}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})||^2$, où $\boldsymbol{W}^{-1/2}$ est la racine carrée de la matrice (diagonale) $\boldsymbol{W}^{-1}$; cet algorithme calcule donc une solution de type *moindre carrés pondérés.*

(iv) SIRT : Repartons de l'algorithme CAV et posons $\boldsymbol{D} = \text{diag}\,\{a_1, \cdots, a_N\}$, avec a_n l'*aire de l'ensemble des pixels impliqués dans tous les rayons contenant le pixel* n. La mise à jour (2.18) devient alors très proche d'une autre technique itérative utilisée en reconstruction tomographique : l'algorithme SIRT introduit par [Gilbert, 1972]. Les itérées générées par cet algorithme sont donc celles d'un algorithme du

gradient[9] appliqué à un critère des moindres carrés pondérés.

(v) SART : Si on pose $J = P$ et qu'on associe à B_j les indices de la j-ième des P projections, on retrouve les bases de l'algorithme SART (*Simultaneous Algebraic Reconstruction Technics*) introduit par [Andersen et Kak, 1984].

Convergence pour $\boldsymbol{Ax} = \boldsymbol{y}$ consistant

Si le système est *consistant*, la convergence vers une solution du système est garantie pour BICAV et ART si $\forall k,\ 0 < \epsilon_1 < \alpha^{(k)} < \epsilon_2 < 2$, avec ϵ_1, ϵ_2 deux constantes indépendantes de k ; voir respectivement [Censor et Elfving, 2001, Th. 7.1] et [Herman, 1980, Sec. 16.8]. Pour CAV et SIRT, les conditions sont celles des algorithmes du gradient à pas fixe sur un critère des moindres carrés, cf. [Demoment et Idier, 2001*b*, p. 43], et on montre finalement que $\epsilon_1 < \alpha^{(\cdot)} < \epsilon_2$ garantit la convergence. Pour ces quatre algorithmes, on garantit la convergence vers la solution de *norme minimale* $\boldsymbol{x}^\dagger$ si l'initialisation est telle que $\boldsymbol{x}^{(0)} \in \mathrm{Ker}(\boldsymbol{A})^\perp$; condition vérifiée par les images uniformes souvent utilisées en pratique comme point initial.

Pour BICAV et ART, la vitesse de convergence dépend de manière *critique* de l'ordre dans lesquels les projections sur les sous-espaces sont effectuées, c.à.d. de l'ordre d'utilisation des blocs $\boldsymbol{A}_j^T$ lors des J mises à jours ; ainsi, la convergence de l'ART est d'autant plus accélérée que les mises à jours sont orthogonales les

[9] Une étude menée par Lakshminarayanan et Lent [1979] montre en effet la similitude du comportement de l'algorithme SIRT avec un algorithme de type *Richardson* qui n'est autre qu'un simple algorithme du gradient sur un critère des moindres carrés.

unes aux autres ; voir par exemple [Herman et Meyer, 1993, Sec. III] et [Guan et Gordon, 1994]. Pour ces algorithmes, il importe également de noter que le facteur de relaxation $\alpha^{(\cdot)}$ a un impact majeur sur la reconstruction : pour α petit (ex. $\alpha = 0,05$), les composantes basses fréquences apparaissent en premier dans la reconstruction, l'inverse étant vrai si α est grand [Natterer, 1999, Sec. 4].

Convergence pour $\boldsymbol{Ax} = \boldsymbol{y}$ inconsistant

Le cas réel *inconsistant* est méthodologiquement et pratiquement plus délicat. On constate par exemple que l'ART a un comportement asymptotique cyclique *autour de la solution de norme minimale* $\boldsymbol{x}^\dagger$ [Censor *et al.*, 2001, Sec. I]. En pratique, on contourne ce problème en se limitant à des valeurs $\alpha^{(\cdot)} \ll 2$ (ex. $\alpha = 0,05$) de manière à réduire l'amplitude du cycle. Les propriétés de convergence pour les algorithmes BICAV restent encore peu étudiées dans le cas inconsistant (cf. section 6 de la référence précédente). On souligne enfin que les algorithmes SIRT et CAV ne présentent pas ce comportement asymptotique cyclique[10].

Aspects de mise en œuvre

La mise en œuvre d'un algorithme ART s'appuie sur une *forte sous-relaxation* [Herman, 1980, p. 196] et sur un *arrêt prématuré*[11] de l'algorithme pour limiter

[10]Les algorithmes CAV et SIRT étant de type gradient à pas fixe, la convergence a lieu vers $\boldsymbol{x}^\dagger$ si $\boldsymbol{x}^{(0)} \in \mathrm{Ker}(\boldsymbol{A})^\perp$.

[11]Cet aspect est peu mis en avant dans la littérature ; nous renvoyons à [Censor *et al.*, 2001, Sec. V.C] qui établit clairement ce fait.

une dégradation de l'image. Cette procédure se transpose aux algorithmes BICAV à la nuance près que le nombre de blocs influe sur la valeur de α « optimale » ; voir aussi à ce sujet [Eggermont et Herman, 1981, p.49]. Comme le montre la figure 2.3, la convergence des algorithmes BICAV et ART est typiquement stoppée après quelques itérations[12]. Ceci contraste avec les mises en œuvre des algorithmes (du gradient) CAV ou SIRT qui présentent une convergence d'autant plus lente que le problème est mal conditionné ; voir par exemple [Bertsekas, 1995, p.67].

En terme de *coût d'implantation*, on notera que la matrice $\boldsymbol{A}$ est *très creuse* pour la tomographie (généralement moins de 1% d'éléments non nuls) ce qui permet d'utiliser de algorithmes adaptés pour les produits matriciels et vectoriel. En pratique, il semble que le coût de calcul d'une itération complète soit comparable à celui d'une RPC [Herman, 1980, p. 204] ; cependant ce propos est très peu abordé dans la littérature.

Épilogue

Si l'on s'interroge sur la qualité des reconstructions produites par ces algorithmes, un premier élément de réponse est apporté par l'étude de *robustesse* de $\boldsymbol{x}^{\dagger}$, solution vers laquelle on cherche à faire converger ces méthodes itératives. Il est bien établi que la stabilité de cette solution dépend du *nombre de condition* de $\boldsymbol{R}$

[12] On souligne que l'erreur tracée sur cette courbe est une distance calculée directement sur le fantôme et *n'est pas* le résidu $||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{(k)}||$ qui décroît lui de manière monotone ; cf. [Censor *et al.*, 2001, Sec. IV]

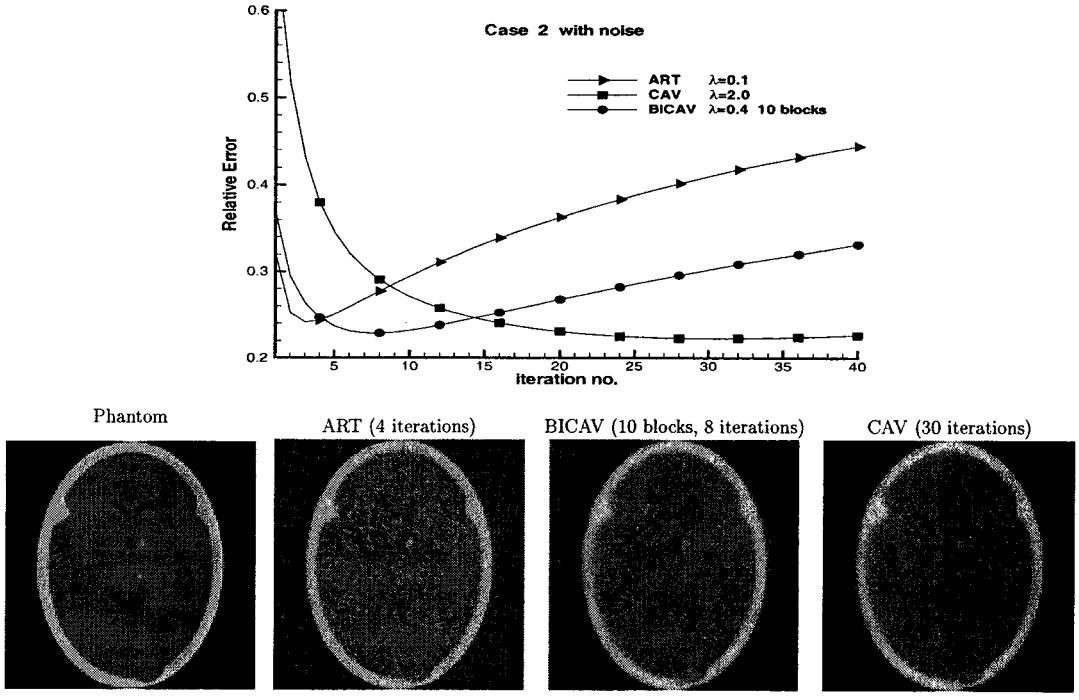


Figure 2.3: Mise en œuvre extraite de [Censor *et al.*, 2001] des algorithmes ART, BICAV et CAV sur données bruitées ; (haut) : évolution de l'erreur relative mesurée sur la reconstruction du fantome ; bas : fantôme synthétique utilisé, et meilleurs reconstructions obtenues pour les trois algorithmes.

qu'on définit par

$$\text{Cond}\ (\boldsymbol{R}) = ||\boldsymbol{R}||\,||\boldsymbol{R}^{\dagger}|| \tag{2.19}$$

où $||\boldsymbol{R}||$ est une norme matricielle induite par une norme, par exemple par la norme euclidienne. Dans ce cas, on a

$$\text{Cond}\ (\boldsymbol{R}) = \sigma_{\max}/\sigma_{\min} \geq 1,$$

avec $\sigma_{\min}$ et $\sigma_{\max}$, respectivement, la plus petite et la plus grande des valeurs singulières de $\boldsymbol{R}$. Pour la reconstruction tomographique, le caractère mal posé du problème implique que Cond $(\boldsymbol{R}) \gg 1$, le problème inverse numérique est *mal conditionné* et la robustesse de $\boldsymbol{x}^\dagger$ est alors *très mauvaise.* Or la qualité des images que fournissent ces algorithmes est en pratique à peu près du niveau de celles obtenues par la RPC, ce qui paraît étonnant à première vue puisque :

(*i*) la sous-relaxation permet une convergence cyclique proche de la solution de norme minimale $\boldsymbol{x}^\dagger$ du système (inconsistant),

(*ii*) et la solution $\boldsymbol{x}^\dagger$ n'étant pas régularisée, son comportement est typiquement de très mauvaise qualité et présente une amplification excessive des composantes hautes fréquence, donc du bruit.

La situation devient en fait claire si on souligne que, moyennant une forte sous relaxation, les composantes basses fréquences de l'image apparaissent prématurément au cours de la convergence. *Arrêter l'algorithme dès les premières itérations produit donc une solution lissée, ce qui constitue une forme de régularisation*[13]. Finalement, compte tenu de leur coût d'implantation plus élevé, ces algorithmes ont été unanimement remplacés par la RPC dans les tomographes commerciaux.

[13]Cette méthode est d'ailleurs connue comme régularisante pour un problème d'inversion numérique mal conditionné ; cf. [Demoment et Idier, 2001*b*, Sec 2.1.1]

2.3 En résumé

Garantir l'unicité et une certaine stabilité en reconstruction d'image tomographique passe par la *régularisation* du problème d'inversion. En pratique, si l'excursion angulaire est suffisante et les sources de bruit faibles, un simple filtrage des composantes HF suffit pour aboutir à des solutions pertinentes pour le clinicien. C'est finalement ce que font la RPC et les approches algébriques « Standard » : la première de manière explicite *via* le filtre de reconstruction, les secondes plutôt implicitement en arrêtant très tôt le cours des itérations.

Comme on l'a déjà précisé, la reconstruction précise des interfaces entre tissus passe par une reconstruction des composantes HF de cette image. Ceci est impossible avec ces approches standard et il est nécessaire de développer des alternatives qui permettent une restriction plus pertinente de l'espace des solutions. Sur ce point, le formalisme « algébrique » offre un net avantage en comparaison des approches « analytiques » : il permet d'introduire aisément des contraintes supplémentaires dans la formulation du problème.

Bibliographie annotée

Le qualificatif d' « approche algébrique » en tomographie renvoie systématiquement à la *modélisation paramétrique* de la scène, et en général, à une procédure de reconstruction itérative cherchant à *projeter* sur des ensembles (type POCS) ou

à *minimiser* un critère particulier. Le nombre d'algorithmes itératifs possibles est devenu si conséquent qu'en faire un historique un tant soit peu précis est très délicat. Dans cet exposé, on s'est contenté de présenter les algorithmes *non pénalisés* qui ont eu un impact important dans la communauté. Certain d'entre-eux seront d'ailleurs présenté dans une forme pénalisée dans les chapitres suivant.

Comme complément d'information, un bon point d'entrée consiste à consulter [Fiani, 2001, IV.3]. L'exposé de l'ART et de certaines de ses variantes donné par [Herman, 1980, Chap. 11] est une référence que nous considérons maintenant plutôt comme « historique » que comme réellement incontournable (la clarté laisse parfois à désirer et la présentation laborieuse d'une extension pénalisée de l'ART est à éviter). La présentation faite par [Kak et Slaney, 1987, Chap. 7] n'offre pas grand chose de plus et, en général, les monographies traitant spécifiquement de la reconstruction tomographique offrent une présentation lacunaire (voire inexistante) des approches algébriques.

La meilleure démarche consiste certainement à consulter les contributions originales que nous avons utilisées pour ce document. On notera en particulier que [Eggermont et Herman, 1981] est un bon article de synthèse sur les propriétés d'algorithmes de type « Bloc Kaczmarz » qui englobent les versions BICAV (matrice $\boldsymbol{D}_j$ non diagonale en général).

CHAPITRE 3

TOMOGRAPHE HÉLICOÏDAL

On cherche dans ce chapitre à justifier le développement d'une approche alternative de la reconstruction en tomographie hélicoïdale. On tente en particulier de dégager les causes des dégradations parfois sévères (et largement rapportées) que subissent les reconstructions en passant du mode axial au mode hélicoïdal. Cette analyse nous amènera à la conclusion que ces dégradations sont causées, au moins en partie, par l'algorithme de reconstruction.

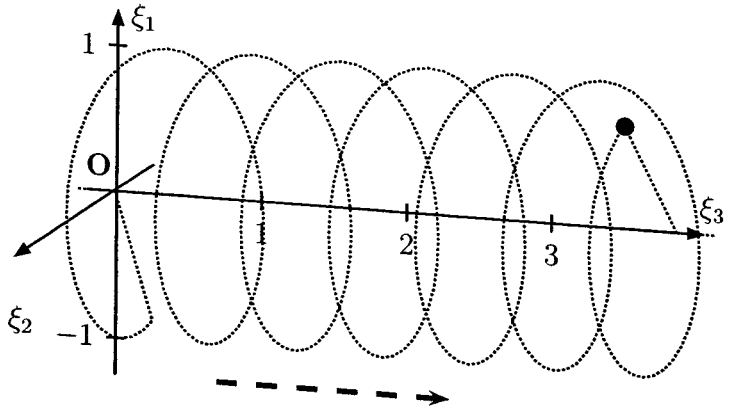


Figure 3.1: Trajectoire de la source en mode hélicoïdal.

3.1 Tomographe hélicoïdal simple coupe

Au cours des dix dernières années, la tomographie en mode axiale a progressivement été remplacée par une tomographie en mode hélicoïdale. Ce mode d'ac-

quisition se distingue principalement du mode axial par une translation à vitesse constante de la table pendant que le volume d'intérêt est imagé. Ce faisant, la trajectoire de la source autour du patient décrit une hélice plutôt que des cercles équidistants ; cf. figure 3.1. Introduisons pour la suite l'angle $\bar{\varphi} = w.t$ définit sur $\mathbb{R}$ où t est le temps et w la *vitesse angulaire de la source* en radian par seconde. On notera qu'une fois le *pas de l'hélice* P donné, l'angle $\bar{\varphi}$ définit de manière unique la position de la source sur l'axe $O\xi_3$

$$\xi_3 = \frac{\bar{\varphi}}{2\pi} P.$$

Enfin, l'angle de projection dans le plan de l'image $\xi_1 O \xi_2$ est relié à la variable angulaire $\bar{\varphi}$ simplement par

$$\varphi = \mathrm{mod}_{2\pi}(\bar{\varphi})$$

où $\mathrm{mod}_{2\pi}(\cdot)$ définit l'opération du modulo à 2π. Ces informations sont illustrées sur la figure 3.2.

Afin de favoriser la clarté de l'exposé, on considérera que l'échantillonnage dans le plan de l'image est à faisceau parallèle ; notons néanmoins que la littérature expose ses travaux sur les algorithmes de reconstruction pour une géométrie en éventail (tomographes de 3ème ou 4ème génération).

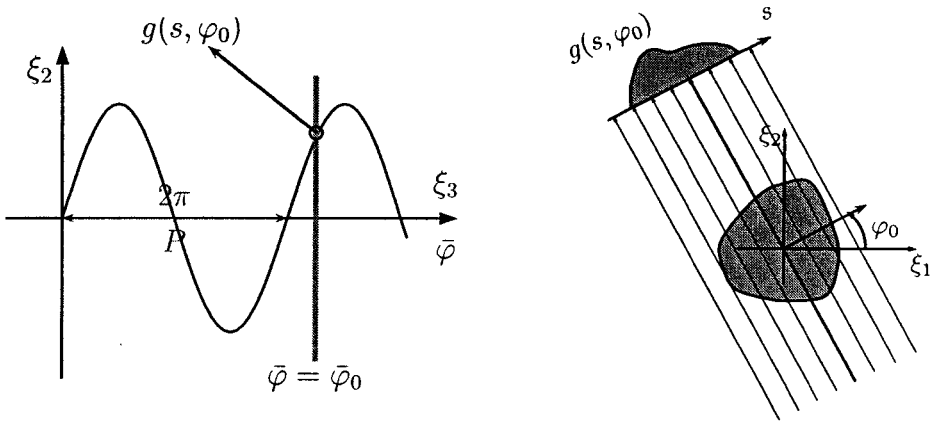


Figure 3.2: Trajectoire de la source en mode hélicoïdal suivant l'axe du tomographe ($\leftarrow$), et dans le plan de l'image $\xi_1 O \xi_2$ associé à l'angle $\bar{\varphi}_0$ sur l'hélice($\rightarrow$). Pour une position angulaire $\bar{\varphi} = \bar{\varphi}_0$ selon l'axe du tomographe, l'angle de projection dans le plan de l'image est $\varphi_0 = \mathrm{mod}_{2\pi}(\bar{\varphi}_0)$.

Intérêt d'un échantillonnage hélicoïdal

C'est dans la deuxième moitié des années 1980 qu'apparaît l'idée d'une tomographie sans arrêter la table sur laquelle le patient est installé. Outre l'élimination des étapes répétées d'accélérations et de décélérations propres au mode « stop and go », l'acquisition en continu permet principalement de réduire les temps d'acquisition souvent trop longs en tomographie axiale.

En effet, certaines parties du corps ou certains organes ne peuvent être imagés dans un état stationnaire : c'est évidemment le cas du cœur, mais c'est aussi le cas de la cage thoracique qui nécessiterait une apnée irréaliste pour être imagée correctement. Il y a donc un réel besoin de diminuer le temps d'acquisition des tomographes si on veut éviter l'apparition d'artefacts sur les reconstructions[1], et le

[1]Les artefacts sont des différences systématiques entre l'image et la réalité.

mode hélicoïdal est un moyen d'y parvenir.

La diminution du temps d'acquisition permet, en parallèle, une diminution de la dose de rayonnement ionisant (diminution du temps d'exposition) nécessaire pour imager un volume donné. Pour un temps d'exposition donné, l'échantillonnage hélicoïdal permet par ailleurs une distribution plus uniforme de la dose de rayonnement administrée puisque le patient est translaté en continu [Wang et Vannier, 1993, p. 1635]. Ces éléments sont souvent mis en avant car ils vont dans le sens d'un examen moins invasif.

Dégradation de la qualité des reconstructions

La diminution du temps d'acquisition ne se fait pas sans contreparties en termes de qualité de reconstruction. Dans un premier temps, la diminution de la dose administrée pour imager un volume va nécessairement de paire avec une diminution du rapport signal à bruit (abrég. RSB) octroyé au volume. L'autre source de dégradation provient plus de la démarche adoptée pour reconstruire les images. En effet, contrairement au mode axial, la reconstruction d'un volume échantillonné sur l'hélice *n'est plus naturellement séparable* en une suite de reconstruction 2D. L'introduction d'une étape *ad hoc* rend la reconstruction 3D de nouveau « séparable ». Cette technique permet d'utiliser la traditionnelle RPC pour reconstruire des plans successifs, mais les distortions introduites peuvent être significatives.

3.1.1 Technique de reconstruction standard en mode hélicoïdal

Les premiers articles traitant quantitativement des méthodes d'acquisition et de reconstruction paraissent au début des années 1990 [Crawford et King, 1990; Kalender *et al.*, 1990]. À cette époque, l'objectif des auteurs est de produire des coupes par RPC en partant des données échantillonnées sur l'hélice, et le mouvement continu de la table est plutôt perçu comme une version perturbée d'un problème stationnaire axial. Formuler le problème de cette manière conduit immédiatement à l'obstacle suivant : l'échantillonnage hélicoïdal empêche une formulation directe du problème de reconstruction à base de plans bien définis. Les données sont alors qualifiées d'*inconsistantes* pour n'importe quelle coupe dans le volume [Crawford et King, 1990, Sec. II].

Création d'un jeu de données consistantes

Une fois les données hélicoïdales en notre possession, on se fixe une série de plans axiaux qui définissent autant de plans de reconstruction. La démarche adoptée consiste alors à créer, pour chacun de ces plans, un nouveau jeu de projection par *interpolation* des données échantillonnées sur l'hélice. La reconstruction 3D est alors ramenée à une succession de problèmes axiaux qu'on résout par RPC. *En terme de méthodologie, cette approche de la reconstruction hélicoïdale repose donc principalement sur l'emploi d'une fonction d'interpolation et de la RPC.*

Jusqu'à très récemment, la fonction interpolatrice employée de manière systématique était la fonction linéaire. Il existe deux méthodes pour mettre en œuvre cette interpolation linéaire :

- dans l'interpolation linéaire *full-scan* (LI-360°), chaque projection du plan de reconstruction est une somme pondérée de deux vues distantes de 2π, l'une située en amont du plan de reconstruction, l'autre située en aval ; cf. figure 3.3-gauche.
- l'interpolation de type *half-scan* (LI-180°) tire parti du fait qu'en mode axial, les projections d'angles opposés (de différence angulaire égale à π) sont censées être identiques[2]. Dès lors, il est possible d'utiliser les données situées à une distance correspondant à la moitié du pas de l'hélice comme l'illustre la figure 3.3-droite.

Artefacts propres au mode hélicoïdal

Le choix de la technique d'interpolation s'est longtemps résumé au cadre linéaire restreint décrit ci-dessus. Depuis 1998, on a commencé à modifier l'interpolatrice de manière à produire un filtrage sur le volume reconstruit et lisser les artefacts, cf. [Hu et Shen, 1998]. En pratique, le pas de l'hélice joue évidemment un rôle majeur dans la « consistances » finale des jeux de projections axiales reconstitués,

[2]Propriété propre à la géométrie parallèle, qui peut néanmoins être étendue aux autres géométries moyennant quelques adaptations [Crawford et King, 1990].

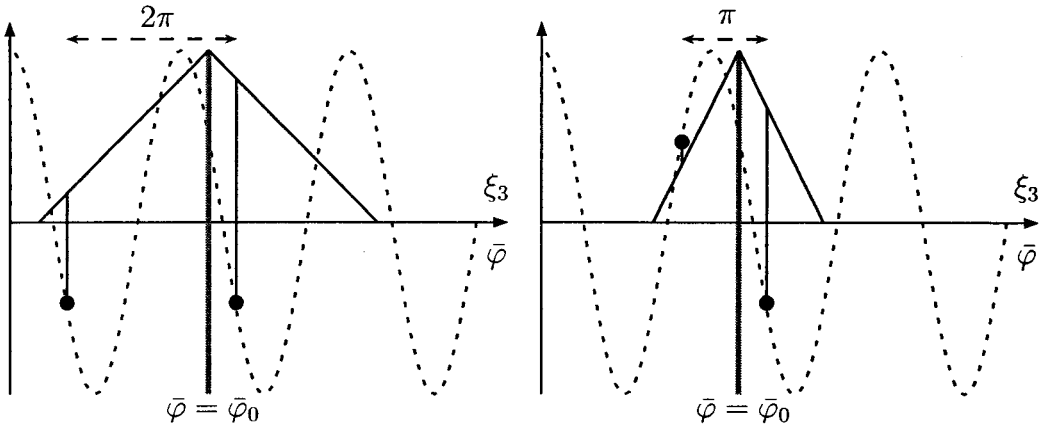


Figure 3.3: Procédés usuels d'interpolation adopté pour recréer un jeu « axial consistant » dans le plan défini par $\bar{\varphi} = \bar{\varphi}_0$. Chaque *pseudo* projection axiale d'angle φ découle de deux projections obtenues dans les plans $\bar{\varphi}_0 + \Delta\varphi$ et $\bar{\varphi}_0 - 2\pi + \Delta\varphi$ pour le mode « full-scan » ($\leftarrow$), ou $\bar{\varphi}_0 + \Delta\varphi$ et $\bar{\varphi}_0 - \pi + \Delta_\varphi$ pour le mode « half-scan » ($\rightarrow$) ; $\Delta_\varphi = \varphi - \mathrm{mod}_{2\pi}(\bar{\varphi}_0)$.

et l'interpolation ne peut évidemment pas recréer l'information perdue par un pas d'hélice trop important. Il se produit alors des déformations au niveau des images (2D et 3D) qui sont propres à ce mode hélicoïdal. Ces artefacts peuvent prendre une forme très accentués et être facilement reconnaissables ; ils peuvent aussi induire des déformations plus subtiles dans l'image et leurrer le diagnostic ; les publications [Hu et Shen, 1998] et [Ogata, 1999] illustrent respectivement ces deux types de comportement.

La situation vue du praticien...

Le praticien se retrouve bien souvent à régler des paramètres pour lesquels il n'a pas de critère explicite de choix. Ces paramètres sont de deux natures :

(i) ceux ayant trait à l'algorithme de reconstruction : quelle fonction interpola-

trice utiliser (linéaire ou autre) et quel type d'interpolation (LI-180°, LI-360°, mixte, *etc*) ;

(ii) ceux qui définissent le contexte de la mesure : le réglage du pas de l'hélice et de l'épaisseur du faisceau...

Des différences mineures d'ajustement peuvent produire des différences significatives sur les reconstructions, et le réglage du tomographe reste une tâche délicate si on en croit la pléthore de travaux cliniques étudiant la « meilleure » configuration *en fonction* de la partie du corps humain à imager ; voir par exemple [Parodi *et al.*, 1997] pour le diagnostic des lésions du ménisque ou [Ogata, 1999] pour le diagnostic des sténoses artérielles.

3.1.2 Analyse des dégradations

Le paragraphe précédent montre que les techniques de reconstruction en mode hélicoïdal sont fortement heuristiques dans leurs approches. À notre connaissance, la seule contribution posant correctement le problème de reconstruction dans son cadre 3D reste un rapport technique[3] de F. NATTERER [Natterer, 1994]. Une étude bibliographique menée sur les dix dernières années montre que cette contribution est largement passée inaperçue, et il a fallu attendre des travaux récents pour que

[3]Dans cette courte note, l'auteur décrit un algorithme permettant la reconstruction d'une fonction sans aliasing de support fréquentiel donné pour un coût équivalent à une RPC. Ce rapport n'a malheureusement pas été publié et son accès demeure difficile sans une lecture détaillée de certain de ces travaux antérieurs – notamment de [Natterer, 1993].

le phénomène de repliement spectral soit abordé de manière appropriéé ; voir [Yen *et al.*, 1999] ainsi que [Wang et Vannier, 1999].

D'autre part, l'analyse des difficultés rencontrées par les approches standard nécessite, à notre sens, de s'interroger sur le recours systématique à la RPC comme outil de reconstruction hélicoïdale. En effet, cet algorithme souffre de deux lacunes maintenant bien documentées,

(*i*) tout d'abord, cet algorithme coupe les hautes fréquences dans les reconstructions et s'avère donc peut adapté à la reconstruction de régions homogènes séparées par des interfaces franches ;

(*ii*) d'autre part, *cet algorithme est généralement très peu robuste dans les problèmes tomographiques à angles de vues limités*[4] ; cf. [Delaney et Bresler, 1998; Jaffe, 1990; Peng et Stark, 1989].

Si le premier point a déjà été largement abordé au chapitre 2, le second mérite quelques commentaires. Indiquons tout d'abord que le problème à angles de vue limités se caractérise par une instabilité accrue comparativement au problème « complet »— ex. [Faridani, 1999, Sec. 5]. Or, il paraît sensé de considérer que l'échantillonnage hélicoïdal du volume associe pour chaque plan de reconstruction un faible nombre de projections obtenus dans un faible cône angulaire. Reconstruire une image par RPC à partir de ces jeux incomplets conduirait évidemment à des

[4] *On considère qu'un problème est à « angle de vue limité » quand l'échantillonnage angulaire ne couvre pas l'intervalle* $[0;\pi[$ *nécessaire pour obtenir un jeu de projection complet.*

résultats inexploitables, et en supposant implicitement une certaine continuité *longitudinale* dans le volume, l'interpolation « complète » ces ensembles incomplets de projection. Il est tout à fait remarquable que l'interpolation agit comme une connaissance *a priori* de continuité sur l'objet. Comme le montrerons les simulations du chapitre 7, cette démarche a tendance à faire apparaître de la continuité au niveau des interfaces, et ainsi contribue à dégrader la résolution de la méthode.

Finalement, nous croyons que l'emploi de la RPC en tomographie hélicoïdale introduit *de facto* un *a priori* fort, mais pas toujours pertinent, de manière à stabiliser la solution. Nous proposons alors de développer une méthode intrinsèquement plus robuste permettant l'introduction d'une connaissance *a priori* plus adaptée au problème tomographique ; c'est la démarche que nous développons dans la suite de ce document.

3.2 Le cas du tomographe hélicoïdal multicoupes

Une nouvelle génération de tomographes hélicoïdaux émerge depuis deux ans : la génération des tomographes multicoupes. Le procédé diffère matériellement du tomographe hélicoïdal usuel par l'introduction de plusieurs rangées de détecteurs juxtaposées (voir figure 3.4). Si le pas de l'hélice est correctement ajusté, l'introduction de plusieurs « barrettes » de détecteurs permet un échantillonnage sensiblement plus important suivant l'axe du tomographe. Les artefacts de reconstruction

dûs à l'avancement de la table sont alors moins prononcés. Notons cependant que le problème de reconstruction est formellement plus compliqué à résoudre. En effet, la géométrie exacte de projection produite par ces tomographe se rapproche d'une projection de type conique (connue sous l'appellation anglosaxonne « cone beam »). Or, les algorithmes de reconstruction pour ce type de géométrie sont notoirement plus complexes et lourds à implanter. Leur mise en œuvre n'est également pas exempte d'artefacts propres (c'est particulièrement vrai dans les plans d'images éloignés du plan d'émission). Dans le cas des tomographes multicoupes, il est néanmoins possible de se passer de cette formulation contraignante du problème : le nombre de barrettes est suffisamment réduit (typiquement quatre) pour considérer que les plans de projection sont parallèles (l'angulation maximale est inférieure au degré). Les algorithmes de reconstruction employés sont relativement similaires à ceux décrits dans la section précédente. L'interpolation linéaire sur l'hélice est uniquement remplacée par une convolution mettant en jeu les données issues des différentes rangées de capteurs (des différentes trajectoires d'hélice, en somme). Le lecteur intéressé pourra se référer à l'article [Hu, 1999] qui fait un tour relativement synthétique de toutes les questions d'échantillonnage et de reconstruction ; voir également [Wang et Vannier, 1999].

Dans ce qui va suivre, nous allons nous intéresser à la tomographie axiale ou hélicoïdale simple coupe. Notons qu'il semble possible cependant d'adapter la formulation hélicoïdale présentée dans ce document au cas du tomographe multicoupe,

soit en faisant l'approximation usuelle de plans de rayonnement parallèles, soit en reformulant le modèle de création des données de manière à tenir compte de l'angulation de chacun des plans.

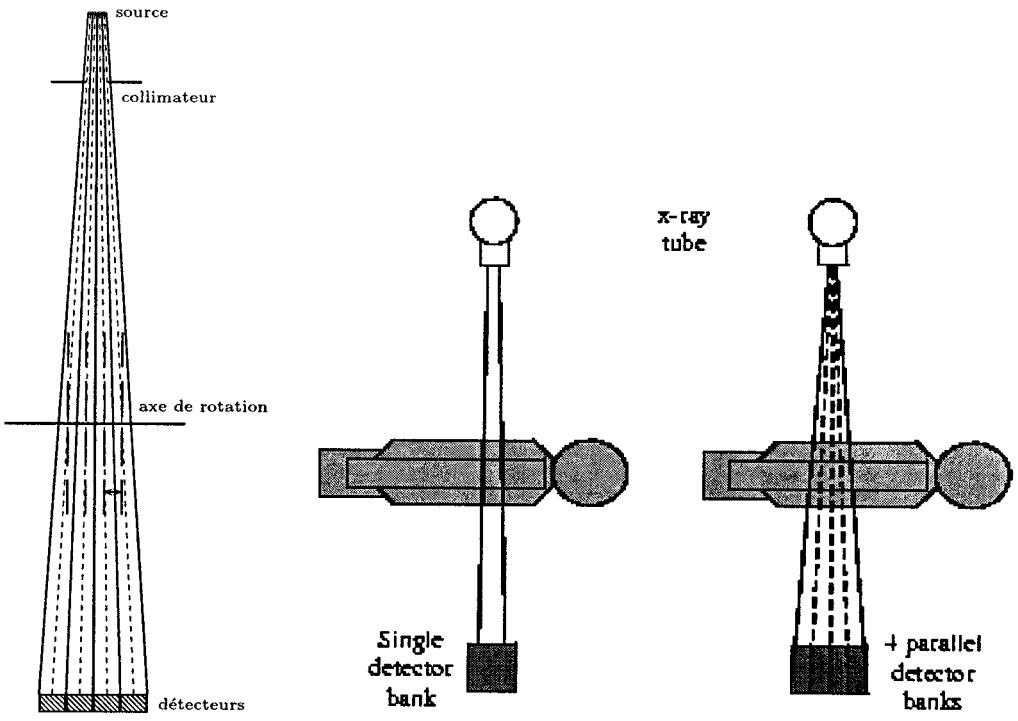


Figure 3.4: Échantillonnage hélicoïdal multicoupe : (←) ensemble de quatres rangées de capteurs montées en parallèle ; (→) illustration du mode de saisie *simple coupe* vs. *multicoupes* ; illustration tirée de http ://www.impactscan.org/.

3.3 En résumé...

Une revue de littérature montre sans ambiguïté que les techniques de reconstruction en tomographie hélicoïdale reposent toutes sur une méthodologie assez commune. Plutôt que de rentrer dans des détails techniques laborieux, nous avons

cherché dans ce chapitre à dégager l'essentiel des démarches de reconstruction.

À la question « Que pouvons nous réellement escompter en terme de précision des reconstructions en mode hélicoïdal ? », une piste indicatrice évidente consiste à comparer, toutes choses égales par ailleurs, la reconstruction *d'un même plan de coupe* reconstruit suivant les deux modalités : on s'aperçoit alors rapidement que la précision du mode hélicoïdal ne dépasse pas celle obtenue dans le mode axiale ; à cela on peut donner deux raisons majeures :

(*i*) pour vraiment bénéficier des avantages du mode hélicoïdal (diminution conjointe de la dose administrée et du temps d'examen), la quantité totale d'information produite en mode hélicoïdal se retrouve nécessairement inférieure à celle du mode axial ;

(*ii*) le mode hélicoïdal se ramène, par le biais d'une interpolation ou d'un simple filtrage dans le domaine des projections, au cas axial afin d'utiliser la RPC standard.

Le second point nous suggère d'abandonner la RPC (et les *heuristiques* associées) au profit d'une approche plus robuste et permettant de restituer des informations haute fréquence sur l'image. Le premier point caractérise le mode de saisie du tomographe et nous suggère d'adopter une méthode de reconstruction robuste à un échantillonnage lacunaire du volume à imager.

Par la suite, les approches pénalisées serviront de ligne directrice pour conce-

voir une technique de reconstruction hélicoïdale : en effet, ces approches ont fait la preuve de leur efficacité en *mode axial* en alliant précision et robustesse des reconstruction — cf. par exemple [Delaney et Bresler, 1998] ; nous leur consacrons la prochaine partie de ce mémoire.

PARTIE II

Inversion régularisée en imagerie

CHAPITRE 4

APPROCHE RÉGULARISÉE EN IMAGERIE : LE CADRE DÉTERMINISTE

Ce chapitre s'éloigne quelque peu de la tomographie pour traiter plus généralement des méthodes déterministes d'inversion en imagerie. Plutôt que de faire un inventaire exhaustif des différentes approches possibles, on cherchera plutôt à présenter les outils qui paraissent les plus adaptés à notre problématique. Ce chapitre constitue donc une introduction aux approches pénalisées qui sont assez largement employées en tomographie axiale et que nous étendrons à la tomographie hélicoïdale. Parallèlement, cet exposé présente certains éléments nécessaires à la motivation de choix de mise en œuvre ultérieurs.

4.1 Régularisation de l'inversion numérique

On supposera dans ce qui suit que le problème d'observation numérique que nous avons à inverser s'écrit sous une forme linéaire suivante

$$\boldsymbol{y} = \boldsymbol{R}\boldsymbol{x} + \boldsymbol{\varepsilon} \tag{4.1}$$

avec $\boldsymbol{R}$ une matrice de $\mathbb{R}^{M\times N}$ décrivant le phénomène physique (ex. la projection tomographique), $\boldsymbol{x} \in \mathbb{R}^N$ le vecteur des inconnues (l'image) et $\boldsymbol{y} \in \mathbb{R}^M$ le vecteur des mesures. La relation (4.1) peut s'interpréter comme une discrétisation d'une relation linéaire continue ; ce modèle intervient largement en imagerie, au travers notamment des opérateurs discrets de convolution ou de projection tomographique — c.à.d. le formalisme « algébrique » du chapitre 2.

Instabilité numérique

Revenons à la solution non régularisée $\boldsymbol{x}^\dagger$ et tentons une analyse de son comportement *via* sa décomposition en valeurs singulière. À partir de (4.1), on écrit :

$$\boldsymbol{x}^\dagger = \boldsymbol{R}^\dagger \boldsymbol{R}\boldsymbol{x} + \sum_{j=1}^{r} \frac{\langle \boldsymbol{u}_j, \boldsymbol{\varepsilon} \rangle}{\sigma_j} \boldsymbol{v}_j \tag{4.2}$$

où $\{\sigma_j, \boldsymbol{u}_j, \boldsymbol{v}_j\}_{j=1}^r$ est le système singulier de $\boldsymbol{R}$, les valeurs singulières σ_j étant classées de manière décroissante. On a déjà souligné que le mauvais conditionnement du problème d'inversion est associé à des valeurs singulières très proches de zéros. Il est également instructif d'ajouter que les valeurs singulières *les plus faibles* sont associées aux composantes spectrales *les plus coupées* par le système d'imagerie. Les systèmes d'imagerie étant généralement « passe bas » (c'est par exemple le cas du tomographe), ce sont les dernières valeurs singulières — c.à.d. celles associées aux composantes spectrales hautes fréquences (abrég. HF) — qui sont les plus faibles.

Le facteur σ_j^{-1} dans le membre de droite de (4.2) révèle une importante amplification des composantes HF du bruit dans $\boldsymbol{x}^\dagger$, ce qu'on constate effectivement en pratique.

Régularisation de l'inversion

Les problèmes d'inversion mal conditionnés sont intrinsèquement instables et leur inversion naïve conduit systématiquement à des difficultés. En particulier, l'exemple précédent suggère qu'une *inversion basée exclusivement sur les données observées est vouée à l'échec.* S'affranchir de cet obstacle nécessite en général de changer de *méthodologie.* La notion de « régularisateur », initialement introduit dans un cadre continu [Nashed, 1981, p.223], fournit une première étape vers la définition d'une solution stable :

Définition 2 *Un régularisateur de l'équation linéaire* $\boldsymbol{y} = \boldsymbol{R}\boldsymbol{x}$ *est une famille d'opérateurs* $\{L_\alpha; \alpha \in \Lambda\}$ *telle que*

$$\begin{cases} \forall \alpha, & L_\alpha : \mathbb{R}^M \to \mathbb{R}^N \text{ est continue} \\ \forall \boldsymbol{y} \in \mathbb{R}^M, & \lim_{\alpha \to 0} L_\alpha(\boldsymbol{y}) = \boldsymbol{R}^\dagger \boldsymbol{y}; \end{cases}$$

Dans cette définition, $\boldsymbol{R}^\dagger$ est l'inverse généralisée de $\boldsymbol{R}$ introduite en section 2.2.2 et α est le *paramètre de régularisation* qui prend ses valeurs dans un ensemble qui reste à définir. L'introduction d'un régularisateur n'est néanmoins pas suffi-

sant pour assurer que la solution sera « satisfaisante » vis-à-vis de l'*utilisateur*. En particulier, la solution continue peut toujours manquer de robustesse[1] ou encore produire des solutions que nous jugeons inappropriées. Finalement, l'utilisateur doit intervenir dans la conception et l'évaluation d'une approche régularisante[2] : en pratique, la ligne directrice consiste à introduire une forme de connaissance *a priori* sur la solution recherchée afin de stabiliser l'inversion dans le sens « attendue » par l'utilisateur ; le paramètre de régularisation α est alors ajusté (suivant un critère objectif ou non) de manière à produire une solution jugée intéressante.

Deux salles, deux ambiances

Il est usuel de classer les approches régularisantes suivant deux grandes familles distinctes : celles qui procèdent par *réduction de l'espace des solutions*, et celles qui cherchent à *minimiser un certain objectif*, éventuellement sous contrainte. Notons néanmoins qu'un tel découpage reste formel puisque certaines de ces approches peuvent être classées dans les deux familles suivant l'interprétation qu'on en fait.

---

[1] C'est évident puisqu'on on peut choisir $\forall\alpha$, $L_\alpha(\boldsymbol{y}) \equiv \boldsymbol{R}^\dagger$ et définir ainsi un « régularisateur » toujours mal conditionné.

[2] C'est lui qui juge de la qualité du résultat final, et c'est également lui le mieux placé pour introduire dans L_α des *a priori* pertinents (douceur, forme impulsionnelle, continuité par morceaux, *etc.*) sur la solution qu'il recherche.

4.2 Contrôle de dimension

Une première approche consiste à opérer par *contrôle de dimension* : on décompose alors la solution dans des sous espaces de dimension réduits (ex. décomposition spectrale ou en ondelette, projection sur des ensembles convexes) et on construit une solution en éliminant la contribution des sous espaces dominés par le bruit.

Un exemple révélateur : la TSVD

La plus connue de ces techniques est certainement la *décomposition en valeurs singulières tronquée* (abrég. TSVD) qui est devenue un outil de référence pour l'inversion de problèmes mal conditionnés. Le principe de la TSVD repose simplement sur l'utilisation des composantes « dominantes » de la SVD de $\boldsymbol{x}^{\dagger}$ à l'exclusion de toutes autres. Ainsi, en s'appuyant sur (2.14), la solution obtenue par TSVD s'écrit formellement

$$\boldsymbol{x}_k^{\mathrm{TSVD}} = \boldsymbol{T}_k \boldsymbol{y}$$

où l'opérateur régularisant $T_k(\cdot)$ est un opérateur linéaire $\boldsymbol{T}_k$ de taille $N \times M$ obtenue en tronquant la SVD de $\boldsymbol{R}$:

$$\boldsymbol{T}_k = \sum_{j=1}^{k} \left(\frac{\boldsymbol{v}_j \boldsymbol{u}_j^T}{\sigma_j} \right).$$

L'*ordre de troncature* $k \leq r$ est la dimension du sous espace engendré par les valeurs singulières supérieures à un paramètre $\alpha > 0 : k = \max\{j \in \mathbb{N} \mid \sigma_j \geq \alpha\}$. Le paramètre α est supposé fourni initialement ; en ce sens, il constitue la connaissance *a priori* introduite par l'utilisateur. Pour peu que α soit choisi correctement, la TSVD traite efficacement le mauvais conditionnement du problème numérique. Cependant, cette méthode reste d'un intérêt limité en imagerie principalement pour deux raisons :

1. le coût de calcul rend la TSVD définitivement peu attractive pour traiter un problème d'imagerie réaliste[3] ;
2. comme toutes les approches procédant par contrôle de dimension, la TSVD renonce *de facto* à restaurer les composantes HF trop bruitées du spectre de l'image [Demoment et Idier, 2001*b*].

Contrôle de dimension et restauration des bords francs

Si l'utilisateur souhaite obtenir des images visuellement acceptables sans compromettre la restauration des bords francs, la suppression des HF inhérente aux contrôle de dimension est généralement considérée comme rédhibitoire. Les approches pénalisées que nous présentons maintenant sont capables de fournir une

[3]La mise en œuvre d'une SVD exacte nécessite $O(N^3)$ opérations élémentaires : une telle complexité ne permet pas de traiter efficacement des images de plusieurs centaines de millier de pixels. Il est néanmoins possible de contourner partiellement le problème en *approchant* les composantes de la SVD ; pour plus de détails, nous renvoyons à [Hansen, 1992*b*, 2.8.2] et à [Golub et Van Loan, 1996, 5.2.5].

solution intéressante à ce type de problème.

4.3 Approches pénalisées (régularisation de Tikhonov généralisée)

Repartons de l'équation d'observation (4.1) : puisque la solution exacte $\boldsymbol{x}$ est inaccessible à partir de la seule connaissance des données $\boldsymbol{y}$ et du modèle physique $\boldsymbol{R}$, une démarche naturelle consiste à chercher une solution qui permet une certaine *adéquation* entre les données $\boldsymbol{y}$ et la sortie du modèle $\boldsymbol{Rx}$. De manière formelle, on cherchera un élément de l'ensemble

$$S_{\mathcal{Q}} = \{\boldsymbol{x} \in X : \min \mathcal{Q}(\boldsymbol{y}, \boldsymbol{x})\} \tag{4.3}$$

avec $X \subseteq \mathbb{R}^N$ un ensemble *convexe* et *fermé*, et $\mathcal{Q} : \mathbb{R}^M \times \mathbb{R}^N \to \mathbb{R}$ une *fonction d'adéquation aux données* qu'on choisira souvent strictement convexe et coercive par rapport à $\boldsymbol{x}$ — dans ce cas, le rang de $\boldsymbol{R}$ détermine à lui seul l'unicité du problème d'optimisation. Ici, on suppose que $\boldsymbol{y}$ est élément de $\mathbb{R}^M$ afin de privilégier la clareté de l'exposé ; en pratique, les mesures peuvent prendre leurs valeurs dans un autre ensemble dicté par le contexte (ex. $\mathbb{N}^M$, l'ensemble des M-uplets entiers positifs).

Pour le traitement d'images, $X = \mathbb{R}^N$ (cas non contraint), $X = \mathbb{R}_+^N$ (contrainte

de positivité), ou $X = \{1, \cdots, 256\}$ (valeurs discrètes) constituent les choix les plus répandus. D'autre part, on se tourne naturellement vers la norme euclidienne (où ℓ_2) pour mesurer l'adéquation aux données

$$\forall \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^M \qquad \mathcal{Q}(\boldsymbol{a}, \boldsymbol{b}) = ||\boldsymbol{a} - \boldsymbol{b}||_2^2.$$

On note néanmoins que la norme[4] ℓ_p peut être employée dans un certaine nombre d'applications relevant plus de l'estimation robuste que du traitement d'image ; cf. [Burrus *et al.*, 1994; Yarlagadda *et al.*, 1985].

Finalement, on constate sans peine que l'ensemble des solutions $S_{\mathcal{Q}}$ (4.3) généralise l'ensemble des solutions de « moindres carrés pondérée » défini par :

$$S_{\mathrm{w}} = \left\{ \boldsymbol{x} \in \mathbb{R}^N : \min ||\boldsymbol{W}^{1/2}(\boldsymbol{y} - \boldsymbol{R}\boldsymbol{x})||^2 \right\}$$

rencontré par exemple au chapitre 2, page 47. Du point de vue de la *régularisation*, la différence entre ces deux formulations repose sur les éventuelles contraintes (dites « contraintes dures ») qu'impose l'ensemble X sur la solution. En pratique, ce type de contrainte est séparable (ex. $x_n \geq 0$ pour $n = 1, \cdots, N$) et reste insuffisante pour stabiliser correctement une inversion qui demeure principalement déterminée

[4]Pour tout $\infty \geq p \geq 1$, on définit la norme ℓ_p de $\boldsymbol{x} \in \mathbb{R}^N$ par $||\boldsymbol{x}||_p = \left(\sum_{n=1}^N |x_n|^p \right)^{1/p}$ où $|a|$ désigne la valeur absolue de a élément de $\mathbb{R}$.

par les mesures.

Pénalisation de l'inversion

Une stabilisation correcte de l'inversion passe par l'ajout d'un *a priori* sur la solution. Cet objectif peut être atteint avec facilité par l'ajout d'un terme de *pénalisation* $\mathcal{P}(\boldsymbol{x})$ dans (4.3). Ainsi, on introduit un *critère composite* $J : \mathbb{R}^N \to \mathbb{R}$ tel que

$$\forall \boldsymbol{y} \in \mathbb{R}^M,\ \alpha \geq 0, \qquad J(\boldsymbol{x}) = \mathcal{Q}(\boldsymbol{y}, \boldsymbol{x}) + \alpha \mathcal{P}(\boldsymbol{x}) \tag{4.4}$$

avec $\mathcal{P} : \mathbb{R}^N \to \mathbb{R}$ une fonction dont les caractéristiques restent à définir. Le paramètre de régularisation α pondère l'influence de la pénalisation dans le critère composite ; son ajustement est évidemment nécessaire pour obtenir une solution adéquate pour l'utilisateur. Pour α fixé, on définit alors la solution régularisée $\boldsymbol{x}_\alpha$ comme un élément de l'*ensemble des solutions pénalisées* S_α :

$$S_\alpha = \{\boldsymbol{x} \in X : \min J(\boldsymbol{x})\}; \tag{4.5}$$

l'unicité étant garantie si X est un ensemble convexe et J une fonction strictement convexe. Pour des raisons qui vont devenir claire par la suite, cette formulation est également appelée *régularisation de* TIKHONOV *généralisée*. La section suivante présente les pénalisations usuellement introduites en traitement d'image (reconstruction tomographique comprise) ; on notera que, dans la mesure où la pénalisation

modélise un certain comportement *attendu* de l'image, $\mathcal{P}(\boldsymbol{x})$ est également qualifiée de *modèle d'image.*

4.3.1 Modèles d'images introduit *a priori*

Nous présentons maintenant trois classes de modèles d'images : les modèles quadratiques, les modèles non quadratiques mais convexes, et les modèles non convexes. Si le choix d'un modèle d'image est guidé par nos connaissances *a priori* sur la solution, les contraintes de mise en œuvre sont généralement suffisantes en imagerie pour nécessiter la recherche d'un *compromis* entre complexité du modèle et coût d'implantation. De manière à refléter cette démarche, la présentation faite ci-dessous suit un ordre croissant en terme de coût d'implantation.

4.3.1.1 Pénalisation quadratiques

Pénalisation de Tikhonov

Le modèle d'image certainement le plus simple pénalise les trop fortes *valeurs* de solution par le biais d'une mesure de type ℓ_2 :

$$\mathcal{P}_{\ell_2}(\boldsymbol{x}) = ||\boldsymbol{x} - \boldsymbol{m}||_2^2, \qquad \boldsymbol{m} \in \mathbb{R}^N. \tag{4.6}$$

Pour $\boldsymbol{m} \equiv \boldsymbol{0}$, on retrouve la pénalisation utilisée par TIKHONOV dans ses premiers travaux datant du début des années 60 [Tikhonov, 1963] ; cet auteur est générale-

ment considéré comme l'inventeur de ces approches pénalisées. On notera que la pénalisation (4.6) s'annule uniquement pour $\boldsymbol{x} \equiv \boldsymbol{m}$; cette caractéristique a parfois conduit à qualifier cette pénalisation de terme de « rappel » à $\boldsymbol{m}$.

Mesure d'irrégularité locale

De manière plus générale, on peut être intéressé par une pénalisation des *irrégularités locales.* Une technique largement répandue, qui trouve là encore ses fondements dans les travaux de TIKHONOV [Tikhonov et Arsénine, 1976, p. 60], consiste à pénaliser la norme ℓ_2 d'un opérateur de différentiation numérique d'ordre k appliqué à la solution :

$$\begin{aligned} \mathcal{P}_{\ell_2}^{(k)}(\boldsymbol{x}) &= ||\nabla^{(k)}(\boldsymbol{x}) - \boldsymbol{w}||_2^2, \qquad \boldsymbol{w} \in \mathbb{R}^C,\ k \in \mathbb{N} \\ &= ||\boldsymbol{D}^{(k)}\boldsymbol{x} - \boldsymbol{w}||_2^2 \end{aligned} \tag{4.7}$$

avec $\nabla^{(k)}(\cdot) = \boldsymbol{D}^{(k)}$ une matrice $C \times N$ de différentiation numérique d'ordre k. Le paramètre $\boldsymbol{w}$ joue là encore le rôle d'un terme de rappel puisque la mesure (4.7) est nulle si $\boldsymbol{D}^{(k)}\boldsymbol{x} \equiv \boldsymbol{w}$. En pratique, le choix le plus répandu est sans conteste la dérivée première ($k = 1$) et $\boldsymbol{w} \equiv \boldsymbol{0}$, ceci principalement pour favoriser l'apparition de *zones uniformes* dans l'image 2D ou 3D. Dans ce cas, $\boldsymbol{D}^{(1)}$ se réduit souvent à une matrice des différences finies du première ordre entre paire de pixels voisins dans le plan — si $\boldsymbol{x}$ est une image 2D — ou dans l'espace — si $\boldsymbol{x}$ est une image

3D.

Aspects de mise en œuvre

L'intérêt des pénalisations quadratiques est qu'elles conduisent aux algorithmes de plus faible coût calculatoire. Afin d'illustrer ce propos, on s'intéresse au cas typique suivant :

$$\begin{aligned} \text{Adéquation quadratique} \quad &: \quad \mathcal{Q}(\boldsymbol{y}, \boldsymbol{x}) = ||\boldsymbol{y} - \boldsymbol{R}\boldsymbol{x}||_2^2, \\ \text{Modèle quadratique} \quad &: \quad \mathcal{P}(\boldsymbol{x}) = ||\boldsymbol{D}^{(k)}\boldsymbol{x} - \boldsymbol{w}||_2^2, \\ \text{Cas non contraint} \quad &: \quad X = \mathbb{R}^N. \end{aligned}$$

Le critère composite défini par (4.4) est alors une forme *quadratique* de type « moindres carrés pénalisés »

$$J_{\ell_2}(\boldsymbol{x}) = ||\boldsymbol{y} - \boldsymbol{R}\boldsymbol{x}||_2^2 + \alpha||\boldsymbol{D}^{(k)}\boldsymbol{x} - \boldsymbol{w}||_2^2, \tag{4.8}$$

minimisée par toutes les solutions de l'équation normale $\left(\boldsymbol{R}^T\boldsymbol{R} + \alpha\boldsymbol{B}\right)\boldsymbol{x} = \boldsymbol{R}^T\boldsymbol{y} - \boldsymbol{w}$ où a été posé $\boldsymbol{B} \equiv (\boldsymbol{D}^{(k)})^T\boldsymbol{D}^{(k)}$. Comme on vérifie souvent $\mathrm{Ker}(\boldsymbol{R}) \cap \mathrm{Ker}(\boldsymbol{D}^{(k)}) = \emptyset$, l'équation normale admet une unique solution qui s'écrit :

$$\boldsymbol{x}_\alpha^{\ell_2} = \left(\boldsymbol{R}^T\boldsymbol{R} + \alpha\boldsymbol{B}\right)^{-1}\left(\boldsymbol{R}^T\boldsymbol{y} - \boldsymbol{w}\right). \tag{4.9}$$

Ce type d'inversion passe donc par la résolution du système linéaire défini par (4.9), résolution qui peut être effectuée en $O(N^3)$ opérations par le biais de la factorisation LU et du pivot de GAUSS ; voir [Golub et Van Loan, 1996, Chap. 3]. Dans le cas où $\boldsymbol{R}$ est une matrice de *convolution*, les algorithmes opérant dans le domaine de FOURIER peuvent faire descendre cette complexité à $O(N \log N)$. Ces algorithmes peuvent également être utilisés dans certains cas pour le problème de reconstruction tomographique (*i.e.* avec $\boldsymbol{R}$ l'opérateur de RADON discrétisé).

En traitement d'image, le terme d'adéquation aux données est quasi systématiquement de type « moindres carrés ». On souligne néanmoins que si on souhaite abandonner l'adéquation euclidienne l'inversion conduit à un problème d'optimisation non contraint de la forme

$$\min_{\boldsymbol{x}} \mathcal{Q}(\boldsymbol{y}, \boldsymbol{x}) + \alpha ||\boldsymbol{D}^{(k)}\boldsymbol{x} - \boldsymbol{w}||_2^2 \qquad \text{pour} \qquad \boldsymbol{x} \in \mathbb{R}^N.$$

Sous nos hypothèses, ce problème possède une solution unique *généralement sans forme explicite* ; dans ce cas, le recours à un algorithme *itératif* devient incontournable. D'autre part, si on impose des *contraintes* sur la solution, l'ensemble X n'est plus $\mathbb{R}^N$, et le programme mathématique à résoudre prend la forme générale

$$\min_{\boldsymbol{x}} \mathcal{Q}(\boldsymbol{y}, \boldsymbol{x}) + \alpha ||\boldsymbol{D}^{(k)}\boldsymbol{x} - \boldsymbol{w}||_2^2 \qquad \text{sujet à} \qquad \boldsymbol{x} \in X.$$

Ce problème de programmation convexe possède une solution unique là encore non explicite en générale, cependant sa résolution est rendue plus délicate que dans le cas non contraint[5] et sans solution explicite (bien que la convexité de X assure, si J est strictement convexe, l'unicité de la solution).

Compromis résolution/rapport signal à bruit

En vu d'estimer les performances des pénalisations quadratiques, on se propose d'étudier le comportement de la solution $\boldsymbol{x}_\alpha^{\ell_2}$ pour la pénalisation quadratique définie par (4.6) avec $\boldsymbol{m} \equiv \boldsymbol{0}$; *i.e.* on considère l'expression (4.9) pour $\boldsymbol{B} \equiv \boldsymbol{I}$ et $\boldsymbol{w} \equiv \boldsymbol{0}$. Dans ce cas particulier, on peut exprimer explicitement la solution *via* le système singulier $\{\boldsymbol{u}_j, \sigma_j, \boldsymbol{v}_j\}_{j=1}^r$ de $\boldsymbol{R}$:

$$\boldsymbol{x}_\alpha^{\ell_2} = \sum_{j=1}^{r} \frac{\sigma_j}{\alpha + \sigma_j} \frac{1}{\sigma_j} \boldsymbol{v}_j \boldsymbol{u}_j^T \boldsymbol{y};$$

expression qui nous apprend que, pour l'essentiel, $\boldsymbol{x}_\alpha^{\ell_2}$ est simplement une « version filtrée » de l'inverse généralisée $\boldsymbol{x}^\dagger$. Pour s'en convaincre, on constatera simplement

[5] Il est néanmoins remarquable que les contraintes de type *séparables* largement employées en traitement d'image — ex. contrainte de positivité ($x_n \geq 0$, pour $n = 1, \cdots, N$) ou de type « boite » ($a_n \leq x_n \leq b_n$, avec $b_n \geq a_n$) — sont très facilement intégrées aux algorithmes de *relaxation sur les coordonnées* [Bertsekas, 1995, Sec. 2.7] ou de *projection sur l'ensemble des contrainte* [Gilbert, 1999, Chap. 11] ou [Bertsekas, 1995, Sec. 2.3].

que

$$\boldsymbol{x}_{\alpha}^{\ell_2} \approx \begin{cases} \sum_{j=1}^{r} \frac{\boldsymbol{v}_j \boldsymbol{u}_j^T}{\sigma_j} \boldsymbol{y} & \text{pour} \quad \sigma_j \gg \alpha, \\ \sum_{j=1}^{r} \frac{\boldsymbol{v}_j \boldsymbol{u}_j^T}{\alpha} \boldsymbol{y} & \text{pour} \quad \sigma_j \ll \alpha, \end{cases}$$

c.à.d. que les composantes *basse fréquence* de cette solution sont approximativement celles de $\boldsymbol{x}^\dagger$ alors que les composantes *haute fréquence* sont atténuées comparativement à celles de $\boldsymbol{x}^\dagger$.

Conclusion

La minimisation non contrainte d'un critère pénalisé complètement quadratique conduit à des solutions formellement simples à la mise en œuvre rapide. En contrepartie, les solutions obtenues ont un comportement comparable à une régularisation par contrôle de dimension — c.à.d. que la réduction du bruit dans l'image se fait au prix d'un lissage systématique des contours. Il faut souligner qu'en traitement d'image, la simplification algorithmique obtenue dans ce cadre peut justifier qu'on adopte cette solution si les impératifs de mise en œuvre sont prépondérants[6]. Dans le cas contraire, on est alors souvent amené à abandonner le caractère quadratique de la pénalisation pour obtenir des solutions plus « conformes » à nos attentes.

[6]Ceci est moins vrai en tomographie puisque la RPC produit des résultats similaires pour un coût informatique inférieur.

4.3.1.2 Pénalisation non quadratique

La perte de résolution introduite par la pénalisation quadratique est soulignée depuis maintenant deux décennies par la communauté du traitement d'image ; voir par exemple [Geman et Reynolds, 1992, Sec. I.B]. Pour suggérer une alternative, récrivons la pénalisation $\mathcal{P}_{\ell_2}^{(1)}$ sous une forme légèrement différente de celle déduite par (4.7) :

$$\mathcal{P}_{\ell_2}^{(1)}(\boldsymbol{x}) = \sum_{c=1}^{C} |\langle \boldsymbol{d}_c^{(1)}, \boldsymbol{x} \rangle - w_c|^2 \tag{4.10}$$

où $\boldsymbol{d}_c^{(1)}$ est la c-ème ligne de l'opérateur des différences finies du première ordre $\boldsymbol{D}^{(1)}$ et w_c est le c-ème élément du vecteur $\boldsymbol{w} \in \mathbb{R}^C$. Pour l'essentiel, le produit scalaire $\langle \boldsymbol{d}_c^{(1)}, \boldsymbol{x} \rangle$ représente la différence inter pixels sur la paire c de pixels voisins dans l'image. On constate alors sans difficulté que la pénalisation quadratique pénalise *sans distinction* les variations inter pixels et conduit à une solution complètement « douce ». Une alternative assez naturelle consiste alors à *adapter la pénalisation en fonction de l'amplitude des variations.* Ainsi, on se propose de considérer à la place de (4.7) la pénalisation

$$\mathcal{P}_{\phi}^{(k)}(\boldsymbol{x}) = \sum_{c=1}^{C} \phi(\langle \boldsymbol{d}_c^{(k)}, \boldsymbol{x} \rangle - w_c) \tag{4.11}$$

où $\phi : \mathbb{R} \to \mathbb{R}$ est une fonction appelée *fonction de coût.* Pour préserver les bord francs, ϕ a par exemple un comportement quadratique proche de l'origine mais

croît ensuite moins « vite » que la parabole. Finalement, on propose de résoudre le programme mathématique

$$\begin{aligned} \min \quad & J_\phi(\boldsymbol{x}) \\ \text{sujet à} \quad & \boldsymbol{x} \in X \end{aligned} \tag{4.12}$$

avec

$$J_\phi(\boldsymbol{x}) = \mathcal{Q}(\boldsymbol{y}, \boldsymbol{x}) + \alpha \sum_{c=1}^{C} \phi(\langle \boldsymbol{d}_c^{(k)}, \boldsymbol{x} \rangle - w_c) \tag{4.13}$$

un critère pénalisé faisant intervenir une fonction de coût ϕ conduisant à une solution aux contours « préservés » ; solution qu'on notera $\boldsymbol{x}_\alpha^\phi$. On présente maintenant quelques-unes de ces fonctions ϕ non quadratiques en les classant suivant deux catégories distinctes suivant qu'elles sont convexes ou non. Cette séparation, on le verra bientôt, n'a rien de formel : en pratique, elle conditionne la convexité du problème (4.12) et, à ce titre, a un impact très sensible sur le résultat et sur les modalités de mises en œuvre.

Pénalisation convexes non quadratiques

Il s'agit de fonctions paires, strictement convexes et coercives. Ces fonctions présentent toutes un accroissement asymptotique moins important que la parabole tout en restant convexes ; pour des raisons qui vont bientôt être claires, ces fonctions de coût sont qualifiées de « compromis convexes ».

Compromis convexes

La classe des fonctions $\ell_2\ell_1$ constitue un sous ensemble notable et rassemble tous les compromis convexes se comportant de manière quadratique proche de l'origine et de manière linéaire à l'infini :

$$\phi(u) \sim \begin{cases} u^2 & \text{pour} \quad u \downarrow 0 \qquad \text{« régime } \ell_2 \text{ »} \\ u & \text{pour} \quad u \to \infty \qquad \text{« régime } \ell_1 \text{ ».} \end{cases}$$

Les fonctions $\ell_2\ell_1$ les plus connues sont certainement la *fonction de* HUBER introduite initialement en statistique robuste [Huber, 1981] et la *fonction hyperbolique* introduite par [Charbonnier *et al.*, 1997] ; voir la figure 4.1.a. Ces deux fonctions sont au moins deux fois continûment différentiables (abrég. C^2) et s'écrivent respectivement,

$$\begin{aligned} &\forall s > 0, \\ &\phi(u;s) = \begin{cases} u^2 & \text{pour } |u| < s, \\ 2s|u| - s^2 & \text{sinon} \end{cases} \\ &\phi(u;s) = \sqrt{u^2 + s^2}. \end{aligned} \tag{4.14}$$

Introduites par [Bouman et Sauer, 1993], les fonctions de coût de type ℓ_p conduisent à des solutions assez similaires sans toutefois appartenir à la famille $\ell_2\ell_1$. Ces fonc-

tions définies par

$$\phi(u;p) = |u|^p, \qquad 1 \leq p < 2$$

sont moins régulières que les fonctions $\ell_2\ell_1$ puisqu'elles sont C^1 pour $1 < p < 2$ et non différentiables en $u = 0$ pour $p = 1$. On notera à ce propos que la fonction hyperbolique (4.14) peut être vue comme une version « perturbée » [7] de la norme ℓ_1 $\phi(u) = |u|$. On pourra se reporter à [Li, 1998, TABLE 1] pour trouver d'autres exemple de potentiels convexes. Enfin, on notera que toutes ces pénalisations dépendent d'un paramètre qu'il est nécessaire de régler pour obtenir une solution satisfaisante.

Robustesse et simplicité de mise en œuvre

Le principal attrait de ces compromis convexes est qu'ils conservent au critère composite J_ϕ son caractère strictement convexe et coercif : cette propriété est largement souhaitable puisqu'elle assure simultanément,

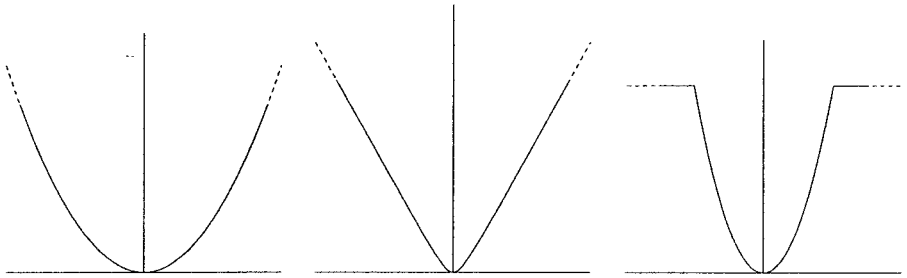
(*i*) l'*existence* et l'*unicité* de $\boldsymbol{x}_\alpha^\phi$, solution de (4.12), permettant ainsi le recours à des techniques d'optimisation usuelles (contraintes ou non) qui seront détaillées au chapitre 8 ;

(*ii*) la *continuité* de la $\boldsymbol{x}_\alpha^\phi$ vis-à-vis des données et des paramètres de réglages,

[7] On remarquera d'ailleurs qu'une technique largement répandue pour obtenir une inversion pénalisée par $\phi(u) = |u|$ est de considérer une suite d'inversions pour $\phi(u;s) = \sqrt{u^2 + s^2}$ et de faire décroître $s \downarrow 0$; [Chan et Chiu-Kwong, 1998].

garantissant ainsi la robustesse de la méthode, cf. [Li *et al.*, 1995].

La pénalisation est maintenant largement utilisée non seulement en imagerie mais également dans d'autres champs d'application nécessitant d'inverser un problème numériquement mal conditionné — ex. : l'*analyse spectrale* [Ciuciu, 2000], et le *contrôle non destructif* par déconvolution impulsionnelle [Gautier *et al.*, 2001].



(a) u^2 (b) $\sqrt{u^2+s^2}-s$ (c) $\min\{u^2; s^2\}$

Figure 4.1: Exemples de fonctions coût proposées dans la littérature : (a) quadratique, (b) compromis convexe $\ell_2\ell_1$, (c) non convexe $\ell_2\ell_0$.

Pénalisation non convexes $\ell_2\ell_0$

Au début des années 1980, il est devenu clair que la restauration de bords francs nécessitait de discerner le niveau de pénalisation suivant l'amplitude des variations. De nombreux auteurs ont alors préconisé d'utiliser des fonctions de coût *non convexse* dont les plus répandues sont de type $\ell_2\ell_0$.

Le potentiel qu'on qualifie de $\ell_2\ell_0$ est paire et monotone sur $\mathbb{R}^+$, quadratique

proche de l'origine et tendant vers une valeur finie à l'infini

$$\phi(u) \sim \begin{cases} u^2 & \text{pour} \quad u \downarrow 0 \qquad \text{« régime } \ell_2 \text{ »} \\ < M & \text{pour} \quad u \to \infty \qquad \text{« régime } \ell_0 \text{ ».} \end{cases}$$

L'un des premiers potentiels de ce type a été introduit par [Blake et Zisserman, 1987] : il s'agit de la *quadratique tronquée*

$$\phi(u; s) = \min\{u^2, s^2\}$$

qui dépend d'un paramètre de seuil s et dont on fournit une illustration sur la figure 4.1.c. Cette fonction n'étant pas différentiable en $u = s$, on peut s'intéresser à des variantes plus régulières comme la fonction introduite dans l'article de [Geman et McClure, 1987*a*]

$$\forall s > 0, \quad \phi(u) = \frac{u^2}{s^2 + u^2};$$

pour d'autres exemples, le lecteur pourra consulter par exemple [Teboul *et al.*, 1998, TABLE I]. Comme les compromis convexes, ces potentiels dépendent d'un paramètre qu'on peut assimiler, en première approximation, à un seuil entre le régime quadratique (ℓ_2) et le régime asymptotique borné (ℓ_0). Il est enfin essentiel de souligner que l'emploi de ces pénalisations soulève des difficultés sur le plan méthodologique et pratique ; c'est ce que nous exposons maintenant.

Contraintes de mise en œuvre

Coercivité et continuité de J_ϕ sont généralement garanties et assurent l'existence d'un minimiseur global. Cependant, l'emploi de potentiels $\ell_2\ell_0$ conduit systématiquement à un critère J_ϕ *non convexe* et *multimodal* —cf. [Li, 1995, Sec. IV.B] ou [Blake, 1989, p. 3]. La minimisation de tels critères est d'autant plus ardu que les problèmes d'imagerie sont de grande taille. Pour s'affranchir des minima locaux, il est nécessaire d'abandonner les techniques itératives de descente au profit d'approches au coût calculatoire bien plus élevé ; on citera en particulier les plus répandues comme *le recuit simulé* [Geman et Geman, 1984], *les algorithmes génétiques* [Haupt, 1995], ou *la non convexité graduelle* [Blake et Zisserman, 1987; Nikolova *et al.*, 1998]. Faire une présentation de ces techniques sort largement du cadre de ce document et nous renvoyons aux références précédentes et à leurs bibliographies.

Contraintes méthodologiques

Sur le plan méthodologique, l'emploi d'une pénalisation à potentiel non convexe relève plus de l'approche *détection-estimation*[8] : c'est en particulier ce que montre [Blake et Zisserman, 1987] en reliant un critère « mixte » *continu/combinatoire* de

[8]En entend par approche de « détection estimation » une approche traitant de manière conjointe l'estimation des inconnues et de variables booléenne « cachées » étiquetant (détectant) des discontinuités dans la solution ; voir [Idier et Blanc-Féraud, 2001] pour une revue de ces approches.

la forme

$$J(\boldsymbol{x}, \boldsymbol{b}) = ||\boldsymbol{y} - \boldsymbol{R}\boldsymbol{x}||_2^2 + \alpha \sum_{c=1}^{C} (1 - b_c) \langle \boldsymbol{d}_c^{(1)}, \boldsymbol{x} \rangle^2 + s b_c, \quad s > 0 \tag{4.15}$$

avec $\boldsymbol{b} \in \mathbb{R}^C$ un vecteur de variables *binaires* — c.à.d. $b_c \in \{0; 1\}$, à un critère *continu* composite faisant intervenir la *quadratique tronquée* :

$$J_\phi(\boldsymbol{x}) = ||\boldsymbol{y} - \boldsymbol{R}\boldsymbol{x}||_2^2 + \alpha \sum_{c=1}^{C} \min\{s; \langle \boldsymbol{d}_c^{(1)}, \boldsymbol{x} \rangle^2\}.$$

En pratique, ces potentiels $\ell_2\ell_0$ conduisent à des solutions qui peuvent différer sensiblement de celles obtenues par compromis convexes (par ex. $\ell_2\ell_1$). En particulier, la nature classificatrice de la méthode peut conduire à des images *sursegmentées* qui suppriment une part des détails qu'un utilisateur serait susceptible d'apprécier. Ce lien entre processus de lignes et pénalisation non convexe est rendu explicite dans [Geman et Reynolds, 1992, Sec. III] ; voir également [Li, 1995, Sec. II.A].

Finalement, il faut noter que le caractère non convexe de J_ϕ réintroduit une forme d'*instabilité* de la solution $\boldsymbol{x}_\alpha^\phi$ vis-à-vis des données et des paramètres de réglage. Cette instabilité, soulignée notamment dans [Bouman et Sauer, 1993, Sec. I.B], est à nouveau liée à l'aspect « décision » de la méthode ; voir à ce propos [Idier et Blanc-Féraud, 2001, Sec. 6.3.2].

En conclusion

Le choix du modèle *a priori* nécessite un examen *lucide* de l'objectif qu'on se propose d'atteindre. En restauration ou reconstruction d'image, cet objectif relève souvent du débruitage avec rehaussement de résolution. Or, si les compromis convexes ont montré une réelle capacité à produire de telles solutions, l'emploi de modèles non convexes est généralement *incompatible* avec cet objectif : ils produisent une segmentation et donc une perte des détails qui pourraient faire sens pour l'utilisateur.

On s'en doute, ce point est généralement décisif pour l'imagerie médicale puisque la prise de décision (*i.e. le diagnostique*) doit finalement rester le privilège du praticien. Cet aspect est à notre sens trop peu souligné dans la littérature, et en générale le choix d'un compromis convexe s'appuie sur deux autres arguments qui sont, il est vrai, d'une importance centrale pour la mise en œuvre :

(*i*) la solution est *continue* vis-à-vis des *paramètres du modèle* [Vogel, 1997, Sec. 2], ce qui simplifie la procédure de réglage (automatique ou non) ;

(*ii*) la convexité permet d'utiliser des *techniques itératives « simples »* comme, par exemple, les algorithmes à direction de descente dans un problème non contraint.

Nous reviendrons sur ces deux aspects par la suite dans le cadre de notre application.

Bibliographie annotée

La monographie de [Tikhonov et Arsénine, 1976] et l'article de revue de [Nashed, 1981] constituent des références largement citées pour les approches régularisantes. Leur lecture demande néanmoins une certaine « culture » mathématique, principalement du fait de leur traitement de la régularisation dans les espaces fonctionnels. Pour une introduction de la régularisation au sens de TIKHONOV généralisée ou par contrôle de dimension, les présentations de [Vogel, 2002, Chap. 2] et de [Demoment et Idier, 2001*b*] seront appréciées pour une approche plus pédagogique.

Pour une présentation de la TSVD, on pourra se reporter à [Hansen, 1990]. Bien que la TSVD soit traditionnellement considérée comme inadaptée pour reconstruire des images avec des bords francs, [Hansen *et al.*, 2000] ont récemment introduit une version dérivée, la PP-TSVD, qui permet d'obtenir des solutions de ce type. Sur le plan des performances, les capacités réelles de cette méthode restent néanmoins largement inconnues. *A contrario*, la bibliographie sur les approches pénalisées en inversion constitue une part non négligeable des publications en traitement d'image. Leur essor se produit dans les années 1980 à la suite des travaux fondateurs des frères GEMAN [Geman et Geman, 1984] qui ont (entre autre) démontrés que les potentiels non convexes pouvaient restaurer les discontinuités dans les images. Des potentiels non convexes ont notamment été utilisés en imagerie médicale d'émission dans [Geman et McClure, 1987*a*] ou dans [Hebert et Leahy,

1989], et en tomographie de transmission par [Dinten, 1990]. L'emploi de potentiels quadratiques en imagerie est initié aux années 1970 avec notamment [Hunt, 1973] ; la première utilisation en reconstruction tomographique semble être [Herman et Lent, 1976]. L'emploi des compromis convexes s'est aujourd'hui largement développé en traitement d'image pour les raisons que nous avons évoqué plus haut ; historiquement, les contributions significatives datent du début des années 1990 avec notamment les contributions [Green, 1990], [Bouman et Sauer, 1993] et [Charbonnier, 1994]. On notera que les pénalisations « entropiques » manquent à notre présentation des compromis convexe : ces pénalisations présentent la particularité de forcer la positivité des composantes tout en favorisant, elles aussi, les portions régulières dans l'image ; [O'Sullivan, 1995] constitue un bon point d'entrée sur le sujet. Nous fournirons à la fin du chapitre 6 un certain nombre de références sur les applications au problème tomographique des approches pénalisées.

Tous ces auteurs posent principalement l'inversion pénalisée dans un espace de dimension finie ; démarche que nous avons d'ailleurs adoptée pour la présentation. On doit néanmoins noter qu'une partie non négligeable des auteurs de la communauté de physique mathématique traitent un problème similaire *dans un cadre fonctionnel*, et discrétise par la suite le problème d'inversion pénalisée pour la mise en œuvre numérique. En particulier, on peut citer les contribution de [Rudin *et al.*, 1992], [Dobson et Santosa, 1996], [Aubert et Vese, 1997]. Cette variété de point de

vue révèle une interaction croissante entre les communautés de *physique mathématique* et du *traitement de signal et d'image*; les contributions de [Vogel et Oman, 1998], et de [Teboul *et al.*, 1998] sont à ce propos révélateurs.

On souligne finalement qu'il est possible de définir le problème de reconstruction comme un problème d'estimation dans un cadre probabiliste. Cette interprétation probabiliste offre certaines possibilités qui sont, il est vrai, assez séduisantes; nous lui consacrerons le prochaine chapitre.

CHAPITRE 5

INTERPRÉTATION PROBABILISTE ET INFÉRENCE BAYÉSIENNE

L'inversion pénalisée peut être interprétée dans un cadre probabiliste par l'intermédiaire de l'inférence bayésienne. Cependant, l'intérêt d'un tel lien reste surtout formel. L'apport du cadre probabiliste réside finalement surtout dans sa capacité à introduire certains outils méthodologiques qui n'ont pas d'équivalent dans le cadre « déterministe ». En ce sens, ce cadre permet d'aborder certaines questions que le cadre déterministe ne peut traiter.

Pour notre étude, ce chapitre permet d'introduire les *modèles d'observation poissonniens* largement employés en tomographie d'émission et en tomographie de transmission à faible dose, et de présenter quelques méthodes d'*estimation des hyperparamètres* qui conduisent à des mises en œuvre non supervisées.

Note : l'exposé ci-dessous s'affranchit pour partie des définitions et propriétés qu'un exposé rigoureux du cadre probabiliste ne saurait ignorer. Un tel exposé sortirait néanmoins du cadre de ce mémoire, et nous renvoyons le lecteur vers un des nombreux ouvrages de référence, ex. [Picinbono, 1993] ou [Bass, 1974].

5.1 Vraisemblance et adéquation aux données

Repartons du modèle d'observation additif tel que nous l'avons rappelé au début du chapitre précédent

$$\boldsymbol{y} = \boldsymbol{R}\boldsymbol{x} + \boldsymbol{\varepsilon}. \tag{5.1}$$

Jusqu'à présent, rien n'avait été précisé concernant la composante additive de « bruit » $\boldsymbol{\varepsilon}$. On suppose maintenant que $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_M)^T$ représente une *réalisation d'un vecteur aléatoire* $\boldsymbol{\mathcal{E}} = (\mathcal{E}_1, \cdots, \mathcal{E}_M)^T$ qui admet une *loi à densité* $f_{\boldsymbol{\mathcal{E}}}(\boldsymbol{\varepsilon})$. En considérant l'image $\boldsymbol{x}$ *déterministe*, on déduit aisément que le vecteur des observées $\boldsymbol{y}$ constitue une réalisation d'un vecteur aléatoire $\mathbf{Y}$ dont la densité de probabilité se déduit de celle du bruit :

$$f_{\mathbf{Y}}(\boldsymbol{y}; \boldsymbol{x}) = f_{\boldsymbol{\mathcal{E}}}(\boldsymbol{y} - \boldsymbol{R}\boldsymbol{x})$$

Un exemple classique est de considérer que $\mathcal{E}$ admet une loi à densité *gaussienne* qu'on choisira centrée par simplicité, c.à.d.

$$\forall \boldsymbol{\varepsilon} \in \mathbb{R}^M, \qquad f_{\boldsymbol{\mathcal{E}}}(\boldsymbol{\varepsilon}) = \frac{1}{Z} \exp\left\{-\frac{1}{2}\boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}\right\}$$

où Z et $\boldsymbol{\Sigma} \in \mathbb{R}^{M \times M}$ sont respectivement, la *constante de normalisation* et la *matrice de covariance* qui, par hypothèse, est définie positive. Dans ce cas, le vecteur

aléatoire $\mathbf{Y}$ admet une loi à densité gaussienne de même covariance que $\boldsymbol{\mathcal{E}}$ et de moyenne $\boldsymbol{Rx}$:

$$f_{\mathbf{Y}}(\boldsymbol{y};\boldsymbol{x}) = \frac{1}{Z}\exp\left\{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{Rx})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{Rx})\right\} \tag{5.2}$$

Par la suite, on identifiera la *fonction de vraisemblance des données* que nous noterons $V(\boldsymbol{y};\boldsymbol{x})$ à la densité de probabilité des observations paramétrée par l'image $f_{\mathbf{Y}}(\boldsymbol{y};\boldsymbol{x})$, *i.e.*

$$V(\boldsymbol{y};\boldsymbol{x}) = f_{\mathbf{Y}}(\boldsymbol{y};\boldsymbol{x}) \qquad \boldsymbol{x}\in X,\ \boldsymbol{y}\in Y$$

où, de manière générale, $X \subseteq \mathbb{R}^N$ et $Y \subseteq \mathbb{R}^M$; X et Y étant simplement $\mathbb{R}^N$ et $\mathbb{R}^M$ pour le cas du modèle additif gaussien. Cette fonction de vraisemblance joue un rôle central dans la construction d'estimateurs : elle résume à elle seul toute l'information contenue dans les données sur l'objet déterministe $\boldsymbol{x}$.

Estimation au sens du maximum de vraisemblance

Formuler un modèle statistique du comportement des données permet d'envisager l'*estimation* de l'image dans un sens qui reste à déterminer. En statistique

orthodoxe, l'estimateur du *maximum de vraisemblance* (abrég. MV)

$$\boldsymbol{x}^{\mathrm{MV}} = \arg\max_{\boldsymbol{x}\in X} V(\boldsymbol{y};\boldsymbol{x}) \tag{5.3}$$

est largement utilisé en particulier pour ses bonnes propriétés asymptotiques — biais nul et variance minimale ; voir par ex. [Fourgeaud et Fuchs, 1972, Chap. 14]. On souligne néanmoins le caractère formel de l'expression (5.3) puisque, pour le moment, on ne dispose d'aucune garantie nous assurant que cette solution soit définie (ex. ce qui nécessite en particulier que $V(\boldsymbol{y};\cdot)$ soit bornée supérieurement) et qu'elle soit unique. D'autre part, on aura souvent intérêt à aborder le problème d'optimisation *via* une transformation monotone qu'on choisira logarithmique :

$$V(\boldsymbol{y};\boldsymbol{x}) \neq 0, \qquad \boldsymbol{x}^{\mathrm{MV}} = \arg\min_{\boldsymbol{x}\in X} \; -\log V(\boldsymbol{y};\boldsymbol{x})$$

cette approche simplifiant généralement les expressions puisque la fonction exponentielle intervient couramment dans l'expression des densités de probabilité.

Hypothèses gaussiennes et moindres carrés pondérés

Dans le cas d'un problème d'inversion (5.1) et sous l'hypothèse d'un modèle gaussien du bruit, l'opposé du logarithme de la vraisemblance définie par (5.2)

s'écrit

$$-\log V(\boldsymbol{y};\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{y}-\boldsymbol{R}\boldsymbol{x})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{R}\boldsymbol{x}); \tag{5.4}$$

cette forme quadratique est convexe puisque $\boldsymbol{\Sigma}^{-1}$ est définie positive. L'ensemble des solutions du MV s'écrit alors

$$S_{\text{MV}} = \{\boldsymbol{x} \in \mathbb{R}^N : \min ||\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y}-\boldsymbol{R}\boldsymbol{x})||^2\}. \tag{5.5}$$

Ces solutions correspondent donc aux *solutions de moindres carrés pondérés* définies au début du chapitre précédent ; voir page 82. De manière plus générale, on peut associer une vraisemblance à une *fonction d'adéquation aux données* $\mathcal{Q}(\boldsymbol{y},\boldsymbol{x})$ pour peu que la loi de probabilité ci-dessous ait effectivement un sens

$$\forall \boldsymbol{x} \in X, \quad f_{\mathbf{Y}}(\boldsymbol{y};\boldsymbol{x}) = \frac{1}{Z_1} e^{-\mathcal{Q}(\boldsymbol{y},\boldsymbol{x})/T_1}$$

avec Z_1 la constante de normalisation et T_1 un paramètre de « température ». Réciproquement, on associera une adéquation aux données $\mathcal{Q}(\boldsymbol{y},\boldsymbol{x})$ à toute vraisemblance.

Ces conditions sont généralement réalisées en pratique ; c'est en particulier le cas des vraisemblance gaussiennes ou poissoniennes[1] utilisées couramment en imagerie

[1]La section 5.3 donnera l'occasion de revenir sur ce parallèle : on y construira une fonction d'adéquation aux données $\mathcal{Q}$ dérivée d'une loi d'observation poissonienne. Quand la physique du phénomène sous jacente se décrit par un décompte corpusculaire (ex. tomographie à faible dose),

— tomographie comprise. On établira donc souvent que les solutions du maximum de vraisemblance correspondent aux solutions *non régularisées* minimisant le terme d'adéquations aux données tel que

$$-\log V(\boldsymbol{y};\boldsymbol{x}) \leftrightarrow \frac{\mathcal{Q}(\boldsymbol{y},\boldsymbol{x})}{T_1}.$$

Conclusion

Pour un problème inverse numérique instable comme le nôtre, le MV n'apporte donc pas de solution pertinente et la recherche d'une formulation stable reste nécessaire. Dans ce cadre, l'inférence bayésienne, que nous présentons ci-dessous, peut fournir des alternatives intéressantes.

5.2 Inférence bayésienne

L'inférence bayésienne se distingue de l'inférence classique par l'apport, dans la formulation initiale du problème, d'une connaissance *a priori* sur la grandeur à estimer. Dans un contexte bayésien, cette information sur l'objet prend la forme d'une loi à densité donnée *a priori* $f_{\mathbf{X}}(\boldsymbol{x})$: dans ce cadre, $\boldsymbol{x} \in X \subseteq \mathbb{R}^N$ constitue une *réalisation* d'un vecteur aléatoire $\mathbf{X} = (\mathrm{X}_1, \cdots, \mathrm{X}_N)^T$.

---

de tels modèles sont sensés représenter plus fidèlement la mesure que les lois gaussiennes qui conduisent à une adéquation $\mathcal{Q}$ de type « moindres carrés ».

5.2.1 Vraisemblance *a posteriori* et estimateurs bayésiens

Introduisons la notation $f_{\mathbf{A}|\mathbf{B}}(\boldsymbol{a}|\boldsymbol{b})$ (ou $f(\boldsymbol{a}|\boldsymbol{b})$ en raccourci) pour désigner la loi de probabilité à densité du vecteur aléatoire $\mathbf{A}$ « conditionnellement » à l'événement $\mathbf{B} = \boldsymbol{b}$. Le cadre bayésien conduit à modifier la notation de la densité des observées étant donnée l'image ; celle-ci sera notée désormais

$$f_{\mathbf{Y}|\mathbf{X}}(\boldsymbol{y}|\boldsymbol{x}) = f_{\boldsymbol{\mathcal{E}}}(\boldsymbol{y} - \boldsymbol{R}\boldsymbol{x});$$

cette densité correspondant toujours à la *vraisemblance des données* notée $V(\boldsymbol{y};\boldsymbol{x})$. La règle de Bayes fournit alors le lien existant entre les lois *a priori* et *a posteriori* :

$$f_{\mathbf{X}|\mathbf{Y}}(\boldsymbol{x}|\boldsymbol{y}) = V(\boldsymbol{y};\boldsymbol{x})\frac{f_{\mathbf{X}}(\boldsymbol{x})}{f_{\mathbf{Y}}(\boldsymbol{y})}$$

Le terme au numérateur $f_{\mathbf{Y}}(\boldsymbol{y})$ est la densité *marginale*

$$f_{\mathbf{Y}}(\boldsymbol{y}) = \int_X f_{\mathbf{Y}|\mathbf{X}}(\boldsymbol{y}|\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

qui assure la normalisation de la loi *a posteriori* ; la loi *a posteriori* s'écrit donc à un facteur près :

$$f_{\mathbf{X}|\mathbf{Y}}(\boldsymbol{x}|\boldsymbol{y}) \propto f_{\mathbf{X}}(\boldsymbol{x})\,V(\boldsymbol{y};\boldsymbol{x}).$$

Le membre de droite de cette relation est appelé *fonction de vraisemblance a posteriori* et sera notée

$$V_{\mathcal{P}}(\boldsymbol{y};\boldsymbol{x}) = f_{\mathbf{X}}(\boldsymbol{x})\, V(\boldsymbol{y};\boldsymbol{x}).$$

Dans un sens strictement bayésien, la vraisemblance *a posteriori* résume toute l'information disponible sur l'image ; à ce titre, elle permet la construction d'un certain nombre d'estimateurs aux caractéristiques diverses. L'estimateur du *maximum a posteriori* (abrég. MAP), développé plus bas, est certainement le plus connu, néanmoins d'autres estimateurs comme la *moyenne a posteriori* ou le MAP *marginal* sont parfois utilisés.

Maximum a posteriori et approche pénalisée

En pratique, on se tourne souvent vers l'estimateur du *maximum a posteriori* pour résoudre un problème d'inférence dans le cadre bayésien. De manière équivalente, on est alors amené à considérer l'ensemble des minimiseurs de l'inverse de la log vraisemblance *a posteriori*

$$S_{\text{MAP}} = \{\boldsymbol{x} \in X : \min J_{\text{MAP}}(\boldsymbol{x})\}$$

où

$$\begin{aligned} J_{\text{MAP}}(\boldsymbol{x}) &= -\log V_{\mathcal{P}}(\boldsymbol{y};\boldsymbol{x}) \\ &= -\log V(\boldsymbol{y};\boldsymbol{x}) - \log f_{\mathbf{X}}(\boldsymbol{x}); \end{aligned}$$

l'un des exemples les plus répandus correspond à un critère J_{MAP} de type « moindres carrés pénalisé »

$$J_{\text{MAP}}(\boldsymbol{x}) = \|\boldsymbol{\Sigma}^{1/2}(\boldsymbol{y} - \boldsymbol{R}\boldsymbol{x})\|^2 - \log f_{\mathbf{X}}(\boldsymbol{x})$$

associé au modèle d'observation additif gaussien, modèle que nous introduisions au début de ce chapitre.

Comme pour l'estimation au sens du MV, *un lien fort peut généralement être établi* entre l'estimation du MAP et la régularisation par pénalisation dans un cadre déterministe : sous réserve que l'on puisse établir des liens tels que

$$\begin{aligned} -\log V(\boldsymbol{y};\boldsymbol{x}) &\leftrightarrow \mathcal{Q}(\boldsymbol{y},\boldsymbol{x}), \\ \exists \alpha \geq 0 \ : \ -\log f_{\mathbf{X}}(\boldsymbol{x}) &\leftrightarrow \alpha \mathcal{P}(\boldsymbol{x}) \end{aligned}$$

alors l'ensemble des solutions S_{MAP} correspond à l'ensemble des solutions produites par la régularisation de TIKHONOV généralisée ; cf. relation (4.4) du chapitre pré-

cédent. Le lien entre la vraisemblance $V(\boldsymbol{y};\boldsymbol{x})$ et la fonction d'adéquation $\mathcal{Q}(\boldsymbol{y},\boldsymbol{x})$ ayant été abordé plus haut, reste à déterminer si on peut donner un sens au second lien exprimé ci-dessus, et en particulier si les modèles d'images $\mathcal{P}$ introduits en section 4.3.1 ont un « équivalent » probabiliste.

5.2.2 Modèles d'image probabilistes à base de champ de MARKOV

Dans l'approche bayésienne, le choix de la densité $f_{\mathbf{X}}(\boldsymbol{x})$ est le problème de modélisation qui comporte le plus de subjectivité. L'objectif est d'employer une classe de modèles pertinente vis-à-vis de l'application (ex. l'imagerie médicale) et souple à mettre en œuvre (de charge calculatoire réduite). Les *champs aléatoires de Markov* définissent une classe de modèles qui satisfont en pratique ces deux impératifs souvent contradictoires.

Champs de Markov

Un champ de MARKOV (abrég. MRF) est un champ aléatoire dont les propriétés sont régies par des interactions locales. Plus précisément, la probabilité conditionnelle d'un point connaissant tous les autres points ne dépend que des valeurs des points *voisins*. En supposant que le support de l'objet est un ensemble de site $\mathcal{S} = \{1, \cdots, S\}$ — ex. un maillage régulier de l'espace ; on peut donner la définition suivante [Brémaud, 1998, Sec. 7.1].

Définition 3 (Champ aléatoire de Markov) *On appelle « champ aléatoire de*

MARKOV » *associé à* $\mathcal{S}$ *et à un système de voisinage* η *tout champ* $\mathbf{X}$ *de support* $\mathcal{S}$ *tel que les densités de probabilité conditionnelles de ses éléments* X_i *de coordonnées* i *vérifient la relation suivante,*

$$f(x_i|x_j, j \in \Omega) = f(x_i|x_j, j \in \eta_i), \tag{5.6}$$

pour tout sous-ensemble Ω *de* $\mathcal{S}$ *contenant le voisinage* η_i *de* i *et ne contenant pas* i.

Pour que cette définition soit parfaitement rigoureuse, il reste néanmoins à définir la notion de « système de voisinage ». Un système de voisinage η sur $\mathcal{S}$ est un ensemble

$$\eta = \{\eta_i \subset \mathcal{S}, i \in \mathcal{S}\}$$

où le « voisinage du pixel » i noté η_i doit vérifier les deux propriétés suivantes,

$$\begin{cases} i \notin \eta_i, \\ j \in \eta_i \quad \Leftrightarrow \quad i \in \eta_j; \end{cases}$$

la première propriété indique que le voisinage du site i ne contient pas le site i, la seconde indique que si le site i est un voisin du site j, alors le site j est également voisin du site i (propriété de réciprocité).

Notons que les MRF ne sont utiles que si le nombre de voisins est restreint car, dans ce cas, la description locale des interactions permet d'appliquer des méthodes de résolution à charge calculatoire réduite.

Théorème de Hammersley-Clifford et champs de Gibbs

L'intérêt d'une formulation à base de MRF serait finalement très réduite si on ne pouvait écrire la probabilité *a priori* $f_{\mathbf{X}}(\boldsymbol{x})$ de l'objet sous une forme explicite et simple. Ceci est possible en faisant intervenir les potentiels de GIBBS : on introduit l'ensemble $\mathcal{C}$ constitué de C sous-ensembles de $\mathcal{S}$, chaque élément de $\mathcal{C}$ étant appelé une *clique* ; on donne alors la définition suivante [Brémaud, 1998, Sec. 7.2].

Définition 4 (Champ aléatoire de Gibbs) *Sur un support fini $\mathcal{S}$, on appelle « champ aléatoire de* GIBBS *» (GRF) associé à l'ensemble de clique $\mathcal{C}$, un champ* $\mathbf{X}$ *dont la densité de probabilité est de la forme,*

$$f_{\mathbf{X}}(\boldsymbol{x}) = \frac{1}{Z_2} \exp\left\{-\frac{U(\boldsymbol{x})}{T_2}\right\} \tag{5.7}$$

où, Z_2, T_2 sont, respectivement, les constantes de normalisation et de température. $U(\boldsymbol{x})$ est appelée la fonction d'énergie et s'écrit,

$$U(\boldsymbol{x}) = \sum_{c \in \mathcal{C}} V_c(\boldsymbol{x})$$

avec V_c une fonction potentiel associée à la clique c.

Une formulation par champs de GIBBS présente le gros avantage de donner directement accès à la densité *a priori*, $f_{\mathbf{X}}(\boldsymbol{x})$, par l'intermédiaire de (5.7). Sous des hypothèses réalistes en traitement d'image, le *théorème de* HAMMERSLEY-CLIFFORD [Winkler, 1995, Th. 3.3] établi néanmoins l'équivalence entre champs de MARKOV et champs de GIBBS. En pratique, on définit alors souvent un MRF par sa densité de GIBBS équivalente, l'ensemble des cliques C découlant directement du système de voisinage considéré par le MRF.

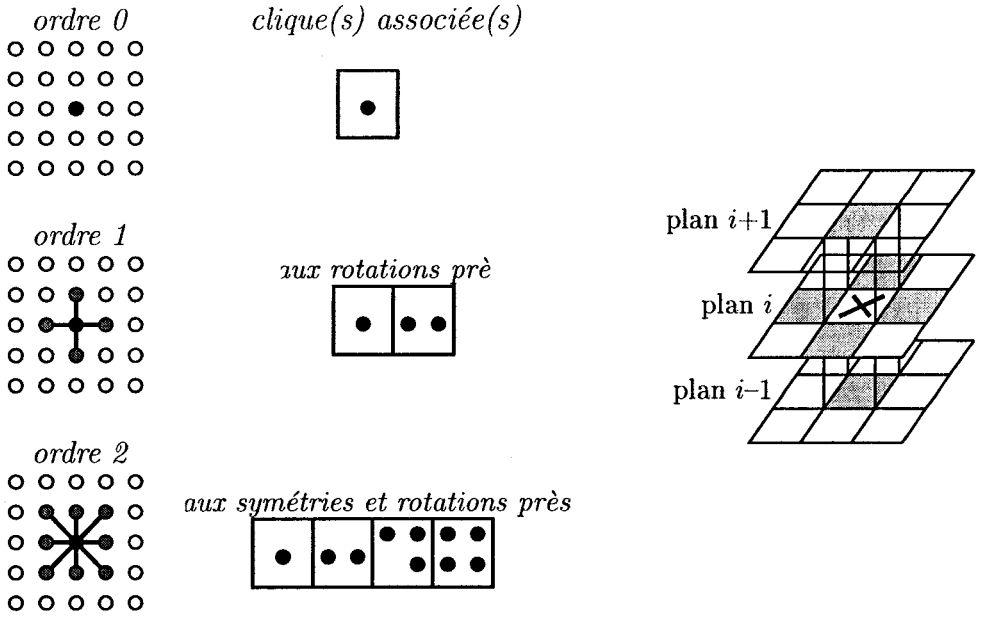


Figure 5.1: Voisinages d'ordre zéro, un et deux dans le plan avec les cliques associées (←) ; voisinage du premier ordre en 3D (→) ; cette figure est reproduite avec l'aimable autorisation de Nicolas Villain.

La figure 5.1 présente l'exemple des voisinages du *premier* et du *second ordre* ainsi que les cliques qui leurs sont associées. Notons qu'un voisinage d'ordre n

est toujours contenu dans le voisinage d'ordre immédiatement supérieur — ex. le voisinage d'ordre 0, soit le singleton que représente le pixel courant, est contenu dans le voisinage d'ordre 1. Ceci est à la base de la notion de « hiérarchie » dans les voisinages.

Nous sommes à présent en mesure de formaliser le lien existant entre les inversions pénalisées déterministes de la section 4.3.1 et certains estimateurs du MAP à base de champs de GIBBS.

MAP markovien et inversion déterministe

L'emploi de GRF comme modèle *a priori* est très répandu en traitement d'image où les potentiels $V_c(\boldsymbol{x})$ sont choisis afin de favoriser l'apparition de zones homogènes. En utilisant les notations du chapitre précédent — cf. section 4.3.1, ces potentiels s'écrivent pour la clique c :

$$V_c(\boldsymbol{x}) = \phi(\langle \boldsymbol{d}_c^{(k)}, \boldsymbol{x} \rangle - w_c), \tag{5.8}$$

et de là, un très grand nombre d'auteurs relient le programme mathématique ci-dessous à l'estimation au sens du MAP :

$$\begin{aligned} \min \quad & \mathcal{Q}(\boldsymbol{y}, \boldsymbol{x}) + \frac{T_1}{T_2} \sum_{c=1}^{C} \phi(\langle \boldsymbol{d}_c^{(k)}, \boldsymbol{x} \rangle - w_c) \\ \text{sujet à} \quad & \boldsymbol{x} \in X. \end{aligned}$$

où la fonction d'adéquation $\mathcal{Q}$ est définie *via* la vraisemblance des données $V(\boldsymbol{y};\boldsymbol{x})$. En posant $\alpha = T_1/T_2$, on reconnaît alors l'inversion pénalisée de type TIKHONOV généralisée telle qu'introduite en section 4.3.1 du chapitre précédent.

Concernant cette interprétation probabiliste de la régularisation par pénalisation, on doit néanmoins noter que si le terme d'adéquation $\mathcal{Q}$ découle généralement d'une loi de vraisemblance normalisable, un certain nombre de cas pratiques conduisent à des lois *a priori* qui sont *non normalisables*. C'est par exemple[2] ce qui se produit pour un choix (très répandu) d'une « loi » pénalisant la différence entre paire de pixels, *i.e.*

$$U(\boldsymbol{x}) = \sum_{c\in\mathcal{C}^2} \phi(\langle \boldsymbol{d}_c^{(1)}, \boldsymbol{x}\rangle - w_c),$$

avec $\mathcal{C}^2$ l'ensemble des cliques à deux éléments ; voir à ce sujet [Idier, 2001, Sec. 7.3.1].

[2]Il est néanmoins possible de « perturber » légèrement le modèle de manière à le rendre de nouveau normalisable ; par exemple, un moyen simple pour y arriver consiste à introduire une faible pénalisation de l'amplitude absolue de chaque pixel :

$$U(\boldsymbol{x}) = \sum_{c\in\mathcal{C}^2} \phi(\langle \boldsymbol{d}_c^{(1)}, \boldsymbol{x}\rangle) + \epsilon \sum_{n\in\mathcal{S}} (x_n)^2$$

avec ϵ une constante strictement positive.

Le lien avec la régularisation de TIKHONOV reste finalement assez formel et n'explique pas, à lui seul, l'attrait que portent beaucoup d'auteurs à l'emploi d'un cadre probabiliste pour l'inversion. À notre sens, le réel apport du cadre probabiliste réside dans sa capacité à introduire des outils, des modèles ou des solutions qui n'ont pas d'interprétation dans le cadre « déterministe ».

5.3 Quelques apports du cadre probabiliste

Des opérations comme la *marginalisation* ou des outils comme l'*échantillonnage pseudo aléatoire* trouvent leur pleine signification et utilité dans un cadre probabiliste bayésien. D'autre part, le cadre probabiliste permet de construire des modèles d'observation basés sur une description statistique du phénomène physique. Pour des problèmes liés à l'imagerie — restauration, reconstruction — nous illustrons maintenant deux contributions du cadre probabiliste largement représentées dans la littérature, soit

1. la construction d'un modèle d'observation poissonien qui s'avère dans certains cas préférable au modèle additif gaussien,
2. l'estimation des paramètres du modèle markovien qui permettent la mise en œuvre non supervisée d'inversion.

5.3.1 Modèle aléatoire d'observation : le cas poissonien

Dans certains cas, la grandeur physique d'intérêt peut avoir une nature corpusculaire ; c'est par exemple le cas en tomographie d'émission, en tomographie de transmission à faible dose ou dans certaines applications en astronomie. Ce caractère discret conduit à assimiler la mesure $\boldsymbol{y}$ à un décompte d'événements[3], on supposera alors que

$$\boldsymbol{y} = \boldsymbol{n} = (n_1, \cdots, n_M)$$

prend ses valeurs dans $\mathbb{N}^M$ (l'ensemble des M-uplets entiers et positifs) et constitue une réalisation d'un vecteur aléatoire $\mathbf{N} = (\mathrm{N}_1, \cdots, \mathrm{N}_M)^T$ à état discret. Dans ce cadre, on considère souvent qu'une *suite poissonienne* indépendante décrit assez fidèlement les fluctuations du vecteur des mesures, et la *vraisemblance des données* $V(\boldsymbol{n}; \boldsymbol{x})$ sera associée à la probabilité de l'événement $\mathbf{N} = \boldsymbol{n}$ paramétrée par l'image $\boldsymbol{x}$ et s'écrit :

$$V(\boldsymbol{n}; \boldsymbol{x}) = \prod_{m=1}^{M} \exp\{-\mu_m(\boldsymbol{x})\} \frac{\mu_m(\boldsymbol{x})^{n_m}}{n_m!} \; ;$$

La moyenne de la suite poissonienne $\boldsymbol{\mu}(\boldsymbol{x}) = (\mu_1(\boldsymbol{x}), \cdots, \mu_M(\boldsymbol{x}))$ dépend de l'image inconnue $\boldsymbol{x}$ qu'on choisira éventuellement de probabiliser par la suite pour construire un estimateur bayésien.

[3] On peut également ajouter à ce phénomène une composante *indépendante* de bruit d'instrumentation continue et/ou discrète. Nous ne suivrons pas cette démarche afin de garder une présentation concise ; voir par exemple [Sauer et Thibault, 2001] pour plus de détails.

Modèle poissonien en tomographie de transmission

La forme que prend $\boldsymbol{\mu}(\boldsymbol{x})$ dépend de l'application ; dans le cas de la tomographie de transmission, cette dépendance correspond directement à l'échantillonnage de la transformée de Radon *via* la loi de BERR-LAMBER :

$$\mu_m(\boldsymbol{x}) = N_0 e^{-\langle \boldsymbol{r}_m, \boldsymbol{x} \rangle}, \qquad m = 1, \cdots, M$$

avec N_0 le nombre de photons X émis par la source, et $\boldsymbol{r}_m$ la m-ème ligne de la matrice $\boldsymbol{R}$ découlant de l' « expansion en série » de la scène telle qu'introduite chapitre 2, page 41. On déduit de l'expression ci-dessus l'opposé de la log vraisemblance qui sera strictement convexe et coercive si $\boldsymbol{R}$ est de rang au moins égal à N

$$-\log V(\boldsymbol{n}; \boldsymbol{x}) = \sum_{m=1}^{M} \log(n_m!) - n_m \log(N_0) + n_m \langle \boldsymbol{r}_m, \boldsymbol{x} \rangle + N_0 \, e^{-\langle \boldsymbol{r}_m, \boldsymbol{x} \rangle}. \qquad (5.9)$$

On notera que le mauvais conditionnement de $\boldsymbol{R}$ induit une instabilité prononcée du MV « poissonnien » obtenu par minimisation de (5.9). En ce sens, le modèle poissonien ne peut garantir à lui seul la stabilité numérique de l'inversion et l'introduction d'un *a priori* pertinent reste nécessaire pour obtenir des images $\boldsymbol{x}$ satisfaisantes ; cf. par exemple [Bouman et Sauer, 1996; Fessler, 1994].

Approximation quadratique

Pour des raisons d'étude et de mise en œuvre, il peut être intéressant d'étudier le comportement de la log vraisemblance exprimée ci-dessus en fonction de ses paramètres ; ainsi en posant

$$\begin{aligned} \forall n_m \neq 0, \quad \hat{y}_m &= \log(N_0/n_m), \\ \widehat{\boldsymbol{\Sigma}}^{-1} &= \operatorname{diag}\{\sqrt{n_1}, \cdots, \sqrt{n_M}\}, \end{aligned}$$

BOUMAN et SAUER montrent que l'approximation *quadratique*

$$-\log V(\boldsymbol{n}; \boldsymbol{x}) \approx \frac{1}{2} (\hat{\boldsymbol{y}} - \boldsymbol{R}\boldsymbol{x})^T \widehat{\boldsymbol{\Sigma}}^{-1} (\hat{\boldsymbol{y}} - \boldsymbol{R}\boldsymbol{x}) + c(\boldsymbol{n}), \tag{5.10}$$

décrit assez fidèlement le comportement de la log vraisemblance pour peu que le nombre d'événement n_m dépasse quelques dizaines [Bouman et Sauer, 1996; Sauer et Bouman, 1993] ; dans (5.10), $c(\boldsymbol{n})$ est une fonction indépendante de $\boldsymbol{x}$.

Comme ces auteurs le soulignent, cette approximation quadratique ne permet pas de considérer que le vecteur $\hat{\boldsymbol{y}} = (\hat{y}_1, \cdots, \hat{y}_M)^T$ suit un modèle gaussien puisque le caractère poissonien fait intervenir les mesures n_m dans la matrice diagonale $\widehat{\boldsymbol{\Sigma}}$. Néanmoins, si on considère que les n_m varient faiblement autour d'une valeur $\sigma > 0$ (c.à.d. que $\widehat{\boldsymbol{\Sigma}} \approx \sigma \boldsymbol{I}$), ou si de manière plus générale on décide de négliger leur

influence, alors on écrira

$$-\log V(\boldsymbol{n};\boldsymbol{x}) \approx (\hat{\boldsymbol{y}} - \boldsymbol{R}\boldsymbol{x})^T \boldsymbol{\Sigma}^{-1} (\hat{\boldsymbol{y}} - \boldsymbol{R}\boldsymbol{x}) + c(\boldsymbol{n}). \tag{5.11}$$

avec $\boldsymbol{\Sigma} \in \mathbb{R}^{M \times M}$ indépendante des mesures n_m. Dans ce cas, l'estimateur du MV correspond au minimiseur du critère des moindres carrés pondérés donné par (5.11).

Conclusion

Dans un certain nombre d'applications, dont la tomographie fait partie, le modèle poissonien peut conduire à des estimations (restauration, reconstruction) plus précises qu'en employant un simple modèle additif sous hypothèse gaussienne [Bouman et Sauer, 1996]. On notera que le « raffinement » du modèle peut intervenir à plusieurs niveaux en tirant partie, soit de la vraisemblance poissonienne, soit de son approximation quadratique donnée par (5.10). Notons enfin qu'en tomographie de transmission, le rapport signal à bruit est en général suffisant pour que le modèle quadratique le plus « simpliste » (*i.e.* découlant de l'hypothèse de bruit additif gaussien) produise des résultats satisfaisants.

5.3.2 Estimation des hyperparamètres du modèle

L'intérêt des modèles markoviens en imagerie ayant été souligné en section précédente, on considère le modèle d'image paramétré suivant

$$f_{\mathbf{X}}(\boldsymbol{x};\delta) = \frac{1}{Z(\boldsymbol{\lambda})} \exp\left\{-\frac{U(\boldsymbol{x};\delta)}{T_2}\right\}$$

où on a posé $\boldsymbol{\lambda} = (T_2, \delta)$; dans la littérature, les paramètres du modèle $\boldsymbol{\lambda}$ sont souvent qualifiés d' « hyperparamètres » par opposition au vecteur de paramètres que constitue l'image $\boldsymbol{x}$. L'énergie de GIBBS $U(\boldsymbol{x};\delta)$ dépend dans la plupart des cas d'un paramètre[4]. En utilisant les notations introduites précédemment, la règle de BAYES conduit à exprimer la loi jointe

$$\begin{aligned} f_{\mathbf{X},\mathbf{Y}}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\lambda}) &= V(\boldsymbol{y};\boldsymbol{x})\, f_{\mathbf{X}}(\boldsymbol{x};\boldsymbol{\lambda}) \\ &= \frac{1}{Z(\boldsymbol{\lambda})} \exp\left\{\log V(\boldsymbol{y};\boldsymbol{x}) - \frac{U(\boldsymbol{x};\delta)}{T_2}\right\} \end{aligned} \tag{5.12}$$

qu'on supposera bien définie. Cette loi jointe constitue le point de départ d'un certain nombre d'estimateurs sur l'ensemble des paramètres $\boldsymbol{\lambda}$ du modèle markovien. Le *maximum de vraisemblance marginal* et l'approche *bayésienne hiérarchique* sont certainement les deux méthodes les plus répandues ; avant de les décrire suc-

[4]Comme exemple, on peut considérer le modèle construit à partir du compromis convexe hyperbolique $\phi(u_c;\delta) = \sqrt{u_c^2 + \delta^2}$ qui conduit à une énergie gibssienne de la forme $U(\boldsymbol{x};s) = \sum_c \sqrt{u_c^2 + \delta^2}$, avec $u_c = \langle \boldsymbol{d}_c^{(1)}, \boldsymbol{x} \rangle$ et $\delta \in \mathbb{R}$.

cinctement on souhaite fortement souligner que la dépendance de la constante de normalisation

$$Z(\boldsymbol{\lambda}) = \int_X \exp\{-U(\boldsymbol{x};\delta)/T_2\}\,\mathrm{d}\boldsymbol{x};$$

conduit à de sérieux obstacles de mise en œuvre, ceci quelque soit la méthode retenue pour estimer ces hyperparamètres.

Maximum de vraisemblance marginal

Étant donné les observation $\boldsymbol{y}$, l'estimation au sens du MV de $\boldsymbol{\lambda}$ se déduit de la loi jointe en marginalisant par rapport à $\boldsymbol{x}$, c.à.d.

$$\begin{aligned}
\hat{\boldsymbol{\lambda}}_{\mathrm{MV}} &= \arg\max_{\boldsymbol{\lambda}} f_{\mathbf{Y}}(\boldsymbol{y};\boldsymbol{\lambda}) \\
&= \arg\max_{\boldsymbol{\lambda}} \int_X f_{\mathbf{X},\mathbf{Y}}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\lambda})\,\mathrm{d}\boldsymbol{x} \\
&= \arg\max_{\boldsymbol{\lambda}} \frac{1}{Z(\boldsymbol{\lambda})} \int_X \exp\left\{\log V(\boldsymbol{y};\boldsymbol{x}) - \frac{U(\boldsymbol{x};s)}{T_2}\right\}\mathrm{d}\boldsymbol{x}
\end{aligned}$$

pour cet estimateur, le nombre d'inconnues est très inférieur au nombre de données et on admet en général que les performances de cet estimateur sont asymptotiquement bonnes. Il semble pourtant qu'en pratique et appliqué à des problèmes d'imagerie non synthétiques, l'estimation par MV ne conduise pas nécessairement au « meilleur réglage » possible des hyperparamètres ; voir par exemple [Descombes *et al.*, 1999] et [Idier, 2000, p. 63]. D'autre part, le calcul de cet estimateur est sou-

vent délicat : l'intégrale ci-dessus n'admet pas de solution analytique en général. La mise en œuvre repose par exemple sur une procédure de type *Expectation-Maximization* (abrég. EM) stochastique qui alterne étape d'optimisation et étape d'échantillonnage pseudo aléatoire — par une technique de relaxation stochastique, ex. échantillonnage de GIBBS, [Geman et Geman, 1984]. Les algorithmes qui en découlent sont d'un coût calculatoire important même si des efforts récents ont été menés en ce sens ; voir par exemple [Saquib *et al.*, 1998], [Jeffs et Pun, 1996] et [Zhou *et al.*, 1997].

Approche bayésienne hiérarchique

L'approche bayésienne hiérarchique peut être qualifiée de méthode « complètement bayésienne » : les hyperparamètres $\boldsymbol{\lambda}$ sont traités comme une réalisation d'un vecteur aléatoire $\boldsymbol{\Lambda}$ auquel on associe une loi *a priori* $f_{\boldsymbol{\Lambda}}(\boldsymbol{\lambda})$. La loi jointe (5.12) s'exprime alors

$$f_{\mathbf{X},\mathbf{Y},\boldsymbol{\Lambda}}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda}) = V(\boldsymbol{y};\boldsymbol{x})\, f_{\mathbf{X}|\boldsymbol{\Lambda}}(\boldsymbol{x}|\boldsymbol{\lambda})\, f_{\boldsymbol{\Lambda}}(\boldsymbol{\lambda}). \tag{5.13}$$

On choisit usuellement d'*échantillonner* la loi ci-dessus par un algorithme de relaxation stochastique — généralement l'échantillonneur de Gibbs. Ce choix est préféré à une simple maximisation de la loi jointe par rapport à $\boldsymbol{x}$ et à $\boldsymbol{\lambda}$ qui peut conduire à des problèmes de nature méthodologique ; cf. [Descombes et Goussard,

2001, Sec. 8.3.6]. Une fois la loi correctement échantillonnée, on a souvent recours à la moyenne empirique afin d'estimer $\boldsymbol{x}$ au sens de la moyenne *a posteriori*.

Le choix des lois *a priori* sur les hyperparamètres est délicat :

(*i*) d'une part, l'information à disposition sur les hyperparamètres est souvent limitée ou inexistante[5] ; on se replie alors vers des lois *a priori* « non informatives », c.à.d. les plus uniformes possibles.

(*ii*) d'autre part, la mise en œuvre de l'échantillonneur de GIBBS nécessite de pouvoir exprimer explicitement les lois conditionnelles complètes ; pour cela on se restreint souvent aux familles de lois *conjuguées*, voir par ex. [Cheng *et al.*, 1996] ou [Dunmur et Titterington, 1997].

Enfin, il est nécessaire de signaler que certains auteurs considèrent que cette démarche hiérarchique apporte finalement peu à la question du réglage des hyperparamètres ; voir en particulier [Idier, 2000, §5.4].

[5]En pratique, on dispose souvent d'une seule information de support.

PARTIE III

Approches pénalisées en tomographie : du mode axial au mode hélicoïdal

CHAPITRE 6

TOMOGRAPHIE EN GÉOMÉTRIE AXIALE

Une présentation des approches pénalisées en tomographie axiale permet maintenant d'exposer dans un cadre simple des objectifs — robustesse des reconstructions, localisation précise des interfaces, maîtrise des coûts d'implantation — et un argumentaire qui gardera toute sa pertinence en tomographie hélicoïdale, au chapitre suivant.

Le choix d'une méthodologie de reconstruction s'inscrivant dans un cadre applicatif particulier, il paraît utile de rappeler que les algorithmes implantés dans les tomographes commerciaux sont conçus dans une logique assez distincte de celle que soustend notre projet. En effet, les méthodes de reconstruction standard fournissent, dans les meilleurs délais, une information de nature qualitative permettant le diagnostic, alors que notre objectif est de permettre l'extraction d'une information quantitative — la localisation des interfaces osseuses — quitte à nécessiter une charge calculatoire plus conséquente. On précise néanmoins que, dans un contexte d'imagerie, le coût de mise en œuvre constitue un élément particulièrement sensible qui rentre nécessairement en compte.

L'objectif et le cadre étant fixé, il nous a paru intéressant de décomposer ce chapitre en trois sections : une première section motive l'emploi d'une inversion pé-

nalisée pour notre problème de reconstruction d'images ; la seconde section illustre les capacités de la méthode retenue sur un problème synthétique mais néanmoins réaliste ; enfin, la troisième section est consacrée aux limitations des approches pénalisées lors de leur emploi en tomographie, ou plus généralement en imagerie.

Note : Ce chapitre est dévolue aux outils méthodologiques ; les considérations algorithmiques ainsi que les techniques d'implantation ne seront pas traitées dans ce chapitre. À titre informatif, on indique néanmoins que les reconstructions axiales produites dans ce chapitre ont été obtenues par un algorithme similaire à celui développé dans le cadre hélicoïdale — cf. section 7.4.

6.1 Choix d'une approche pénalisée en tomographie

On motive maintenant l'emploi d'une approche pénalisée pour notre problème de reconstruction tomographique. Cette section permet également d'exposer certains choix de structure du critère pénalisé adopté pour effectuer l'inversion en géométrie axiale comme en géométrie hélicoïdale.

6.1.1 Motivation

Au cours des trois premiers chapitres de ce manuscrit, on a largement souligné que les approches de reconstruction standard (RPC, synthèse de Fourier, méthodes de type POCS, ...) supprimaient la contribution des hautes fréquences dans la so-

lution afin de stabiliser l'inversion[1]. Cette démarche permet d'implanter des algorithmes de reconstruction pour un coût relativement réduit, néanmoins elle induit une perte de résolution qui entrave, en particulier, la localisation précise des bords francs dans l'image.

Approches standard et contrôle de dimension

Sans exception, les approches régularisantes exposées au chapitre 4 procèdent par filtrage des hautes fréquences :

- c'est le cas des algorithmes POCS de la section 2.2 qui, par un arrêt prématuré de l'algorithme sous relaxé, ont recours au principe du « contrôle de dimension » en ne conservant que les contributions des sous espaces associés aux basses fréquences de la solution non régularisée [Natterer, 1999, Sec. 4.1] ;
- c'est également le cas de la RPC qui élimine la contribution des composantes hautes fréquences par le filtre de reconstruction.

Dans la mesure où il importe pour notre application de reconstruire une image avec des bords francs, le caractère lissant de telles méthodes est *de facto* incompatible avec nos objectifs. On s'oriente donc vers les approches pénalisées qu'on introduit finalement assez aisément en repartant du formalisme « algébrique » présenté à la fin du chapitre 2.

[1]Le lecteur pourra se reporter au chapitre 2 pour une présentation de la *rétroprojection convoluée* (abrég. RPC), de la *synthèse de* FOURIER, et de certaines méthodes de *projection sur des ensembles convexes* (abrég. POCS) dont l'ART fait partie.

La paramétrisation de la scène développée section 2.2 permet d'aboutir à l'expression suivante liant le vecteur des observations $\boldsymbol{y}$ à celui de l'image $\boldsymbol{x}$:

$$\boldsymbol{y} = \boldsymbol{R}\boldsymbol{x} + \boldsymbol{\varepsilon} \tag{6.1}$$

où on a introduit une perturbation $\boldsymbol{\varepsilon} \in \mathbb{R}^M$ afin de tenir compte de différents « bruits », et $\boldsymbol{R} \in \mathbb{R}^{M \times N}$ constitue l'opérateur de transformée de RADON discrétisé produit *via* l' « expansion en série » de la scène ; partant de (6.1), les différentes approches permettant de définir une solution au problème de reconstruction sont rassemblées sous le terme générique d'*approches algébriques.*

Ensemble de solutions non pénalisées

Comme nous le soulignions à la fin du chapitre 2, les approches **standard** s'appuyant sur ce formalisme algébrique sont d'un attrait finalement très restreint pour notre application : elle produisent une version « filtrée » d'une reconstruction appartenant à l'ensemble — numériquement instable — des solutions au sens des moindres carrés

$$S = \left\{ \boldsymbol{x} \in \mathbb{R}^M : \min ||\boldsymbol{y} - \boldsymbol{R}\boldsymbol{x}||^2 \right\}.$$

6.1.2 Reconstruction pénalisée en tomographie

Le formalisme algébrique est particulièrement attractif puisqu'il permet d'ajouter un terme de pénalisation au terme d'adéquation aux données. Ainsi, conformément à la discussion de la section 4.3 sur les approches pénalisées, on se propose d'aborder la reconstruction tomographique en cherchant un élément de l'ensemble

$$S_\alpha = \{\boldsymbol{x} \in X : \min J(\boldsymbol{x})\}$$

avec $X \subseteq \mathbb{R}^N$ un ensemble considéré convexe et fermé ; X se réduit souvent à $\mathbb{R}^N$ — cas non contraint — ou constitue un ensemble de contraintes *séparables* — ex. $\mathbb{R}^N_+$. Le critère $J : \mathbb{R}^N \to \mathbb{R}$ est de la forme

$$J(\boldsymbol{x}) = \mathcal{Q}(\boldsymbol{y}, \boldsymbol{x}) + \alpha \mathcal{P}(\boldsymbol{x}) \qquad \alpha \geq 0,$$

avec $\mathcal{Q} : \mathbb{R}^M \times \mathbb{R}^N \to \mathbb{R}$ une *fonction d'adéquation aux données* et $\mathcal{P} : \mathbb{R}^N \to \mathbb{R}$ une *fonction de pénalisation* qui doit permettre de préserver les interfaces franches dans l'image.

Construction d'une pénalisation en tomographique

En tomographie, la fonction de pénalisation $\mathcal{P} : \mathbb{R}^N \to \mathbb{R}$ est choisie de manière à favoriser l'apparition de zones localement douces dans l'images 2D ou 3D ; plus

particulièrement on posera

$$\mathcal{P}(\boldsymbol{x}\,;\,s) = \mathcal{P}_\phi(\boldsymbol{x}\,;\,s) = \sum_{c=1}^{C} \phi(\langle \boldsymbol{d}_c^{(1)}, \boldsymbol{x}\rangle\,;\,s),$$

où $\phi : \mathbb{R} \to \mathbb{R}$ est une *fonction de coût* éventuellement paramétrée par s qui pénalise la différence $\langle \boldsymbol{d}_c^{(1)}, \boldsymbol{x}\rangle$ entre les éléments de la c-ème paire de pixels voisins. Conformément à l'exposé de la section 4.3.1, le choix de la fonction de coût ϕ a un impact majeur sur l'aspect des solutions. En rappelant qu'une pénalisation *quadratique* conduit à une solution lissée et empêche de localiser précisément les interfaces, nous motivons maintenant le choix d'une pénalisation de type « compromis convexe ».

Choix d'un compromis convexe

Le choix d'un *compromis convexe* est étayé par des considérations méthodologiques et pratiques déjà évoquées dans le cadre de ce mémoire (voir page 98) ; nous les rappelons néanmoins brièvement :

1. *les fonctions ϕ non convexes produisent une segmentation marquée sur les reconstructions* ;
2. *la convexité permet d'utiliser des techniques itératives de minimisation « simples »* ;
3. *la convexité du critère permet d'assurer la continuité de la solution vis-à-vis*

de l'ensemble des paramètres de régularisation ;

Si les deux derniers points relèvent clairement de considérations de mise en œuvre, le premier point mérite le commentaire suivant : notre objectif est de permettre un traitement efficace de l'image par un expert (ex. un chirurgien orthopédiste pour le projet d'implant personnalisé) ; en ce sens, nous rejetons la segmentation de l'image puisqu'elle prend finalement la décision « à la place » de l'expert. Finalement, notre choix d'une fonction de coût se porte vers le compromis convexe de type $\ell_2\ell_1$

$$\phi(\cdot\,;\,s) : u \mapsto \sqrt{u^2 + s^2},$$

qui est au moins deux fois continûment différentiable (abrég. C^2) et qui conduira dans une certaine mesure à simplifier la procédure d'implantation algorithmique.

Choix d'un système de voisinage

Effets de bords exclus, chaque élément de l'image rassemble de 4 à 8 voisins pour une image bidimensionnelle (abrég. 2D) et de 6 à 24 voisins pour une image tridimensionnelle (abrég. 3D) ; c'est ce qu'illustre la figure 6.1 pour une image 2D. En dépit d'un support relativement restreint, l'emploi de ces « voisinages » conduit à des améliorations souvent sensibles des reconstructions en 2D ou en 3D.

Dans le cadre de ce projet, on adoptera pour la tomographie axiale 2D un voisinage à huit éléments pour chaque pixel (effets de bord exclus). Notons qu'à



Figure 6.1: Effets de bords exclus, chaque élément dans l'image 2D est mutuellement voisin de quatre (←) à huit pixels (→) ; les différences premières entre éléments voisins correspondent donc aux différences verticales et horizontales auxquelles il faut éventuellement ajouter les différences entre éléments diagonaux.

l'occasion du passage en géométrie hélicoïdale, le cadre intrinsèquement 3D nous conduira à définir un voisinage étendu dans le volume.

Terme d'adéquation aux données en tomographie

Si le rapport signal à bruit considéré est faible, il peut être intéressant d'adopter pour la tomographie de transmission une adéquation dérivée d'un modèle d'observation *poissonien* ; ainsi, en adoptant les notations et le vocabulaire de la section 5.3.1, la fonction $\mathcal{Q}$ s'écrit

$$\mathcal{Q}(\boldsymbol{n};\boldsymbol{x}) = \sum_{m=1}^{M} \log(n_m!) - n_m \log(N_0) + n_m \langle \boldsymbol{r}_m, \boldsymbol{x} \rangle + N_0\, e^{-\langle \boldsymbol{r}_m, \boldsymbol{x} \rangle}.$$

où $\boldsymbol{r}_m$ et N_0 sont respectivement la m-ème ligne de la matrice $\boldsymbol{R}$ et le nombre total de photons X émis par la source. Le vecteur $\boldsymbol{n} = (n_1, \cdots, n_M)$ constitue le décompte de photons X ayant atteint chaque capteur. Ce modèle est néanmoins assez peu employé en tomographie de transmission : le dosage administré dans le

contexte hospitalier suffit souvent largement pour qu'un simple modèle quadratique (éventuellement pondérée) puisse être utilisée sans dégradation notable de la qualité des reconstructions ; c.à.d.

$$\mathcal{Q}(\boldsymbol{y}, \boldsymbol{x}) = (\boldsymbol{y} - \boldsymbol{R}\boldsymbol{x})^T \boldsymbol{W} (\boldsymbol{y} - \boldsymbol{R}\boldsymbol{x}); \tag{6.2}$$

dans ce cas, l'inversion est de type « moindres carrés pénalisés ». On notera que la matrice $\boldsymbol{W}$ constitue un ensemble de degrés de liberté laissé à l'utilisateur ; elle peut néanmoins se déduire d'une approximation quadratique du modèle d'observation poissonien[2], ce qui constituerait un raffinement du modèle quadratique suceptible d'améliorer la précision des reconstructions.

Dans le cadre de ce projet, nous utiliserons simplement une adéquation quadratique de la forme (6.2) ; ce choix a été motivé principalement par deux considérations : d'une part, notre contexte applicatif bénéficie d'un dosage suffisant et d'autre part, le choix d'une adéquation purement poissonienne ne permettrait pas certains développements algorithmiques qui s'avèreront intéressants par la suite. Dans la suite, on choisira néanmoins par soucis de simplicité de l'exposé $\boldsymbol{W} = \boldsymbol{I}$.

Pour la géométrie axiale, on utilisera principalement l'opérateur $\boldsymbol{R}$ qui se déduit d'une paramétrisation de la scène à « base de disque »— cf. section 2.2.1 de ce

[2]Voir par exemple la relation (5.10) du chapitre précédent.

manuscrit. En tomographie hélicoïdale, le changement de géométrie d'acquisition conduit à une modification de la structure de l'opérateur reliant les observations aux éléments de l'image ; ce travail de reformulation dans le cadre hélicoïdal est présenté au chapitre suivant.

Formulation du critère pénalisé

À la lumière des choix exposés dans cette section, l'approche pénalisé retenue pour la reconstruction tomographique axiale s'écrit sous la forme

$$J_\phi(\boldsymbol{x}\,;\,\boldsymbol{\lambda}) = ||\boldsymbol{y} - \boldsymbol{R}\boldsymbol{x}||_2^2 + \alpha \sum_{c=1}^{C} \sqrt{\delta_c^2 + s^2} \tag{6.3}$$

où on a posé $\delta_c = \langle \boldsymbol{d}_c^{(1)}, \boldsymbol{x} \rangle^2$ et $\boldsymbol{\lambda} = (\alpha, s)$ qui représente l'ensemble des paramètres libres qui doivent être ajustés ; on justifiera par la suite un réglage manuel (ou mise en œuvre « supervisée ») de ces paramètres — cf. section 6.3.

Le problème de minimisation du critère pénalisé (6.3) reste *formellement simple* : l'ensemble des contraintes est très souvent séparable (ex. $X = \mathbb{R}_+^N$) et les propriétés de stricte convexité et de coercivité garantissent l'existence et l'unicité d'une solution bornée. D'autre part, la continue différentiabilité du critère garantit qu'une large classe de méthodes d'optimisation *itératives* sont utilisables.

Ce contexte méthodologique favorable masque néanmoins de réels obstacles de mise en œuvre : pour des images 2D de taille normale (ex. quelques centaines de

pixels de coté), la taille du problème d'optimisation auquel il faut faire face pose déjà des difficultés d'implantation. Ce problème apparaît évidemment sous une forme aigüe pour la reconstruction d'image 3D ; se sera en particulier le cas dans le prochain chapitre où les approches pénalisées sont étendues à la tomographie hélicoïdal.

Finalement, on attire l'attention du lecteur sur le fait que cette structure de critère sera reprise au prochain chapitre pour étendre la reconstruction pénalisée à la géométrie hélicoïdale.

6.2 Une première mise en œuvre

Cette section propose une comparaison *qualitative* des performances de la RPC et des approches pénalisées à base de compromis convexe $\ell_2\ell_1$. On se place dans le cadre axial bidimensionnel, ce contexte permettant une illustration relativement simple de notre propos.

Protocole expérimental et reconstructions

Les projections et le fantôme ont été générés par le logiciel SNARK93 [Browne *et al.*, 1993] développé par le *Medical image processing group* (MIPG) ; avec RECLBL [Huesman *et al.*, 1977], ce logiciel constitue une référence dans le domaine. Nous décrivons maintenant plus en détails le contexte qui a permis de produire ces

données.

Fantôme synthétique

Le fantôme utilisé pour ces premières mises en œuvre est illustré sur la figure 6.2 ; les niveaux d'atténuation ayant été codés de manière croissante du plus foncé au plus clair. Dans sa conception initiale, ce fantôme introduit par SHEPP et LOGAN [Shepp et Logan, 1974] représentait une coupe tomographique du crâne et les niveaux d'atténuation sur les différents motifs étaient attribués en ce sens. On notera que la version que nous utilisons est plutôt représentative des écarts importants associés à des interfaces entre tissus mous et tissus osseux ; ce choix correspond plus justement au contexte orthopédique dans lequel ce projet s'inscrit.

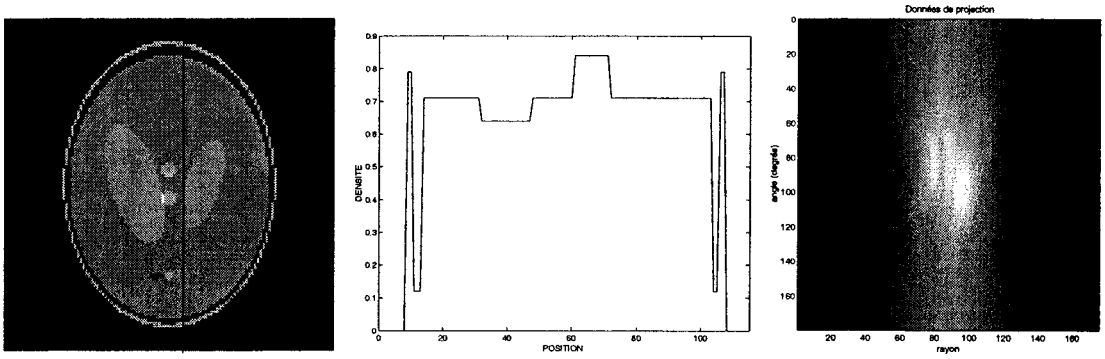


Figure 6.2: Fantôme de SHEPP et LOGAN utilisé pour les mises en œuvre (←) ; tracé de la 63-ème colonne de ce fantôme (↑) ; Sinogrammes composé de 151 projections échantillonnées sur 175 rayons obtenu pour un faisceau d'épaisseur nulle et sans bruit d'observation (→).

Création des données de projection

Les projections que nous avons simulées sous SNARK93 ont toutes été produites en géométrie parallèle pour une source monochromatique. Tous les problèmes de reconstruction ont également bénéficié du même nombre de vues et de la même fréquence d'échantillonnage : soit 151 projections prisent uniformément dans $[0; \pi]$ radians, chaque projection bénéficiant de 175 rayons. Par la suite, nous avons tenu compte de l'épaisseur de faisceau et fait varier le niveau et la nature du bruit. Dans le cas de données non bruitées et pour un faisceau d'épaisseur nulle, le sinogramme du fantôme est représenté sur la figure 6.2.

Pour donner un aperçu des capacités et des limitations de chaque méthode de reconstruction, nous avons considéré quatre contextes de mise en œuvre distincts,

contexte ① : on s'intéresse dans un premier temps aux reconstructions obtenues dans un contexte qu'on pourrait qualifier de « parfait » : bruit et épaisseur de faisceau nuls ;

contexte ② : on s'intéresse ensuite aux reconstructions obtenues pour un contexte se rapprochant plus des conditions réelles de mise en œuvre : l'épaisseur de faisceau est considérée finie et les données seront perturbées par l'adjonction d'un bruit soit *additif* (décorrélé, gaussien de moyenne $\mu = 0$ et d'écart type $\sigma = 0,005$), soit *multiplicatif* (décorrélé, gaussien de moyenne $\mu = 1$ et d'écart type $\sigma = 0,01$),

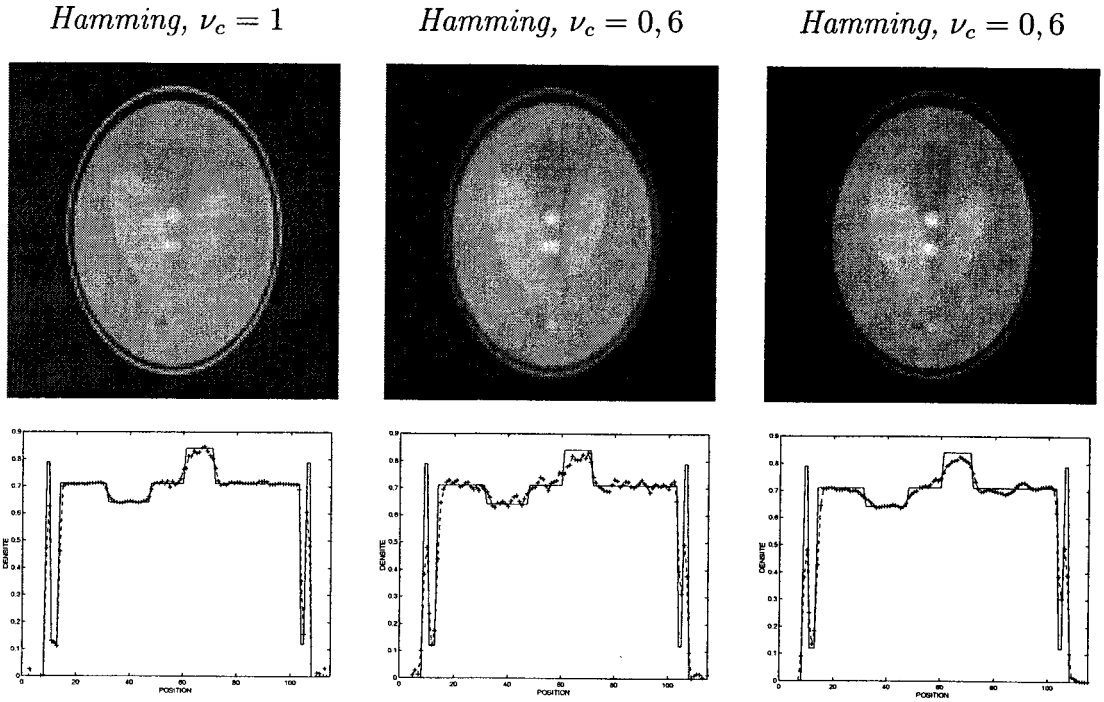


Figure 6.3: Cas ① et ② : reconstruction par RPC du fantôme à partir de projections obtenues pour : une épaisseur de faisceau nulle et sans perturbation (①) ; une épaisseur de faisceau égale à la largeur d'un pixel et un bruit additif ou multiplicatif (②) sur les projections — cf. texte pour les caractéristiques du bruit. La fréquence de coupure du filtre a été ajusté manuellement de manière à obtenir le meilleur compromis visuel entre la résolution et le niveau de bruit.

contexte ③ : on testera les limites de robustesse au bruit de chacune de ces méthodes,

contexte ④ : enfin, on testera la robustesse face au problème à angle de vue restreint — échantillonnage angulaire ne couvrant pas $[0; \pi[$.

Reconstructions par RPC

Le logiciel SNARK93 a également été utilisé pour produire les reconstructions par RPC ; cf. [Browne *et al.*, 1993, Sec. 7.2]. Le filtre de convolution et sa fréquence de coupure ν_c (réduite à la fréquence d'échantillonnage des projections) ont été choisis qualitativement de manière à fournir un bon compromis entre la résolution et le rapport signal à bruit. On notera finalement que la reconstruction d'une de ces images par SNARK se fait en environ 0,5 secondes.

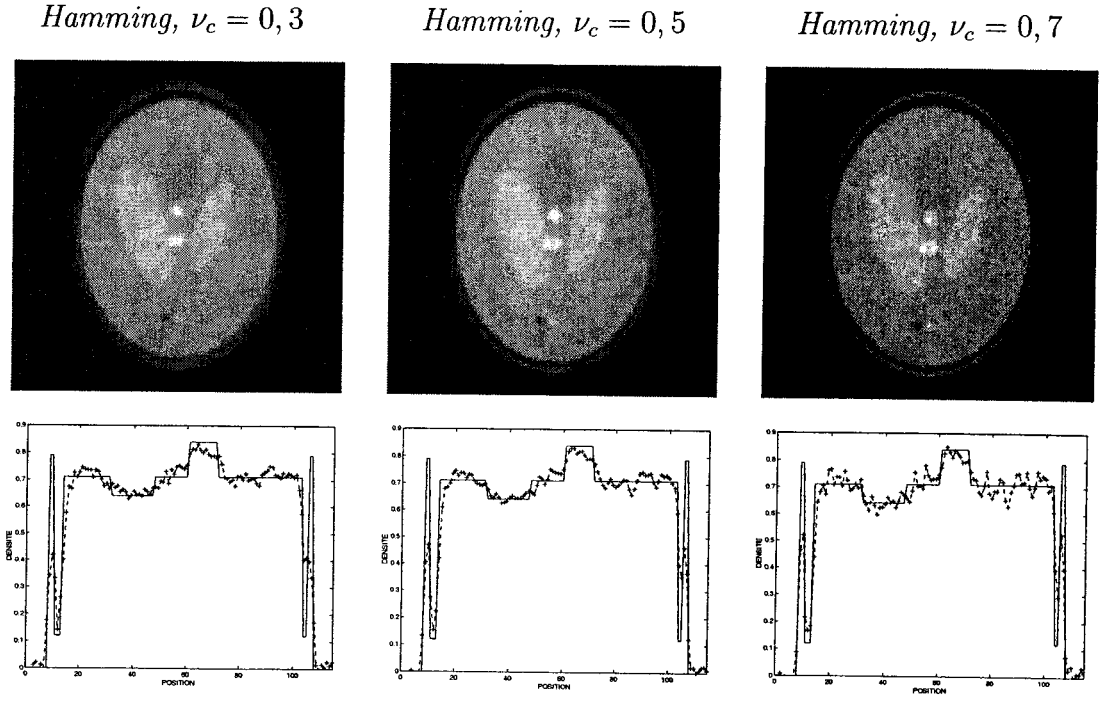
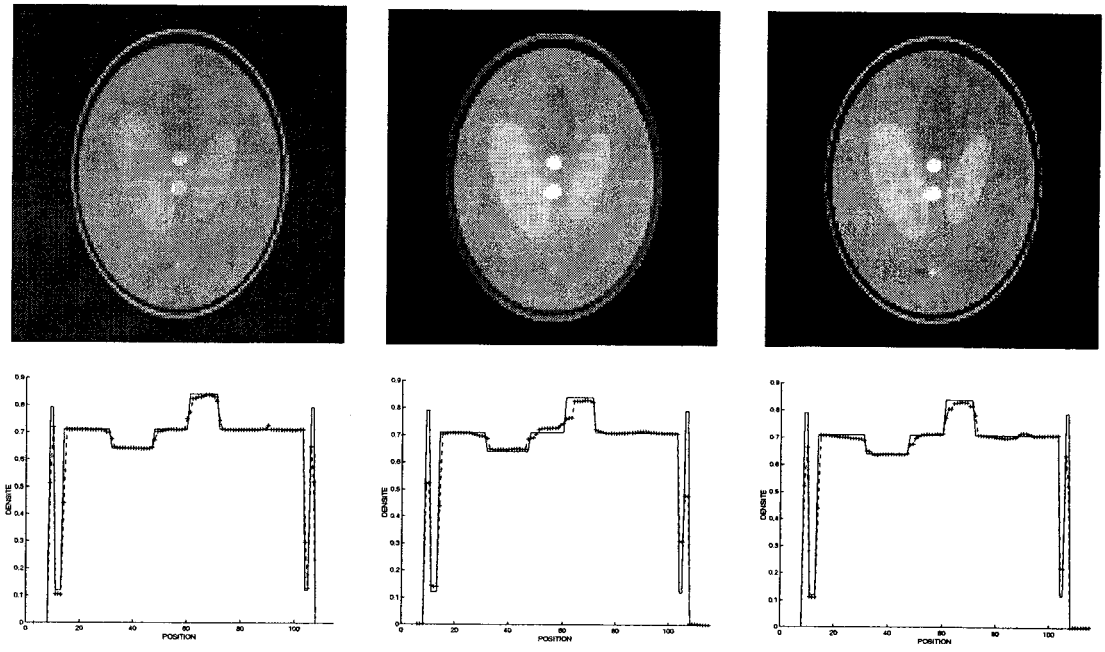


Figure 6.4: Cas ③ : reconstructions par RPC du fantôme à un niveau de perturbation « limite » pour la robustesse de la méthode (bruit additif gaussien décorrélé, centré et d'écart type $\sigma = 0,008$).

Pour un contexte expérimental « parfait » ou faiblement perturbé, les reconstructions par RPC présentées sur la figure 6.3 sont assez satisfaisantes et permettent



① (non bruité) ② – bruit additif ② – bruit multiplicatif

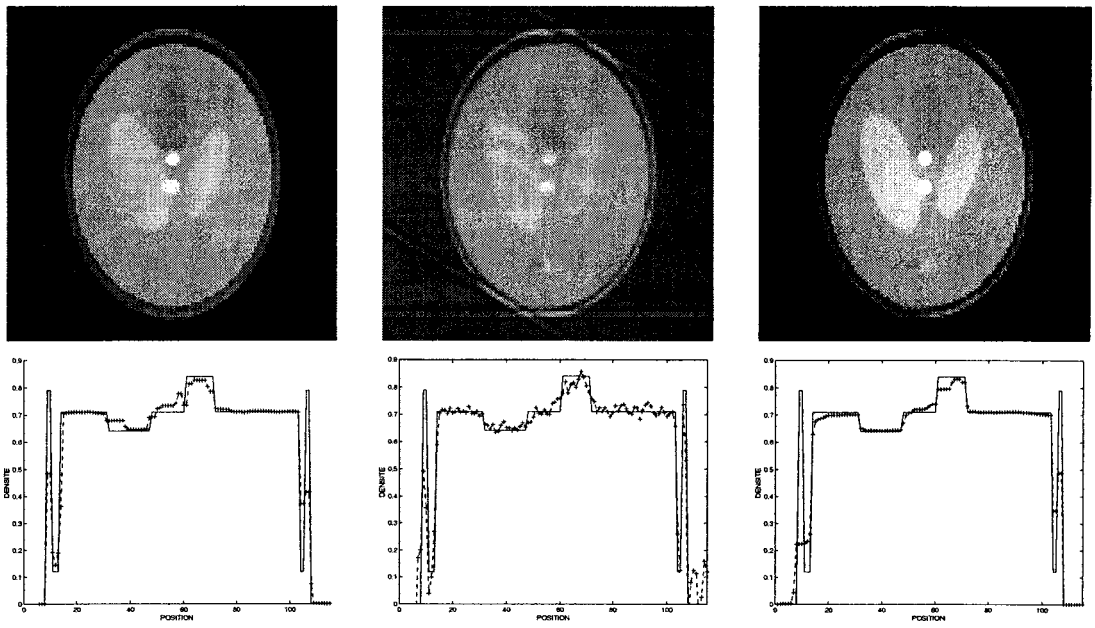
Figure 6.5: Cas ① et ② : reconstruction par approche pénalisée du fantôme à partir de projections obtenues pour une épaisseur de faisceau nulle et sans bruit (①) ; une épaisseur de faisceau égale à la largeur d'un pixel et un bruit additif (②) ou multiplicatif (②). Les paramètres α et s ont été ajustées manuellement de manière à obtenir le meilleur compromis visuel entre la résolution et le niveau de bruit.

de localiser les interfaces à un ou deux pixels près. Cependant, ces performances se dégradent sensiblement même pour des niveaux qui restent limités. Ainsi, les reconstructions reproduites sur la figure 6.4 ont été obtenues pour un bruit additif de moyenne nulle et d'écart type $\sigma = 0,008$ (l'effet d'un bruit multiplicatif augmenté dans les mêmes proportions produit un effet comparable). Sur la reconstruction du centre, il devient difficile de localiser les détails « intracraniens », et l'extraction d'une information quantitative est encore plus délicate. Si on souhaite accentuer les détails, l'augmentation de la bande passante accentue la contribution

des hautes fréquences au prix d'une image « sur bruitée » ; cf. reconstruction de droite. *A contrario*, la diminution de la bande passante produit une image plus régulière mais introduit inévitablement un lissage des contours combiné à un effet de Gibbs (un comportement oscillant) lié à la réponse du filtrage linéaire dans les transitions de fortes amplitudes ; cet effet n'est néanmoins pas visible sur ces exemples. On doit également souligner que l'instabilité s'accroît si l'échantillonnage angulaire ne couvre pas l'ensemble $[0;\pi[$. Cet effet est illustré sur la reconstruction par RPC de la figure 6.6 : pour 151 projections faiblement bruitées saisies dans $[0;5\pi/6]$ (soit 30 degrés manquants), la RPC conduit à des artefacts de reconstruction prononcés. *Ce problème doit attirer notre attention pour la suite dans la mesure où la reconstruction en géométrie hélicoïdale doit faire face à un problème assez comparable.*

Approche pénalisée de type $\ell_2\ell_1$

En guise de comparaison, une inversion pénalisée a été mise en œuvre dans les mêmes conditions expérimentales ; les résultats sont visibles sur la partie supérieure de la figure 6.5. Ces solutions ont été obtenues par minimisation *non contrainte* du critère pénalisé (6.3) ; les paramètres $\boldsymbol{\lambda} = (\alpha, s)$ ont été ajustés manuellement. Pour le contexte expérimental ①, l'opérateur $\boldsymbol{R}$ adopté correspond à une paramétrisation de la scène à base d'indicatrices sur les pixels ; pour les autres contextes expérimentaux, la paramétrisation à base de disques à été retenue afin de permettre la prise



Cas ③ ($\ell_2\ell_1$) **Cas ④ (RPC)** **Cas ④ ($\ell_2\ell_1$)**

Figure 6.6: Cas ③ : limite de robustesse au bruit de la reconstruction pénalisée (bruit additif décorrélé, gaussien centré d'écart type $\sigma = 0,01$) ; cas ④ test de robustesse à une excursion angulaire incomplète : 151 projections produites uniformément dans $[0; 5\pi/6[$ perturbées par un bruit additif décorrélé, gaussien centré d'écart type $\sigma = 0,001$. La colonne du centre illustre les performances de la RPC (filtre de Hamming, $\nu_c = 0,8$), celle de droite illustre celles de l'approche pénalisée (modèle $\ell_2\ell 1$, $\alpha = 2,5.10^{-3}$; $s = 8.10^{-5}$).

en compte de l'épaisseur du faisceau.

Dans tous ces cas de figure, ces reconstructions se comparent positivement à celles obtenues par RPC. D'autre part, la robustesse de la méthode au bruit et à une excursion angulaire limitée est bien supérieure à ce qu'on peut attendre d'une reconstruction par RPC ; c'est en particulier ce qu'illustre les résultats présentés sur la figure 6.5 et 6.6. Néanmoins, on souligne que les temps d'exécution des deux méthodes ne sont absolument pas du même ordre : alors que la reconstruction par

SNARK d'une de ces images prenait environ une demi seconde, notre mise en œuvre de l'approche pénalisée nécessite une vingtaine de seconde.

6.3 Limitations de la méthode pénalisée

Toutes les méthodes possèdent leurs avantages et leurs inconvénients. Concernant les approches pénalisées, les dernières mises en œuvre nous montrent qu'un gain substantiel de qualité peut être obtenu au prix de certaines limitations ; nous en avons dénombrées principalement trois,

1. L'erreur de discrétisation qu'entraîne le modèle,
2. le coût de mise en œuvre,
3. le réglage des paramètres de régularisation.

Nous examinons maintenant ces trois difficultés dans cet ordre.

Limitation liées au modèle d'observation

On considère maintenant le fantôme introduit par HERMAN dans sa monographie [Herman, 1980] et qui est représenté sur la figure 6.7. Ce fantôme représente une coupe tomographique au niveau du crâne, les niveaux d'atténuation ayant été attribués en ce sens : il est constitué d'une fine zone elliptique de forte atténuation (la partie osseuse du crâne) qui entoure une zone centrale d'atténuation beaucoup plus faible où apparaissent des inhomogenéités au support relativement restreint ;

l'objectif est évidemment de pouvoir reconstruire correctement ces détails apparaissant au centre car ils correspondent aux caractéristiques médicalement significatives. Une représentation lisible de ces détails est fournie par l'image du centre de la figure 6.7 qui représente le même fantôme après un seuillage des atténuations supérieures à 0,22 cm^{-1} et inférieures à 0,1945 cm^{-1}.

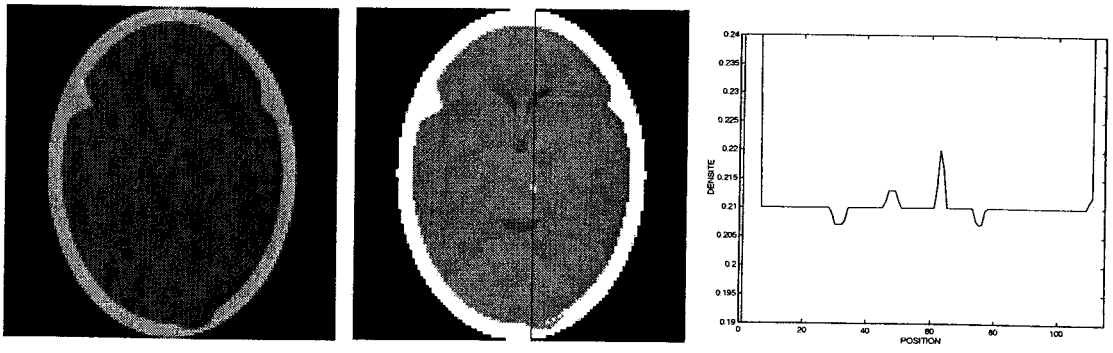


Figure 6.7: Fantôme de HERMAN utilisé pour les mises en œuvre (←) et sa version seuillée de manière à laisser apparaître les détails intracraniens (↑) ; tracé de la soixante troisième colonne de ce fantôme (→).

Mise en lumière de l'erreur de pixelisation

Pour une mise en œuvre dans le contexte de simulation ① (*i.e.* faisceau d'épaisseur nulle, données non bruitées), les reconstructions par RPC et pénalisées sont visibles sur la figure 6.8 ; la RPC a été mise en œuvre avec un filtre cosinus et une fréquence de coupure réduite de $\nu_c = 1$; la reconstruction pénalisée a utilisé un modèle $\boldsymbol{R}$ déduit d'une paramétrisation de la scène à base d'indicatrice sur les

pixels — c.à.d. des « pixels carrés ».

Le constat est flagrant : la reconstruction pénalisée est d'une qualité très inférieure à ce que produit la RPC avec les même données. En fait, l'explication de ce phénomène relève d'une inadéquation entre les données et le modèle $\boldsymbol{R}$ construit pour l'inversion. Il faut tout d'abord préciser que le fantôme utilisé sous SNARK est défini de manière *continue* ; les projections qui en découle résultent donc de l'échantillonnage d'une fonction continue. L'opérateur $\boldsymbol{R}$ adopté pour l'inversion pénalisée découle lui d'une modélisation de la scène à base de pixels : en comparant les données — *i.e.* les projections issues de SNARK — avec la sortie du modèle $\boldsymbol{R}$ obtenue avec l'*image pixelisée* 115×115 du fantôme, on constate de légères différences qui produisent des distorsions sur l'image reconstruite. De manière à confirmer cette conclusion, on a produit une reconstruction pénalisée à partir de projections obtenue pour un fantôme décrit par des pixels et non plus de manière continue. La reconstruction obtenue est alors de très bonne qualité comme illustré par l'image de droite de la figure 6.8.

Nous attirons néanmoins l'attention du lecteur sur le fait que ce type d'erreur n'a pas d'effet sensible sur les reconstructions du fantôme de SHEPP *et* LOGAN. *Ce constat est important dans la mesure où ce type d'image synthétique se rapproche nettement plus des images articulaires que nous souhaitons reconstruire.*

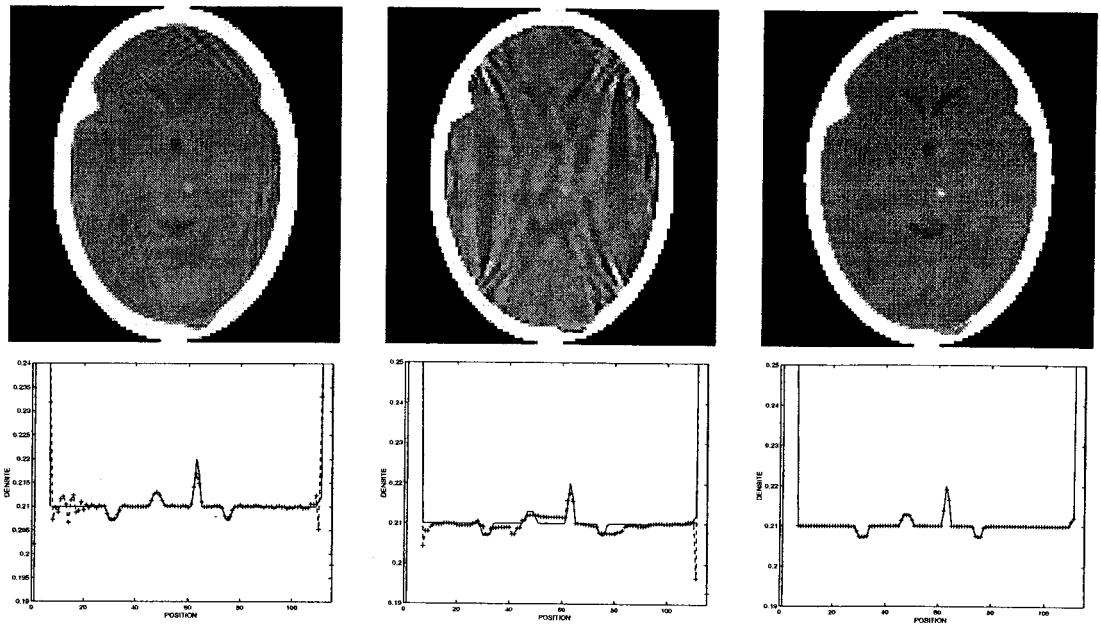


Figure 6.8: Cas ① : reconstruction du fantôme de la figure 6.7 par RPC (←), par approche pénalisée à partir des même projections (↑) et pour des projections produites pour le fantôme pixelisé.

Éléments de solution

Ce problème provenant d'une erreur de discrétisation, son effet doit devenir moins sensible à mesure que le nombre de pixels à reconstruire dans l'objet augmente. En ce sens, un résultat plus intéressant devrait être obtenu en utilisant le même jeu de données mais en reconstruisant une image plus fine. Cette démarche peut néanmoins être considérée inadéquate puisqu'elle conduit à une augmentation de la taille du problème de reconstruction. Sans appronfondir le sujet, on suggère néanmoins deux alternatives qui nous semble intéressantes :

1. en s'inspirant de [Soussen, 2000, Chap. II], on peut penser introduire une

grille irrégulière construite en fonction d'informations anatomiques connues au préalable, ou à défaut de façon adaptative ; l'idée directrice étant d'augmenter *localement* le nombre de pixels autour des *interfaces* de manière à limiter l'erreur de discrétisation qu'elles introduisent sans augmenter inconsidérément le nombre de paramètres à estimer.

2. Une autre approche s'inspire plutôt des travaux de [Gautier, 1996, Sec. 4.3] ou de [Fiani, 2001, Sec. III.4.2]) : elle consiste à modifier légèrement le critère (6.3) pour y ajoutant un *terme de rappel non quadratique* à une valeur de référence w, c.à.d. à minimiser

$$J(\boldsymbol{x}) = ||\boldsymbol{y} - \boldsymbol{R}\boldsymbol{x}||_2^2 + \alpha_1 \sum_{c=1}^{C} \sqrt{\delta_c^2 + s_1^2} + \alpha_2 \sum_{n=1}^{N} \phi_2(x_n - w); \qquad (6.4)$$

Dans (6.4), les paramètres $\alpha_1, \alpha_2 \geq 0$ ajustent la contribution des différentes parties de la pénalisation dans le résultat final ; la fonction ϕ_2 est par exemple un *compromis convexe* paramétré de la forme

$$\phi_2(x_n - w) = \sqrt{(x_n - w)^2 + s_2^2}, \qquad s_2 > 0$$

ou une autre fonction non quadratique (convexe ou non) présentée dans le en section 4.3.1 de ce mémoire. En réglant w à sa valeur de référence « intracranienne » de 0,21 cm^{-1}, et après ajustement des autres paramètres libres, on

peut espérer améliorer la qualité de la reconstruction.

D'autre part, on note que l'erreur de discrétisation est moins prononcée en considérant un problème de reconstruction plus réaliste qui prend en compte une épaisseur *non nulle* de faisceau. On rappelle que dans ce cadre expérimental, nous adoptons un modèle paramétrique de la scène « à base de disque » qui permet une prise en compte aisée de l'épaisseur du faisceau. Dès lors, la sortie de notre modèle (c.à.d. le produit de $\boldsymbol{R}$ et du fantôme décrit sous forme de pixels) se compare plus favorablement avec les données de projections utilisées pour l'inversion (c.à.d. les projections générées par SNARK pour un faisceau épais). Ce constat suggère que le raffinement du modèle $\boldsymbol{R}$ pourrait amener une réduction de ce type d'erreur.

Finalement, les modèles pixeliques largements répandus ne permettent pas toujours de construire des solutions pénalisées de meilleure qualité que la RPC. À cet égard, il est assez surprenant que ces erreurs de discrétisation ne soient pas vraiment documentées dans la littérature ; incontestablement, la compréhension de ce phénomène nécessite un examen plus complet.

Limitations liées à la taille du problème numérique

En reconstruction d'image, l'inversion pénalisée conduit à des problèmes d'optimisation de grande dimension et la méthode est finalement peu compétitive en regard des temps d'exécution d'une simple RPC. En fait, il faut bien avoir à l'esprit

que ces deux techniques sont plus complémentaires que concurrentes et que le choix de l'une ou l'autre dépend des impératifs de mise en œuvre et de précision.

La méthodologie suivie pour minimiser le critère a évidemment un impact non négligeable, et on va le voir par la suite, la conception d'un algorithme traitant efficacement la minimisation d'un critère pénalisé loin d'être aisée. *Cette difficulté est particulièrement préoccupante en géométrie hélicoïdale où le caractère intrinsèquement tridimensionnel du problème conduit à une augmentation sensible de la taille du problème de minimisation.*

La question du réglage des paramètres de régularisation

Régulariser nécessite de gérer un certain *compromis* entre la stabilisation du problème initial et l'introduction d'un certain biais dans la solution finale. En pratique, la recherche de ce compromis passe par le réglage des paramètres de régularisation (également appelés *hyperparamètres*), réglage qu'on peut aborder suivant deux angles distincts : l'approche *supervisée* nécessite l'intervention d'un opérateur qui, en s'appuyant sur son expérience propre, ajuste les paramètres par « essai-erreur » ; l'approche *non supervisée* estime la valeur à partir des données en s'appuyant sur des méthodes empiriques ou non.

Paramètres de régularisation associés à la RPC

Pour un algorithme d'inversion comme la RPC, les paramètres de réglages sont associés[3] au *filtre de reconstruction*, c.à.d. au choix du filtre et de sa fréquence de coupure. Bien que ces choix aient un impact significatif sur la qualité des reconstructions, leur ajustement ne constitue pas une question sensible en pratique :

1. tout d'abord, le contexte de mesure est suffisamment bien maîtrisé en milieu hospitalier pour permettre un calibrage de la méthode en fonction de la zone à imager ;
2. ensuite, le temps de reconstruction plutôt restreint permet d'adopter une stratégie de type essai/erreur.

Finalement, le problème du réglage des hyperparamètres de la RPC a donc été résolu assez efficacement par l'intervention d'un *opérateur* formé en partie à cet effet.

Hyperparamètres d'un critère pénalisé

La mise en œuvre d'une inversion pénalisée nécessite de fixer les paramètres « libres » intervenant dans l'expression du critère composite découlant de l'approche

[3]Notons également que la mise en œuvre d'une RPC nécessite de définir un type d'interpolation qui permet de ramener l'opération de rétroprojection sur la grille cartésienne que constitue l'image.

de TIKHONOV généralisé qu'on écrira sous forme générique

$$J(\boldsymbol{x};\boldsymbol{\lambda}) = \mathcal{Q}(\boldsymbol{y},\boldsymbol{R}\boldsymbol{x}) + \alpha\mathcal{P}(\boldsymbol{x};s). \tag{6.5}$$

de manière à simplifier les notations, on rassemble tous les hyperparamètres sous la notation $\boldsymbol{\lambda} = (\alpha, s)$, soit

(*i*) les paramètres du modèle $\mathcal{P}$ sont rassemblés sous la notation s ; le nombre de ces paramètres croît généralement avec la complexité du modèle introduit ;

(*ii*) le paramètre α ajuste le poids du modèle $\mathcal{P}$ dans la solution.

La solution de l'inversion sera alors notée $\boldsymbol{x}_{\boldsymbol{\lambda}}$ de manière à faire apparaître explicitement cette dépendance dans la solution.

En pratique, la qualité des reconstructions varie souvent sensiblement an fonction des hyperparamètres. Même si le critère est convexe[4], le coût de mise en œuvre peut être conséquent et rendre une procédure de réglage par « essai/erreur » lourd et fastidieux. Il apparaît donc légitime de rechercher une méthode non supervisée pour mettre en œuvre une reconstruction pénalisée.

[4]On rappelle que les critères convexes permettent d'assurer simultanément (1) la continuité de la solution en fonction des hyperparamètres, et (2) la convergence vers le minimiseur pour des techniques algorithmiques « simples » comme les algorithmes à direction de descente ; cf. page 98.

Approche non supervisée et cadre déterministe

Dans le cadre déterministe de ce chapitre, il existe finalement peu de méthodes permettant d'estimer les hyperparamètres à partir des données, et à notre connaissance, leur intérêt se limite au cas des critères quadratique dépendant d'un seul hyperparamètre

$$J_{\ell_2}(\boldsymbol{x}) = ||\boldsymbol{y} - \boldsymbol{R}\boldsymbol{x}||_2^2 + \alpha||\boldsymbol{D}\boldsymbol{x} - \boldsymbol{a}||_2^2 \qquad \boldsymbol{a} \in \mathbb{R}^M. \tag{6.6}$$

avec $\boldsymbol{D}$ un opérateur linéaire — ex. une matrice calculant les différences finies d'ordre k dans l'image.

Dans ce cadre quadratique, la *validation croisée généralisée* donne généralement de bons résultats ; nous renvoyons le lecteur à [Golub *et al.*, 1979] pour un exposé de la méthode ainsi qu'à [Fortier *et al.*, 1993] pour une utilisation en traitement d'images. Toujours pour le cas quadratique, *la méthode de la courbe en L* permet également d'obtenir des résultats satisfaisants : cette méthode simple préconise de sélectionner la valeur de λ produisant la courbure maximale sur un tracé en échelle log-log de la courbe paramétrée en λ définie par

$$(||\boldsymbol{D}\boldsymbol{x}_\lambda - \boldsymbol{a}||_2, ||\boldsymbol{R}\boldsymbol{x}_\lambda - \boldsymbol{y}||_2)\,;$$

on pourra se reporter à [Hansen, 1992*a*], et à [Calvetti *et al.*, 2001] pour des détails

de mise en œuvre.

Approche non supervisée et cadre probabiliste

Comme nous l'avons vu au chapitre 5, le cadre de l'inférence bayésienne permet souvent de construire une loi de probabilité dite *loi a posteriori* à partir du critère pénalisé (6.5). Ce cadre rend possible la construction d'estimateurs sur les paramètres de la loi. On soulignera néanmoins que les approches adoptées dans ce cadre soulèvent des difficultés méthodologiques et pratiques assez conséquentes ; voir par exemple [Descombes et Goussard, 2001, Sec. 8.4]. D'une part, les algorithmes mis en œuvre sont en général d'un coût algorithmique prohibitif, ce qui rend délicat leur emploi en imagerie 3D. D'autre part, les résultats d'estimation de paramètres qu'ils produisent pour des images réelles ne font pas l'unanimité dans la communauté ; cf. [Descombes *et al.*, 1999], voir également à ce propos [Idier, 2000, Sec. 5.3]. Des travaux récents méritent néanmoins d'être signalés et des références complémentaires sont données en section 5.3.2 de ce mémoire.

Le choix de la mise en œuvre supervisée

Dans le cadre de ce projet, une inversion pénalisée de type *supervisé* nous paraît plus opportune ; ce choix mérite néanmoins d'être commenté et justifié :

1. Le coût calculatoire élevé des méthodes non supervisées rend leur utilisation en imagerie 3D encore délicate. La pertinence d'un réglage non supervisé

n'étant pas garantit, l'investissement nécessaire pour concevoir et mettre en œuvre de telles méthodes reste difficile à justifier.

2. Pour une application cherchant à localiser précisément les contours osseux, nous avons de bonne raisons de penser qu'il est possible de *calibrer* les paramètres de régularisation. Cette assertion mérite néanmoins d'être confirmée de manière expérimentale.

3. En milieu hospitalier, la mise en œuvre de la RPC fait déjà intervenir un opérateur. En supposant que le réglage supervisé peut s'appuyer sur une calibration préalable (cf. point 2), la mise en œuvre d'une reconstruction pénalisée peut bénéficier d'un opérateur supervisant le réglage.

4. Pour l'inversion pénalisée que nous choisirons de mettre en œuvre, le nombre d'hyperparamètres à ajuster est limité à deux, ce qui reste relativement restreint et permet d'envisager un ajustement manuel à partir des valeurs calibrées.

Le second argument repose sur un raisonnement intuitif qui nous paraît néanmoins assez solide. On notera d'abord que les valeurs « correctes » des hyperparamètres correspondent, dans notre cas, à un lissage correcte du bruit et à une extraction correcte des contours osseux. Or, les tomographes hospitalier sont utilisés dans un contexte rendu très uniforme[5] (rapport signal à bruit, nombre de

[5]Ceci même en tomographie hélicoïdale puisqu'il est possible de diminuer le pas de l'hélice si celui-ci ne permet pas d'obtenir des reconstructions probantes.

vue, *etc*). D'autre part, l'application orthopédique dans laquelle on s'inscrit — reconstruction précise de l'interface tissus mous/tissus osseux — contribue à cette uniformisation puisque les atténuations de l'os sont assez stationnaires en général et toujours bien supérieurs aux atténuations des autres tissus ; cf. le tableau de la figure 2. Finalement, le contexte global de mise en œuvre doit contribuer à rendre « stationnaires » les valeurs des hyperparamètres entre différents patients.

6.4 En résumé

L'apport des approches pénalisées en reconstruction tomographiques est maintenant largement admis : elles permettent un gain sensible de résolution et de rapport signal à bruit comparativement aux approches standard comme la RPC. En dépit de certaines limitations soulignées en section 6.3 de ce chapitre, leur emploi dans le cadre du projet d'implant personnalisé nous semble justifié. Ce point de vu est en particulier étayé par les deux remarques suivantes :

1. l'expérience que nous avons accumulée semble montrer que l'*erreur de discrétisation* ne dégrade pas significativement les images constituées de larges zones uniformes ; ce sont justement les images que nous rencontrerons dans notre cadre applicatif à vocation orthopédique — cf. section 6.2 ;
2. les contraintes de temps d'exécution ne sont pas prioritaires pour notre application.

Dans le chapitre suivant, nous allons donc étendre cette méthodologie à la tomographie hélicoïdale et tenter une analyse de la contribution de cette démarche vis-à-vis des méthodologies standard.

Bibliographie annotée

L'emploi de potentiels quadratiques en imagerie est initié dans les années 1970 avec notamment [Hunt, 1973] ; la première utilisation en reconstruction tomographique semble être [Herman et Lent, 1976]. Les pénalisations non convexes en imagerie médicale ont été introduites au cours des années 1980 avec notamment [Geman et McClure, 1987*b*] ; par la suite, diverses fonctions de coût non convexes ont été employées en reconstruction tomographique de transmission ou d'emission dans un cadre 2D ou éventuellement 3D [Charbonnier *et al.*, 1997; Delaney et Bresler, 1998; Geman et Yang, 1995; Green, 1990; Hebert et Leahy, 1989].

L'emploi des compromis convexes date approximativement du début des années 1990 avec notamment [Bouman et Sauer, 1993] qui traitent d'une application en tomographie de transmission ; les mêmes auteurs ayant également travaillés au raffinement des modèles d'observation en tomographie d'emission et de transmission [Bouman et Sauer, 1996; Sauer et Bouman, 1993]. Dans ce domaine, les compromis convexes ont fait leurs preuves et ont notamment démontrés une bonne robustesse au problème à angles de vue restreints ; voir par exemple [Delaney et Bresler, 1998].

Par conséquent, une bonne partie des contributions récentes travaillent sur les problématiques d'implantation ; nous aurons l'occasion de donner un certain nombre de contributions allant en ce sens dans les prochains chapitres.

Enfin, on souligne que d'autres approches peuvent être adoptées pour obtenir une image tomographique à partir d'un critère pénalisé. Une démarche très similaire consiste à considérer le tomographe comme un système linéaire invariant et à effectuer une *restauration* des images produites par la RPC ; cette approche a été mise en œuvre par [Villain *et al.*, 2001] avec un certain succès. Indiquons également qu'un certain nombre d'auteurs choisissent une approche par « contours » plutôt que par pixel ; on pourra se reporter à [Soussen, 2000] pour une étude comparative entre les deux types d'approches.

CHAPITRE 7

TOMOGRAPHIE EN GÉOMÉTRIE HÉLICOÏDALE

Note : une partie du contenu de ce chapitre a été publiée dans [Allain *et al.*, 2001].

7.1 Introduction

En tomographie 3D, un volume d'intérêt est traditionnellement produit par empilement d'une série de plans préalablement reconstruites. En *géométrie axiale*, ces coupes sont indépendamment reconstruites à partir de jeux de projections saisies dans des plans perpendiculaires à l'axe du tomographe. Bien que généralement satisfaisante pour le diagnostic médical, la précision des volumes reconstruits est souvent trop faible pour les applications métrologiques. Cette lacune, largement constatée dans la littérature [Green, 1990; Sauer et Bouman, 1993; Villain *et al.*, 2001], est en partie imputable à la méthode de reconstruction axiale utilisée : l'algorithme de *rétroprojection convoluée* (abrég. RPC).

Omniprésent il y a encore 10 ans en milieu hospitalier, le mode axial a été massivement remplacé par un échantillonnage *hélicoïdal* du volume d'intérêt : le volume est imagé plus rapidement, ce qui permet d'améliorer la résolution temporelle du système d'imagerie et de diminuer la dose de rayonnement ionisant administrée au patient. Ces gains s'accompagnent souvent d'une dégradation significative de la

précision des reconstructions et de l'apparition éventuelle d'artefacts. Une analyse détaillée des méthodologies de reconstruction standard en tomographie hélicoïdale nous conduisent à penser que cette perte de qualité des reconstructions est en partie liée à la présence d'heuristiques[1] introduites à la seule fin de tirer partie de la RPC. Le lecteur trouvera au chapitre 3 de ce document un certain nombre d'arguments étayant cette conjecture.

À la suite des développements du chapitre précédent, ce chapitre montre qu'une inversion pénalisée en tomographie hélicoïdale peut être développée de manière à s'affranchir le plus possible des heuristiques et de la RPC. Ces approches pénalisées ont été comparées sur la base de données synthétiques avec les approches standard : des résultats intéressant laissent penser que cette démarche conduit à une amélioration sensible de la qualité des reconstructions en géométrie hélicoïdale.

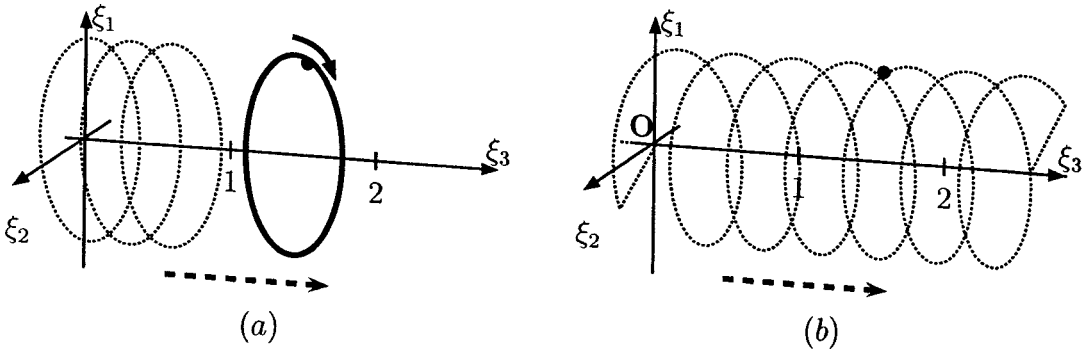


Figure 7.1: Trajectoires de la source en mode axiale (←) et en mode hélicoïdal (→).

[1]On interpole les projections sur l'hélice de manière à composer des « pseudo jeux » de projections dans des plans axiaux prédéfinis ; ce sont ces plans qui seront reconstruits.

7.2 Méthodologie

Inscrivons le volume d'intérêt dans un système d'axes 3D (ξ_1, ξ_2, ξ_3), l'axe du tomographe étant l'axe $O\xi_3$. On s'appuiera sur une modélisation paramétrique du volume d'intérêt x ; on pose alors

$$x(\xi_1, \xi_2, \xi_3) \approx x_N(\xi_1, \xi_2, \xi_3) = \sum_{n=1}^{N} x_n b_n(\xi_1, \xi_2, \xi_3),$$

avec $x_n \in \mathbb{R}$ pour tout $n \in \{1, \cdots, N\}$. L'ensemble $\{b_n\}_{n=1}^{N}$ des « fonctions de base » consiste en l'extension 3D des exemples présentées en section 2.2.1 dans le plan. Dans notre cas, nous choisissons une base d'indicatrice rectangulaire ou cylindrique sur les voxels — des voxels « rectangulaires » ou « cylindriques ». Dans ce cadre, le problème de reconstruction de l'image 3D se ramène à l'estimation du vecteur $\boldsymbol{x} \in \mathbb{R}^N$ rassemblant les N paramètres de la modélisation de la scène. Par la suite, on posera $\boldsymbol{x} = \{\boldsymbol{x}_t \in \mathbb{R}^L\}_{t=1}^{T}$ avec T le nombre de plans de voxels décrivant le volume d'intérêt.

Modèle d'observation

Dans la suite, on note $\boldsymbol{p}_h \in \mathbb{R}^M$ le vecteur qui rassemble l'ensemble des projections en géométrie hélicoïdale. Notre démarche s'inscrit dans une approche « problème inverse » de la reconstruction tomographique ; en ce sens, on cherchera à inverser un modèle d'observation $\mathcal{H} : \mathbb{R}^N \to \mathbb{R}^M$ de tomographie hélicoïdale qui

permet d'écrire $\boldsymbol{p}_h = \mathcal{H}(\boldsymbol{x})$, et on supposera que les mesures $\boldsymbol{y}_h \in \mathbb{R}^M$ qui sont effectivement à notre disposition s'écrivent

$$\boldsymbol{y}_h = \mathcal{H}(\boldsymbol{x}) + \boldsymbol{\varepsilon}.$$

où $\boldsymbol{\varepsilon} \in \mathbb{R}^M$ représente les différentes erreurs liées aux bruits d'instrumentation ou de modèle.

Méthodologie de reconstruction

La reconstruction tomographique étant un problème d'inversion mal posé, on prendra soin de *régulariser* le problème initial par une pénalisation (ou modèle d'image) $\mathcal{P}$ soigneusement choisi. Dans ce cadre, le volume reconstruit est obtenu en résolvant un problème d'optimisation non contraint,

$$\boldsymbol{x}_{\boldsymbol{\lambda}} \in \{\boldsymbol{x} \in X : \min J(\boldsymbol{x}\,;\,\boldsymbol{\lambda})\} \tag{7.1}$$

où $X \subseteq \mathbb{R}^N$ est supposé fermé et convexe, et où $J : \mathbb{R}^N \to \mathbb{R}$ est un critère pénalisé de la forme

$$J(\boldsymbol{x}\,;\,\boldsymbol{\lambda}) = ||\mathcal{H}(\boldsymbol{x}) - \boldsymbol{y}_h||_2^2 + \alpha\mathcal{P}(\boldsymbol{x}\,;\,s) \tag{7.2}$$

où $\boldsymbol{\lambda} = (\alpha \geq 0, s)$ rassemble les paramètres libres qui doivent être ajustés pour la mise en œuvre. On note que $\boldsymbol{x}_{\boldsymbol{\lambda}}$ résulte d'un compromis entre un modèle d'image

régularisant l'inversion et une mesure d'adéquation aux données $\boldsymbol{y}_h$. Comme pour le cadre axial, une adéquation aux données différente de la mesure quadratique $||\mathcal{H}(\boldsymbol{x}) - \boldsymbol{y}_h||_2^2$ pourrait être adoptée — voir par exemple la section 6.1 ; de manière à privilégier la clarté de l'exposé, on adoptera néanmoins un critère de la forme (7.2).

Les difficultés méthodologiques rencontrées lors de la résolution de (7.1) dépendent intimement des spécificités du modèle d'observation et du modèle d'image retenus. Les deux prochaines sections montrent que, sous des hypothèses usuelles, ces deux modèles peuvent être structurellement simples.

7.2.1 Régularisation $\ell_2\ell_1$ 3D

Le modèle d'image choisi doit permettre de stabiliser l'inversion tout en favorisant l'apparition de caractéristiques attendues dans le volume reconstruit ; ainsi, la pertinence d'une pénalisation $\mathcal{P}$ relève en partie du domaine d'application. Dans le cadre de ce projet, les éléments à imager sont typiquement constitués de larges zones 3D homogènes séparées par des interfaces franches. D'autre part, puisqu'on cherche avant tout *une reconstruction précise des frontières* entre tissus mous et tissus osseux, on adoptera un modèle d'image 3D qui pénalise les irrégularités locales tout en préservant généralement bien les discontinuités. Plus spécifiquement, on adoptera une extension simple du modèle convexe 2D introduit au chapitre

précédent :

$$\mathcal{P}(\boldsymbol{x}) = \sum_{c=1}^{C} \phi(\delta_c\,;\, s) \quad \text{avec} \tag{7.3a}$$

$$\phi(\cdot\,; s) : u \to \sqrt{u^2 + s^2}\;; \tag{7.3b}$$

où $s > 0$ est un paramètre du modèle. δ_c représente la différence entre voxels constituant la c-ème paire d'éléments mutuellement voisins. Comme l'illustre la figure 7.2, on supposera que chaque voxel possède (effets de bord exclus) 10 voisins : soit 8 voxels dans le plan courant t, et un voxel dans chacun des plans $t-1$ et $t+1$.

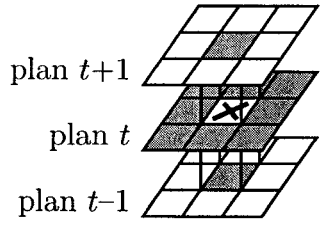


Figure 7.2: Système de voisinage 3D retenu : le voxel courant (croix) interagit avec 8 voisins dans le plan de l'image et 2 voisins dans l'axe du tomographe.

De tels voisinages 3D ont été employés en reconstruction tomographie axiale de transmission [Villain *et al.*, 2001] ou en tomographie d'émission [Hebert et Leahy, 1989]. Soulignons qu'une telle pénalisation est convexe, coercive et continûment différentiable (C^1) ; ces propriétés permettent de ne pas alourdir inconsidérément le volume de calcul nécessaire pour minimiser le critère (7.2).

7.2.2 Modèle d'observation en géométrie hélicoïdale

En géométrie *axiale*, on montre aisément que la paramétrisation de la scène permet la construction d'un opérateur linéaire $\boldsymbol{R}$ creux et de grande taille reliant les projections dans le plan et les pixels de l'image 2D [Herman, 1980, Chap. 6] ; ainsi, à une projection d'angle φ^j donnée correspond une sous matrice $\boldsymbol{R}^j$ de $\boldsymbol{R}$. Nous montrons maintenant que ce modèle peut être facilement étendu à la géométrie hélicoïdale. Ce faisant, *l'opérateur d'observation* $\mathcal{H}$ *sera une matrice* $\boldsymbol{H}$ *dont la taille et le caractère creux sont sensiblement accentués par rapport au modèle en géométrie axiale* $\boldsymbol{R}$. *On montrera également qu'une forme partiellement paramétrée de ce modèle peut être adoptée sans perte significative de généralité : le coût de stockage de* $\boldsymbol{H}$ *peut alors être réduit dans des proportions souvent importantes.*

Formulation du modèle d'observation

Introduisons l'angle de projection $\bar{\varphi}$ défini sur $\mathbb{R}$ dont le *modulo* 2π est φ, angle de projection dans le plan de l'image $\xi_1 O \xi_2$. Il importe de souligner qu'un angle $\bar{\varphi} = \bar{\varphi}^i$ définit de manière univoque un couple $(\varphi = \varphi^i, \xi_3 = \xi_3^i)$; i.e. une projection d'angle $\bar{\varphi}^i$ définit une et une seule projection *axiale* dans le volume. L'ensemble des angles $\{\bar{\varphi}^i \in \mathbb{R}\}_{i=1}^{N_p}$ collectés sur l'hélice est fixé par le contexte de mesure (pas de l'hélice, pas d'échantillonnage) ; la discrétisation du volume d'intérêt (extension du volume, épaisseur d'une tranche) va déterminer les angles de projection associés à chacune des tranches.

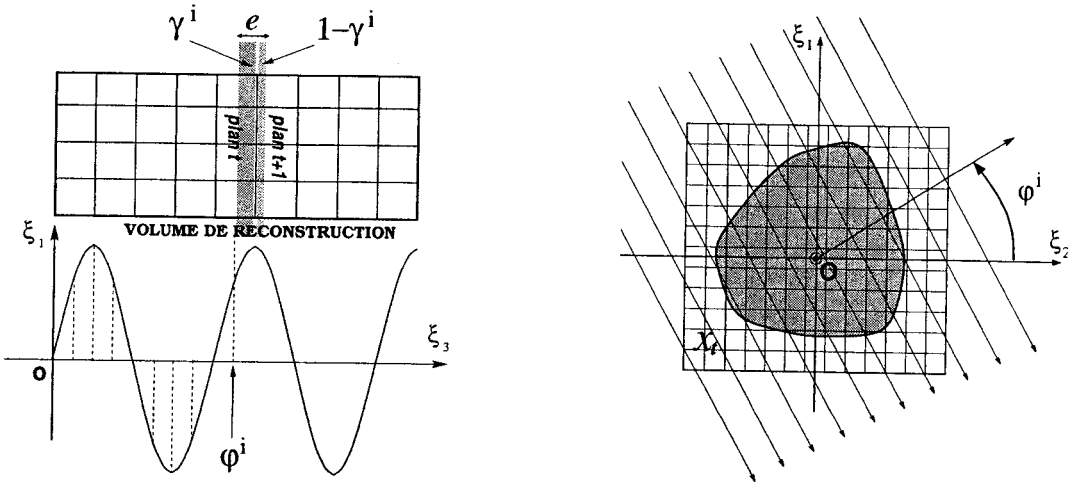


Figure 7.3: Projection en géométrie hélicoïdale dans le plan $\xi_2 O \xi_3$ ($\leftarrow$) et dans le plan $\xi_1 O \xi_2$ ($\rightarrow$).

Pour ce modèle, nous prenons en considération l'extension du faisceau e suivant l'axe $O\xi_3$ dans le modèle ; cette extension e est usuellement de l'ordre du côté d'un voxel ou légèrement inférieure. On notera alors qu'une projection d'angle $\bar{\varphi}^i$ met en jeu *au plus* deux plans de voxels mitoyens $\boldsymbol{x}_t$ et $\boldsymbol{x}_{t+1}$; ces deux plans contribuant pour une proportion de $\gamma^i \in (0;1]$ et $1-\gamma^i \equiv \bar{\gamma}^i$ dans la projection — *i.e.* γ^i représente la proportion d'épaisseur de faisceau chevauchant le t-ème plan de voxel du volume. La figure 7.3 illustre la situation que nous venons de décrire.

Rassemblons les projections associées aux plans $(\boldsymbol{x}_t, \boldsymbol{x}_{t+1})$ sous le vecteur $\boldsymbol{p}_t$, et posons $\boldsymbol{\varphi}_k = \{\varphi_t^j \in [0;2\pi)\}_{j=1}^{N_t}$; il est alors possible d'écrire

$$\boldsymbol{p}_t = \boldsymbol{H}_t \begin{pmatrix} \boldsymbol{x}_t \\ \\ \boldsymbol{x}_{t+1} \end{pmatrix} \quad \text{avec} \quad \boldsymbol{H}_t = \begin{pmatrix} \gamma_t^1 \boldsymbol{R}_t^1 & \bar{\gamma}_t^1 \boldsymbol{R}_t^1 \\ \vdots & \vdots \\ \gamma_t^{N_t} \boldsymbol{R}_t^{N_t} & \bar{\gamma}_t^{N_t} \boldsymbol{R}_t^{N_t} \end{pmatrix} \tag{7.4}$$

où γ_t^j représente la proportion de faisceau engagé dans le plan de voxels $\boldsymbol{x}_t$ lors de la projection d'angle φ_t^j. $\boldsymbol{R}_t^j \in \mathbb{R}^{R\times L}$ représente[2] l'opérateur de *projection axial* d'angle φ_t^j. On notera néanmoins que $\boldsymbol{H}_T$ doit être adapté pour tenir compte de « l'effet de bord ». La relation (7.4) conduit à modéliser le problème d'observation en tomographie hélicoïdale par la relation linéaire suivante :

$$\boldsymbol{p}_h = \mathcal{H}(\boldsymbol{f}) = \boldsymbol{H}\boldsymbol{x}, \tag{7.5}$$

où

$$\boldsymbol{p}_h = \begin{pmatrix} \boldsymbol{p}_1 \\ \vdots \\ \boldsymbol{p}_T \end{pmatrix}, \qquad \boldsymbol{H} = \begin{pmatrix} \boxed{\boldsymbol{H}_1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boxed{\boldsymbol{H}_2} & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \boxed{\boldsymbol{H}_T} \end{pmatrix} \in \mathbb{R}^{M\times N}. \tag{7.6}$$

Un des intérêts de ce modèle est qu'il est principalement construit à partir des sous matrices extraites d'un opérateur de projections dans le plan ; il est donc possible d'utiliser un code développé en géométrie axiale pour le générer[3]. Comparativement au modèle axial, ce modèle hélicoïdal est néanmoins intrinsèquement tridimensionnel : le recouvrement entre blocs contigus est de L colonnes (le nombre

[2]Pour une modélisation de la scène à base d'indicatrices sur les pixels, l'entrée (r, l) de $\boldsymbol{R}_t^j$ représente la contribution du pixel $l \in \{1, \dots, L\}$ dans le rayon $r \in \{1, \dots, R\}$ de la projection courante.

[3]C'est effectivement la démarche que nous avons adoptée : après avoir développé un code permettant de construire des opérateurs d'observation $\boldsymbol{R}$ dans le cadre axial, nous avons utilisé ce code pour générer des opérateurs dans le cadre hélicoïdal.

de voxels dans un plan d'image) et les matrices $\boldsymbol{H}_t$ sont en général *distinctes* les unes des autres.

Le stockage de l'opérateur $\boldsymbol{H}$ devient vite délicat à mesure que le nombre d'inconnues et de mesures augmente. En guise d'indicateur, considérons une situation assez fréquemment rencontrée en pratique : si le pas de l'hélice est égale à l'épaisseur d'un plan de voxel[4], le nombre d'entrées non nulles dans $\boldsymbol{H}$ est deux fois le nombre T de plans de voxels multipliant le nombre d'éléments non nuls dans un opérateur axial $\boldsymbol{R}$ qui serait associé à la rotation complète. Dans ce cas, la reconstruction d'une image de quelques dizaines de plans de 512×512 voxels conduit à un opérateur $\boldsymbol{H}$ qui rassemble typiquement plusieurs centaines de millions d'éléments non nuls. Dans la plupart des cas, le stockage *tel quel* de $\boldsymbol{H}$ est donc difficilement envisageable, et nous proposons maintenant un moyen de contourner cette difficulté en montrant qu'une forme *paramétrée* de $\boldsymbol{H}$ peut être adoptée sans approximation du modèle.

Modèle d'observation paramétré

En général, le modèle d'observation présenté ci-dessus ne comporte aucune invariance spatiale permettant de réduire l'espace de stockage requis : les blocs $\boldsymbol{H}_t$ sont distincts et ne présentent pas de structure paramétrique (comme par exemple

[4]Remarquons que dans cette configuration, la dose administrée au patient est identique à celle d'un examen axial de type « stop and go » car chaque plan d'image se voit associé à une rotation complète.

les matrices circulantes ou Toeplitz). Il est néanmoins possible de retrouver une certaine invariance suivant l'axe $O\xi_3$ si on suppose que *le pas P de l'hélice est un multiple entier de l'épaisseur e d'un plan de voxel.* Dans ce cas, seul les P/e première matrices[5] $\boldsymbol{H}_t$ suffissent à décrire complètement $\boldsymbol{H}$ qui s'écrit alors

$$\boldsymbol{H} = \begin{pmatrix} \boxed{\widetilde{\boldsymbol{H}}_r} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boxed{\boldsymbol{H}_r} & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \boxed{\widehat{\boldsymbol{H}}_r} \end{pmatrix}, \tag{7.7}$$

où $\boldsymbol{H}_r$ regroupe sous une structure identique à celle de l'égalité (7.6) un ensemble de $K = P/e$ matrices $\{\boldsymbol{H}_t\}_{t=1}^{K \leq T}$ définissant l'opération de projection sur un tour ; $\widetilde{\boldsymbol{H}}_r$ et $\widehat{\boldsymbol{H}}_r$ diffèrent de $\boldsymbol{H}_r$ pour tenir compte de l'extension finie du volume de reconstruction[6]. Cette structure « quasi » circulante par blocs (le recouvrement étant toujours de L éléments) permet en général une diminution sensible du coût de stockage de l'opérateur $\boldsymbol{H}$ puisque seul le modèle $\boldsymbol{H}_r$ décrivant l'observation sur un tour est nécessaire (effets de bords mis à part). *Enfin, on souligne qu'une*

[5]Ce nombre peut encore être réduit d'un facteur 8 par le jeu des symétries : un facteur 2 s'obtient facilement par le caractère impaire de la transformée de RADON, et la symétrie du plan de l'image composée par une grille rectangulaire régulière permet de déduire encore d'un facteur 4. Finalement, seul le stockage des projections obtenues sur 1/8-ème de tour est nécessaire. Par la suite, nous nous limitons uniquement à la forme paramétrée sans tenir compte de ces gain potentiels.

[6]En particulier, $\widehat{\boldsymbol{H}}_r$ tient de l'effet de bord dû à la dernière tranche ; d'autre part, si T n'est pas un multiple de P (i.e. la dernière rotation n'est pas complète), $\widehat{\boldsymbol{H}}_r$ comporte moins de P blocs matriciels.

fois les données de projection disponibles, la discrétisation du volume d'intérêt peut être ajustée de manière à faire apparaître cette invariance du modèle d'observation ; en ce sens, cette simplification ne porte pas réellement atteinte à la généralité du modèle.

Par la suite, nous avons largement tiré parti de cette paramétrisation de $\boldsymbol{H}$ pour mettre en œuvre des reconstructions hélicoïdales pénalisées sur des images 3D importantes ; en particulier, ceci a permis d'obtenir les reconstructions du volume $127 \times 127 \times 40$ de la section 7.3 de ce document.

7.2.3 Critère pénalisé pour la tomographie hélicoïdale

L'inversion pénalisée en tomographie hélicoïdale peut être abordée au travers d'une simple extension du cadre tomographique axial. Plus précisément, on cherchera à minimiser un critère pénalisé

$$J(\boldsymbol{x}\,;\,\boldsymbol{\lambda}) = \|\boldsymbol{y}_h - \boldsymbol{H}\boldsymbol{x}\|^2 + \alpha \sum_{c=1}^{M} \sqrt{\delta_c^2 + s^2}. \tag{7.8}$$

de structure identique à celui développé dans le cadre du chapitre précédent ; cf. relation (6.3). Cet objectif est au moins C^1, convexe et coercif, et moyennant la convexité de X, le minimiseur de (7.8) est unique et borné

$$\boldsymbol{x}_{\boldsymbol{\lambda}} = \underset{\boldsymbol{x} \in X}{\arg\min}\, J(\boldsymbol{x}\,;\,\boldsymbol{\lambda}). \tag{7.9}$$

Bien que ce minimiseur n'ait pas de forme explicite, on assure aisément la convergence d'algorithmes s'appuyant sur les propriétés locales de J — ex. si $X = \mathbb{R}^N$, un simple algorithme de « plus profonde descente » avec recherche linéaire inexacte suffit.

Finalement, le passage en géométrie hélicoïdale se distingue nettement par sa difficulté d'implantation : alors que la tomographie axiale 3D peut bénéficier du caractère séparable de son opérateur d'observation, le cadre hélicoïdal est intrinsèquement 3D et conduit à une très forte augmentation de la taille du problème à considérer. Dans ces conditions, l'implantation d'algorithmes possédant de bonnes propriétés de convergence (ex. quasi NEWTON) devient vite difficile, et l'objectif consiste à trouver un juste compromis entre une maîtrise du coût d'implantation et une bonne vitesse de convergence ; c'est dans cet esprit que nous avons développé nos algorithmes de reconstruction dont le détail sera exposé plus loin dans ce chapitre.

Mise en œuvre approchée : reconstruction séparée des plans 2D

Avant de présenter les résultats d'inversion obtenus sur données synthétiques, nous développons brièvement une approximation de l'inversion pénalisée (7.8) qui permet de reconstruire séparément les différents plans du volume d'intérêt. L'objectif est ici de sacrifier en partie la précision propre à l'inversion régularisée pour favoriser l'efficacité d'implantation.

En partant de (7.4) et (7.6), on en déduit l'approximation suivante

$$\overline{\boldsymbol{p}}_t \approx \overline{\boldsymbol{H}}_t\, \overline{\boldsymbol{x}}_t \qquad \text{avec} \qquad \overline{\boldsymbol{x}}_t = \begin{pmatrix} \boldsymbol{x}_{t-1} \\ \boldsymbol{x}_t \\ \boldsymbol{x}_{t+1} \end{pmatrix} \tag{7.10}$$

et où $\overline{\boldsymbol{H}}_t$ rassemble sous une structure telle que (7.6) les blocs $\{\boldsymbol{H}_{t-1}, \boldsymbol{H}_t\}$, et $\overline{\boldsymbol{p}}_t$ rassemble les projections impliquant directement le t-ème plan d'intérêt — c.à.d. $\boldsymbol{p}_{t-1}$ et $\boldsymbol{p}_t$:

$$\overline{\boldsymbol{H}}_t = \begin{bmatrix} \boxed{\boldsymbol{H}_{t-1}} & \mathbf{0} \\ \mathbf{0} & \boxed{\boldsymbol{H}_t} \end{bmatrix} \qquad \text{et} \qquad \overline{\boldsymbol{p}}_t = \begin{pmatrix} \boldsymbol{p}_{t-1} \\ \boldsymbol{p}_t \end{pmatrix}. \tag{7.11}$$

Les trois plans constituant $\overline{\boldsymbol{x}}_t$ peuvent alors être reconstruits en minimisant éventuellement sous contrainte séparable le *critère pénalisé réduit* $J_t : \mathbb{R}^{3L} \to \mathbb{R}$ suivant,

$$J_t(\overline{\boldsymbol{x}}_t\,;\, \boldsymbol{\lambda}) = \|\overline{\boldsymbol{y}}_t - \overline{\boldsymbol{H}}_t \overline{\boldsymbol{x}}_t\|^2 + \alpha \sum_c \phi([\delta_c]_t\,;\, s) \tag{7.12}$$

où $\overline{\boldsymbol{y}}_t$ représente les mesures effectivement à notre disposition qui impliquent directement la tranche t du volume — c.à.d. $\overline{\boldsymbol{y}}_t = \overline{\boldsymbol{p}}_t +$ « bruit » dans un modèle d'observation additif. On note $[\delta_c]_t$ la différence entre éléments constituant la paire c de voisins dans $\overline{\boldsymbol{x}}_t$ — l'ensemble des $[\delta_c]_t$ représente simplement un sous ensemble des δ_c introduit dans le cadre volumique complet. Pour la fonction ϕ, on choisira une fonction C^1 strictement convexe et coercive de manière à garantir là encore

l'unicité et le caractère borné du minimiseur

$$\widehat{\overline{\boldsymbol{x}}}_t = \arg\min_{\overline{\boldsymbol{x}}_t} J_t(\overline{\boldsymbol{x}}_t\,;\,\boldsymbol{\lambda}) \qquad \text{avec} \qquad \widehat{\overline{\boldsymbol{x}}}_t = \begin{pmatrix} \widehat{\boldsymbol{x}}_{t-1} \\ \widehat{\boldsymbol{x}}_t \\ \widehat{\boldsymbol{x}}_{t+1} \end{pmatrix}. \tag{7.13}$$

L'image $\widehat{\boldsymbol{x}}_t$ correspond à une approximation du plan t de $\boldsymbol{x}_{\boldsymbol{\lambda}}$ reconstruit par la méthode complète (7.9). Une reconstruction d'un ou de plusieurs plans peut alors être effectué en effectuant successivement une reconstruction correspondant à (7.13). On souligne que ce problème individuel est $T/3$ fois moins important que le problème initial (7.9) et, à ce titre, qu'on peut espérer le résoudre en s'appuyant sur un algorithme à convergence rapide — quasi NEWTON, gradient conjugué préconditionné, *etc.*

Dans cette étude, nous mettrons en œuvre principalement cette technique dans l'objectif de s'assurer qu'elle conduit à des résultats acceptables en pratique ; en particulier, la résolution des sous problèmes sera effectuée par le même algorithme que celui adopté pour le problème exact — c.à.d. un algorithme SOR. En terme de qualité de la reconstruction, le test effectué sur un fantôme 3D de grande taille est concluant même si une légère dégradation a été constatée.

7.3 Résultats de simulations

Cette section cherche à fournir des éléments de comparaison qualitatifs et quantitatifs entre approches standard et pénalisée en tomographie hélicoïdale. Nous montrons en particulier que sur données synthétiques, la reconstruction pénalisée du volume d'intérêt conduit à un gain sensible de précision sur les reconstructions.

Création des données synthétiques

À notre connaissance, il n'existe pas de code permettant de générer des données synthétiques en géométrie hélicoïdale. Le modèle que nous avons construit en section 7.2.2 a donc été utilisé pour produire des projections hélicoïdales ; ces données ont alors servi pour l'inversion pénalisée et pour produire les reconstructions « standard ».

7.3.1 Fantômes synthétiques de petite taille

Les différentes méthodes sont testées dans un premier temps sur deux fantômes de petites tailles (quelques milliers de voxels) et de forme géométrique simple. Ces premiers essais on le mérite de permettre de distinguer clairement certaines propriétés ou certains travers des méthodes de reconstruction.

Fantôme hémisphérique

Le premier fantôme 3D est une demi sphère binaire pixelisée constituée par 30 plans de voxels d'épaisseur 1 millimètre (abrég. mm) ; cf. figure 7.4-(a). Cet objet 3D a servi à produire un jeu de projections en géométrie hélicoïdale en utilisant la sortie du modèle construit en section précédente.

L'échantillonnage du volume d'intérêt débute en $\xi_3 = 0$; le pas de l'hélice est fixé à 10 mm, le nombre de projections sur une rotation est fixé à 10, chaque projection rassemblant 39 rayons. Le volume de reconstruction est composé de 6 plans de 30 pixels de coté (soit un ensemble de 5400 paramètres). Chacun des 6 plans représente dans le fantôme un disque binaire de diamètre différent que l'on cherche à reconstruire. Le contexte de mesure ainsi que les plans de reconstruction sont illustrés sur la figure 7.4.

En dépit de sa symétrie par rotation suivant l'axe du tomographe, nous considérons que ce fantôme constitue une « figure de mérite » intéressante : d'une part, la reconstruction d'un disque permet de jauger aisément des distorsions introduites par les méthodes de reconstruction, et d'autre part, les diamètres des disques dans chacun de ces plans subissent une dynamique de variation importante — accentuée pour les premiers plans et moins marquées par la suite. On notera également à ce propos que le premier plan de reconstruction devrait être vide puisque tangent à la sphère.

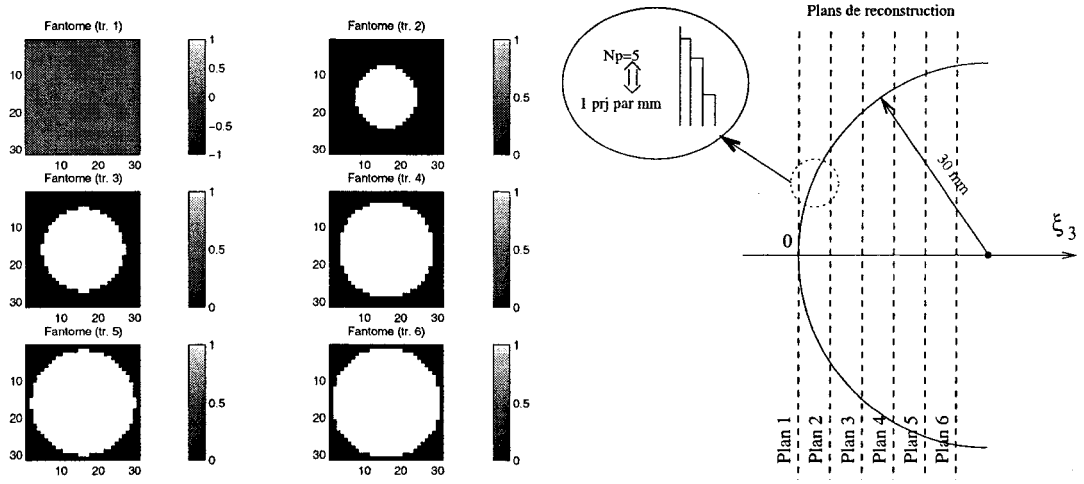


Figure 7.4: Fantôme utilisé pour créer le jeux de projections hélicoïdales (haut) ; plans du fantôme choisis pour la reconstruction (←) et emplacement par rapport au fantôme (→).

Profil de sensibilité de tranche

Un des moyens privilégiés pour quantifier la résolution axiale est de produire un *profil de sensibilité de tranche* (abrég. SSP pour « slice sensitive profil »). Ce procédé consiste à mesurer ou à simuler la réponse axiale à un créneau du système d'imagerie. Nous avons simulé ce test en nous appuyant sur un fantôme en forme de demi cylindre dont l'axe correspond à celui du tomographe (voir figure 7.8). Le jeu de projections hélicoïdales est constitué de 40 projections non bruitées obtenues pour une épaisseur de faisceau nulle ; et le nombre de plans de reconstruction a été fixé à 10, ce qui réserve 4 projections par plan d'image. Une fois les 10 plans reconstruits, l'énergie obtenue dans chaque plan donne une bonne idée de la réponse axiale du tomographe. La figure 7.8 montre les plans reconstruits par les méthodes standard ainsi que l'énergie de chaque plan normalisé par l'énergie d'une tranche

non vide du fantôme.

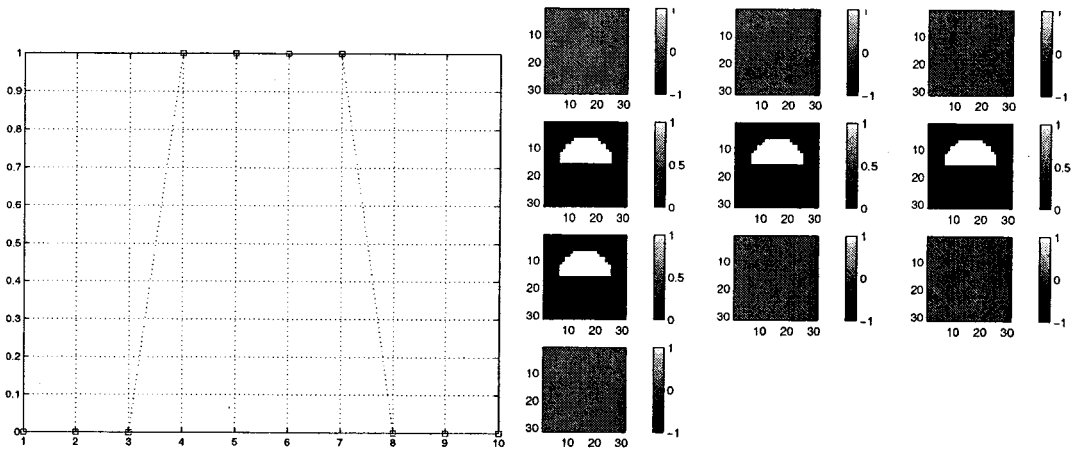


Figure 7.5: Fantôme synthétique utilisé pour l'estimation de la SSP (←) ; plans de reconstruction du fantôme choisis (↑) et énergie normalisée contenue dans chacun des plans du fantôme destinés à être reconstruit.

Simulation d'algorithmes de reconstruction standard

Nous avons utilisé le modèle d'observation développé en section 7.2.2 pour produire un ensemble de projections en géométrie hélicoïdale pour chacun des deux fantômes. Conformément à la démarche décrite en section 3.1.3.1.1, une interpolation LI-180° et LI-360° a été menée de manière à constituer autant des jeux « axiaux complets » que de plan de reconstruction ; chacun de ces jeux étant ensuite utilisé pour reconstruire une image 2D à partir de la RPC implantée dans le logiciel SNARK. Ce faisant, nous disposons d'une base comparative qui nous permettra ultérieurement de juger de l'intérêt d'une approche pénalisée de la reconstruction hélicoïdale.

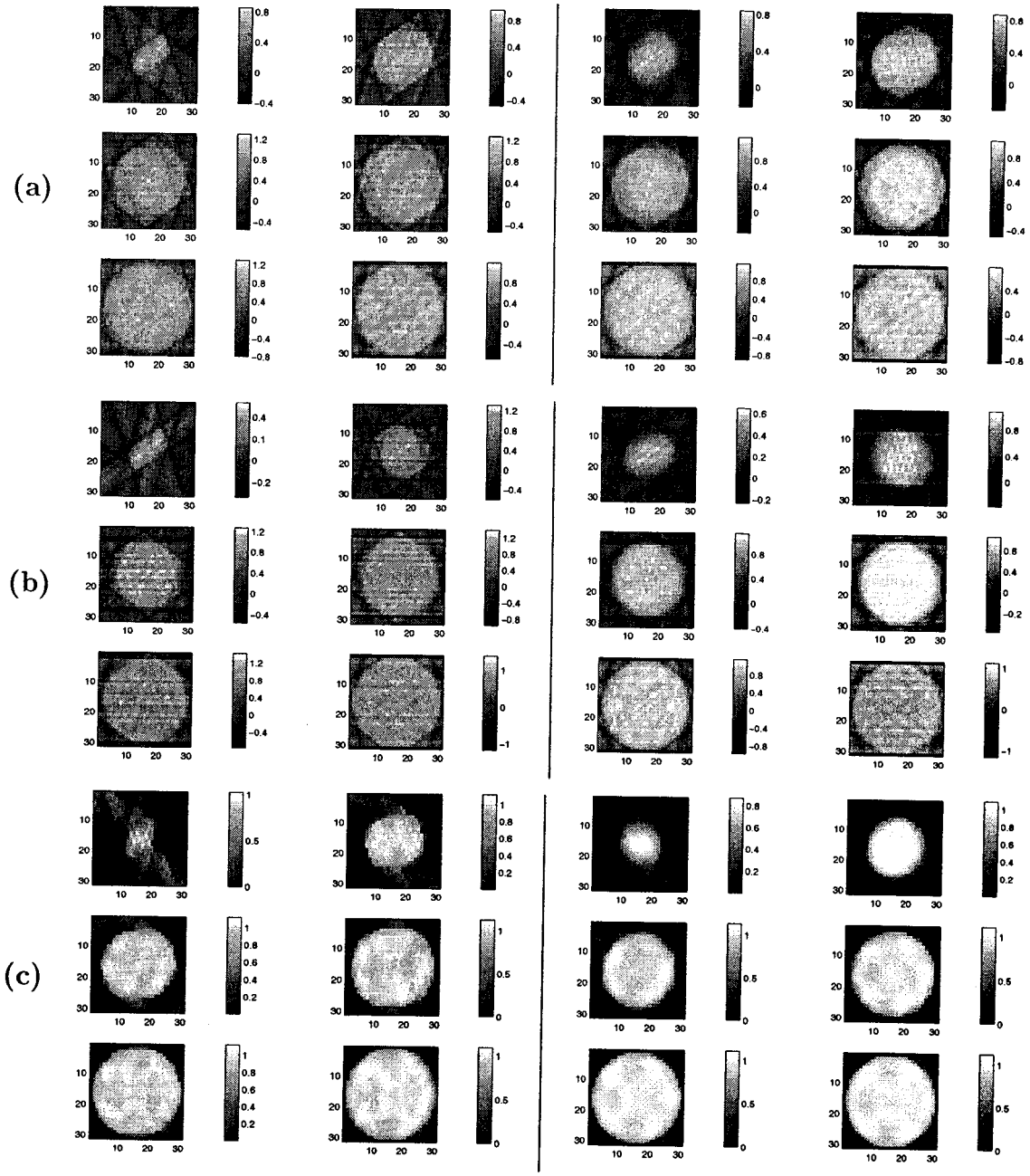


Figure 7.6: Reconstruction du fantôme de la figure 7.4 pour un faisceau d'épaisseur nulle (←) ou égale au coté d'un voxel (→) : par RPC+LI-360° **(a)**, RPC+LI-180° **(b)**, approche pénalisée $\ell_2\ell_1$ **(c)**. RPC mise en œuvre avec un filtre de HAMMING et une fréquence de coupure réduite $\nu_c = 0,2$; les hyperparamètres de l'approche pénalisée ont été ajustés manuellement.

Reconstructions pour des données non bruitées

Les reconstructions obtenues sans bruit pour les deux interpolations LI-180° et LI-360° sont représentées sur la figure 7.6-**(a,b)**pour deux épaisseurs de faisceau distinctes. On notera que chaque problème de reconstruction est fortement sous déterminé : pour chacun des 6 plans, on constitue par interpolation sur l'hélice un jeu de 390 données (10 projections $\times$ 39 rayons), nombre à comparer avec les 5400 paramètres de l'image 3D à estimer. En ce sens, on comprend que les résultats de la RPC soient, même sans bruit, de piètre qualité. Dans ces reconstructions, il est néanmoins possible de retrouver certaines caractéristiques usuelles des images hélicoïdales standard. En particulier, on notera que la perte de résolution suivant l'axe du tomographe varie suivant le type d'interpolation *et* l'épaisseur du faisceau. On note par exemple que l'interpolation LI-180° permet un gain de résolution suivant l'axe du tomographe, et que l'épaisseur du faisceau produit un effet de flou comparable à celui produit par l'introduction d'une réponse impulsionnelle axiale supplémentaire dans le système de mesure. Soulignons également que la forme des cercles subit une déformation qui est d'autant plus accentuée que la variation de diamètre est rapide (i.e. l'effet s'amoindrit à mesure qu'on se rapproche du centre de la sphère) ; cet effet sera également constaté pour l'approche pénalisée.

Immunité au bruit

Pour un contexte de mesure identique, un bruit additif gaussien non corrélé, centré, de variance égale à 1 a été ajouté aux mesures avant interpolation ; ceci représente un rapport signal à bruit d'environ 48 décibels. La figure 7.7-**(a,b)** présente les reconstructions obtenues pour l'interpolation LI-180° et la RPC ; celles produites pour l'interpolation LI-360° présentent des caractéristiques très similaires. Comparativement aux reconstructions standard sans bruit de la figure 7.6-(a,b), on constate une dégradation significative de la qualité des images, effet déjà constaté dans le cadre axial au chapitre précédent.

Profil de sensibilité de tranche (SSP)

La mise en œuvre d'une SSP donne un élément de caractérisation quantitatif de la résolution axiale des algorithmes standard ; La figure 7.8-**(a,b)** présente les plans reconstruits pour chacune des deux interpolations ainsi que l'énergie de chaque plan normalisé par l'énergie d'une tranche non vide du fantôme. Pour chacune de ces deux interpolations, on constate un effet d'étalement assez marqué, sensiblement plus prononcé pour l'interpolation LI-360°.

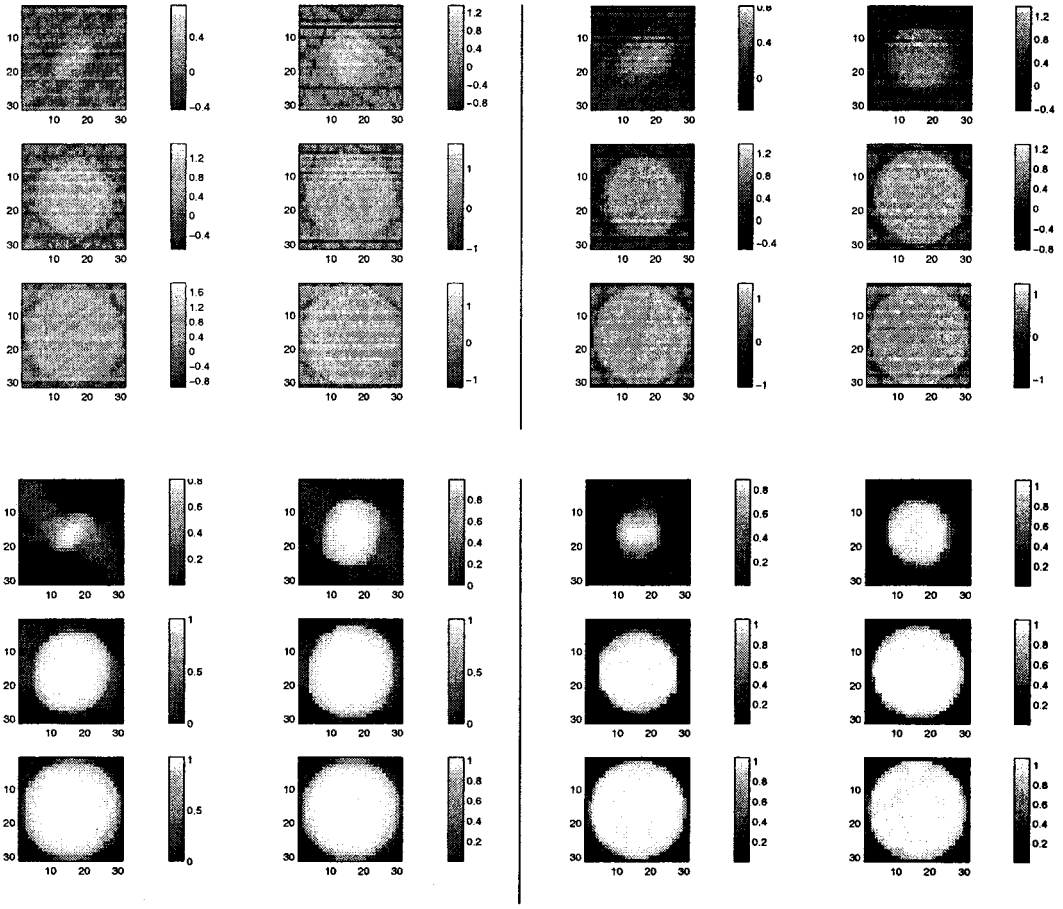


Figure 7.7: Reconstruction du fantôme 7.4 pour des données bruitées ($\sigma = 1$) et une épaisseur de faisceau nulle ($\leftarrow$) ou égale à la largeur d'un plan de voxels ($\rightarrow$) : RPC+LI-180° (haut) ou approche pénalisée $\ell_2\ell_1$ (bas). La RPC a été mise en œuvre avec un filtre de reconstruction de type HAMMING et une fréquence de coupure réduite $\nu_c = 0,2$.

Approches pénalisées $\ell_2\ell_1$

La minimisation du critère pénalisé (7.8) sous contrainte de positivité ($X = \mathbb{R}_+^N$) pour des projections non bruitées conduit aux reconstructions de la figure 7.6-(c) ; les différents hyperparamètres ont été réglés manuellement. Ces images sont à rapprocher de celles obtenues par les méthodes standard visibles sur la même

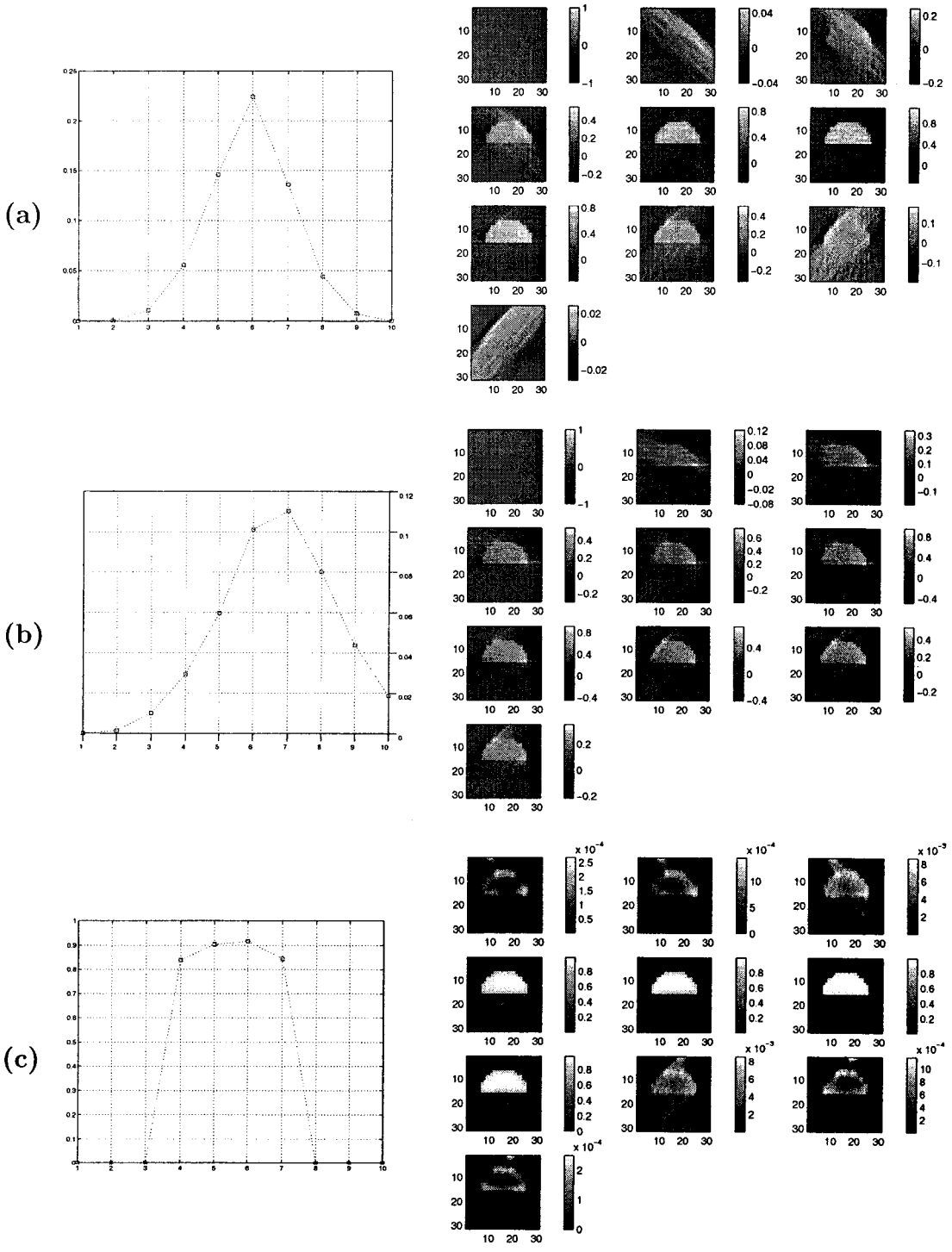


Figure 7.8: Tracées des SSP : RPC+LI-180° **(a)**, RPC+LI-360° **(b)**, app. pénalisée $\ell_2\ell_1$ **(c)**.

figure. L'amélioration de la qualité des images est sensible : on notera en particulier la nette diminution des stries typiques des reconstructions par RPC pour un nombre de vues limité. On notera néanmoins que les contours des disques dans chaque image subissent des déformations surtout marquées là où la géométrie de l'objet varie le plus rapidement — *i.e.* proche de l'origine. On souligne ici qu'aucune des méthodes présentées n'est capable de reconstruire correctement le premier plan d'image qui est normalement vide. Pour les algorithmes standard, on le comprend aisément : on cherche à reconstruire un plan vide à partir de données provenant des plans suivants. Pour l'approche pénalisée, le modèle d'observation développé en section 7.2.2 s'appuie sur une valeur uniforme de l'image sur chaque voxel ; or, c'est entre les deux premiers plans de reconstruction que la géométrie de la sphère est la moins stationnaire (là se produit la plus forte augmentation de diamètre) : la première image reflète ce phénomène.

La figure 7.7-(**c**) présente les reconstructions pénalisées pour les données bruitées ; les reconstructions standard produites à partir du même jeu de données sont présentées sur la même figure. Comme le suggère cet exemple, l'inversion pénalisée permet généralement un gain de robustesse appréciable comparativement à une reconstruction standard. Finalement, afin d'apprécier la résolution suivant l'axe du tomographe, la SSP a été tracée et représentée sur la figure 7.8-(**c**) : ce résultat se compare très favorablement aux SSP des méthodes standard tracées sur la même figure ; on notera en particulier que le niveau des artefacts présents dans les parties

normalement vides du fantôme est de l'ordre du centième de l'unité pour l'inversion pénalisée.

7.3.2 Fantôme synthétique de grande taille

Nous nous intéressons maintenant à un problème de reconstruction de taille plus importante : les projections sont produites à partir de 40 coupes de 127×127 voxels extraites uniformément sur toute la longueur du fantôme représenté sur la figure 7.9. Ce fantôme est constitué de formes géométriques variées rassemblant des parties cylindriques (ex. coupe 36), coniques (ex. coupe 19) et sphériques ; d'autre part, un méplat et une rainure apparaissent sur les parties les plus volumineuses du fantôme (cf. coupe 19).

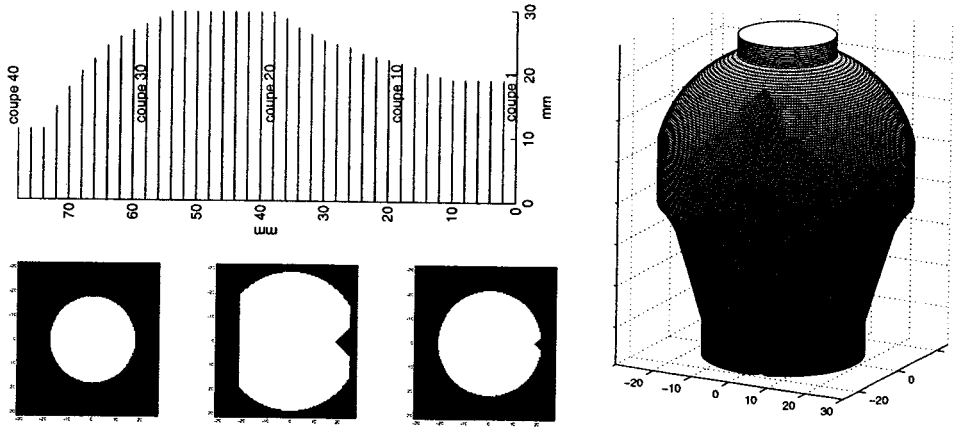


Figure 7.9: Fantôme 3D de grande taille : (→) vue en perspective, (← haut) variation du profile suivant l'axe du tomographe, (← bas) coupes 10, 19 et 36 extraite du fantôme.

Le contexte de mesure est tel que chaque plan de mesure bénéficie de 5 projec-

tions saisies uniformément sur une demi rotation, l'épaisseur du faisceau correspondant à l'épaisseur d'une tranche. Un bruit blanc centré gaussien a été ajouté aux données afin de simuler l'effet de différentes sources de bruit (l'écart type est de $\sigma = \sqrt{2}$, soit approximativement un rapport signal à bruit de 30 décibels). Ces projections bruitées sont ensuite utilisées pour reconstruire les 40 plans correspondant aux plans de mesure, soit par approche standard, soit par approche pénalisée sous contrainte de positivité ($X = \mathbb{R}_+^N$). La figure 7.10 présente trois coupes extraites des volumes reconstruits par ces deux approches[7]. Là encore, la reconstruction pénalisée se distingue nettement de la méthode standard. Dans cette même figure, les contours extraits dans chacune de ces trois coupes (seuillage à 0,7) ont été comparés aux vrais contours issus du fantôme : la localisation des contours du fantôme est assez précise hormis pour la pointe de la rainure en « V ». En fait, nous avons constaté expérimentalement que la distance maximale séparant un contour segmenté du bord du fantôme intervient *systématiquement* au niveau des « points anguleux » dans les contours. Dans la mesure où de tels points anguleux ne sont pas représentatifs des contours réguliers des interfaces osseuses, ce type d'erreur ne constitue pas un handicap majeur pour notre application.

Un critère quantitatif de type *erreur quadratique moyenne* (abrég. EQM) a été calculé sur chaque tranche de manière à suivre l'évolution de l'erreur suivant l'axe

[7]Nous ne présentons ici que le résultat standard associé à l'interpolation LI-180° ; le résultat pour l'interpolation LI-360° n'apporte pas d'élément supplémentaire de discussion.

du tomographe :

$$\mathrm{EQM}(t) = \frac{1}{L}||(\boldsymbol{x}_{\boldsymbol{\lambda}})_t - \boldsymbol{x}_t||^2$$

où $\boldsymbol{x}_t$ et $(\boldsymbol{x}_{\boldsymbol{\lambda}})_t$ représente respectivement le t-ème plan du fantôme et du volume reconstruit ; le suivi de cet indicateur est tracé sur la figure 7.11. Une EQM sur le volume *complet* a également été calculée pour cette reconstruction pénalisée et une reconstruction standard (LI-180°) : la seconde est approximativement 14 fois plus importante que la première.

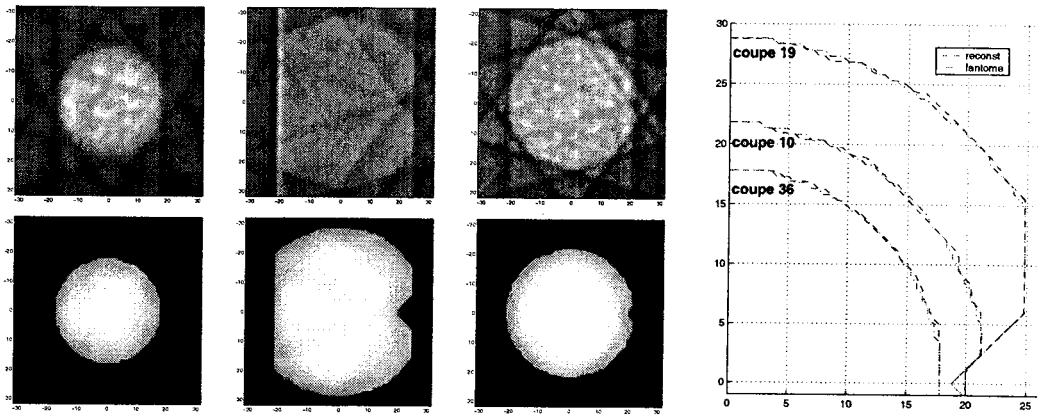


Figure 7.10: Reconstruction des coupes 36, 19 et 10 par LI-180° + RPC (haut) ou par approche pénalisée $\ell_2\ell_1$ (bas) et tracé des courbes segmentées à 0,7 pour le volume reconstruit par approche pénalisée. La RPC a été mise en œuvre avec un filtre de reconstruction de type cosinus et une fréquence de coupure réduite $\nu_c = 0,3$; les hyperparamètres de l'approche pénalisée ont été ajustés manuellement.

Enfin, la figure 7.12 présente la coupe 19 reconstruite par l'*approche pénalisée approchée* définie par (7.12) : on constate que la reconstruction reste de bonne qualité en dépit d'une légère diminution du niveau d'intensité au niveau des bords ; les contours segmentés extraits pour les trois plans 10, 19 et 36 confirment ce constat.

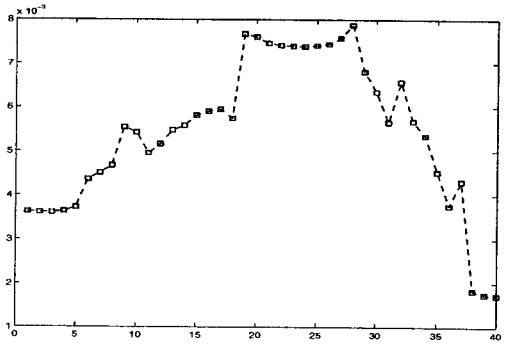


Figure 7.11: Tracé de l'EQM dans chaque tranche pour la reconstruction exacte.

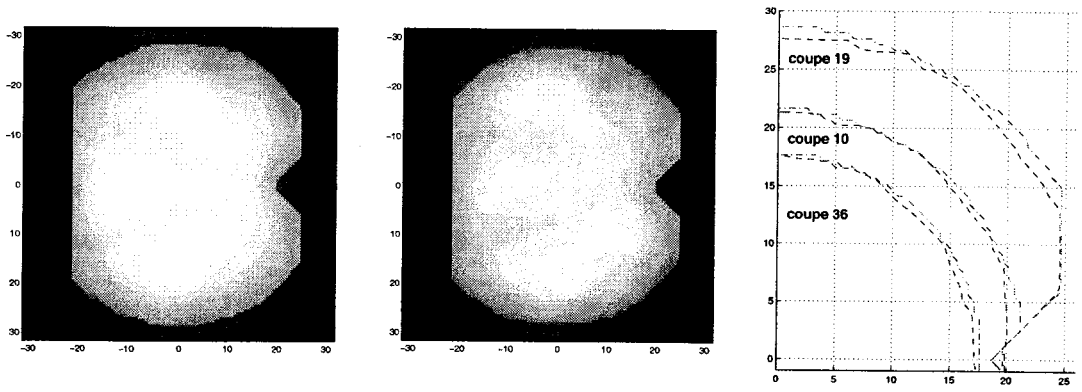


Figure 7.12: Reconstruction de la coupe 19 par l'approche pénalisée $\ell_2\ell_1$ exacte ($\leftarrow$) et inexacte ($\uparrow$) ; contours segmentés pour les plan 10, 19 et 36 pour la reconstruction du volume de manière inexacte ($\rightarrow$)

L'EQM obtenu sur le volume complet reconstruit par cette approche pénalisée « inexacte » a également été calculée : elle est 12 fois moins importante que celle produite par une reconstruction standard.

7.4 Choix algorithmiques en tomographie hélicoïdale

En pratique, le choix d'une stratégie d'optimisation conditionne en grande partie la rapidité de l'algorithme de reconstruction mis en œuvre. Cette section met donc

l'accent sur les problématiques algorithmiques et sur les solutions développées dans le cadre de la minimisation d'un critère pénalisé en tomographie hélicoïdale.

Formulation générique du problème

Dans la mesure où l'implantation des approches pénalisées en tomographie *axiale* a conduit à un certain nombre de conclusions intéressantes, il est instructif d'examiner la structure du problème en géométrie *hélicoïdale* en la comparant à cette référence. On considère donc le problème de minimisation

$$\boldsymbol{x}_\lambda = \arg\min_{\boldsymbol{x}\in X} J(\boldsymbol{x}\,;\,\boldsymbol{\lambda}) \tag{7.14}$$

avec J strictement convexe qui s'écrit sous forme générique

$$J(\boldsymbol{x}\,;\,\boldsymbol{\lambda}) = \langle \boldsymbol{x}, \boldsymbol{Q}\boldsymbol{x}\rangle - 2\langle \boldsymbol{q}, \boldsymbol{x}\rangle + \alpha \sum_{c=1}^{C} \phi(\delta_c); \tag{7.15}$$

avec $\delta_c = \langle \boldsymbol{d}_c, \boldsymbol{x}\rangle$ et $\boldsymbol{Q} \in \mathbb{R}^{N\times N}$ une matrice symétrique et définie non négative (abrég. DNN). La forme (7.15) décrit simultanément — et à une constante additive près — le problème d'inversion en géométrie *axiale* et *hélicoïdale*, ces deux problèmes ne diffèrent que par la partie quadratique de J :

(géométrie axiale) $\boldsymbol{Q} = \boldsymbol{R}^T\boldsymbol{R}$, $\boldsymbol{q} = \boldsymbol{R}^T\boldsymbol{y}$ — *(géométrie hélicoïdale)* $\boldsymbol{Q} = \boldsymbol{H}^T\boldsymbol{H}$, $\boldsymbol{q} = \boldsymbol{H}^T\boldsymbol{y}_h$,

et éventuellement par le vecteur $\boldsymbol{d}_c$ qui doit tenir compte du caractère 2D ou 3D du modèle d'image considéré. La pénalisation ϕ est par exemple la fonction $\ell_2\ell_1$ hyperbolique définie par (7.3), cependant les éléments de cette section s'appliquent à de nombreuses fonctions convexes adoptées en traitement d'image[8].

Caractéristiques de $\boldsymbol{R}$ et $\boldsymbol{R}^T\boldsymbol{R}$

En géométrie *axiale*, $\boldsymbol{R}$ est creux et de grande taille ; son taux de remplissage est approximativement $\rho_{\boldsymbol{R}} = 2/\sqrt{M}$ pour une discrétisation à base de « bandes » de la transformée de RADON [Fessler et Booth, 1999, p. 690]. Le nombre de mesures étant typiquement de l'ordre du nombre de paramètres à estimer, une image de $N = 512 \times 512$ pixels conduit approximativement à 7×10^{10} éléments dans $\boldsymbol{R}$; le taux d'occupation est néanmoins très faible $\rho_{\boldsymbol{R}} \approx 2/\sqrt{N} \ll 1\%$.

Le produit $\boldsymbol{R}^T\boldsymbol{R}$ n'est cependant pas creux en général (cf. figure 7.13) ce qui rend en principe délicat son stockage et les opérations arithmétiques (produit matrice-vecteur, inversion, *etc*). Cet obstacle reste néanmoins peu contraignant en *géométrie à rayon parallèle* puisque $\boldsymbol{R}^T\boldsymbol{R}$ est approximativement *Toeplitz-bloc-Toeplitz* [Delaney et Bresler, 1996; Sauer et Bouman, 1993] : on peut alors tirer partie de la transformée de FOURIER rapide (abrég. TFR) pour diminuer sensiblement les coûts d'implantation.

[8] Pour l'algorithme que nous allons utiliser, il est néanmoins nécessaire que le rapport $\phi'(u)/u$ soit borné pour tout $u \in \mathbb{R}$.

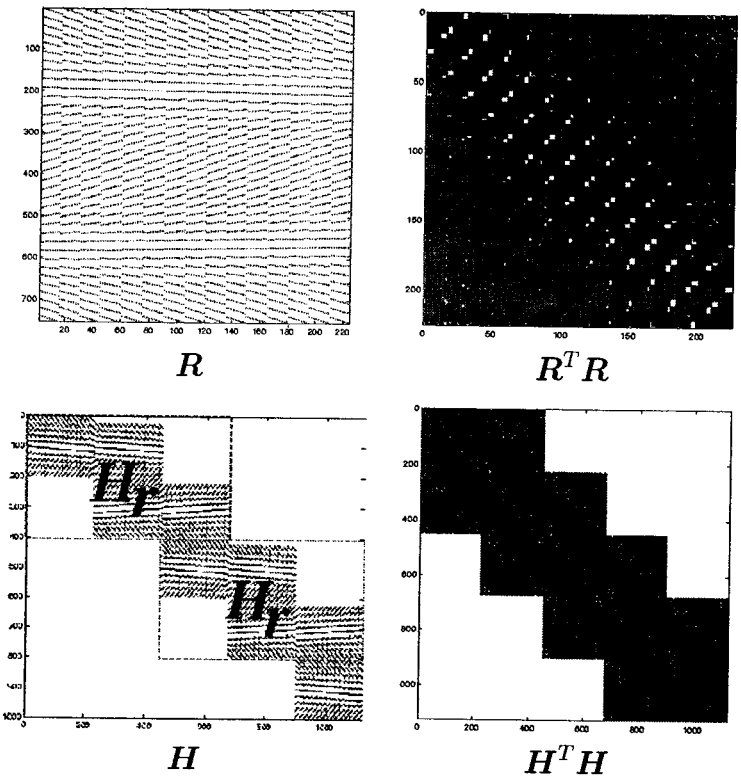


Figure 7.13: Opérateur d'observation et matrice $\boldsymbol{Q}$ obtenue en géométrie axiale (haut) et en géométrie hélicoïdale (bas) sur un problème de petite taille (15×15 pixels en axial et 5 plans de 15×15 voxels en hélicoïdal) ; dans $\boldsymbol{H}$, l'opérateur $\boldsymbol{H}_r$ a été délimité par des pointillés afin de simplifier l'interprétation de sa structure.

Caractéristiques de $\boldsymbol{H}$ et $\boldsymbol{H}^T\boldsymbol{H}$

Comparativement à l'opérateur d'observation en géométrie axial, la taille de $\boldsymbol{H}$ et son caractère creux sont sensiblement augmentés. Si on considère une image de 30 plans de 512×512 voxels, un pas d'hélice égale à 3 plans de voxels et 300 projections par rotation, le nombre d'éléments dans $\boldsymbol{H}$ est approximativement 1000 fois supérieur à celui de l'opérateur axial $\boldsymbol{R}$ considéré ci-dessus. On note que $\boldsymbol{H}$ possède une structure bi-diagonale par blocs, chaque bloc étant constitué de L

colonnes ; cf. figure 7.13. Si $\boldsymbol{H}$ peut être décrite à partir d'un bloc élémentaire $\boldsymbol{H}_r$ (cf. page 170), alors sa structure est proche d'un arrangement bloc-circulant ; d'autre part, le taux d'occupation $\rho_{\boldsymbol{H}_r}$ de $\boldsymbol{H}_r$ est égale à :

$$\rho_{\boldsymbol{H}_r} = \begin{cases} \dfrac{2}{3}\rho_{\boldsymbol{R}} & \text{pour K=2} \\ \left[1 - 2\dfrac{1 + \cdots K - 1}{(K+1)K}\right]\rho_{\boldsymbol{R}} & \text{pour K>2,} \end{cases}$$

avec $\rho_{\boldsymbol{R}}$ le taux de remplissage associé à l'opérateur axial $\boldsymbol{R}$ utilisé pour construire la matrice $\boldsymbol{H}_r$.

Le produit $\boldsymbol{H}^T\boldsymbol{H}$ a une structure de blocs diagonaux *pleins* se recouvrant sur L colonnes. Ce recouvrement empêche *a priori* d'utiliser la base de FOURIER pour diagonaliser $\boldsymbol{H}^T\boldsymbol{H}$; si ce recouvrement peut être négligé (*i.e.* si le pas de l'hélice est important) on peut néanmoins penser que la base de FOURIER constitue une bonne approximation de la base diagonalisant $\boldsymbol{H}^T\boldsymbol{H}$.

7.4.1 Stratégies de minimisation

La minimisation itérative d'un critère pénalisé en tomographie passe principalement par des méthodes à *direction de descentes* non contraintes ou des algorithmes de *relaxation sur les coordonnées de J*. Afin de motiver l'emploi de la relaxation en tomographie hélicoïdale, nous examinons maintenant l'implantation de ces deux familles algorithmiques dans le cadre de la tomographie axiale et hélicoïdale. Dans

ce qui suit, on pose afin de simplifier les notations

$$\boldsymbol{D} = (\boldsymbol{d}_1|\cdots|\boldsymbol{d}_C) \qquad \text{et} \qquad \boldsymbol{\phi}'(\delta_\bullet) = (\phi'(\delta_1), \cdots, \phi'(\delta_C))^T$$

où on rappelle que $\delta_c = \langle \boldsymbol{d}_c, \boldsymbol{x} \rangle$ est la différence entre les valeurs de la c-ème paire d'éléments voisins dans l'image.

Algorithmes à direction de descente

Les algorithmes à direction de descente sont largement utilisés en tomographie d'émission : la mise à jour $k \rightarrow k+1$ s'écrit alors

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \theta^{(k)} \boldsymbol{\xi}^{(k)} \tag{7.16}$$

où $\boldsymbol{\xi}^{(k)}$ est une *direction de descente* calculée à partir du gradient courant

$$\boldsymbol{\nabla} J(\boldsymbol{x}^{(k)}) = 2\boldsymbol{Q}\boldsymbol{x}^{(k)} + \alpha \boldsymbol{D}^T \boldsymbol{\phi}'(\delta_\bullet^{(k)}), \tag{7.17}$$

où $\theta^{(k)}$ est obtenue par *recherche linéaire*. Ces algorithmes sont en majorité *non contraints* et on se concentrera sur les méthodes de ce type. Parmi les choix les plus répandus, citons le *gradient conjugué* éventuellement préconditionné et certaines formes d'algorithmes « semi quadratiques » (abrég. SQ) qui seront développées au chapitre suivant ; certains algorithmes dérivés de la technique EM (*Expectation-*

Maximisation) tombent également dans cette classe d'algorithmes et sont pour leur part largement employés en tomographie d'émission.

Si on exclut le gradient conjugué *sans préconditionneur* et l'algorithme de plus profonde descente, le calcul de $\boldsymbol{\xi}^{(k)}$ nécessite habituellement de résoudre une équation de la forme

$$\boldsymbol{B}^{(k)}\boldsymbol{\xi}^{(k)} = -\boldsymbol{\nabla} J(\boldsymbol{x}^{(k)}) \tag{7.18}$$

avec $\boldsymbol{B}^{(k)}$ faisant intervenir $\boldsymbol{Q}$. Par exemple, les algorithmes SQ dont la forme suit (7.16) nécessitent de résoudre (7.18) où $\boldsymbol{B}^{(k)}$ correspond à l'une des deux matrices suivantes :

$$\boldsymbol{B}_{\text{GR}}^{(k)} = 2\boldsymbol{Q} + \alpha \boldsymbol{D}^T \operatorname{diag}\big(\phi'(\delta_c^{(k)})/\delta_c^{(k)}\big)\boldsymbol{D} \qquad \text{ou} \qquad \boldsymbol{B}_{\text{GY}} = 2\boldsymbol{Q} + \alpha \boldsymbol{D}^T\boldsymbol{D}; \tag{7.19}$$

ces deux opérateurs définissant deux formes SQ distinctes : la première est de GEMAN et REYNOLDS (abrég. GR), et la seconde de GEMAN et YANG (GY). On note enfin que pour le gradient conjugué *préconditionné*, le préconditionnement du gradient courant fait en principe intervenir $\boldsymbol{Q}$; voir par ex. [Fessler et Booth, 1999] pour une application en tomographie axiale.

Un contexte de mise en œuvre difficile

Pour un même algorithme de descente, le coût de calcul et de stockage requis par itération dépend fortement de la taille et de la structure de la matrice $\boldsymbol{Q}$ et de l'opérateur $\boldsymbol{D}$. En pratique, l'efficacité d'implantation est souvent limitée par les opérations impliquant $\boldsymbol{Q}$. Le calcul du gradient de J défini par (7.17) constitue un exemple révélateur : alors que la structure de $\boldsymbol{D}$ permet très souvent de calculer le terme $\boldsymbol{D}^T\boldsymbol{\phi}'(\delta_\bullet^{(k)})$ sans effectuer explicitement de produit matrice-vecteur, le produit $\boldsymbol{Q}\boldsymbol{x}^{(k)}$ ne peut être effectué simplement dans beaucoup de situations pratiques. La *tomographie axiale à rayons parallèles* constitue néanmoins une exception notable : $\boldsymbol{Q} = \boldsymbol{R}^T\boldsymbol{R}$ est Toeplitz-bloc-Toeplitz et le produit $\boldsymbol{R}^T\boldsymbol{R}\boldsymbol{x}^{(k)}$ peut être calculé par TFR. Cette technique permet par exemple de mettre en œuvre un algorithme du gradient conjugué à faible coût d'implantation [Delaney et Bresler, 1996].

Outre le calcul du gradient, la résolution d'une équation normale augmente encore le volume de calcul. La possibilité du recours à la TFR est assez marginale, et le gradient conjugué est largement employé pour résoudre *itérativement* et *approximativement* l'équation normale courante [Charbonnier *et al.*, 1997; Delaney et Bresler, 1998]. Si la matrice normale est *indépendante* de l'itération (ex. $\boldsymbol{B}_{\text{GY}}$), son inverse peut éventuellement être calculée et stockée au préalable ; cependant cette solution est rarement retenue imagerie puisque l'inverse est de grande taille

et n'est pas creuse en général[9].

La mise en œuvre de ce type d'algorithme pour la tomographie hélicoïdale reste donc difficile à envisager : la dimension du problème est bien supérieure au cadre axial et la structure de l'opérateur $\boldsymbol{Q} = \boldsymbol{H}^T\boldsymbol{H}$ ne permet pas le recours à la TFR. Dans ce contexte, il n'est évidemment pas question de résoudre une équation normale à chaque itération ; d'autre part, adopter une équation normale indépendante de l'itération (ex. $\boldsymbol{B}_{\text{GY}}$) simplifie le problème sans le régler : si $\boldsymbol{H}^T\boldsymbol{H}$ est creuse, son inverse ne l'est plus en général et son stockage devient vite impossible. Finalement, les préconditionneurs du gradient conjugué développés pour la géométrie axiale [Fessler et Booth, 1999] ne peuvent être utilisés directement en géométrie hélicoïdale ; une adaptation de ces préconditionneurs reste éventuellement à construire.

Algorithmes de relaxation

Pour s'affranchir des délicates questions d'implantation associés à la reconstruction d'image de grande taille, certains auteurs choisissent de mettre en œuvre un algorithme de *relaxation sur les composantes* de J ; la mise à jour de la n-ème

[9]Une approche intermédiaire consiste à *factoriser* cette matrice — ex. factorisation de CHOLESKY si celle-ci est définie positive — et à stocker ses facteurs : à chaque itération de l'algorithme, la résolution de l'équation normale se « réduit » à celle de deux systèmes triangulaire effectuée en $O(N^2)$ opérations; cf. [Golub et Van Loan, 1996, Ch. 4]. Si la matrice normale est creuse, l'intérêt de cette approche est que la factorisation conduit encore à des matrices creuses; cette démarche reste néanmoins irréaliste pour un problème de très grande taille puisque la résolution de systèmes triangulaires reste trop coûteuse.

composante s'écrit alors :

$$x_n^{(k+1)} = (1-\theta)x_n^{(k)} + \bar{x}_n^{(k+1)} \qquad \text{avec} \qquad \bar{x}_n^{(k+1)} = \underset{u \in X_n}{\arg\min}\, J_n(u) \tag{7.20}$$

où θ est un *paramètre de relaxation* qui peut être ajusté de manière à accélérer la convergence asymptotique de l'algorithme [Ortega et Rheinboldt, 1970, Sec. 10.5]. $J_n : \mathbb{R} \to \mathbb{R}$ est le critère *monovarié* nécessairement strictement convexe qui s'écrit :

$$J_n(u) \equiv J(\boldsymbol{x}^{(k)} + u\boldsymbol{e}_n)$$

où $\boldsymbol{e}_n$ est le n-ème vecteur canonique de $\mathbb{R}^N$. Dans un schéma itératif de type « *successive over relaxation* » (abrég. SOR) la mise à jour des inconnues à lieu immédiatement, c.à.d.

$$\text{(SOR)} \qquad \boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + (x_n^{(k+1)} - x_n^{(k)})\boldsymbol{e}_n; \tag{7.21}$$

un schéma voisin mais distinct consiste à mettre à jours les composantes une fois effectué le balayage complet : c'est la méthode de JACOBI qui est généralement plus économe en volume de mémoire mais d'une vitesse de convergence moins intéressante [Ciarlet, 1988]. Pour ces deux algorithmes, une itération correspond à un balayage complet des N composantes. On souligne que la contrainte convexe $X_n \subseteq \mathbb{R}$ est aisément prise en compte : elle conduit principalement à projeter le mi-

nimiseur scalaire *non contraint* sur X_n si celui-ci est à l'extérieur ; pour $X_n = \mathbb{R}^+$, il suffit de ramener à zéro les valeurs (7.20) négatives. Par la suite, on suppose le problème (7.20) non contraint de manière à simplifier l'exposé.

Dans la littérature du traitement d'image, ces méthodes de relaxation sont également appelées « *coordinate ascent* » (abrég. CA) en référence au problème équivalent de maximisation de l'opposée de J. Pour l'application en tomographie, ces algorithmes de relaxation présentent l'intérêt de faire converger rapidement les *hautes fréquences* de l'image[10] ; la *rétroprojection convoluée* (abrég. RPC) est alors souvent utilisée comme initialisation de l'algorithme puisqu'elle représente souvent une image « lissée » de la solution pénalisée. Notons que l'ordre des mises à jour a un impact sensible sur la convergence : pour la tomographie, un balayage aléatoire de l'ensemble des pixels semble préférable au simple balayage lexicographique ; cf. [Bowsher *et al.*, 1998] et [Zheng *et al.*, 2000, Sec. III.D].

Ces algorithmes de relaxation sur les composantes s'avèrent finalement attractifs pour la tomographie axiale : leur vitesse de convergence reste souvent intéressante (quelques dizaines de balayages complets suffisant souvent en tomographie axiale [Bouman et Sauer, 1996; Erdogan et Fessler, 1999; Zheng *et al.*, 2000]) et ils sont faciles à mettre en œuvre. Notons néanmoins que dans le cas général — et notamment pour les fonctions ϕ de type $\ell_2\ell_1$ — *le minimiseur (7.20) n'a pas de forme*

[10] Ce phénomène, analysé dans [Sauer et Bouman, 1993, Sec. IV] pour J *quadratique*, a été largement constaté expérimentalement pour les critères pénalisés utilisés en tomographie ; cf. [Bouman et Sauer, 1996, p. 484] et [Erdogan et Fessler, 1999, p. 801].

explicite et la minimisation de J_n s'apparente à une recherche linéaire itérative et incomplète. Les détails de mise en œuvre diffèrent assez sensiblement suivant la stratégie adoptée pour effectuer chacune des recherches linéaires. L'une d'elle est maintenant détaillée dans le cadre de la tomographie hélicoïdale : elle conduit à un algorithme particulièrement économe en volume de calcul.

7.4.2 Relaxation semi quadratique et tomographie hélicoïdale

L'algorithme que nous avons implanté pour la reconstruction tomographique hélicoïdale est une adaptation directe de travaux effectués dans le cadre de la restauration d'images [Brette et Idier, 1996; Villain *et al.*, 2001]. Cet algorithme de type SQ « *single site update* » (abrég. SSU) met à jour une seule composante de l'image à la fois, mais il importe de souligner que cette méthode est *a priori* distincte d'un simple algorithme SOR sur J. Pour l'essentiel néanmoins, cette distinction n'est que formelle et il est légitime — et intéressant — de présenter les algorithmes de relaxation SQ comme des schémas itératifs SOR sur J.

Dans ce qui suit, on notera $[\boldsymbol{A}]_{n,m}$ l'élément situé en (ligne n, colonne m) d'une matrice $\boldsymbol{A}$, et $[\boldsymbol{A}]_{n,\bullet}$ et $[\boldsymbol{A}]_{\bullet,m}$ respectivement la n-ème ligne et m-ème colonne de $\boldsymbol{A}$.

SOR semi quadratique de GR

Le schéma itératif ci-dessous est référencé dans la littérature sous le nom d'algorithme semi quadratique SSU. Pour la mise à jour courante k, la composante de l'image $n = \mathrm{mod}_k(N)$ est relaxée et les $N-1$ autres composantes recopiées :

$$\begin{aligned} & & x_n^{(k+1)} &= x_n^{(k)} + \theta\, \mu_n^{(k)} \\ \forall\, m \neq n & & x_m^{(k+1)} &= x_m^{(k)} \end{aligned}$$

où $\mu_n^{(k)}$ s'écrit :

$$\mu_n^{(k)} = \frac{[\boldsymbol{H}^T(\boldsymbol{y}_h - \boldsymbol{H}\boldsymbol{x}^{(k)})]_n - [\boldsymbol{D}\boldsymbol{L}^{(k)}\boldsymbol{D}^T\boldsymbol{x}^{(k)}]_n}{b_{\mathrm{GR}}^{(k)}} \tag{7.22}$$

avec

$$b_{\mathrm{GR}}^{(k)} = [\boldsymbol{H}^T\boldsymbol{H}]_{n,n} + \alpha[\boldsymbol{D}\boldsymbol{L}^{(k)}\boldsymbol{D}^T]_{n,n}. \qquad \text{et} \qquad \boldsymbol{L}^{(k)} = \mathrm{diag}\big(\phi'(\delta_c^{(k)})/2\delta_c^{(k)}\big). \tag{7.23}$$

Dans la construction SQ de GR, les composantes de $\boldsymbol{L}^{(k)}$ correspondent à des « variables duales ». On trouvera dans [Idier *et al.*, 2001] des conditions suffisantes de convergence globale pour la famille des algorithmes SQ ; dans le cas de l'algorithme SSU définit ci-dessus, la convergence pour $\phi(u) = \sqrt{u^2 + s^2}$ et $\theta \in]0; 2[$ est une application directe de ces conditions de convergence.

Rattachement à l'existant

Le comportement de ce schéma itératif peut recevoir un éclairage complémentaire intéressant en remarquant que la mise à jour courante $x_n^{(k+1)}$ s'écrit :

$$x_n^{(k+1)} = x_n^{(k)} - \theta\big(2b_{\mathrm{GR}}^{(k)}\big)^{-1} \left.\frac{\mathrm{d}J_n}{\mathrm{d}\,u}\right|_{u=0} \tag{7.24}$$

avec J_n le critère monovarié associé à la composante courante x_n ; ce résultat illustre dans une forme scalaire le lien fort existant entre formulation itérative SQ et algorithme à direction de descente sur le critère J ; ce thème sera repris plus en détail au chapitre suivant — page 256 pour le résultat ci-dessus.

L'intérêt majeur de l'expression (7.24) est de permettre d'associer l'algorithme SSU semi quadratique de GR à de simples algorithmes de relaxation largement étudiés et utilisés notamment en tomographie. Ainsi, *l'algorithme SSU semi quadratique de GR*[11] *est un algorithme SOR sur J mis en œuvre par une recherche linéaire réduite à une seule sous itération (7.24).* D'autre part, la relation (7.24) correspond, pour l'essentiel, à l'algorithme introduit récemment dans [Erdogan et Fessler, 1999, eq. (25)][12]. Ce constat peut paraître étonnant dans la mesure où le cadre SQ adopté par [Brette et Idier, 1996] est *a priori* très distinct du prin-

[11] *Un résultat similaire tient évidemment pour l'algorithme de GY ; cf. page 256.*

[12] Cet article traitant de reconstruction en tomographie de transmission à partir d'un modèle poissonien, la relation (25) dans cette référence ne correspond pas, au sens strict, à (7.24), cependant dans le cadre d'un simple modèle quadratique, il n'est pas difficile de se convaincre que ces deux relations sont identiques.

cipe constructif de « fonction majorante » adopté par [Erdogan et Fessler, 1999]. Au prochain chapitre, on montrera néanmoins que les constructions SQ définissent *effectivement* des algorithmes de type majorant.

De manière assez inattendue, ce dernier lien conforte le choix de l'algorithme SSU semi quadratique de GR : l'étude menée en tomographie axiale dans [Erdogan et Fessler, 1999, Sec. IV] montre que cet algorithme est d'un coût informatique faible et qu'il bénéficie d'une convergence rapide vers la solution pénalisée du problème de reconstruction. Il est alors légitime d'espérer que cet algorithme ait de bonnes performances pour le problème de reconstruction hélicoïdale.

Mise en œuvre

Avant de conclure, nous présentons le détail du coût informatique de cet algorithme et rapportons certains éléments constatés expérimentalement sur sa mise en œuvre sur le problème de grande taille traité lors des tests sur données synthétiques.

Coût de mise en œuvre

Le coût de calcul pour une itération de cet algorithme (*i.e.* pour un balayage complet) est particulièrement faible : il correspond à N fois le nombre d'opérations élémentaires associées à chaque relaxation. Pour le système à dix voisins adopté

dans cette étude, le coût de calcul d'une relaxation est principalement conditionné[13] par les opérations matricielles impliquant $\boldsymbol{H}$. Dans l'implantation, seule $\boldsymbol{H}_r$ a été effectivement stockée de manière à limiter le coût de stockage. Le calcul de chacune des relaxations se fait alors par une indexation à l'intérieur de $\boldsymbol{H}_r$ et une gestion des effets de bord.

Pour la relaxation du n-ème voxel, le calcul du premier terme du dénominateur équivaut au produit scalaire $\langle \boldsymbol{h}_n, \boldsymbol{h} \rangle$, $\boldsymbol{h}_n$ étant la n-ème colonne de $\boldsymbol{H}$. $\boldsymbol{h}_n$ est *très creux* : le nombre d'éléments non nuls étant $2M_n$, avec M_n le nombre de rayons intersectant le voxel n au cours de l'échantillonnage hélicoïdal, ce produit scalaire représente $2 \times 2M_n$ opérations élémentaires. Le calcul du premier terme du dénominateur s'écrit :

$$\langle \boldsymbol{h}, \boldsymbol{\epsilon}^{(k)} \rangle \qquad \text{avec} \qquad \boldsymbol{\epsilon}^{(k)} = (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^{(k)}),$$

pour $\boldsymbol{\epsilon}^{(k)}$ connu, le produit scalaire représente là encore $2 \times 2M_n$ opérations élémentaires. Une fois la mise à jour de x_n effectuée, on met à jour $\boldsymbol{\epsilon}^{(k)}$ simplement par deux additions matricielles de M termes en se rappelant que $\boldsymbol{x}^{(k+1)}$ et $\boldsymbol{x}^{(k)}$ diffèrent

[13]Le coût associé à la partie pénalisation est finalement assez faible : en guise d'illustration, intéressons nous au coût introduit par $[\boldsymbol{D}\boldsymbol{L}^{(k)}\boldsymbol{D}^T]_{n,n}$ dans $b_{\text{GR}}^{(k)}$: la ligne n de $\boldsymbol{V}$ étant constituée d'autant d'entrées non nulles que le voxel n possède de voisins et chacune de ces entrées étant ± 1, le calcul du second terme de $b_{\text{GR}}^{(k)}$ se réduit à 10 multiplications et autant d'additions. Reste à mettre à jour les 10 valeurs de $\phi'(u)/u$ impliquant la composante x_n : ceci représente pour $\phi(u) = \sqrt{u^2 + s^2}$ quelques opérations élémentaires supplémentaire et une racine carrée.

en un seul site :

$$\boldsymbol{\epsilon}^{(k+1)} = \boldsymbol{\epsilon}^{(k)} - \boldsymbol{h}(x_n^{(k+1)} - x_n^{(k)}).$$

Finalement, si on suppose que l'ensemble des voxels bénéficient tous du même nombre de rayons, on en déduit que le coût total de l'algorithme par balayage dans l'image est de l'ordre de :

$$N(8M_n + 2N) \qquad \textit{(opérations élémentaires).}$$

L'espace mémoire requis pour la mise en œuvre est également réduit dans la mesure où on a supposé que la matrice $\boldsymbol{H}_r$ permet de décrire complètement $\boldsymbol{H}$. Le nombre d'éléments non nuls dans $\boldsymbol{H}_r$ est identique à celui de deux opérateurs de projection axiaux identiques. Si on ajoute l'espace nécessaire au stockage de $\boldsymbol{\epsilon}^{(k)}$, de $\boldsymbol{x}^{(k)}$ et des C rapport $\phi'(\delta_c)/\delta_c$, l'espace requis correspond à :

$$M + C + N + N_{nz} \qquad \textit{(variables)}$$

où M_{nz} est le nombre d'éléments non nuls dans $\boldsymbol{H}_r$. À l'usage, le stockage de $\boldsymbol{H}_r$ s'avère néanmoins délicat pour des images de plus grande taille. Une solution possible pour économiser d'avantage d'espace mémoire est de faire appel au jeux des symétries comme nous le suggérions page 170 de manière à éliminer un maximum de redondance dans $\boldsymbol{H}_r$.

Reconstruction du fantôme $127 \times 127 \times 40$

La matrice d'observation hélicoïdale ayant été construite à partir d'opérateurs axiaux, on s'attend à ce que l'algorithme hérite de la bonne vitesse de convergence reconnu à ces algorithmes. Les relaxations dans l'image sont effectuées de manière lexicographique jusqu'à convergence[14] ; le critère d'arrêt adopté est un simple test sur la décroissance du critère J après un balayage complet :

$$J(\boldsymbol{x}^{(k-N)}) - J(\boldsymbol{x}^{(k)}) < tol \qquad \text{pour } k/N \text{ entier positif;}$$

Le test implanté pour obtenir les reconstructions de la figure 7.14 correspond à $tol = 10^-2$. Pour une initialisation par un vecteur uniformément nul et une mise en œuvre sur relaxée ($\theta \approx 1,7$), le test d'arrêt est activé après plus de 400 balayages complets dans l'image, ce qui représente approximativement 20 minutes de temps de calcul sur notre machine[15]. Nous avons néanmoins constaté expérimentalement qu'à peine plus d'une soixantaine d'itérations étaient nécessaires pour obtenir une solution de bonne qualité sur cet exemple (ce qui correspond à un seuil $tol \approx 100$ atteint en un peu plus de 160 secondes d'exécution). En guise d'illustration, la figure 7.14 présente les éléments du volume extraits à la 70-ème itération de l'algorithme : la coupe 19 ainsi que les trois contours sont d'une qualité comparable aux éléments

[14]Nous sommes néanmoins conscient de la faible efficacité de ce balayage récemment constatée en tomographie axiale, et un balayage aléatoire devrait être testé à court terme.

[15]Processeur AMD ATHLON XP 1900+ cadencé à 1600 MHZ et muni de 256 Koctets de cache.

de la figure 7.10. Ce résultat nous suggère que les composantes haute fréquence (en particulier les bords) convergent rapidement comme nous l'espérions. On indique enfin que nous n'avons pas constaté d'amélioration de vitesse de convergence en initialisant l'algorithme par une RPC : il semble que les artefacts produits par le faible nombre de projections disponibles dans la RPC soient difficilement résorbés par la mise à jour SOR — cf. figure 7.14.

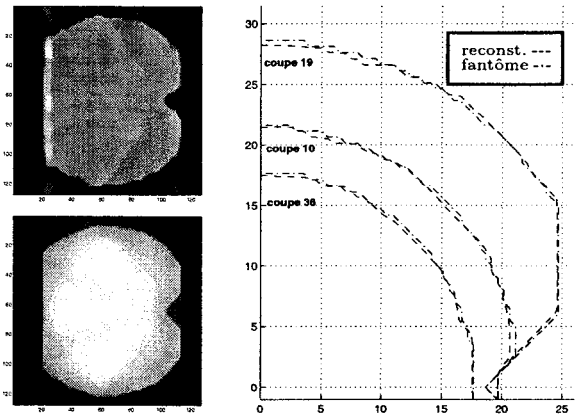


Figure 7.14: Reconstruction de la coupe 19 par approche pénalisée $\ell_2\ell_1$ après 70 itération de l'algorithme ($\leftarrow$) pour une initialisation par RPC (haut) et par une image uniformément nulle (bas) ; tracés segmentées à 0,7 extraits des coupes 10, 19 et 36 obtenues après 70 itération pour l'initialisation par un vecteur nul.

7.5 Conclusion

Dans ce chapitre, nous avons construit un modèle d'observation de type « algébrique » pour la géométrie hélicoïdale à partir de modèles déjà existant en géométrie axiale. Nous avons également montré qu'une forme paramétrique de ce modèle pouvait être adoptée pour restreindre les impératifs de stockage sans toutefois porter

atteinte à l'intégrité du modèle.

Finalement, nous avons montré que l'inversion pénalisée en géométrie hélicoïdale a une formulation similaire à celle retenue en tomographie axiale. Des résultats prometteur sur données synthétiques permettent de penser qu'un gain sensible en robustesse et en résolution peut être atteint tant dans le plan de l'image que suivant l'axe du tomographe. Ces résultats de simulation méritent néanmoins d'être étayés par des essais complémentaires sur données réelles.

Sur le plan de la mise en œuvre, le problème de reconstruction en géométrie hélicoïdale se révèle nettement plus épineux qu'en géométrie axiale. Le cadre intrinsèquement 3D de l'échantillonnage conduit en effet à un problème de minimisation de très grande taille qui rend prohibitive toute tentative standard de minimisation — ex. algorithme à direction de descente. Dans cette étude, nous avons mis en œuvre un algorithme SSU semi quadratique de GEMAN et REYNOLDS qui s'identifie à un schéma itératif SOR sur le critère initial J particulièrement économe en volume de calcul et en espace mémoire. D'autre part, la vitesse de convergence de l'algorithme reste intéressante sur l'exemple de grande taille traité dans ce chapitre ; il semble que la mise en œuvre hélicoïdale « hérite » du bon comportement constaté en tomographie axiale des schémas de relaxation. Une étude plus poussé reste néanmoins nécessaire pour permettre de confirmer que cet algorithme est concurrentiel par rapport aux autres schémas SOR adoptés dans la littérature ; voir à ce sujet [Erdogan et Fessler, 1999; Sotthivirat et Fessler, 2002].

Avant de clore ce chapitre, nous souhaitons souligner qu'en dépit du faible coût de cet algorithme, le traitement de « vraies » images 3D médicales reste encore difficile. L'exemple 3D de « grande taille » traité ici reste encore peu représentatif de la réalité : la taille des images adoptées est typiquement plus importante et le nombre de mesure est lui aussi bien plus élevé. La difficulté majeure réside dans le stockage de $\boldsymbol{H}_r$ qui peut à lui seul saturer la mémoire disponible : à titre d'exemple, pour une image de 256×256 et telle que $N \approx M$, l'opérateur axial $\boldsymbol{R}$ permettant de construire $\boldsymbol{H}_r$ occupe plus de 250 Méga-octets d'espace. Dans un tel contexte, deux solutions peuvent être envisagées :

1. conformément à la remarque faite page 170, $\boldsymbol{H}_r$ reste une description redondante de $\boldsymbol{H}$, et le nombre d'éléments non nuls stocké peut encore être réduit d'un facteur huit ;
2. l'inversion pénalisée approchée décrite page 173 « découpe » le problème 3D en une série de petit problèmes de volume réduit ; cette démarche peut être adoptée pour estimer le volume d'intérêt.

PARTIE IV

Minimisation des critères pénalisés

CHAPITRE 8

MINIMISATION DES CRITÈRES PÉNALISÉS

Les deux chapitres précédents ont montré que les approches pénalisées étaient des alternatives intéressantes pour la reconstruction d'image en tomographie hélicoïdale. Le problème d'imagerie se retrouve ainsi transformé en un problème de minimisation éventuellement sous contrainte et généralement de grande taille. De manière plus générale, les approches pénalisées en restauration ou en reconstruction d'image conduisent à un problème similaire et l'obtention d'une solution pénalisée passe par une étape d'optimisation qui est un problème à part entière. Même en l'absence d'obstacle méthodologique (ex. problème convexe et continûment différentiable), les dimensions du problème numérique rendent délicate l'implantation de méthodes d'optimisation standard, et une étude minutieuse de la structure du problème est généralement nécessaire pour aboutir à des mises en œuvre finalement réalistes.

Le présent chapitre est dédié à l'étude du problème d'optimisation des critères pénalisés. Le nombre de variantes associées à l'implantation d'une seule méthode étant généralement important, nous n'aborderons que très superficiellement les questions d'implantation dans ce chapitre[1]. À côté des méthodes *primales* qui

[1]Dans la mesure du possible, ceci permet également de dissocier la méthode d'optimisation adoptée de l'algorithme numérique employé pour la mettre en œuvre.

sont les plus connues et qui cherchent à minimiser directement le critère J, nous présentons largement les formulations semi quadratiques également appelées formulations « augmentées ». Ces dernières donnent une formulation primale/duale du problème initial qui est, du moins formellement, plus aisée à résoudre. Nous terminons ce chapitre en montrant que ces algorithmes semi quadratiques sont principalement identiques à des algorithmes de minimisation à direction de descente et à pas fixe sur le critère primal J. Cette équivalence peut alors servir d'outil afin d'étudier le comportement des algorithmes de formes primales/duales.

Note : Ce chapitre traitant des problématiques de minimisation, nous avons décidé d'alléger nos notations en ne faisant plus apparaître la dépendance aux paramètres libres du critère — *i.e.* les hyperparamètres. Nous espérons ainsi ne pas surcharger les expressions et simplifier la lecture.

8.1 Formulation primale du problème

Dans le cadre d'une inversion pénalisée, on rappelle que nous cherchons un minimiseur $\boldsymbol{x}^\star$ tel que

$$\boldsymbol{x}^\star \in \{\boldsymbol{x} \in X : \min J(\boldsymbol{x})\} \tag{8.1}$$

avec $X \subseteq \mathbb{R}^N$ un ensemble de contraintes *séparables* (ex. $X = \mathbb{R}^N_+$) et où $J : \mathbb{R}^N \to \mathbb{R}$ est un critère pénalisé de la forme

$$J(\boldsymbol{x}) = Q(\boldsymbol{x}) + \Phi(\boldsymbol{x}). \tag{8.2}$$

où Q est une forme *quadratique convexe*

$$Q(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{Q}\boldsymbol{x} \rangle - 2\langle \boldsymbol{q}, \boldsymbol{x} \rangle + \mu \tag{8.3}$$

avec $\boldsymbol{Q} \in \mathbb{R}^{N \times N}$ est une matrice définie non négative, $\boldsymbol{q} \in \mathbb{R}^N$ et $\mu \in \mathbb{R}$. D'autre part, on supposera par la suite que la pénalisation Φ prend la forme suivante :

$$\Phi(\boldsymbol{x}) = \sum_{i=1}^{I} \phi_i(\langle \boldsymbol{v}_i, \boldsymbol{x} \rangle - w_i), \qquad \boldsymbol{v}_i \in \mathbb{R}^N,\ w_i \in \mathbb{R} \tag{8.4}$$

avec $\phi_i : \mathbb{R} \to \mathbb{R}$ une fonction dont les propriétés restent à définir. Par la suite, nous supposerons que J est *strictement convexe, coercif, et* C^1, ce qu'on garantit en général sous les hypothèses suivantes :

$$\ker\{\boldsymbol{Q}\} \cap \ker\{\boldsymbol{V}^T\} = \{\boldsymbol{0}\}, \tag{8.5a}$$

$$1 \leq i \leq I, \quad \phi_i \textit{ strictement convexes et } \mathrm{C}^1 \tag{8.5b}$$

où on a posé $\boldsymbol{V} = (\boldsymbol{v}_1|\cdots|\boldsymbol{v}_I)$. Nous souhaitons souligner ici que ce cadre est suffisamment réaliste pour traiter une large classe de problèmes appliqués à l'imagerie ou au traitement de données ; un certain nombre de ces applications sont maintenant illustrées.

Quelques formulations appliquées

Le problème d'inversion en tomographie du chapitre 6 conduit à la minimisation (éventuellement sous contrainte de positivité) d'un critère de type « moindres carrés pénalisé »

$$J(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{R}\boldsymbol{x}\|^2 + \alpha \sum_{c=1}^{C} \phi(\langle \boldsymbol{d}_c, \boldsymbol{x} \rangle), \qquad \alpha \geq 0 \tag{8.6}$$

où $\boldsymbol{R} \in \mathbb{R}^{M \times N}$ est un opérateur d'observation linéaire reliant l'image aux données de projection $\boldsymbol{y}$, et $\langle \boldsymbol{d}_c, \boldsymbol{x} \rangle$ la différence entre valeurs constituant la c-ème paire d'éléments voisins dans l'image. Un tel problème s'inscrit clairement dans la formulation (8.2–4) en posant $Q(\boldsymbol{x}) = ||\boldsymbol{y}_h - \boldsymbol{R}\boldsymbol{x}||_2^2$, $I = C$ et

$$1 \leq c \leq C, \qquad \boldsymbol{v}_c = \boldsymbol{d}_c, \quad w_c = 0, \quad \phi_c = \alpha\phi$$

et on montre sans difficulté que les conditions (8.5) sont réalisées si ϕ est strictement convexe et C^1. Notons que la formulation (8.2–4) ne se réduit pas au cadre des « moindres carrés » : un modèle d'observation poissonnien peut aisément être

adopté pour le problème de reconstruction en posant $Q(\boldsymbol{x}) \equiv 0$, $I = M + C$ et

$$1 \leq i \leq M, \qquad \boldsymbol{v}_i = \boldsymbol{r}_i, \quad w_i = 0, \quad \phi_i : u \mapsto uy_i + N_0 e^u + \text{ constante}$$

$$M+1 \leq i \leq M+C, \qquad \boldsymbol{v}_i = \boldsymbol{d}_i, \quad w_i = 0, \quad \phi_i : u \mapsto \alpha\phi,$$

avec y_i le nombre de photons X ayant atteint le i-ème capteur, et avec $\boldsymbol{r}_m$ et N_0 respectivement la m-ème ligne de $\boldsymbol{R}$ et le nombre total de photons X émis par la source ; cf. section 6.1-6.1.2.

Enfin, la formulation adoptée permet également de prendre en considération certains problèmes liés à la statistique robuste [Alliney et Ruzinsky, 1994; Yarlagadda *et al.*, 1985]. Considérons simplement à titre illustratif le problème de *prédiction linéaire* décrit dans [Yarlagadda *et al.*, 1985] : étant donné des observations $\boldsymbol{y} = (y_1, \cdots, y_M) \in \mathbb{R}^M$, on cherche un minimiseur $\boldsymbol{x}^\star$ de

$$J(\boldsymbol{x}) = ||\boldsymbol{y} - \boldsymbol{Y}\boldsymbol{x}||_p^p, \qquad \infty > p > 1 \tag{8.7}$$

avec $\boldsymbol{Y} \in \mathbb{R}^{M \times N}$ la matrice dont chaque ligne s'écrit $\boldsymbol{y}_m = (y_m, \cdots, y_{m-M+1})$. Ce critère se met sous la forme définie par (8.2–4) avec $Q \equiv 0$, $I = M$, et

$$\forall m, \; w_m = y_m, \quad \boldsymbol{v}_m = \boldsymbol{z}_m, \quad \phi_m : u \mapsto |u|^p.$$

Condition nécessaire et suffisante d'optimalité

On rappelle qu'un minimum local strict $\boldsymbol{x}^\star$ d'une fonction $J : \mathbb{R}^N \to \mathbb{R}$ de classe C^1 vérifie *nécessairement* la condition d'optimalité du première ordre (abrég. CN1), [Bertsekas, 1995, Prop. 2.1.2]

$$\forall \boldsymbol{x} \in X, \qquad \langle \boldsymbol{\nabla} J(\boldsymbol{x}^\star), (\boldsymbol{x} - \boldsymbol{x}^\star) \rangle \geq \mathbf{0}.$$

Un point vérifiant la CN1 est également appelé *point stationnaire.* On note que dans le cas non contraint $X = \mathbb{R}^N$ ou si $\boldsymbol{x}^\star$ est un point strictement intérieur à X, alors la condition ci-dessus équivaut simplement à $\boldsymbol{\nabla} J(\boldsymbol{x}^\star) = \mathbf{0}$. Alors qu'en général on ne dispose pas de condition nécessaire *et* suffisante permettant de garantir qu'un point stationnaire est un optimum de J (c.à.d. un minimum local de J), la convexité de J permet de s'affranchir de cette difficulté : pour ces fonctions, un point stationnaire unique existe (c'est le minimum global) et la CN1 devient une condition nécessaire *et* suffisante. Cette garantie est intéressante puisque la recherche du minimum *global* peut s'appuyer sur les caractéristiques *locales* de la fonction objectif. D'autre part, la nature coercive de l'objectif[2] garantit que le minimiseur $\boldsymbol{x}^\star$ est atteint dans un compact.

[2]Une fonction J est norme-coercive sur $\mathbb{R}^N$ si elle « tend vers l'infini à l'infini » : $\lim_{||\boldsymbol{x}|| \to \infty} J(\boldsymbol{x}) = +\infty$.

Recherche itérative de la solution

Même dans le cas où le problème d'optimisation défini par (8.1–2) possède une solution explicite, la taille du problème numérique empêche généralement toute résolution directe et le calcul du minimiseur passe très souvent par une mise en œuvre *itérative*. Succinctement, on construit une application $\mathcal{M} : \mathbb{R}^N \to \mathbb{R}^N$ appelée *algorithme* ou *schéma itératif* qui, à partir de $\boldsymbol{x}^{(0)}$ donné initialement, génère une suite d'itérée $\{\boldsymbol{x}^{(k)}\}_{k=0,1,\ldots}$ suivant

$$\boldsymbol{x}^{(k+1)} = \mathcal{M}(\boldsymbol{x}^{(k)}).$$

Notre objectif est que la suite $\{\boldsymbol{x}^{(k)}\}_{k=0,1,\ldots}$ *converge globalement* vers le minimiseur $\boldsymbol{x}^\star$ sous des conditions plus ou moins restrictives. Au regard de l'utilisateur, deux points sont alors important :

(1) *l'algorithme amène t-il à la solution en peu d'itérations ?*

(2) *son coût d'implantation informatique par itération est-il avantageux ?*

Le premier point est lié aux propriétés de convergence de l'algorithme : c'est un indicateur intrinsèque à la méthode d'optimisation indépendant des choix pratiques d'implantation. Une fois la méthode choisie, le second point illustre la « consommation » des ressources informatiques associées à l'implantation — ex. mémoire et nombre d'opérations élémentaires. Ces deux aspects sont rarement indépendants et

prennent d'autant plus d'importance que le problème est de grande taille. Nous les gardons donc en mémoire pour juger de la pertinence d'une méthode d'optimisation et d'une implantation.

8.2 Algorithmes de minimisation des critères pénalisés

Cette section se propose de faire un tour d'horizon des différents algorithmes d'optimisation employés pour la minimisation des critères pénalisés. Cette revue ne prétend néanmoins pas à l'exhaustivité et nous sert plus de revue de littérature dans le domaine qui touche à notre activité : le traitement d'image. À ce propos, à coté des algorithmes classiques, d'autres approches, peu ou pas utilisées dans notre domaine, mériteraient d'être étudiées. On pense par exemple aux approches spécialement conçues pour traiter les problèmes de très grande taille, et particulièrement aux différentes techniques de *décomposition* — ex. décomposition de DANTZIG–WOLF ou de BENDERS, voir par exemple [Avriel, 1976, p. 477] et [Bertsekas, 1999, p. 524] — et les méthodes tirant partie de la *séparabilité de la fonction objectif* [Nocedal et Wright, 2000, Sec. 9.4]. Une étude sérieuse de leur utilité dans le domaine reste certainement un objectif à moyen terme.

Indiquons finalement que les méthodes de minimisation présentées ci-dessous sont en général inadaptées si J n'est pas continûment différentiable[3] ; le lecteur

[3] À titre informatif, on précise que certaines contributions du domaine s'intéressent à la minimisation d'un critère non différentiable ; voir par ex. [Alliney et Ruzinsky, 1994; Ciuciu *et al.*,

pourra se reporter à [Bonnans *et al.*, 1997, Sec. 6] pour un exposé de techniques d'optimisation non différentiable.

8.2.1 Algorithmes de relaxation

Dans son principe, la relaxation consiste à fragmenter le problème de minimisation initial suivant ses variables en une série de sous problèmes qui seront faciles à résoudre.

Relaxation par coordonnée ou par bloc [Bertsekas, 1995, Sec. 2.7]

Dans sa forme la plus simple, l'algorithme de relaxation minimise l'objectif en effectuant une récursion sur les composantes de $\boldsymbol{x}$ et en minimisant le critère *monovarié* $J_n(u)$ associé à la variable courante x_n mise à jour :

$$x_n^{(k+1)} = \arg\min_{u \in X} J_n(u) \tag{8.8}$$

où

$$J_n(u) = J(x_1^{(k+1)}, \cdots, x_{n-1}^{(k+1)}, u, x_{n+1}^{(k)}, \cdots, x_N^{(k)});$$

une itération complète $k \longrightarrow k+1$ est alors obtenue après un balayage complet des N composantes de $\boldsymbol{x}$. On notera que pour J vérifiant (8.5), la relation (8.8) a bien un sens et conduit à une valeur finie. Ce schéma itératif est largement connue dans

2001; Li et Santosa, 1996].

la littérature sous le nom de *méthode de* GAUSS SEIDEL (abrég. GS).

En pratique, on fait plutôt appel à une méthode légèrement différente appelée *sur relaxation successive* (abrég. SOR) ; l'opération consiste à s'arrêter en amont ou en aval du point minimisant J_n à l'itération courante :

$$\begin{aligned} \bar{x}_n^{(k+1)} &= \arg\min_{u \in X} J_n(u) \\ x_n^{(k+1)} &= x_n^{(k)} + \theta(x_n^{(k)} - \bar{x}_n^{(k+1)}); \end{aligned} \tag{8.9}$$

où $\theta > 0$ est le *paramètre de relaxation.* Cette modification permet souvent d'accélérer la convergence au prix d'un encombrement mémoire légèrement plus important.

Un intérêt de cette mise à jour coordonnée par coordonnée est de permettre d'imposer des contraintes *séparables* sans problème méthodologique particulier : par exemple, la contrainte de positivité des x_n passe simplement par la mise à zéro d'une composante si sa mise à jour est négative.

Notons que les méthodes de relaxation ne sont pas limitées à la mise à jour coordonnée par coordonnée, mais qu'elles peuvent optimiser par *blocs de coordonnées.* Dans ce cas, chaque itération conduit à un sous problème d'optimisation multivarié qui doit être résolu éventuellement sous contraintes. On notera qu'une condition suffisante de convergence globale de ces algorithmes est la convexité stricte de l'objectif suivant chaque coordonnée (ou chaque bloc de coordonnées).

Charge calculatoire et vitesse de convergence

L'intérêt principal de ce type de méthodes est leur facilité de mise en œuvre : une itération complète requiert N recherches linéaires indépendantes — *i.e.* minimisations d'une fonction monovariée — qui peut être mise en œuvre facilement sans calculer de dérivée (ex. par une technique de type *section dorée* ; cf. [Bertsekas, 1995, App. C]). Incidemment, il conduit à des algorithmes d'un coût d'implantation très réduit appréciés pour l'optimisation des très grands systèmes.

Le principal inconvénient de ces schémas itératifs réside dans leur taux de convergence qui est comparable à celui d'un simple algorithme du gradient [Bertsekas, 1995, p. 144]. Il existe cependant deux configurations dans lesquels ces algorithmes seront particulièrement intéressants [Luenberger, 1973] :

1/ si les variables (ou blocs de variables) sont très fortement *découplées*[4], alors la convergence sera accélérée ;

2/ si la structure du problème permet une optimisation immédiate (*ex.* expression explicite) pour la composante courante, alors la mise en œuvre ne nécessite plus de recherche linéaire, ce qui simplifie le schéma itératif.

Notons que l'ordre dans lequel la relaxation est mise en œuvre a une influence parfois sensible sur la vitesse de convergence. La partition adoptée dans une ver-

[4]C.à.d. que dans le cas où J est deux fois continûment différentiable, le Hessien de J est quasiment diagonal (ou bloc diagonal si on considère la relaxation par bloc de coordonnées).

sion relaxation par « bloc » a également un impact et, intuitivement, il semble souhaitable de suivre une certaine partition « naturelle » du problème — c.à.d. partitionner en suivant la structure du Hessien.

Emploi dans le domaine signal/image

Ces algorithmes restent principalement employés pour les problèmes de grande taille : ainsi dans [Bouman et Sauer, 1996; Erdogan et Fessler, 1999; Sauer et Bouman, 1993; Zheng *et al.*, 2000], les auteurs utilisent la relaxation sur les composantes pour minimiser des critères pénalisés en reconstruction d'images tomographique. Notons également que la relaxation est également adoptée pour résoudre approximativement les systèmes linéaires découlant, par exemple, d'un algorithme à direction de descente (cf. notre exposé en section 8.2.3).

8.2.2 Algorithmes de type POCS

La minimisation du critère pénalisé peut, sous certaines conditions, s'appuyer sur la méthode de BREGMAN : la méthode est de type *projective sur des ensembles convexes* (abrég. POCS) et produit des schémas itératifs de « relaxation sur les données » dont les plus connus sont les algorithmes ART employés en tomographie pour la minimisation non contrainte des critères *quadratiques* avec pénalisation

séparable[5][Fiani, 2001, Annexe D].

L'extension de la méthode de BREGMAN à des critères pénalisés « moins simplistes » reste néanmoins délicate sans le recours à la formulation semi quadratique que nous développerons en section suivante ; nous renvoyons à [Fiani, 2001, Sec. IV.3-4] pour une revue détaillée des algorithmes de type BREGMAN en tomographie et le détail des développements ayant mené l'extension aux critères convexes non séparables.

8.2.3 Algorithmes à directions de descente

Les algorithmes à direction de descente sont des schémas itératifs répandus en optimisation et largement employés pour la minimisation des critères pénalisés. Nous décrivons maintenant leurs principes et certaines des variantes les plus courantes dans le cas *non contraint* ; la prise en compte de contraintes séparables ne sera pas abordée explicitement mais nous signalons que ces techniques peuvent être en partie adaptées pour tenir compte de ces contraintes ; [Kaufman, 1987; Mumcuoglu *et al.*, 1994].

Forme générale du schéma itératif

À l'itération courante k, la mise à jour s'écrit

[5]Une fonction $f : \mathbb{R}^N \to \mathbb{R}$ est séparable si il existe une décomposition du type $f(\boldsymbol{x}) = \sum_n f_n(x_n)$.

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \theta^{(k)}\boldsymbol{\xi}^{(k)}, \quad \theta^{(n)} > 0 \tag{8.10}$$

avec $\boldsymbol{\xi}^{(k)} \equiv \boldsymbol{\xi}\left(\boldsymbol{x}^{(k)}\right)$ et $\theta^{(k)} \equiv \theta\left(\boldsymbol{x}^{(k)}\right)$ respectivement la direction de déplacement et la longueur du pas à l'itération courante. Les directions de déplacement successives $\boldsymbol{\xi}^{(n)}$ sont souvent la solution d'un système linéaire de la forme :

$$\boldsymbol{B}^{(k)}\boldsymbol{\xi}^{(k)} = -\boldsymbol{\nabla} J(\boldsymbol{x}^{(k)}), \tag{8.11}$$

où on pose $\boldsymbol{B}^{(k)} = \boldsymbol{B}\left(\boldsymbol{x}^{(k)}\right)$ avec $\boldsymbol{B} : \mathbb{R}^N \to \mathbb{R}^{N\times N}$ un opérateur défini positif (DP) ; le caractère DP assure d'une part l'existence et l'unicité d'une solution de (8.11) et d'autre part que $\boldsymbol{\xi}^{(k)}$ fera décroître strictement l'objectif ($J(\boldsymbol{x}^{(k)}) > J(\boldsymbol{x}^{(k+1)})$ si $\theta^{(k)}$ est choisi suffisamment petit.

Convergence globale

Garantir la décroissance stricte ne suffit pas à garantir la convergence globale de ces algorithmes. Il faut également s'assurer que les directions produites ne tendent pas à devenir perpendiculaires avec le gradient ou que leur module ne devient pas arbitrairement petit. Ces conditions permettent de définir la classe des algorithmes dits « gradient reliés » [6] dont on peut montrer qu'ils sont globalement convergents si ils sont associés à une recherche linéaire adéquate. Notons que si

[6] Nous reviendrons sur cette notion au début du chapitre suivant.

$\boldsymbol{B}^{(k)}$ est uniformément DP et que ses valeurs propres sont bornées supérieurement, alors les directions produites sont gradient reliées : c'est le cas de l'algorithme de plus profonde descente, de certaines formes quasi newtoniennes et des algorithmes du gradient conjugué sous certaines restrictions — cf. [Nocedal et Wright, 2000, p.121]. Notons également que cette convergence globale peut être garantie sous des conditions techniques supplémentaires pour un pas $\theta^{(n)}$ constant indépendant de l'itération ; nous reviendrons largement sur ce point au chapitre suivant.

Quelques exemples types

La vitesse de convergence et le coût d'implantation varient sensiblement suivant l'algorithme à direction de descente considéré. Dans la mesure où certains d'entre eux sont largement utilisés dans le domaine de l'inversion et du traitement d'image, nous passons en revue les plus répandus, soit les algorithmes

(i) de *plus forte descente* (abrég. SD),

(ii) du *gradient conjugué* (abrég. GC),

(iii) du *gradient conjugué préconditionné* (abrég. GCP),

(iv) de *Newton* (abrég. Nw),

(v) de *quasi Newton* (abrég. qNw).

Algorithme de la plus profonde descente

Cet algorithme produit une mise à jour à l'itération courante suivant la plus forte pente locale

$$\xi^{(k)} = -\nabla J(x^{(k)});$$

sa mise en œuvre nécessite uniquement d'évaluer le gradient de J au point courant ainsi qu'une recherche linéaire adéquate pour déterminer un pas admissible. L'algorithme de SD est connu pour sa convergence asymptotique lente en comparaison du gradient conjugué ou d'une méthode de Newton : l'ordre de convergence est *linéaire* et le taux de convergence peut être particulièrement mauvais si le problème est mal conditionné. Cependant, un bon nombre de techniques de restauration ou de reconstruction d'images s'appuient encore directement ou indirectement sur cet algorithme, ce qui fait encore de cet algorithme une référence. En pratique, on lui préfère souvent un algorithme du gradient conjugué avec ou sans préconditionnement qui permet une convergence plus rapide au prix d'un encombrement mémoire et d'un coût de calcul par itération légèrement supérieur.

Algorithme du gradient conjugué

Cet algorithme, conçu à l'origine pour la programmation quadratique convexe, a été étendu pour la programmation non quadratique au cours des années 1960. Il existe en fait plusieurs versions du GC, les plus connues étant les méthodes de

POLACK–RIBIÈRE et FLETCHER–REEVES. Les propriétés pratiques de ces différentes versions sont restées longtemps mal comprises ; elles le restent encore dans une certaine mesure — cf. [Nocedal et Wright, 2000, p. 127]. En particulier, alors que la convergence globale de la version de FLETCHER–REEVES est garantie et qu'un contre exemple à été trouvé à la convergence de la version de POLACK–RIBIÈRE, cette dernière est de loin la plus efficace et la plus utilisée en pratique. Notons enfin qu'il existe un lien fort entre le gradient conjugué de POLACK–RIBIÈRE et une variante de la méthode l-BFGS que nous présentons plus bas. Nous renvoyons à la monographie de [Bonnans *et al.*, 1997, Sec. 4.6] ou à celle de [Nocedal et Wright, 2000] pour un exposé détaillé et instructif de ces différents aspects. Notons enfin que la mise en oeuvre du GC peut nécessiter un redémarrage périodique afin d'éviter que les erreurs d'arrondi qui s'amplifient au cours des itérations n'empêchent la convergence ; on trouvera un critère permettant de commander ce redémarrage dans [Nocedal et Wright, 2000, p. 123].

On considère généralement le GC comme une alternative intéressante pour les problèmes de grande taille : son coût d'implantation[7] reste faible en regard des approches quasi Newtoniennes et sa vitesse de convergence est souvent intéressante. En particulier l'ordre de convergence peut être *quadratique* si $J(\boldsymbol{x}^\star)$ est

[7]Le coût de stockage correspond principalement à la mise en mémoire du gradient courant et de la direction de descente précédente (soit $2N$ éléments) ; le nombre d'opération élémentaire est majoritairement conditionné par le calcul du gradient courant et deux produits scalaires auquel il faut ajouter le coût de la recherche linéaire.

suffisamment régulière ; cf. [Nocedal et Wright, 2000, p. 123]. Indiquons que les performances de vitesse asymptotique peuvent être sensiblement améliorées si on effectue un *préconditionnement*. L'objectif consiste alors à appliquer le GC au travers d'une transformation linéaire sur les variables du problème afin que le Hessien transformé en la solution soit le plus proche possible d'une matrice diagonale. Il faut néanmoins noter que la construction d'un préconditionneur efficace reste pour une bonne part liée à l'application. Il est en particulier délicat de prévoir si le coût nécessaire à la construction d'un préconditionneur va être finalement compensé et amener à un gain finalement significatif.

Algorithme de NEWTON [Nocedal et Wright, 2000, Chap. 6]

L'algorithme de NEWTON calcule une direction de déplacement à partir du Hessien $\boldsymbol{H}^{(k)}$ de J au point courant :

$$\boldsymbol{H}^{(k)}\boldsymbol{\xi}^{(k)} = -\boldsymbol{\nabla} J(\boldsymbol{x}^{(k)}). \tag{8.12}$$

Pour J suffisamment régulière, ce choix de direction permet une convergence asymptotique *quadratique*. Cette méthode ne permet cependant pas d'assurer dans le cas général que $\boldsymbol{\xi}^{(k)}$ sera une direction de descente. En particulier, on peut légitimement s'interroger sur le cas (nécessairement non strictement non convexe) où l'inverse du Hessien est indéfini, où tout simplement singulier. D'autre part, qu'en est-il de la

convergence de l'algorithme si l'inverse devient singulière à mesure que l'algorithme progresse (objectif nécessairement non fortement[8] convexe) ? Dans toutes ces situations, l'algorithme de NEWTON a souvent un comportement pathologique et peut diverger ou cycler sans converger.

Dans tous ces cas, la méthode de NEWTON doit être modifiée de manière à être rendue « robuste ». En pratique, on cherchera par exemple à détecter les valeurs propres négatives lors de la résolution du système linéaire (8.12) par une décomposition de CHOLESKI ou un gradient conjugué ; une globalisation par région de confiance est également possible. D'autre part, l'algorithme de NEWTON nécessite le calcul du Hessien et la résolution d'un système linéaire à chaque itération. À partir de quelques centaines de variables, le coût informatique requis (charge de calcul et de stockage) est excessif. Les formes quasi newtoniennes ne calculant pas explicitement de dérivées secondes sont à ce titre intéressantes.

Formes quasi newtoniennes

La convergence globale vers un point stationnaire est une propriété incontournable pour un algorithme itératif. Les méthodes de type quasi NEWTON (qNw) offrent un compromis intéressant entre l'efficacité locale de la méthode de Newton et la sécurité des algorithmes de descente. Ces techniques se proposent de

[8] J est *fortement convexe* si les valeurs propres du Hessien sont positives et uniformément bornées au dessus de zéro.

construire une matrice $\boldsymbol{B}^{(k)}$ qui soit une approximation définie positive du Hessien de manière à garantir la décroissance à chaque itération. Les versions BFGS ou SR1 sont les versions les plus utilisées et permettent actuellement de traiter des problèmes allant jusqu'à plusieurs milliers de variables. Ces deux algorithmes construisent une approximation du Hessien à partir de l'information contenue dans les dérivées premières. Les mises à jour étant d'un coût calculatoire réduit, ces algorithmes permettent une réduction notable de la charge informatique en comparaison d'une version newtonienne. Nous proposons au lecteur de se reporter à [Nocedal et Wright, 2000, Sec. 2.8] pour plus de détails. Cette démarche conduit à des algorithmes bien plus robustes que la forme newtonienne pure, et certains résultats de convergence globale ont pu être établis ; en ce qui concerne l'ordre de convergence local, il est généralement super linéaire.

Ces méthodes deviennent néanmoins délicates à mettre en œuvre pour les problème de grande taille ; cependant un certain nombre d'alternatives existent. L'algorithme BFGS à *mémoire limitée* (abrég. l-BFGS) calcule par exemple la direction de descente en tenant implicitement compte d'une estimation du Hessien produite à partir d'un jeu de $p \ll N$ gradients et itérées consécutives. Cet algorithme de complexité réduite se révèle robuste en pratique, mais la vitesse de convergence subit généralement une nette détérioration par rapport à la version BFGS. Une autre approche qui semble prometteuse consiste à tirer parti de la séparabilité de la fonction objectif. Cette méthode peut mener à réduire considérablement le nombre de

variables effectivement impliquées dans la minimisation itérative. Pour une description détaillée de ces deux dernières méthodes, nous renvoyons le lecteur à [Nocedal et Wright, 2000, Sec. 9.4].

Emploi dans le domaine signal/image

Le lien entre l'algorithme *Expectation–Maximization* (abrég. EM) et un simple algorithme du gradient avec mise à l'échelle par une matrice diagonale est maintenant largement connu [Lange et Fessler, 1995, Sec. II.B] ou [Mumcuoglu *et al.*, 1994, Sec. III.C]. De nombreux travaux ayant pour objectif l'amélioration de l'algorithme EM ont également mené à des algorithmes de descente ; nous y reviendrons ultérieurement dans la mesure où un lien fort unit certains d'entre eux avec les algorithmes semi quadratiques que nous étudierons en détail dans ce chapitre et le suivant.

Le gradient conjugué (avec ou sans préconditionnement) est également apprécié dans le domaine pour son faible coût d'implantation [Fessler et Booth, 1999; Mumcuoglu *et al.*, 1994] ; on notera que cette méthode est également utilisée pour résoudre les systèmes linéaires découlant, par exemple, d'une méthode de descente.

Les algorithmes de type BFGS ou l-BFGS semblent peu employés pour les problèmes d'imageries et de traitement de signal. L'emploi de la mise à jour BFGS est évidemment difficile en imagerie puisqu'elle nécessite le stockage d'un opérateur de très grande taille à chaque itération : l'estimation courante de l'inverse du Hessien.

Le cas de la version l-BFGS reste par contre plus nuancé : le calcul de la direction de descente peut être effectuée à faible coût à partir des p paires stockées (de l'ordre de $4pN$ opérations), ce qui rend cet algorithme *a priori* très attractif; cf. [Nocedal et Wright, 2000, Alg. 9.1]. D'autre part, on rappelle que pour $p = 1$, l'algorithme l-BFGS est identique à la forme du gradient conjugué de POLAK-RIBIÈRE [Nocedal et Wright, 2000, p. 228] qui est largement répendue en traitement d'image. L'algorithme l-BFGS pour p allant de quelques unités (ex. 3) à une vingtaine est généralement reconnu comme une alternative de choix pour les problèmes de très grande taille.

8.2.4 Algorithmes semi quadratiques

Une partie importante de notre travail de recherche a consisté à réinterpréter et à étudier ces algorithmes; nous consacrons donc la prochaine section à ces constructions et aux algorithmes qui en découlent.

8.2.5 Autres algorithmes de minimisation

À coté de ces approches principalement développées dans le cadre classique de l'optimisation non contrainte, un certain nombre d'auteurs ont développé des approches distinctes qui méritent d'être citées. En particulier, une méthode de *point intérieur* de type « barrière » est présentées dans [Johnson et Sofer, 2000]

pour résoudre un problème contraint de tomographie d'émission de très grande taille. Une autre approche de type point intérieur est présentée dans [Li et Santosa, 1996] pour minimiser (sous contrainte de positivité) un critère non différentiable.

8.3 Formulations primales/duales « semi quadratiques »

Depuis les années 1990, l'optimisation des critères pénalisés est en partie traitée par les algorithmes « semi quadratiques » découlant des formulations de GEMAN et YANG (abrég. GY) ou de GEMAN et REYNOLDS (abrég. GR). Le principe consiste à minimiser à la place du critère *primal* $J(\boldsymbol{x})$ un critère primal/dual $J^*(\boldsymbol{x}, \boldsymbol{l})$ aux propriétés structurelles remarquables : il est *quadratique* pour les variables primales $\boldsymbol{x}$ et sa minimisation vis-à-vis des variables duales[9] $\boldsymbol{l}$ est *explicite*. De ces constructions découlent toute une famille d'algorithmes de relaxation qui sont globalement convergents sous des hypothèses techniques assez peu contraignantes en pratique.

8.3.1 Constructions « semi quadratiques »

Les approches semi quadratiques (abrég. SQ) se déclinent selon deux versions distinctes suivant qu'on considère la construction primale/duale de GR décrite dans [Geman et Reynolds, 1992] ou celle de GY décrite dans [Geman et Yang, 1995]. Du point de vue historique, il est certainement instructif de noter que ces constructions

[9]Le terme « variables auxiliaires » est également employé dans la littérature du traitement de signal et d'images.

ont été initialement introduites pour résoudre un problème d'optimisation globale sur un critère *non convexe* par un algorithme de « recuit simulé ». Aujourd'hui, ces formulations sont largement adoptées pour optimiser des critères *convexes* : ceci permet l'emploi d'algorithmes d'optimisation déterministes standard d'un coût informatique bien moins élevé.

Principe constructif

Moyennant certaines hypothèses techniques supplémentaires réalistes sur ϕ, ces deux constructions permettent de transformer la formulation *primale* du problème d'optimisation

$$\inf_{\boldsymbol{x}\in X} J(\boldsymbol{x})$$

en une formulation *primale/duale*

$$\inf_{\boldsymbol{x}\in X} \inf_{\boldsymbol{l}} J^*(\boldsymbol{x},\boldsymbol{l})$$

où J^* est un critère « semi quadratique » vérifiant

$$\inf_{\boldsymbol{l}} J^*(\boldsymbol{x},\boldsymbol{l}) = J(\boldsymbol{x})$$

et tel que **(i)** $J^*(\boldsymbol{x}, \boldsymbol{l})$ est *convexe*[10], **(ii)** $J^*(\boldsymbol{x}, \boldsymbol{l})$ est *quadratique* par rapport à $\boldsymbol{x}$, et **(iii)** la minimisation par rapport à $\boldsymbol{l}$ est immédiate. Ces deux constructions s'appuient sur les résultats de l'analyse convexe et sur la notion de *fonction concave conjuguée* [Rockafellar, 1970]. Une présentation complète de ces constructions sort du cadre de ce document, néanmoins un exposé concis s'avère ici nécessaire à la bonne compréhension de certains de nos travaux exposés par la suite ; pour cela, nous nous somme largement appuyé sur [Idier, 2001*b*]. Rappelons avant de poursuivre la forme primale $J(\boldsymbol{x})$ à minimiser éventuellement sous contrainte séparables :

$$J(\boldsymbol{x}) = Q(\boldsymbol{x}) + \Phi(\boldsymbol{x}) \quad \text{avec} \quad \Phi(\boldsymbol{x}) = \sum_{i=1}^{I} \phi_i(\delta_i - w_i), \tag{8.13}$$

où on a posé $\delta_i = \langle \boldsymbol{v}_i, \boldsymbol{x} \rangle$, et avec $Q(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{Q}\boldsymbol{x} \rangle - 2\langle \boldsymbol{q}, \boldsymbol{x} \rangle + \mu$ une forme *quadratique convexe.*

NB. De manière à ne pas nuire à la clarté de l'exposé, les développements qui suivent considèrent le cas *isotrope*, c.à.d. que pour $1 \leq i \leq I$, on pose $\phi_i = \phi$; l'extension au cas non isotrope est néanmoins directe et peut être nécessaire en pratique.

[10]Dans le cas où ϕ est convexe (respectivement. strictement convexe), on peut montrer que le critère augmenté de type GY est convexe (strictement convexe) en $(\boldsymbol{x}, \boldsymbol{l})$ et que le critère augmenté de type GR l'est également à un changement de variable sur $\boldsymbol{l}$ près ; cf. [Idier, 2001*b*].

1 – Construction d'après Geman et Reynolds

Sous réserve que ϕ soit paire, C^1 et que $\phi(\sqrt{\cdot})$ soit *concave* sur $\mathbb{R}^+$, le critère augmenté de GEMAN et REYNOLDS (abrég. GR) s'écrit [Idier, 2001*b*, Sec. IV],

$$J^*_{\mathrm{GR}}(\boldsymbol{x}, \boldsymbol{l}) = Q(\boldsymbol{x}) + \sum_{i=1}^{I} \big(l_i(\delta_i - w_i)^2 + \psi(l_i) \big)$$

avec $\boldsymbol{l} = (l_1, \cdots, l_I)^T$ et où on a posé $\delta_i = \langle \boldsymbol{v}_i, \boldsymbol{x} \rangle$. La fonction

$$\psi : l \mapsto \sup_{u \in \mathbb{R}} \big(\phi(u) - l u^2 \big)$$

est décroissante sur $]0; +\infty[$, et $l_i \geq 0$ est la variable duale associée à δ_i. Les résultats de l'analyse convexe permettent de donner une forme *explicite* de la valeur des variables auxiliaires minimisant J^*_{GR} à $\boldsymbol{x} = \bar{\boldsymbol{x}}$ fixé [Rockafellar, 1970, Chap. 26],

$$1 \leq i \leq I, \qquad l_i = \phi'(\bar{\delta}_i - w_i)/2(\bar{\delta}_i - w_i), \tag{8.14}$$

où $\bar{\delta}_i = \langle \boldsymbol{v}_i, \bar{\boldsymbol{x}} \rangle$; notons que cette quantité est définie quand $\phi''(0)$ existe et est bornée[11] — cf. lemme 4, page 364. Dans le cas *non contraint* $X = \mathbb{R}^N$, la forme quadratique en $\boldsymbol{x}$ à $\boldsymbol{l} = \bar{\boldsymbol{l}}$ permet d'autre part de déduire que le minimiseur, à $\boldsymbol{l}$

[11] On notera que ceci exclut la fonction ℓ_p $\phi = |\cdot|^p$ pour $1 < p < 2$.

fixé, doit vérifier l'équation normale suivante :

$$\left(\boldsymbol{Q}+\boldsymbol{V}\bar{\boldsymbol{L}}\boldsymbol{V}^T\right)\boldsymbol{x}=\boldsymbol{q}+\boldsymbol{V}\bar{\boldsymbol{L}}\boldsymbol{w}, \tag{8.15}$$

avec $\boldsymbol{V}=(\boldsymbol{v}_1|\cdots|\boldsymbol{v}_I)$ et $\bar{\boldsymbol{L}}=\mathrm{diag}(\bar{l}_1,...,\bar{l}_I)$.

2 – Construction d'après Geman et Yang

Sous réserve que ϕ permette de définir une fonction $g(u)\triangleq u^2/2-\phi(u)$ *convexe*[12], la formulation de GEMAN et YANG (abrég. GY) conduit à la construction primale/duale suivante [Idier, 2001*b*, Sec. III]

$$J^*_{\mathrm{GY}}(\boldsymbol{x},\boldsymbol{l})=\boldsymbol{Q}(\boldsymbol{x})+\sum_{i=1}^{I}\left[\frac{1}{2}\left(\delta_i-w_i-l_i\right)^2+\zeta(l_i)\right]$$

avec

$$\zeta: l\mapsto \sup_{u\in\mathbb{R}}\left(\phi(u)+(l-u)^2/2\right).$$

Comme pour la construction de GR, Il est possible de donner une forme explicite des variables auxiliaires minimisant J^*_{GY} à $\boldsymbol{x}$ fixé,

$$1\leq i\leq I,\qquad l_i=(\bar{\delta}_i-w_i)-\phi'(\bar{\delta}_i-w_i), \tag{8.16}$$

[12]On notera que cette condition ne permet pas la construction pour la fonction ℓ_p, $1<p<2$.

et comme pour la construction précédente, la forme du critère augmenté est quadratique en $\boldsymbol{x}$, ce qui implique que le minimiseur non contraint pour les pixels à $\boldsymbol{l}$ fixé doit vérifier l'équation normale suivante :

$$\left(2\boldsymbol{Q} + \boldsymbol{V}\boldsymbol{V}^T\right)\boldsymbol{x} = 2\boldsymbol{q} + \boldsymbol{V}(\bar{\boldsymbol{l}} + \boldsymbol{w}) \tag{8.17}$$

avec $\boldsymbol{w} = (w_1, \cdots, w_I)^T$; soulignons enfin que le membre de gauche ne fait pas intervenir les variables duales $\boldsymbol{l}$: cette particularité peut rendre la construction de GY attractive pour la mise en œuvre.

Intérêt d'un changement d'échelle sur ϕ

En pratique, on choisit généralement une formulation légèrement différente de celle présentée ci-dessus afin de rendre plus flexible l'hypothèse de convexité de g. En effet, on est très souvent amené à choisir ϕ dans une *famille paramétrée* $\phi(u\,;\,s)$ et l'ajustement de s dépend principalement du problème. Or, la convexité de $g(u\,;\,s) = u^2/2 - \phi(u\,;\,s)$ peut n'être assurée que pour certaines valeurs[13] de s, ce qui est clairement gênant en pratique.

Un moyen élégant de contourner cet obstacle est d'introduire un degré de liberté dans la formulation augmentée en construisant la fonction ζ à partir d'une pénalisation $\phi_a(\cdot\,;\,s)$ mise à l'échelle. Pour comprendre le mécanisme, notons que

[13] C'est par exemple le cas si on choisit $\phi(u\,;\,s) = \sqrt{u^2 + s^2}$ puisque $g(x\,;\,s)$ ne sera convexe que pour s pris dans $]0;2]$.

le critère primal $J(\boldsymbol{x})$ peut être récrit de manière équivalente :

$$J(\boldsymbol{x}) = Q(\boldsymbol{x}) + \frac{1}{a}\sum_{i=1}^{I} \phi_a\left(\delta_i - w_i\,;\, s\right),$$

avec $\phi_a(\cdot\,;\,s) \triangleq a\phi(\cdot\,;\,s)$. Sous réserve que

$$\exists a > 0 \;:\; g_a(u) \triangleq u^2/2 - \phi_a(u) \quad \textit{convexe},$$

alors la construction de GY peut être adoptée, non plus sur ϕ, mais sur ϕ_a. L'expression des minimiseurs suivant les variables primales et duales se déduit par une démarche similaire à celle présentée au paragraphe précédent :

$$\begin{aligned} \boldsymbol{l} &= (\boldsymbol{V}^T\bar{\boldsymbol{x}} - \boldsymbol{w}) - a\boldsymbol{\phi}'(\bar{\delta}_\bullet - w_\bullet), \\ \left(2\boldsymbol{Q} + \frac{1}{a}\boldsymbol{V}\boldsymbol{V}^T\right)\boldsymbol{x} &= 2\boldsymbol{q} + \frac{1}{a}\boldsymbol{V}(\bar{\boldsymbol{l}} + \boldsymbol{w}). \end{aligned} \tag{8.18}$$

où nous avons posé la notation $\boldsymbol{\phi}'(\delta_\bullet - w_\bullet) = (\phi'(\delta_1 - w_1), \cdots, \phi'(\delta_I - w_I))^T$. En pratique, un ajustement cohérent de a permet d'assurer la construction d'un critère de GY mis à l'échelle pour de nombreuses familles paramétrées $\phi(\cdot\,;\,s)$; cf. [Idier, 2001*b*, Sec. III.B]. Enfin, dans la mesure où cette mise à l'échelle rend robustes les algorithmes semi quadratiques découlant de la construction de GY, elle devrait toujours être adoptée en pratique. Dans ce qui suit, nous considérerons toujours

qu'un tel facteur d'échelle est intégré à la construction.

Extension multivariée

Alors que dans la construction de GY présentée jusqu'ici, la dualité était appliquée à chacune des fonctions monovariées ϕ, il est possible de l'appliquer à la fonction *multivariée* $\Phi = \sum \phi$. On donne maintenant certains éléments de cette construction qui nous seront utiles par la suite ; les détails peuvent être trouvés dans [Fiani, 2001, Annexe C].

À la condition que la fonction $\Phi(\boldsymbol{x})$ définie par (8.13) soit telle que

$$\exists a > 0 \ : \ h_a(\boldsymbol{x}) \triangleq ||\boldsymbol{x}||^2/2 - a\Phi(\boldsymbol{x}) \quad \text{convexe}$$

alors on peut définir le critère primal/dual semi quadratique $K_{\mathcal{V}} : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ suivant :

$$J^*_{\mathcal{V}}(\boldsymbol{x}, \boldsymbol{l}) = \boldsymbol{Q}(\boldsymbol{x}) + \frac{1}{2a}||\boldsymbol{x} - \boldsymbol{l}||^2 + \zeta_{\mathcal{V}}(\boldsymbol{l}) \tag{8.19}$$

avec

$$\zeta_{\mathcal{V}} : \boldsymbol{l} \in \mathbb{R}^N \mapsto \sup_{\boldsymbol{x} \in \mathbb{R}^N} \left(a\Phi(\boldsymbol{x}) - ||\boldsymbol{l} - \boldsymbol{x}||^2/2 \right).$$

Là encore, une expression explicite peut être donnée pour le minimiseur en $\boldsymbol{x}$ à $\boldsymbol{l}$ fixé, et pour le minimiseur en $\boldsymbol{l}$ à $\boldsymbol{x}$ fixé ; finalement nous donnons les expressions

de ces minimiseurs :

$$\boldsymbol{l} = \boldsymbol{x} - a\boldsymbol{V}\boldsymbol{\phi}'(\bar{\delta}_\bullet - w_\bullet).$$

$$\left(2\boldsymbol{Q} + \frac{1}{a}\boldsymbol{I}\right)\boldsymbol{x} = 2\boldsymbol{q} + \frac{\bar{\boldsymbol{l}}}{a},$$

On notera que comparativement à la construction standard, la forme de l'équation normale s'est simplifiée alors que la remise à jour des variables duales est rendue légèrement plus coûteuse en terme d'opérations élémentaires.

8.3.2 Algorithmes de relaxation semi quadratiques

La forme explicite des minimiseurs pour toutes ces constructions suggère une optimisation *via* des schémas de relaxation sur le critère primal/dual J^* — c.à.d. en alternant la descente successivement suivant les variables primales et duales. Nous décrivons ci-dessous un certain nombre de ces algorithmes appelés « semi quadratiques » dans la littérature du traitement d'images.

Relaxation groupée pour $\boldsymbol{x}$ et $\boldsymbol{l}$

C'est la forme la plus évidente : elle consiste simplement à suivre la partition naturelle du jeu de variables $(\boldsymbol{x}, \boldsymbol{l})$ et à effectuer la mise à jour de l'ensemble des variables primales puis de l'ensemble des variables duales. En supposant une ini-

tialisation $\boldsymbol{x}^{(0)}$, on définit le schéma itératif de type SOR permettant la transition $k \longrightarrow k+1$,

$$\begin{aligned} \boldsymbol{x}^{(k+1)} &= (1-\theta)\boldsymbol{x}^{(k)} + \theta\bar{\boldsymbol{x}}^{(k+1)} \qquad \theta \in]0;2[\\ \boldsymbol{l}^{(k+1)} &= (1-\gamma)\boldsymbol{l}^{(k)} + \gamma\bar{\boldsymbol{l}}^{(k+1)} \qquad \gamma \in]0;1[\end{aligned} \tag{8.20}$$

où γ et θ sont respectivement les *coefficients de relaxation* des variables primales et duales. Les mises à jour (abrég. m.à.j.) $\bar{\boldsymbol{x}}^{(k+1)}$ et $\bar{\boldsymbol{l}}^{(k+1)}$ dépendent alors de la construction choisie ; pour la construction de GR, on a d'après (8.14) et (8.15) :

$$\left(\begin{array}{c} \textit{M.à.j.} \\ \textit{groupée} \\ J^{*}_{\mathrm{GR}} \end{array}\right) \qquad \begin{aligned} \bar{\boldsymbol{x}}^{(k+1)} &= \left(\boldsymbol{B}^{(k)}_{\mathrm{GR}}\right)^{-1}(\boldsymbol{q} + \boldsymbol{V}\boldsymbol{L}^{(k)}\boldsymbol{w}), \\ \boldsymbol{l}^{(k+1)} &= \boldsymbol{\phi}'(\delta^{(k+1)}_{\bullet} - w_{\bullet})/2(\delta^{(k+1)}_{\bullet} - w_{\bullet}) \end{aligned} \tag{8.21}$$

avec $\delta_i^{(k+1)} = \langle \boldsymbol{v}_i, \boldsymbol{x}^{(k+1)} \rangle$, et où on a posé

$$\boldsymbol{B}^{(k)}_{\mathrm{GR}} = \boldsymbol{Q} + \boldsymbol{V}\boldsymbol{L}^{(k)}\boldsymbol{V}^{T}$$

avec $\boldsymbol{L}^{(k)} = \mathrm{diag}(l_1^{(k)}, \cdots, l_1^{(k)})$. Pour les différentes constructions de GY définies précédemment, les expressions définissant la mise à jour groupée s'écrivent respec-

tivement

$$\left(\begin{array}{c} M.à.j. \\ groupée \\ J^*_{\mathrm{GY}} \end{array}\right) \qquad \begin{aligned} \bar{\boldsymbol{x}}^{(k+1)} &= (\boldsymbol{B}^a_{\mathrm{GY}})^{-1}\left[2\boldsymbol{q} + \frac{1}{a}\boldsymbol{V}(\boldsymbol{l}^{(k)} + \boldsymbol{w})\right] \\ \bar{\boldsymbol{l}}^{(k+1)} &= \boldsymbol{V}^T\boldsymbol{x}^{(k+1)} - \boldsymbol{w} - a\boldsymbol{\phi}'(\delta_\bullet^{(k+1)} - w_\bullet) \end{aligned} \tag{8.22}$$

$$\left(\begin{array}{c} M.à.j. \\ groupée \\ J^*_{\mathcal{V}} \end{array}\right) \qquad \begin{aligned} \bar{\boldsymbol{x}}^{(k+1)} &= (\boldsymbol{B}^a_{\mathcal{V}})^{-1}\left[2\boldsymbol{q} + \frac{1}{a}\boldsymbol{l}^{(k)}\right] \\ \bar{\boldsymbol{l}}^{(k+1)} &= \boldsymbol{x}^{(k+1)} - \boldsymbol{w} - a\boldsymbol{V}\boldsymbol{\phi}'(\delta_\bullet^{(k+1)} - w_\bullet), \end{aligned} \tag{8.23}$$

où on rappelle que $a > 0$ est un paramètre de « mise à l'échelle » et où on a posé

$$\boldsymbol{B}^a_{\mathrm{GY}} = 2\boldsymbol{Q} + \frac{1}{a}\boldsymbol{V}\boldsymbol{V}^T \qquad \text{et} \qquad \boldsymbol{B}^a_{\mathcal{V}} = 2\boldsymbol{Q} + \frac{1}{a}\boldsymbol{I}$$

On souligne qu'il est nécessaire que les opérateurs inverses apparaissant dans (8.21-27) existent et soient bornés pour que les algorithmes associés soient correctement définis. $\boldsymbol{Q}$ étant définie non négative par hypothèse, $\boldsymbol{B}^a_{\mathcal{V}}$ admet *de facto* une inverse ; ceci n'est généralement pas vrai pour les autre algorithmes : on montre sans difficulté que $\boldsymbol{B}^{(k)}_{\mathrm{GR}}$ et $\boldsymbol{B}^a_{\mathrm{GY}}$ admettent une inverse si et seulement si $\mathrm{Ker}(\boldsymbol{Q}) \cap \mathrm{Ker}(\boldsymbol{V}\boldsymbol{V}^T) = \{\boldsymbol{0}\}$; cf. section 9.49.4.2.

Mise en œuvre

Dans tous les cas de figure, l'étape de mise à jour des variables primales est la plus lourde : elle nécessite la résolution d'un système linéaire qui dépend éventuellement de l'itération courante si on adopte la construction de GR. Quand le problème est de grande taille, l'inversion directe est inabordable et on doit résoudre approximativement chaque système linéaire par une technique itérative. L'efficacité d'implantation dépend alors de manière critique de la méthode de résolution retenue ; en pratique, on privilégie souvent des méthodes à faible coût tel que le gradient conjugué [Charbonnier *et al.*, 1997; Delaney et Bresler, 1998] dans une version éventuellement préconditionné [Nikolova et Ng, 2001], ou une relaxation de type SOR. On note également que dans [Fiani, 2001], un algorithme de type ART est employé pour résoudre le système linéaire apparaissant dans (8.22), et dans [Geman et Yang, 1995], les auteurs tirent parti du caractère approximativement circulant-bloc-circulant pour inverser le système linéaire par *transformée de* FOURIER *rapide*. Enfin, au meilleur de notre connaissance, aucun auteur n'a mis en œuvre un schéma de relaxation contraint sur les variables primales.

La relaxation SSU : mise à jour alternée x_n et l_{x_n}

Le coût informatique lié à la résolution à chaque itération d'un système linéaire de grande taille peut amener à changer de stratégie dans le schéma de relaxation. À l' « opposé » d'une mise à jour complète primale/duale, la partition suivante

constitue l'alternative qui permet le coût d'implantation le plus léger :

$$x_1, \boldsymbol{l}_{x_1}, \cdots, x_n, \boldsymbol{l}_{x_n}, \cdots, x_N, \boldsymbol{l}_{x_N}$$

où le vecteur $\boldsymbol{l}_{x_n}$ rassemble les variables duales *voisines* de x_n.

Variables duales associées à la minimisation

On notera que ce nombre de variables varie en fonction de la construction adoptée. Ainsi, pour la forme GY *vectorielle* définie par (8.19), le nombre de variables primales et duales est identique ce qui rend la mise à jour particulièrement simple. Pour les autres formes, ces nombres diffèrent en général et les variables duales composant $\boldsymbol{l}_{x_n}$ correspondent aux *indices non nuls* dans la n-ème ligne de l'opérateur $\boldsymbol{V}$, c.à.d.

$$\boldsymbol{l}_{x_n} = \{l_i \ : \ i \in \mathcal{I}_n\}$$

avec $\mathcal{I}_n$ l'ensemble des indices défini par

$$\mathcal{I}_n = \{1 \leq i \leq I \ : \ \langle \boldsymbol{v}_i, \boldsymbol{e}_n \rangle \neq 0\}$$

où $\boldsymbol{e}_n$ est le n-ème vecteur canonique de $\mathbb{R}^N$. Notons que le nombre de variables dans $\boldsymbol{l}_{x_n}$ est souvent très restreint ; effets de bords exclus, l'opérateur $\boldsymbol{V} \equiv \boldsymbol{D}$ des différences finies du première ordre pour un système à 8 voisins tel qu'adopté pour

la reconstruction tomographique conduit à un vecteur $[\boldsymbol{V}]_{n,\bullet}$ non nul en seulement 8 composantes.

Mises à jours SSU

L'itération courante définit les variables mises à jour ; ainsi pour l'itération k, on relaxe la composante primale $n = \mathrm{mod}_N(k)+1$ et les variables duales associées, et on « recopie » la valeur des autres variables, c'est-à-dire

$$\begin{array}{lll} x_n^{(k+1)} & = & (1-\theta)\, x_n^{(k)} + \theta\, \bar{x}_n^{(k+1)} \qquad \theta \in]0;2[\\ \forall\, m \neq n \qquad x_m^{(k+1)} & = & x_m^{(k)} \end{array} \tag{8.24}$$

$$\begin{array}{lll} \boldsymbol{l}_{x_n}^{(k+1)} & = & (1-\gamma)\, \boldsymbol{l}_{x_n}^{(k)} + \gamma\, \bar{\boldsymbol{l}}_{x_n}^{(k+1)} \qquad \gamma \in]0;1[\\ \forall\, m \neq n \qquad \boldsymbol{l}_{x_m}^{(k+1)} & = & \boldsymbol{l}_{x_m}^{(k)} \end{array} \tag{8.25}$$

où les expressions de $\bar{x}_n^{(k+1)}$ et $\bar{\boldsymbol{l}}_{x_n}^{(k+1)}$ dépendent là encore de la construction semi quadratique adoptée. Dans ce qui suit, on notera $[\boldsymbol{A}]_{n,m}$ l'élément situé en (ligne n, colonne m) d'une matrice $\boldsymbol{A}$, et $[\boldsymbol{A}]_{n,\bullet}$ et $[\boldsymbol{A}]_{\bullet,m}$ respectivement la n-ème ligne et m-ème colonne de $\boldsymbol{A}$.

Partant des diverses formes semi quadratiques exposées au début de la section, il est facile de montrer que la minimisation successive de x_n et de ses variables

duales associées conduit aux expressions suivantes :

$$\left(\begin{array}{c} M.\grave{a}.j. \\ SSU \\ J^*_{\mathrm{GR}} \end{array}\right) \quad \begin{array}{lll} & \bar{x}_n^{(k+1)} & = x_n^{(k)} + \dfrac{\langle \boldsymbol{w}, [\boldsymbol{L}^{(k)}\boldsymbol{V}^T]_{\bullet,n}\rangle + q_n - [\boldsymbol{B}^{(k)}_{\mathrm{GR}}\boldsymbol{x}^{(k)}]_n}{b^{(k)}_{\mathrm{GR}}} \\ \forall i \in \mathcal{I}_n, & \bar{l}_i^{(k+1)} & = \phi'(\delta_i^{(k+1)} - w_i)/2(\delta_i^{(k+1)} - w_i) \end{array} \tag{8.26}$$

$$\left(\begin{array}{c} M.\grave{a}.j. \\ SSU \\ J^*_{\mathrm{GY}} \end{array}\right) \quad \begin{array}{lll} & \bar{x}_n^{(k+1)} & = x_n^{(k)} + \dfrac{2q_n + \langle \boldsymbol{w} + \boldsymbol{l}^{(k)}, [\boldsymbol{V}^T]_{\bullet,n}\rangle/a - [\boldsymbol{B}^a_{\mathrm{GY}}\boldsymbol{x}^{(k)}]_n}{b^a_{\mathrm{GY}}} \\ \forall i \in \mathcal{I}_n, & \bar{l}_i^{(k+1)} & = \delta_i^{(k+1)} - w_i - a\phi'(\delta_i^{(k+1)} - w_i) \end{array} \tag{8.27}$$

$$\left(\begin{array}{c} M.\grave{a}.j. \\ SSU \\ J^*_{\mathcal{V}} \end{array}\right) \quad \begin{array}{lll} & \bar{x}_n^{(k+1)} & = x_n^{(k)} + \dfrac{2q_n + l_n^{(k)}/a - [\boldsymbol{B}^a_{\mathcal{V}}\boldsymbol{x}^{(k)}]_n}{b^a_{\mathcal{V}}} \\ & \bar{l}_n^{(k+1)} & = x_n^{(k+1)} - w_n - a\langle \boldsymbol{\phi}'(\delta_\bullet^{(k+1)} - w_\bullet), [\boldsymbol{V}^T]_{\bullet,n}\rangle \end{array} \tag{8.28}$$

où q_n est la n-ème composante de $\boldsymbol{q}$ et où on a posé

$$b^{(k)}_{\mathrm{GR}} = \left[\boldsymbol{Q} + \boldsymbol{V}\boldsymbol{L}^{(k)}\boldsymbol{V}^T\right]_{n,n}, \quad b^a_{\mathrm{GY}} = 2\left[\boldsymbol{Q} + \frac{1}{a}\boldsymbol{V}\boldsymbol{V}^T\right]_{n,n}, \quad b^a_{\mathcal{V}} = 2[\boldsymbol{Q}]_{n,n} + \frac{1}{a}$$

avec a un « facteur d'échelle » associé à la construction de GY standard ou mul-

tivariée. On notera que ces algorithmes sont bien définis uniquement si les dénominateurs apparaissant dans (8.26-31) sont non nuls ; ceci étant garanti sous des conditions similaires à celles données précédemment pour la mise à jour groupée.

Mise en œuvre

On souligne que ces mises à jour sont usuellement très peu coûteuses en terme de volume de calcul : le vecteur $[\boldsymbol{V}^T]_{\bullet,n}$ a un support souvent très réduit et les produits scalaires associés peuvent souvent être réduit à quelques opérations élémentaires — cf. l'exemple en tomographie hélicoidale du chapitre 7. D'autre part, on notera que les diverses constructions conduisent à des coûts d'implantation à peu près identique ; dans la mesure où la construction de GR conduit toujours à des algorithmes plus rapides en pratique, on aura intérêt à choisir cette construction pour une mise en œuvre SSU.

L'autre avantage d'une mise à jour SSU est qu'elle permet une prise en compte aisée des contraintes séparables sur les variables primales. En contrepartie, on s'attend à ce que la vitesse de convergence soit généralement moins intéressante qu'avec un schéma itératif *groupé* ; à notre connaissance, aucune contribution n'aborde vraiment ce point[14]. Le coût d'implantation particulièrement avantageux permet néan-

[14]L'interprétation purement primale de ces algorithmes, développée en section suivante, fournit néanmoins un éclairage intéressant. Les algorithmes SQ avec m.à.j. SSU sont principalement des algorithmes de relaxation sur les composantes de J avec recherche linéaire incomplète : on s'attend donc à ce que la vitesse de l'algorithme soit bonne si les variables primales sont peu couplées, c'est-à-dire si le Hessien de J est à diagonale dominante.

moins souvent de justifier d'implanter cette technique, notamment en restauration d'image [Brette et Idier, 1996; Villain *et al.*, 2001]. D'autre part, Sauer et Bouman [1993] ont montré que la mise à jour SSU permet en tomographie axiale une convergence tout à fait comparable à celle obtenue en adoptant un schéma *groupé* avec résolution de l'équation normale par un algorithme de type gradient conjugué.

Relaxation par blocs

Les deux schémas de relaxation précédent peuvent s'interpréter comme les deux choix extrêmes associés à une *relaxation par bloc.* On choisit donc de partitionner l'ensemble des variables primales en se donnant S ensembles ordonnés d'indices

$$1 \leq s \leq S \qquad \mathcal{N}_s \subseteq \{1, \cdots, N\}$$

tels que chaque indice $\{1, \cdots, N\}$ apparaisse au moins une fois dans $\mathcal{N} = \bigcup_s \mathcal{N}_s$. On choisit alors de minimiser le critère primal/dual adopté suivant le schéma de relaxation suivant

$$\boldsymbol{x}_1, \boldsymbol{l}_{\boldsymbol{x}_1}, \cdots, \boldsymbol{x}_s, \boldsymbol{l}_{\boldsymbol{x}_s}, \cdots, \boldsymbol{x}_S, \boldsymbol{l}_{\boldsymbol{x}_S}$$

où le vecteur $\boldsymbol{x}_s$ rassemble les variables primales dont les indices sont dans $\mathcal{N}_s$

$$\boldsymbol{x}_s = \{x_n \ : \ n \in \mathcal{N}_s\}$$

et avec $\boldsymbol{l}_s$ le vecteur des variables duales qui sont *voisines* d'une variable primale contenue dans $\boldsymbol{x}_s$. Dans le cas de la construction de GY multivariée, une variable duale est associée à chaque variable primale contenue dans $\boldsymbol{x}_s$; pour les autres constructions, on a

$$\boldsymbol{l}_s = \{l_i \ : \ i \in \mathcal{I}_s\}$$

avec $\mathcal{I}_s$ l'ensemble des indices défini par

$$\mathcal{I}_s = \{1 \leq i \leq I \ : \ \langle \boldsymbol{v}_i, \boldsymbol{e}_n \rangle \neq 0, \ \forall n \in \mathcal{N}_s\}.$$

L'itération courante définit le bloc de variables mises à jour ; ainsi pour l'itération k, on relaxe le bloc primal $s = \mathrm{mod}_S(k) + 1$ puis on met à jour les variables duales associées, et enfin on « recopie » la valeur des autres blocs à l'identique :

$$\begin{aligned} \boldsymbol{x}_s^{(k+1)} &= (1-\theta)\, \boldsymbol{x}_s^{(k)} + \theta\, \bar{\boldsymbol{x}}_s^{(k+1)} \qquad \theta \in]0;2[\\ \forall\, r \neq s \qquad \boldsymbol{x}_r^{(k+1)} &= \boldsymbol{x}_r^{(k)} \end{aligned} \tag{8.29}$$

$$\begin{aligned} \boldsymbol{l}_s^{(k+1)} &= (1-\gamma)\, \boldsymbol{l}_s^{(k)} + \gamma\, \bar{\boldsymbol{l}}_s^{(k+1)} \qquad \gamma \in]0;1[\\ \forall\, r \neq s \qquad \boldsymbol{l}_{\boldsymbol{x}_r}^{(k+1)} &= \boldsymbol{l}_{\boldsymbol{x}_r}^{(k)} \end{aligned} \tag{8.30}$$

et il est alors aisé de montrer que la minimisation du critère suivant le bloc s conduit aux mises à jour suivantes,

$$\left(\begin{array}{c} M.à.j. \\ par\ bloc \\ J^*_{\mathrm{GR}} \end{array}\right) \quad \begin{array}{rcl} \bar{\boldsymbol{x}}^{(k+1)} & = & [\boldsymbol{B}^{(k)}_{\mathrm{GR}}]_s^{-1}\big(\boldsymbol{q}_s + [\boldsymbol{V}\boldsymbol{L}^{(k)}]_{\boldsymbol{s},\bullet}\,\boldsymbol{w}_s\big), \\ \forall i \in \mathcal{I}_s, \quad \bar{l}_i^{(k+1)} & = & \phi'(\delta_i^{(k+1)} - w_i)/2(\delta_i^{(k+1)} - w_i) \end{array} \tag{8.31}$$

$$\left(\begin{array}{c} M.à.j. \\ par\ bloc \\ J^*_{\mathrm{GY}} \end{array}\right) \quad \begin{array}{rcl} \bar{\boldsymbol{x}}_{\boldsymbol{s}}^{(k+1)} & = & [\boldsymbol{B}^{a}_{\mathrm{GY}}]_s^{-1}\left[2\boldsymbol{q}_s + \frac{1}{a}[\boldsymbol{V}]_{\boldsymbol{s},\bullet}\big(\boldsymbol{w}_s + \boldsymbol{l}_s^{(k)}\big)\right] \\ \forall i \in \mathcal{I}_s, \quad \bar{l}_i^{(k+1)} & = & \delta_i^{(k+1)} - w_i - a\phi'(\delta_i^{(k+1)} - w_i) \end{array} \tag{8.32}$$

$$\left(\begin{array}{c} M.à.j. \\ par\ bloc \\ J^*_{\mathcal{V}} \end{array}\right) \quad \begin{array}{rcl} \bar{\boldsymbol{x}}_{\boldsymbol{s}}^{(k+1)} & = & [\boldsymbol{B}^{a}_{\mathcal{V}}]_s^{-1}\left[2\boldsymbol{q}_s + \frac{1}{a}\boldsymbol{l}_s^{(k)}\right] \\ \bar{\boldsymbol{l}}_s^{(k+1)} & = & \boldsymbol{x}_s^{(k+1)} - \boldsymbol{w}_s - a[\boldsymbol{V}]_{\boldsymbol{s},\bullet}\,\boldsymbol{\phi}'(\delta_\bullet^{(k+1)} - w_\bullet), \end{array} \tag{8.33}$$

où $[A]_{\boldsymbol{s},\bullet}$ représente la matrice constituée par les *lignes* extraites de $\boldsymbol{A}$ suivant les indices dans $\mathcal{N}_s$; $\boldsymbol{w}_s$ et $\boldsymbol{q}_s$ représentent les vecteurs constitués des éléments extraits respectivement de $\boldsymbol{w}$ et $\boldsymbol{q}$ suivant les indices de $\mathcal{N}_s$. On a également posé dans ces

expressions,

$$[\boldsymbol{B}^{(k)}_{\mathrm{GR}}]_s = \left[\boldsymbol{Q} + \boldsymbol{V}\boldsymbol{L}^{(k)}\boldsymbol{V}^T\right]_{\boldsymbol{s},\boldsymbol{s}}, \quad [\boldsymbol{B}^{a}_{\mathrm{GY}}]_s = 2\left[\boldsymbol{Q} + \tfrac{1}{a}\boldsymbol{V}\boldsymbol{V}^T\right]_{\boldsymbol{s},\boldsymbol{s}},$$
$$[\boldsymbol{B}^{a}_{\mathcal{V}}]_s = 2\left[\boldsymbol{Q} + \frac{\boldsymbol{I}}{a}\right]_{\boldsymbol{s},\boldsymbol{s}}$$

où $[A]_{\boldsymbol{s},\boldsymbol{s}}$ représente la sous matrice constituée des entrées dont les indices de lignes et de colonnes sont simultanément dans $\mathcal{N}_s$.

Mise en œuvre

Le découpage par « bloc » permet une certaine souplesse de mise en œuvre dans le sens où un découpage parcimonieux permet de contrôler l'explosion du coût numérique associé à la résolution du système linéaire à chaque itération. Notons que la littérature ne semble pas faire état de l'utilisation pratique de tels schémas partitionnés pour la mise en œuvre d'algorithme semi quadratiques.

Convergence globale et vitesse

L'étude de convergence globale de ces algorithmes s'inscrit principalement dans le cadre de l'analyse convexe [Charbonnier *et al.*, 1994, 1997] et [Delaney et Bresler, 1998; Idier, 2001*b*]. Les propriétés de convergence de ces schémas de relaxation s'appuient largement sur la convexité des constructions primales/duales. Pour ϕ convexe, on montre sous certaines conditions techniques la convexité du critère de J^*_{GY} dans ses formes *standard* et *mise à l'échelle* [Idier, 2001*b*, Th. 1] ; résultat

qu'on étend sans difficulté à la formulation *multivariée* — cf. Idier *et al.* [2002]. La convexité du critère de J^*_{GR} est plus délicate, on montre cependant que J^*_{GR} est convexe en $(\boldsymbol{x}, \boldsymbol{l})$ *à un changement de variable bijectif près* ; ceci est suffisant pour garantir la convergence des schémas de relaxation sur J^*_{GR} moyennant certaines restrictions techniques standard [Idier, 2001*b*, Th. 3]. Si on écarte le cas ℓ_p $(1 < p < 2)$, la convergence globale d'algorithmes de relaxation sur ces critères semi quadratiques est finalement garantie sous des hypothèses réalistes pour des fonctions ϕ largement utilisées en pratique.

En revanche, on doit noter que la méthode de résolution adoptée pour *mettre en œuvre* la relaxation SQ peut rendre délicate la convergence globale de l'algorithme : en ce sens, la preuve de convergence dépend des choix de mise en œuvre. À titre d'illustration, la résolution approximative de l'équation normale apparaissant dans (8.21) ne permet pas de garantir *de facto* la convergence de la relaxation SQ « groupée » vers le minimiseur global de J^*_{GR}. Un aspect remarquable de la relaxation SQ avec m.à.j. SSU est que chaque minimisation est faite de manière *exacte* permettant ainsi de garantir la convergence globale sans examen supplémentaire.

Jusqu'à présent, la vitesse de convergence de ces algorithmes SQ reste très peu étudiée. En pratique, l'expérience accumulée indique que, pour un schéma de relaxation identique — ex. mise à jour *groupée*, la convergence des algorithmes de GR est meilleure que celles des algorithmes de GY (constructions monovariée et multivariée). Tirer des conclusions sur une base théorique reste par contre plus difficile :

le cadre de l'analyse convexe (cadre dans lequel ces algorithmes ont été construit) permet difficilement une étude de vitesse asymptotique de ces algorithmes ; d'autre part, une comparaison directe sur une base théorique des algorithmes de GY et GR reste délicate dans la mesure où ces algorithmes optimisent un critère *distinct* [Geman et Yang, 1995, p. 937].

Dans la section suivante, on montre que ces algorithmes semi quadratiques sont principalement identiques à des algorithmes de descente sur le critère primal tels que nous les avons présentés en section 8.2.3. Cette correspondance va apporter un éclairage complémentaire fructueux en terme de propriétés de convergence globale et de vitesse asymptotique ; une partie des contributions de cette étude s'appuie donc sur les résultats exposés ci-dessous.

8.4 Formes primales des algorithmes semi quadratiques

Cette section montre dans un premier temps que les algorithmes de relaxation *semi quadratiques* sont, pour l'essentiel, des algorithmes de direction de descente *à pas fixe* sur le critère primal. Ce résultat permettra ensuite de rattacher ces algorithmes à une autre classe de schémas itératifs largement employés en traitement d'image et de signal. Plus précisément, nous montrons que ces algorithmes semi quadratiques sont des algorithmes travaillant par « approximation supérieur » au même titre que les algorithmes introduits comme des généralisations de l'algorithme

EM SAGE [Fessler et Hero, 1995].

Formulation primale des formes m.à.j. « groupées »

On rappelle tout d'abord la forme du critère primal considéré

$$J(\boldsymbol{x}) = Q(\boldsymbol{x}) + \Phi(\boldsymbol{x})$$

avec (8.34)

$$Q(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{Q}\boldsymbol{x} \rangle - 2\langle \boldsymbol{q}, \boldsymbol{x} \rangle + \mu \quad \text{et} \quad \Phi(\boldsymbol{x}) = \textstyle\sum_{i=1}^{I} \phi(\delta_i - w_i);$$

dans la mesure où cette expression nous sera utile par la suite, nous donnons l'expression du gradient de J :

$$\boldsymbol{\nabla} J(\boldsymbol{x}) = 2(\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{q}) + \boldsymbol{V}\boldsymbol{\phi}'(\delta_\bullet - w_\bullet). \tag{8.35}$$

Forme primale à pas fixe

Partant de (8.20) avec $\gamma = 1$ et en utilisant les expressions (8.21–23), il est facile de montrer que les algorithmes semi quadratiques à mise à jour « groupée » sont des algorithmes à pas fixe sur J. Prenons par exemple l'itération SQ de GR (abrég. SQ–GR) : d'après (8.20), la relaxation sur les variables primales s'écrit

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \theta\,(\bar{\boldsymbol{x}}^{(k+1)} - \boldsymbol{x}^{(k)});$$

en remplaçant $\bar{\boldsymbol{x}}^{(k+1)}$ par son expression où on aura pris soin d'intégrer l'expression des variables duales en fonction de $\boldsymbol{x}(k)$, on obtient

$$\begin{aligned}
\boldsymbol{x}^{(k+1)} &= \boldsymbol{x}^{(k)} + \theta\left[\left(\boldsymbol{B}_{\mathrm{GR}}^{(k)}\right)^{-1}(\boldsymbol{q} + \boldsymbol{V}\boldsymbol{L}^{(k)}\boldsymbol{w}) - \boldsymbol{x}^{(k)}\right] \\
&= \boldsymbol{x}^{(k)} + \theta\left(\boldsymbol{B}_{\mathrm{GR}}^{(k)}\right)^{-1}\left[\boldsymbol{q} + \boldsymbol{V}\boldsymbol{L}^{(k)}\boldsymbol{w} - \boldsymbol{Q}\boldsymbol{x}^{(k)} - \boldsymbol{V}\boldsymbol{L}^{(k)}\boldsymbol{V}^T\boldsymbol{x}^{(k)}\right] \\
&= \boldsymbol{x}^{(k)} + \theta\left(\boldsymbol{B}_{\mathrm{GR}}^{(k)}\right)^{-1}\left[\boldsymbol{q} - \boldsymbol{Q}\boldsymbol{x}^{(k)} - \boldsymbol{V}\boldsymbol{L}^{(k)}(\boldsymbol{V}^T\boldsymbol{x}^{(k)} - \boldsymbol{w})\right] \\
&= \boldsymbol{x}^{(k)} + \theta\left(2\boldsymbol{B}_{\mathrm{GR}}^{(k)}\right)^{-1}\left[2\boldsymbol{q} - 2\boldsymbol{Q}\boldsymbol{x}^{(k)} - \boldsymbol{V}\boldsymbol{\phi}'(\delta_\bullet^{(k)} - w_\bullet)\right], \\
&= \boldsymbol{x}^{(k)} - \theta\left(2\boldsymbol{B}_{\mathrm{GR}}^{(k)}\right)^{-1}\boldsymbol{\nabla} J(\boldsymbol{x}^{(k)}); \qquad (8.36)
\end{aligned}$$

cette dernière égalité montre donc que cet algorithme SQ–GR est identique à un algorithme à direction de descente[15] à pas fixe sur le critère primal J ; on notera que le pas fixe de l'itération (8.36) est le paramètre de relaxation adopté pour l'algorithme SQ. Une relation similaire peut être établie pour les divers algorithmes de GY ; ainsi pour la forme standard « mise à l'échelle » (8.22), on a :

$$\begin{aligned}
\boldsymbol{x}^{(k+1)} &= \boldsymbol{x}^{(k)} + \theta\left(\boldsymbol{B}_{\mathrm{GY}}^{a}\right)^{-1}\left[2\boldsymbol{q} + \frac{1}{a}\boldsymbol{V}\boldsymbol{V}^T\boldsymbol{x}^{(k)} - \boldsymbol{V}\boldsymbol{\phi}'(\delta_\bullet^{(k)} - w_\bullet) - \boldsymbol{B}_{\mathrm{GY}}^{a}\boldsymbol{x}^{(k)}\right] \\
&= \boldsymbol{x}^{(k)} + \theta\left(\boldsymbol{B}_{\mathrm{GY}}^{a}\right)^{-1}\left[2\boldsymbol{q} - 2\boldsymbol{Q}\boldsymbol{x}^{(k)} - \boldsymbol{V}\boldsymbol{\phi}'(\delta_\bullet^{(n)} - w_\bullet)\right] \\
&= \boldsymbol{x}^{(k)} - \theta\left(\boldsymbol{B}_{\mathrm{GY}}^{a}\right)^{-1}\boldsymbol{\nabla} J(\boldsymbol{x}^{(k)}); \qquad (8.37)
\end{aligned}$$

[15]Dans la mesure où une condition de convergence de ces algorithmes SQ est que l'opérateur $\boldsymbol{B}_{\mathrm{GR}}^{(k)}$ soit DP, l'itération primale à pas fixe (8.36) génèrent *effectivement* des directions strictement descendantes ; nous reviendrons plus en détail sur ce point au chapitre suivant.

la même démarche mène pour la forme multivariée à :

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \theta\big(\boldsymbol{B}_{\nu}^{a}\big)^{-1}\boldsymbol{\nabla} J(\boldsymbol{x}^{(k)}). \tag{8.38}$$

A priori, ces expressions *primales* des algorithmes SQ ne peuvent pas être obtenues si on choisit de *sur ou sous relaxer* les variables duales ; c.à.d. pour $\gamma \neq 1$. En effet, dans le cas d'une sur ou sous relaxation sur $\boldsymbol{l}$, la minimisation suivant les variables duales de $J^*(\boldsymbol{x}^{(k)}, \boldsymbol{l})$ ne conduit plus à $J(\boldsymbol{x}^{(k)})$. La « relaxation duale » reste néanmoins d'un intérêt marginal : la sur relaxation duale ne permet pas d'assurer la convergence et on a constaté en pratique que la sous relaxation duale ralentit la convergence.

Formulation primale des formes m.à.j. SSU

On pose $w_{\bullet} = 0$ de manière à ne pas alourdir les développements qui vont suivent. Partant de l'expression de J donnée par (8.34), la restriction scalaire de J à la variable x_n conduit à un critère monovarié $J_n : \mathbb{R} \longrightarrow \mathbb{R}$ qui s'écrit :

$$\begin{aligned} J_n(u) &= J(\boldsymbol{x}^{(k)} + u\boldsymbol{e}_n) \\ &= \big\langle (\boldsymbol{x}^{(k)} + u\boldsymbol{e}_n), \boldsymbol{Q}(\boldsymbol{x}^{(k)} + u\boldsymbol{e}_n)\big\rangle - 2\big\langle \boldsymbol{q}, (\boldsymbol{x}^{(k)} + u\boldsymbol{e}_n)\big\rangle + \cdots \\ &\quad \sum_{i \in \mathcal{I}_n} \phi\big(\langle \boldsymbol{v}_i, (\boldsymbol{x}^{(k)} + u\boldsymbol{e}_n)\rangle\big) \end{aligned}$$

où les notations de la page 244 ont été adoptées. La dérivée de J_n s'écrit :

$$\begin{aligned}\frac{\mathrm{d}J_n(u)}{\mathrm{d}\,u} &= 2\big\langle \boldsymbol{e}_n, \boldsymbol{Q}(\boldsymbol{x}^{(k)} + u\boldsymbol{e}_n)\big\rangle - 2q_n + \big\langle \boldsymbol{e}_n, \boldsymbol{V}\boldsymbol{\phi}'(\langle \boldsymbol{v}_\bullet, \boldsymbol{x}^{(k)} + u\boldsymbol{e}_n\rangle)\big\rangle \\ &= -2q_n + \big[2\boldsymbol{Q}(\boldsymbol{x}^{(k)} + u\boldsymbol{e}_n) + \boldsymbol{V}\boldsymbol{\phi}'(\langle \boldsymbol{v}_\bullet, \boldsymbol{x}^{(k)} + u\boldsymbol{e}_n\rangle)\big]_n ,\end{aligned}$$

et il est facile de montrer que les trois couples (8.26–28) exprimant les mises à jour de type semi quadratique SSU sont équivalentes à un pas de descente *scalaire* sur le critère monovarié J_n. Pour s'en convaincre, prenons l'exemple de la mise à jour de la variable x_n par la forme SSU de GR donnée par (8.26) ; la relaxation primale s'écrit par (8.24),

$$x_n^{(k+1)} = x_n^{(k)} + \theta\,(\bar{x}_n^{(k+1)} - x_n^{(k)})$$

et utilisant l'expression de $\bar{x}_n^{(k+1)}$ où les variables duales *non relaxées* $l_i^{(k)} = \bar{l}_i^{(k)}$ ont été remplacées par leur expression en fonction de la variable primale à l'itération k, on obtient

$$\begin{aligned}x_n^{(k+1)} &= x_n^{(k)} + \theta\big(b_{\mathrm{GR}}^{(k)}\big)^{-1}\Big(q_n - \big[\boldsymbol{Q}\boldsymbol{x}^{(k)} + \boldsymbol{V}\boldsymbol{L}^{(k)}\boldsymbol{V}^T\boldsymbol{x}^{(k)}\big]_n\Big) \\ &= x_n^{(k)} + \theta\big(2b_{\mathrm{GR}}^{(k)}\big)^{-1}\big(2q_n - \big[2\boldsymbol{Q}\boldsymbol{x}^{(k)} + \boldsymbol{V}\boldsymbol{\phi}'(\langle \boldsymbol{v}_\bullet, \boldsymbol{x}^{(k)}\rangle)\big]_n\big) \\ &= x_n^{(k)} - \theta\big(2b_{\mathrm{GR}}^{(k)}\big)^{-1}\,\frac{\mathrm{d}J_n}{\mathrm{d}\,u}\bigg|_{u=0}\end{aligned}$$

qui constitue l'égalité que nous souhaitions montrer. On pourra remarquer que cette égalité aurait pu se déduire directement du résultat multivarié par *identification* en

considérant un critère SQ construit à partir de $J_n(u)$; nous espérons néanmoins privilégier la clarté de l'exposé en choisissant de développer explicitement la forme primale à partir de la mise à jour SSU. De la même manière, la formulation purement primale des relaxations SSU semi quadratiques de GY mise à l'échelle et multivariée s'écrit à partir de (8.27-31) :

$$x_n^{(k+1)} = x_n^{(k)} - \theta\big(b_{\mathrm{GY}}^a\big)^{-1} \left.\frac{\mathrm{d}J_n}{\mathrm{d}\,u}\right|_{u=0} \tag{8.39a}$$

$$x_n^{(k+1)} = x_n^{(k)} - \theta\big(b_{\mathrm{V}}^a\big)^{-1} \left.\frac{\mathrm{d}J_n}{\mathrm{d}\,u}\right|_{u=0} \tag{8.39b}$$

En conclusion, on peut donc affirmer que la minimisation SQ avec m.à.j. SSU (abrég. SQ–SSU) du critère J est *équivalente* à une relaxation coordonnée par coordonnée sur J *via* une minimisation scalaire incomplète définie par une des équations ci-dessus.

Dans son principe, l'algorithme SQ–SSU est finalement assez proche de certains algorithmes apparaissant dans la littérature. Citons en particulier la technique *Iterative coordinate descent*–NEWTON RAPHSON (abrég. ICD–NR) [Bouman et Sauer, 1996; Sauer et Bouman, 1993] dont le principe repose sur une relaxation coordonnée par coordonnée associé à une minimisation par une série de pas de NEWTON[16] ou l'algorithme ICD–*Functional substitution* (abrég. ICD–FS) [Zheng *et al.*, 2000,

[16]On indique néanmoins que la version mise en œuvre dans [Bouman et Sauer, 1996] n'est pas exactement une mise en œuvre newtonienne scalaire; nous renvoyons le lecteur à la référence précitée pour plus de détails.

Sec. III] qui remplace le pas de NEWTON par une « approximation de la sécante » [Ortega et Rheinboldt, 1970, Sec. 7.2] de manière à permettre la démonstration de convergence globale.

Une différence rend néanmoins l'algorithme SQ–SSU particulièrement remarquable en regard de ces derniers algorithmes : alors que la recherche linéaire se réduit à une *unique* sous-itération suivant la coordonnée courante, la convergence de cet algorithme est assuré sous des hypothèses peu restrictives — cf. le dernier paragraphe de la section précédente.

Formulation primale des formes m.à.j. par « bloc »

Comme on peut s'en douter à ce niveau de l'exposé, il n'est pas plus difficile de montrer que les algorithmes SQ par « bloc », ont une formulation purement primale. Sans reporter ici le détail des calculs et en adoptant les notations de la page 251, on montre que les mises à jour par bloc SQ, définies par la relaxation (8.29) et les expressions des mises à jour (8.31-33), s'écrivent également

$$\boldsymbol{x}_{\boldsymbol{S}}^{(k+1)} = \boldsymbol{x}_{\boldsymbol{S}}^{(k+1)} - \theta[\boldsymbol{B}_{\bullet}^{(k)}]_{s}^{-1} \boldsymbol{\nabla}_{\boldsymbol{S}} J(\boldsymbol{x}_{\boldsymbol{S}}^{(k)})$$

où $[\boldsymbol{B}_{\bullet}^{(k)}]_s$ correspond à la matrice $[\boldsymbol{B}_{\mathrm{GR}}^{(k)}]_s$, $[\boldsymbol{B}_{\mathrm{GY}}^{a}]_s$ ou $[\boldsymbol{B}_{\nu}^{a}]_s$ suivant la construction SQ adoptée, et avec

$$\boldsymbol{\nabla}_{\boldsymbol{S}} J_{\boldsymbol{S}}(\boldsymbol{x}_{\boldsymbol{S}}^{(k)}) = 2[\boldsymbol{Q}]_{\boldsymbol{S}}\, \boldsymbol{x}_{\boldsymbol{S}}^{(k)} - 2\boldsymbol{q}_{\boldsymbol{S}} + [\boldsymbol{V}]_{\boldsymbol{S},\bullet}\, \boldsymbol{\phi}'(\delta_{\bullet}^{(k)} - w_{\bullet})$$

le gradient de $J_{\boldsymbol{S}}$ obtenu en ne conservant dans J que la dépendance en les variables primales $\boldsymbol{x}_{\boldsymbol{S}}$

$$J_{\boldsymbol{S}}(\boldsymbol{x}_{\boldsymbol{S}}) = \big\langle [\boldsymbol{Q}]_{\boldsymbol{S},\boldsymbol{S}}\, \boldsymbol{x}_{\boldsymbol{S}},\, \boldsymbol{x}_{\boldsymbol{S}} \big\rangle - 2\langle \boldsymbol{q}_{\boldsymbol{S}}, \boldsymbol{x}_{\boldsymbol{S}} \rangle + \sum_{i \in \mathcal{I}_{\boldsymbol{S}}} \phi(\delta_i - w_i).$$

Approximation supérieure...

Tous ces algorithmes s'identifiant formellement aux schémas de relaxation sur le critère primal/dual, la convergence de ces itérations primales à pas fixe est assurée à partir de l'étude développée dans le cadre de l'analyse convexe. Dans un cadre purement primal, on peut néanmoins s'interroger sur la propriété permettant la convergence globale sans le recours à une recherche linéaire. Une première réponse consiste à remarquer que ces algorithmes s'appuient tous sur le principe de l'*approximation quadratique supérieure* ou *majorante.*

Modèle quadratique majorant

Avant d'introduire la notion d'approximation quadratique *majorante*, il est nécessaire de définir ce qu'on entend par la notion plus large d'*approximation quadratique de J tangente en $\boldsymbol{u}$* :

Définition 5 *Soit $J : \mathbb{R}^N \to \mathbb{R}$ une fonction C^1. On pose :*

$$\tilde{J}(\boldsymbol{u}, \boldsymbol{v}) \stackrel{\triangle}{=} J(\boldsymbol{u}) + \langle \boldsymbol{\nabla} J(\boldsymbol{u}), (\boldsymbol{v} - \boldsymbol{u}) \rangle + \frac{1}{2} \langle \boldsymbol{B}(\boldsymbol{u})(\boldsymbol{v} - \boldsymbol{u}), (\boldsymbol{v} - \boldsymbol{u}) \rangle, \tag{8.40}$$

avec $\boldsymbol{B} : \mathbb{R}^N \to \mathbb{R}^{N \times N}$ un opérateur défini positif. $\tilde{J}(\boldsymbol{u}, \boldsymbol{v})$ est appelée approximation quadratique de J tangente en $\boldsymbol{u}$.

Cette définition permet de réinterpréter le schéma itératif défini par

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \theta \big(\boldsymbol{B}_{\bullet}^{(k)}\big)^{-1} \boldsymbol{\nabla} J(\boldsymbol{x}^{(k)}) \tag{8.41}$$

dans la mesure où la récurence ci-dessus peut être réécrite de manière équivalente

$$\boldsymbol{x}^{(n+1)} = (1 - \theta)\boldsymbol{x}^{(n)} + \theta \arg\min_{\boldsymbol{v}} \tilde{J}(\boldsymbol{x}^{(n)}, \boldsymbol{v}). \tag{8.42}$$

où

$$\tilde{J}(\boldsymbol{x}^{(k)}, \boldsymbol{v}) = J(\boldsymbol{x}^{(k)}) + \left\langle \boldsymbol{\nabla} J(\boldsymbol{x}^{(k)}), (\boldsymbol{v} - \boldsymbol{x}^{(k)}) \right\rangle + \frac{1}{2} \left\langle \boldsymbol{B}_{\bullet}^{(k)}(\boldsymbol{v} - \boldsymbol{x}^{(k)}), (\boldsymbol{v} - \boldsymbol{x}^{(k)}) \right\rangle, \tag{8.43}$$

est une approximation quadratique tangente en $\boldsymbol{x}^{(k)}$ au sens de la définition 5.

À partir de ce cadre, une condition de convergence globale peut être formulée assez simplement en exigeant que les $\boldsymbol{B}_{\bullet}^{(k)}$ successifs définissent un modèle quadratique $\tilde{J}(\boldsymbol{x}^{(k)}, \boldsymbol{x})$ *majorant* $J(\boldsymbol{x})$ pour tout $\boldsymbol{x}$: dans ce cas, on assure la convergence de l'itération (8.41) pour un pas fixe $0 < \theta < 2$ sous des hypothèses standard ; cf. notre exposé section 9.3, page 284.

Or, si les hypothèses imposées sur ϕ par les construtions primales/duales sont vérifiées, il n'est pas difficile de montrer que les trois formes primales des algorithmes SQ (8.36–41) définissent effectivement une approximation majorante $\tilde{J}(\boldsymbol{x}, \boldsymbol{x}^{(k)})$ à chaque itération ; nous invitons le lecteur à consulter la preuve en annexe de ce chapitre. Ce résultat s'étend directement aux itérations primales de mises à jour SSU ou par « bloc » en considérant les critères « réduits » associés à chaque schéma de relaxation SQ.

Appartenance à une large classe existante

La mise en évidence d'une approximation quadratique supérieure à chaque itération permet de rattacher les algorithmes SQ à la famille « nombreuse » des algorithmes de type SAGE basés sur le*transfert d'optimisation* ou la construction d'une *fonction de substitution*[17] qui sont connus depuis le milieu des années 1995 par les adeptes de l'algorithme EM et de ses (nombreuses) améliorations successives ; voir à ce propos [Fessler et Hero, 1995; Lange, 1995]. Cette classe d'algorithme connaît actuellement un regain d'intérêt de la part de certains auteurs, avec notamment des contributions algorithmiques dans le cadre de l'inversion pénalisée en imagerie, et plus particulièrement en tomographie [Erdogan et Fessler, 1999; Fessler *et al.*, 1997; Zheng *et al.*, 2000].

Avant de conclure cette section, il est important de souligner un point justifiant les développements techniques du prochain chapitre. Les hypothèses des constructions SQ garantissent le caractère majorant des modèles quadratiques associées et ainsi la convergence globale des itérations. Dans certains cas, ces itérations restent néanmoins globalement convergentes hors des hypothèses de construction initiales. Dans le chapitre suivant, on choisit plutôt de réexaminer les propriétés de convergence globale par une étude de l'admissibilité du pas fixe au sens de la règle d'*Armijo*. On montrera en particulier que cette démarche conduit à des conditions

[17]Ceci constitue une traduction littérale de l'appellation *surrogate fonctions* [Fessler *et al.*, 1997, page 168] aujourd'hui adoptées par la communauté.

de convergence affaiblies par rapport à une étude *via* l' « approximation supérieure ».

8.5 En résumé

Les approches semi quadratiques permettent de construire une large famille de schémas de relaxation spécialement dédiés à la minimisation des critères pénalisés. Formellement simple, ces algorithmes peuvent être mis en œuvre sur des problèmes de très grande taille par des mises à jours de type SSU ou en utilisant une mise à jour en bloc associée à un algorithme tel que le gradient conjugué.

Nous avons montré que ces algorithmes SQ, dans une formulation sans sur relaxation des variables duales, étaient équivalents à des algorithmes de descente sur le critère initial J, ou sur une partition de ce critère. Nous avons également montré que ces algorithmes faisaient partie des algorithmes de type approximation supérieure, et qu'à ce titre, ils se rattachaient à de nombreux algorithmes globalement convergents déjà employés en traitement d'image et de signal. Le prochain chapitre tire largement partie de ce point de vue « primal » des algorithmes SQ : il permettra en particulier d'étudier la vitesse asymptotique et d'affaiblir en partie les conditions de convergence globale obtenues principalement dans le cadre de l'analyse convexe.

Avant de conclure, il nous paraît important d'insister sur l'aspect suivant : le cadre de l'analyse convexe et celui uniquement primal sont finalement assez

complémentaire pour étudier les algorithmes SQ. Pour s'en convaincre, on pourra par exemple considérer la relaxation SQ–SSU du critère de GY : dans une analyse primale/duale, la convergence de l'algorithme (8.27) découle *trivialement* de la propriété de convexité du critère J^*_{GY}, alors qu'une formulation purement primale (8.39a) ne permet pas de conclure sans analyse complémentaire. En pratique, on s'aperçoit souvent que le passage d'un cadre à l'autre s'avère fructueux.

Bibliographie annotée

Les algorithmes semi quadratiques de GR (8.21) et de GY (8.22) sont également connus sous le nom de *Iterative reweighted least squares* (abrég. IRLS) et de *Residual Steepest Descent*, respectivement. Ces formes itératives sont apparues au cours des années 1980 dans la communauté de la statistique robuste sans faire de référence à une quelconque notion de dualité ; voir par exemple [Yarlagadda *et al.*, 1985] et les références qui y sont citées.

Les constructions SQ de GY et GR a été formellement introduite dans le cadre de l'estimation bayésienne avec des pénalisations non convexes par [Geman et Reynolds, 1992] et [Geman et Yang, 1995], respectivement. À la même période, des contributions introduisent ces formulations primales/duales dans le cadre de la régularisation de problèmes inverses mals posés pour des critères convexes ou non [Brette et Idier, 1996; Charbonnier *et al.*, 1994, 1997; Delaney et Bresler, 1998] ;

depuis un certain nombre de publications ont tirés parti d'algorithmes SQ pour minimiser des critères pénalisés, voir par exemple [Çetin et Karl, 2001; Ciuciu *et al.*, 2001] pour les plus récentes. La formulation multivariée de GY apparaît pour la première fois dans [Ciuciu, 2000] pour traiter des signaux complexes en estimation spectrale ; cette construction a ensuite été formalisée dans Fiani [2001] pour permettre de minimiser des critères *convexes non séparables* avec des algorithmes ART développés pour des pénalisations *quadratiques* et *séparables* ; cf. [Fiani, 2001, Sec. IV.4].

Le lien entre les algorithmes SQ et leur formulation complètement primale est rarement rendu explicite dans la littérature ; on notera néanmoins que dans [Vogel et Oman, 1998], l'algorithme de point fixe obtenu à partir de l'équation de EULER–LAGRANGE conduit à la formulation primale à pas fixe. On notera également [Chan et Mulet, 1999] qui constitue, à notre sens, la référence la plus intéressante dans la mesure où, outre la formulation primale, le principe majorant de l'algorithme de GR est démontré et clairement rattaché à des travaux antérieurs issus de l'analyse numérique.

CHAPITRE 9

AUTOUR DE LA CONVERGENCE DES ALGORITHMES SQ

Le lien fort existant entre les schémas de relaxation semi quadratiques (SQ) et des formes itératives purement primales à pas fixe, permet une étude de ces algorithmes dans un cadre distinct de celui habituellement adopté. Ce chapitre se propose donc de réexaminer les propriétés de convergence globale et d'étudier le comportement asymptotique de ces algorithmes SQ. Dans la mesure où le cadre *non contraint* simplifie sensiblement l'analyse, nous l'adopterons pour tout ce chapitre. D'autre part, nous nous limiterons ici à l'étude des algorithmes SQ dans leurs formes à « mises à jours groupée » qui permettent une simplification formelle de l'analyse de convergence globale et de vitesse asymptotique.

Ce chapitre est organisé en deux parties traitant respectivement des propriétés de convergence globale et de la vitesse asymptotique. L'étude de convergence globale fait largement appel à des notions bien connues de programmation mathématique non linéaire et non contrainte, et c'est dans ce cadre que nous développons un nouvel outil permettant d'étudier la convergence des algorithmes à pas fixe : l'admissibilité du pas fixe au sens de la règle d'ARMIJO ; au meilleur de notre connaissance, cette méthode d'analyse est originale. Nous montrons en particulier que cet outil permet d'aboutir à des conditions de convergence moins restrictives

que celles obtenues en imposant le caractère *majorant* de l'approximation quadratique locale produite par les algorithmes standard de descente. Pour les algorithmes SQ, cette analyse permettra en particulier d'étendre sensiblement le domaine de convergence des algorithmes de GY.

L'intérêt pratique de cette extension réside principalement dans la constatation expérimentale suivante : « les algorithmes de GY convergent plus rapidement sur le bord de leur domaine de convergence ». Cette constatation a motivé une étude de vitesse asymptotique des algorithmes SQ. Cette étude s'est avérée globalement instructive même si il semble difficile de tirer des conclusions définitives et qu'un certain nombre de questions demeurent en suspend.

Note : une partie du contenu de ce chapitre a été publié dans [Allain *et al.*, 2002].

9.1 Définition du schéma itératif considéré

On rappelle que J est une fonction *strictement convexe, coercive* et au moins C^1. Décrivons synthétiquement les bases de l'algorithme que nous allons étudier. Étant donné un point initial $\boldsymbol{x}^{(0)}$, on génère une suite $\{\boldsymbol{x}^{(n)}\}_{n=1,\ldots}$ telle que :

$$\boldsymbol{x}^{(n+1)} = \boldsymbol{x}^{(n)} + \theta^{(n)}\boldsymbol{\xi}^{(n)}, \qquad \theta^{(n)} > 0 \tag{9.1}$$

avec $\boldsymbol{\xi}^{(n)} \equiv \boldsymbol{\xi}\left(\boldsymbol{x}^{(n)}\right)$ et $\theta^{(n)} \equiv \theta\left(\boldsymbol{x}^{(n)}\right)$ respectivement la direction de déplacement et la longueur du pas à l'itération courante. On impose aux directions de déplacement successives $\boldsymbol{\xi}^{(n)}$ d'être la solution d'un système linéaire de la forme,

$$\boldsymbol{B}^{(n)}\boldsymbol{\xi}^{(n)} = -\boldsymbol{\nabla} J(\boldsymbol{x}^{(n)}), \tag{9.2}$$

où on a posé $\boldsymbol{B}^{(n)} = \boldsymbol{B}\left(\boldsymbol{x}^{(n)}\right)$ avec $\boldsymbol{B} : \mathbb{R}^N \to \mathbb{R}^{N\times N}$ un opérateur symétrique et définie positive (abrég. DP). On adoptera enfin les conventions suivantes afin de simplifier les écritures :

$$\mathcal{J} \equiv \left\{J\left(\boldsymbol{x}^{(n)}\right)\right\}_{n=0,1\ldots}, \quad \mathcal{X} \equiv \left\{\boldsymbol{x}^{(n)}\right\}_{n=0,1\ldots}, \quad \Xi \equiv \left\{\boldsymbol{\xi}^{(n)}\right\}_{n=0,1\ldots}, \quad \text{et}$$
$$\Theta \equiv \left\{\theta^{(n)}\right\}_{n=0,1\ldots};$$

d'autre part, le *schéma itératif* constitué par les suites $\mathcal{X}$, Ξ, et Θ sera noté $(\mathcal{X}, \Xi, \Theta)$. Dans la mesure où celui-ci peut vérifier seulement (9.1) ou bien à la fois (9.1) et (9.2), nous le préciserons à chaque fois que nécessaire de manière à éviter les confusions.

Approximations quadratiques successives

On rappelle également que si la direction de déplacement courante $\boldsymbol{\xi}^{(n)}$ vérifie (9.2), alors l'itération (9.1) s'écrit également :

$$\boldsymbol{x}^{(n+1)} = (1-\theta^{(n)})\boldsymbol{x}^{(n)} + \theta^{(n)} \arg\min_{\boldsymbol{v}} \tilde{J}(\boldsymbol{x}^{(n)}, \boldsymbol{v})$$

où $\tilde{J}(\boldsymbol{x}^{(n)}, \boldsymbol{v})$ est l'unique forme quadratique *strictement convexe* définie par

$$\tilde{J}(\boldsymbol{x}^{(n)}, \boldsymbol{v}) = J(\boldsymbol{x}^{(n)}) + \langle \boldsymbol{\nabla} J(\boldsymbol{x}^{(n)}), (\boldsymbol{v} - \boldsymbol{x}^{(n)}) \rangle + \frac{1}{2} \langle \boldsymbol{B}^{(n)}(\boldsymbol{v} - \boldsymbol{x}^{(n)}), (\boldsymbol{v} - \boldsymbol{x}^{(n)}) \rangle.$$

En ce sens, le point à l'itération $n+1$ est une combinaison linéaire de l'itérée n et du minimiseur en $\boldsymbol{v}$ de $\tilde{J}\left(\boldsymbol{x}^{(n)}, \boldsymbol{v}\right)$; la valeur $\theta^{(n)}$ étant choisie suivant une méthode qui reste encore à définir.

Algorithme de descente

Notre attention se porte dans ce chapitre sur les méthodes dites « de descente », c'est-à-dire sur les schémas itératifs pour lesquels on a :

$$J(\boldsymbol{x}^{(n+1)}) \leq J(\boldsymbol{x}^{(n)}), \qquad n = 0, \cdots.$$

Pour J bornée inférieurement, cette propriété garantit évidemment que la suite $\mathcal{J}$ est convergente. Cependant, il est important de réaliser que cette propriété n'en-

traîne en rien la convergence de la suite $\mathcal{X}$ vers l'optimum de J, et il devient incontournable de rechercher des conditions réalistes permettant d'assurer en pratique la *convergence globale* d'un schéma itératif $(\mathcal{X}, \Xi, \Theta)$.

9.2 Convergence globale en optimisation non contrainte

L'étude de convergence globale s'appuie sur un examen des suites Ξ et Θ qui doivent posséder « certaines propriétés ». On est amené à s'intéresser aux *suites de directions gradient-reliées*, et à l'*admissibilité de la suite* Θ *au sens d'une certaine règle*. On rappelle ici brièvement ce que recouvre ces deux notions, puis on donne les principales conditions de convergence globale des algorithmes à directions gradient-reliées.

Algorithme à directions gradient-reliées

Intuitivement, l'intérêt d'une suite de direction *gradient-reliée* est d'assurer qu'à chaque itération, la direction courante est susceptible de faire décroître « suffisamment » l'objectif. En termes plus formels : *une suite* Ξ *est gradient-reliée à* $\mathcal{X}$ *si on garantit d'une part le caractère borné de* Ξ*, et d'autre part que les directions* $\boldsymbol{\xi}^{(n)}$ *sont non orthogonales au gradient et qu'elles ne tendent pas à le devenir au cours des itérations*. Pour obtenir une définition précise et un exposé complet sur la manière dont la convergence s'articule autour de cette notion, nous renvoyons le lecteur à [Ortega et Rheinboldt, 1970, 14.3].

Pour un schéma itératif $(\mathcal{X}, \Xi, \Theta)$ vérifiant (9.1) et (9.2), on peut souvent garantir *a priori* que Ξ est gradient-reliée à $\mathcal{X}$ en s'assurant de l'existence d'une borne uniforme inférieure et supérieure sur les valeurs propres de l'opérateur défini positif $\boldsymbol{B}(\cdot)$:

Proposition 1 *Soit* $J : \mathbb{R}^N \to \mathbb{R}^{N\times N}$ *une fonction* C^1 *sur un compact* $D_0 \subset \mathbb{R}^N$ *et* $\boldsymbol{B} : D_0 \to \mathbb{R}^{N\times N}$ *un opérateur défini positif pour lequel* $\exists\ \gamma_2 \geq \gamma_1 > 0$ *tels que,*

$$\forall \boldsymbol{u} \in D_0, \boldsymbol{v} \in \mathbb{R}^N, \qquad \gamma_2||\boldsymbol{v}||^2 \geq \langle \boldsymbol{B}(\boldsymbol{u})\boldsymbol{v}, \boldsymbol{v}\rangle \geq \gamma_1||\boldsymbol{v}||^2.$$

Si la suite de directions Ξ *est produite par*

$$\boldsymbol{B}(\boldsymbol{x}^{(n)})\boldsymbol{\xi}^{(n)} = -\boldsymbol{\nabla} J(\boldsymbol{x}^{(n)}), \tag{9.3}$$

avec $\mathcal{X}$ *la suite des itérées, alors* Ξ *est gradient-reliée à* $\mathcal{X}$.

Preuve La preuve de cette proposition est une adaptation directe de la preuve [Ortega et Rheinboldt, 1970, 14.4.1] —voir également la remarque (NR 14.4.1) dans le même ouvrage. ■

Le caractère uniformément borné sur un compact (et non sur $\mathbb{R}$) n'est pas restrictif dans la mesure où on s'assurera que le choix de $\theta^{(n)}$ permet, une fois

l'algorithme initialisé en $\boldsymbol{x}^{(0)}$, de rester dans l'ensemble de niveau :

$$D_0 = \left\{ \boldsymbol{x} \mid J(\boldsymbol{x}) \leq J(\boldsymbol{x}^{(0)}) \right\}.$$

qui sera effectivement un compact si J est continue et coercive.

Soulignons enfin que, sous réserve d'existence de la borne uniforme inférieure sur les valeurs propres, la proposition 1 permet également de garantir le caractère gradient-relié de Ξ si l'opérateur $\boldsymbol{B}(\cdot)$ est continu, puisque cette hypothèse garantit pour sa part l'existence de la borne uniforme supérieure sur tout compact.

Admissibilité du pas au sens d'Armijo

En pratique, garantir le caractère gradient-relié d'une suite Ξ n'est pas suffisant pour garantir la convergence de $\mathcal{X}$ vers un point stationnaire. La convergence dépend également de la sélection des pas $\theta^{(n)}$ successifs : disposer d'un « bon » pas à chaque itération permet d'assurer un régime de décroissance suffisant pour atteindre un point stationnaire.

Une règle de sélection du pas pour la convergence

Pour un schéma itératif $(\mathcal{X}, \Xi, \Theta)$, on peut introduire la notion d'*admissibilité* de la suite Θ au sens d'une certaine *règle de sélection* ; les principales sont les règles d'ARMIJO, de GOLDSTEIN, et les conditions de WOLF. Pour Ξ gradient-reliée à $\mathcal{X}$,

ces règles assurent toutes la *décroissance stricte* de l'objectif à chaque itération[1]. Cette propriété étant insuffisante pour assurer la convergence vers un *point stationnaire*, ces règles ont également en commun de garantir $\langle \nabla J(\boldsymbol{x}^{(n)}), \boldsymbol{\xi}^{(n)} \rangle / ||\boldsymbol{\xi}^{(n)}|| \to 0$ pour $n \to \infty$. C'est cette dernière propriété qui, combinée au caractère gradient-relié de Ξ, permet d'assurer que $\mathcal{X}$ converge vers un point où le gradient s'annule [Ortega et Rheinboldt, 1970, 14.2].

*L'admissibilité au sens d'*ARMIJO...

Le rôle central joué par la règle d'ARMIJO dans la suite de cette étude nous amène à définir l'admissibilité au sens de cette règle. Dans cet exposé, on distinguera *l'admissibilité indépendamment de toute considération itérative*, et *l'admissibilité dans le cadre d'un schéma itératif* ; cette distinction sera utile par la suite pour formuler clairement certains de nos résultats.

La définition ci-dessous précise ce qu'on entend par « un pas θ vérifiant la règle d'ARMIJO au point $\boldsymbol{u}$ dans la direction $\boldsymbol{\xi}$ ».

Définition 6 *Soit $J : \mathbb{R}^N \to \mathbb{R}$ une fonction C^1. On dira que $\theta > 0$ vérifie la règle d'*ARMIJO *pour $\boldsymbol{u} \in \mathbb{R}^N$ et $\boldsymbol{\xi}(\boldsymbol{u}) \in \mathbb{R}^N$ si il existe $\omega \in]0;1[$ tel que,*

$$J\left(\boldsymbol{u} + \theta \boldsymbol{\xi}(\boldsymbol{u})\right) - J(\boldsymbol{u}) - \omega\theta \langle \nabla J(\boldsymbol{u}), \boldsymbol{\xi}(\boldsymbol{u}) \rangle \leq 0. \tag{9.4}$$

[1]Pour les fonctions coercives, cette propriété garantit que $\mathcal{X}$ reste dans un compact D_0 défini par l'initialisation : c'est la raison pour laquelle ces règles de sélection sont parfois qualifiées de « techniques de stabilisation ».

Cette règle est formellement simple et importante : elle conduit à des conditions suffisantes de convergence globale pour les schémas itératifs $(\mathcal{X}, \Xi, \Theta)$ à suite Ξ gradient-reliée à $\mathcal{X}$ (cf. proposition 2). L'interprétation graphique de la règle d'ARMIJO est illustrée par la figure 9.1 : si la direction ξ est effectivement strictement descendante, alors la règle désigne comme admissible l'ensemble des $\theta > 0$ qui permettent au moins de décroître l'objectif d'une fraction ω du modèle linéaire en $\boldsymbol{u}$, c.à.d. d'au moins $\omega\theta\langle \boldsymbol{\nabla} J(\boldsymbol{u}), \boldsymbol{\xi}(\boldsymbol{u})\rangle$.

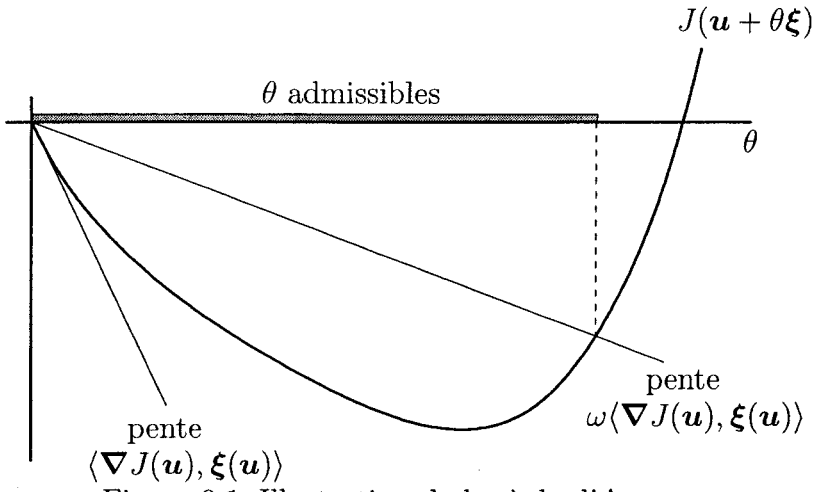


Figure 9.1: Illustration de la règle d'ARMIJO.

L'emploi de la règle d'ARMIJO pour un schéma itératif nous conduit à définir *l'admissibilité au sens d'*ARMIJO *d'une suite* Θ.

Définition 7 *Soit* $J : \mathbb{R}^N \to \mathbb{R}$ *une fonction* C^1, $(\mathcal{X}, \Xi, \Theta)$ *un schéma itératif vérifiant (9.1). La suite* Θ *est dite admissible au sens d'*ARMIJO *pour* $\mathcal{X}$ *et* Ξ *si il*

existe $\omega \in]0;1[$ *tel que,*

$$n = 0, 1 \cdots \qquad J(\boldsymbol{x}^{(n)} + \theta^{(n)}\boldsymbol{\xi}^{(n)}) - J(\boldsymbol{x}^{(n)}) - \omega\theta^{(n)} \left\langle \boldsymbol{\nabla} J(\boldsymbol{x}^{(n)}), \boldsymbol{\xi}^{(n)} \right\rangle \leq 0. \quad (9.5)$$

Notons que l'admissibilité de la suite Θ au sens d'ARMIJO consiste à garantir l'inégalité (9.5) pour *chaque* élément de Θ, ceci pour un *même* ω pris dans $]0;1[$. Si J est suffisamment régulière, on garantit l'existence pour tout n d'un ensemble non vide de pas vérifiant (9.5) pour $\omega \in]0;1[$ *fixé à l'initialisation* [Ortega et Rheinboldt, 1970, p. 491].

...est-elle suffisante ?

Remarquons d'emblée qu'à chaque itération, un pas θ arbitrairement petit vérifiera toujours l'inégalité (9.5) si ξ est effectivement une direction de descente. Étant donné une suite Ξ gradient-reliée prise arbitrairement, on peut donc toujours construire une suite Θ admissible au sens d'ARMIJO mais tendant suffisamment vite vers 0 pour « forcer » la convergence de $\mathcal{X}$ vers un point non stationnaire. L'admissibilité d'une suite de pas au sens d'ARMIJO ne représente donc pas une condition suffisamment restrictive pour assurer la convergence globale des suites gradient-reliées.

Intéressons nous au cas instructif de la *règle de* WOLF ; en plus de l'inéga-

lité (9.4), cette règle incorpore une inégalité supplémentaire qui permet justement d'éviter la sélection « inopportune » de pas arbitrairement petits ; cette seconde inégalité s'écrit [Bonnans *et al.*, 1997, 2.4] :

$$\langle \nabla J\left(\boldsymbol{u}+\theta\boldsymbol{\xi}(\boldsymbol{u})\right), \boldsymbol{\xi}(\boldsymbol{u})\rangle - \omega_2 \langle \nabla J(\boldsymbol{u}), \boldsymbol{\xi}(\boldsymbol{u})\rangle \;\geq\; 0, \tag{9.6}$$

où ω_2 est choisi tel que $0 < \omega < \omega_2 < 1$, ω étant le paramètre introduit en définition 6. Comme au paragraphe précédent, *l'admissibilité d'une suite* Θ *au sens de* WOLF peut être définie pour deux suites Ξ et $\mathcal{X}$, en s'appuyant sur les inégalités (9.4) et (9.6). Dans ce cas, on peut garantir que la propriété

$$\left\langle \nabla J\left(\boldsymbol{x}^{(n)}\right), \boldsymbol{\xi}^{(n)} \right\rangle / ||\boldsymbol{\xi}^{(n)}|| \to 0$$

a bien lieu [Rheinboldt, 1998, 8.1], ce qui est suffisant pour assurer la convergence vers un point stationnaire des schémas itératifs gradient-reliés.

En fait, si on exclut la règle d'ARMIJO, la majorité des règles de sélection classiques se présentent sous la forme de deux inégalités distinctes qui dépendent de deux paramètres. Ces paramètres étant fixés à l'initialisation, une suite Θ est admissible (au sens de cette règle) pour $\mathcal{X}$ et Ξ si, pour tout n, $\theta^{(n)}$ vérifie conjointement les deux inégalités avec les valeurs $\boldsymbol{x}^{(n)}$ et $\boldsymbol{\xi}^{(n)}$ correspondantes. En substance, la première de ces inégalités permet de minorer la décroissance de l'objectif en intro-

duisant une borne supérieure sur le pas θ, la seconde permet de minorer le déplacement $||\boldsymbol{\xi}(\boldsymbol{u})||$ en introduisant une borne inférieure sur θ. La première des deux est justement l'inégalité d'ARMIJO (9.4), la seconde diffère d'une règle à l'autre. Le lecteur pourra s'assurer du bien fondé de ces remarques en consultant notamment [Bonnans *et al.*, 1997, 2.4-2.5].

ARMIJO + *technique du rebroussement* = *une solution.*

La *construction* d'une suite Θ admissible au sens d'une certaine règle s'appuie généralement sur une opération de *recherche linéaire.* Pour la règle d'ARMIJO, on emploie usuellement *la technique du rebroussement* [Gilbert, 1999, Sec.5.3] pour obtenir un $\theta^{(n)}$ tel que la suite Θ soit admissible au sens de la définition 7 ; son fonctionnement est maintenant brièvement rappelé.

On se donne à l'initialisation les paramètres $\omega \in]0;1[$, $\tau \in]0;1[$ et $s > 0$. À l'itération courante n, la technique du rebroussement consiste à choisir le *plus petit entier* $k = l$ qui permet de satisfaire :

$$J(\boldsymbol{x}^{(n)} + s\tau^k \boldsymbol{\xi}^{(n)}) - J(\boldsymbol{x}^{(n)}) - \omega s\tau^k \left\langle \boldsymbol{\nabla} J(\boldsymbol{x}^{(n)}), \boldsymbol{\xi}^{(n)} \right\rangle \leq 0, \tag{9.7}$$

$k \in \mathbb{N}$. En pratique, la démarche consiste à « tester » d'abord un pas initial $\theta^{(n)} = s$ puis, si celui-ci est rejeté (i.e. l'inégalité ci-dessus est fausse pour $k = 0$), à tester successivement $\theta^{(n)} = s\tau^k$ pour $k = 1, 2, \cdots$; on s'arrête à la première valeur

$k = l$ qui vérifie l'inégalité, et on choisit évidemment $\theta^{(n)} = s\tau^l$. Cette procédure définissant une suite strictement décroissante et inférieure à s, on comprend l'origine du mot *rebroussement.*

On remarquera que cette technique fait intervenir un mécanisme intrinsèque qui empêche le choix d'un pas courant « trop petit ». On montre alors que la construction d'une suite admissible au sens d'ARMIJO par technique du rebroussement permet d'assurer la convergence des suites gradient-reliées Ξ et $\mathcal{X}$ correspondantes (voir proposition 2).

Convergence globale

La preuve de convergence globale des algorithmes gradient-reliés peut se décomposer en deux étapes. La première consiste à garantir que tout point d'accumulation de l'algorithme est un point stationnaire de J :

Proposition 2 *Soit un schéma itératif $(\mathcal{X}, \Xi, \Theta)$ vérifiant (9.1). Si les suites Ξ et Θ sont respectivement de type gradient-relié et admissible au sens de la règle d'*ARMIJO*, et si la suite Θ est construite par la technique du rebroussement, alors tout point d'accumulation de $\mathcal{X}$ est un point stationnaire.*

Preuve La preuve de cette proposition peut être trouvée dans [Bertsekas, 1999, prop.1.2.1]. On notera que la proposition issue de cette référence ne mentionne pas l'emploi de la technique du rebroussement, celle-ci y étant introduite implicitement

dans la présentation de la règle d'ARMIJO. ■

Une fois ce résultat obtenu, la convergence vers le minimiseur de J pour tout point d'initialisation de l'algorithme peut être montrée ; le théorème suivant sera utile pour assurer la convergence globale des formes algorithmiques étudiées dans ce document :

Théorème 1 *Soit $J : \mathbb{R}^N \to \mathbb{R}$ une fonction au moins C^1, strictement convexe et coercive. Soit un schéma itératif $(\mathcal{X}, \Xi, \Theta)$ vérifiant les hypothèses de la proposition 2. Soit $\boldsymbol{x}^*$ l'unique minimiseur (global) de J, alors on a $\mathcal{X} \to \boldsymbol{x}^*$.*

Preuve La preuve est donnée en annexe II. ■

Pour les critères *non convexes*, notons qu'en dépit de l'impossibilité d'obtenir un résultat de convergence aussi fort que le théorème 1, on peut néanmoins assurer la convergence dès qu'une itérée est « assez proche » d'un minimum local isolé [Bertsekas, 1999, prop. 1.2.5].

Convergence sans recherche linéaire

Plutôt que de construire la suite Θ au cours des itérations, on peut s'intéresser à la classe des schémas itératifs gradient-reliés pour lesquels une certaine suite Θ *donnée a priori* suffit à assurer la convergence. On pense évidemment ici au cas

particulier des suites Θ *constantes*, i.e. la classe des schémas itératifs *à pas fixe*[2].

Étudier la convergence des schémas itératifs à pas fixe se justifie par plusieurs points de vue. Tout d'abord, elle nous intéresse spécifiquement pour l'étude des algorithmes semi quadratiques de GY ou de GR qui font partie de cette « famille ». De manière plus générale, il nous semble d'un intérêt évident de savoir si la mise en œuvre d'une recherche linéaire est nécessaire à la convergence globale de l'algorithme.

9.3 Convergence globale à pas fixe *via* la règle d'Armijo

On étudie dans cette section la convergence globale d'un schéma itératif gradient relié à pas fixe *via* la règle d'ARMIJO. La simplicité formelle de la règle d'ARMIJO permet d'aboutir rapidement à des résultats fructueux ; en particulier, nous montrerons que son emploi est potentiellement moins restrictif qu'une étude basée sur d'autres « critères » telle que l' « approximation supérieure » déjà rencontrée dans le cadre du chapitre précédent. Commençons tout d'abord par préciser ce qu'on entend par « schéma itératif à pas fixe » dans ce document.

[2]Les suites $\Theta \longrightarrow 0$ à somme divergente et à somme des carrés convergentes constitue un autre cas « classique » de suite Θ donnée *a priori* qui ne sera pas développé dans ce document.

1 – Admissibilité du pas fixe au sens d'Armijo

Un schéma itératif $(\mathcal{X}, \Xi, \Theta)$ pour lequel le pas $\theta^{(n)}$, susceptible de varier au cours des itérations, est remplacé par un pas θ constant est appelé *schéma itératif à pas fixe.* On notera alors que ces algorithmes ne nécessitent aucune procédure de recherche linéaire et que (9.1) prend la forme particulière,

$$\boldsymbol{x}^{(n+1)} = \boldsymbol{x}^{(n)} + \theta \boldsymbol{\xi}^{(n)}, \quad \theta > 0. \tag{9.8}$$

Dans ce cas de figure, Θ devient une suite *constante* $\{\theta\}_{n=0,\ldots}$ que nous noterons, pour des questions de commodités, $\Theta = \theta$; le schéma itératif à pas fixe sera alors noté $(\mathcal{X}, \Xi, \theta)$.

Convergence globale pour le pas fixe

Fixer la longueur du pas ne permet généralement pas d'assurer la convergence globale des suites Ξ gradient-reliées à $\mathcal{X}$. *Cependant, à la lumière des définitions 6 et 7 et du fonctionnement de la technique du rebroussement, la proposition suivante fournit une CS assurant la convergence globale pour un schéma itératif à pas fixe :*

Proposition 3 *Soit un schéma itératif $(\mathcal{X}, \Xi, \Theta)$ vérifiant (9.8), si la suite Ξ est de type gradient-relié et si :*

$$\exists\, \omega \in]0;1[: \ \forall \boldsymbol{u} \in \mathbb{R}^N, \qquad J\left(\boldsymbol{u} + \theta \boldsymbol{\xi}(\boldsymbol{u})\right) - J(\boldsymbol{u}) - \omega\theta \left\langle \boldsymbol{\nabla} J(\boldsymbol{u}), \boldsymbol{\xi}(\boldsymbol{u}) \right\rangle \ \leq 0 \tag{9.9}$$

alors tout point d'accumulation de $\mathcal{X}$ est un point stationnaire. D'autre part, si J est strictement convexe et coercive, alors $\mathcal{X}$ converge vers le minimum global $\boldsymbol{x}^*$ de J.

Preuve Il devrait apparaître clairement que les conditions introduites dans cette proposition implique les hypothèses de la proposition 2. En effet, le caractère gradient-relié de Ξ est posé comme hypothèse commune; pour sa part, la condition (9.9) permet d'assurer que (9.5) tient avec $\theta^{(n)} = \theta$ pour $n = 0$ et pour toutes les itérations suivantes : la suite constante $\Theta = \theta$ est donc admissible au sens d'ARMIJO pour tout $(\mathcal{X}, \Xi)$ vérifiant (9.8). D'autre part, il est également évident que la condition (9.9) garantit, pour toutes les itérations, l'acceptation du pas initial $s = \theta$ par la technique du rebroussement. Finalement, si J est strictement convexe et coercive, le théorème 1 assure la convergence vers le minimum global pour le schéma itératif à pas fixe $(\mathcal{X}, \Xi, \theta)$ correspondant. ■

Note : Avant de poursuivre, on introduit pour le reste de cette étude la désignation suivante : on dira que *la suite constante $\Theta = \theta$ (ou le pas fixe θ) est admissible au sens d'*ARMIJO si la condition (9.9) est vérifiée.

L'admissibilité du pas fixe au sens d'ARMIJO ne constitue pas l'unique démarche possible pour étudier la convergence d'un schéma itératif à pas fixe. À cet égard, la condition d' « approximation supérieure » conduit également à des CS de conver-

gence pour les algorithmes à pas fixe.

2 – Comparaison avec l' « approximation supérieure »

Cette sous-section montre qu'un lien fort existe entre la condition d'approximation supérieure et la propriété d'admissibilité de la suite $\Theta = \theta$ énoncée par (9.9) : la première implique la seconde, qui est donc moins restrictive et constitue un outil d'étude de convergence plus puissant.

Algorithme de WEISZFELD *généralisé (forme relaxée)*

Posons tout d'abord $\tilde{J}(\boldsymbol{u}, \boldsymbol{v})$ une approximation quadratique tangente de J au sens de la définition 5 (page 261) et vérifiant l'hypothèse supplémentaire suivante :

$$\forall \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^N, \qquad \tilde{J}(\boldsymbol{u}, \boldsymbol{v}) \geq J(\boldsymbol{v}); \tag{9.10}$$

On construit alors la suite $\mathcal{X}$ à partir de la relation suivante[3] :

$$\boldsymbol{x}^{(n+1)} = (1-\theta)\, \boldsymbol{x}^{(n)} + \theta \arg\min_{\boldsymbol{v}} \tilde{J}(\boldsymbol{x}^{(n)}, \boldsymbol{v}); \tag{9.11}$$

expression de remise à jour similaire à (8.42) pour un pas θ fixe. Conformément aux développements de la section 9.1, l'algorithme défini par (9.11) s'identifie à

[3]L'écriture qui suit a bien un sens puisque la définition 5 assure l'existence et l'unicité pour tout $\boldsymbol{u}$ de $\arg\min_{\boldsymbol{v}} \tilde{J}(\boldsymbol{u}, \boldsymbol{v})$.

un schéma itératif $(\mathcal{X}, \Xi, \theta)$ vérifiant (9.8) et (9.2) pour lequel $\boldsymbol{B}(\boldsymbol{u})$ satisfait la condition (9.10).

La famille d'algorithmes qui vient d'être construite est connue de la communauté de l'*analyse numérique* depuis les années 1930 sous le nom d'*algorithmes de* WEISZFELD *généralisés* [Weiszfeld, 1937]. Comme nous l'avons déjà évoqué au chapitre précédent, cette famille a été redécouverte par la communauté du *traitement d'images et de signaux* à la suite des travaux sur l'algorithme EM ; dans cette littérature, l'approximation majorante est qualifiée de « *surrogate function* » et le principe algorithmique de « *optimization transfert* ». Pour notre part, nous préférons conserver les désignations *approximation supérieure* (ou *majorante*) et *algorithme de* WEISZFELD *généralisé* dans la mesure où elles renvoient directement aux contributions originales.

Résultats de convergence globale

Moyennant des hypothèses classiques sur J (coercive, strictement convexe et C^2), la convergence globale de cette famille d'algorithmes est établie dans [Voss et Eckhardt, 1980] pour le pas unitaire, i.e. $\theta = 1$, sous les conditions techniques suivantes :

$$\forall \boldsymbol{u} \in \mathbb{R}, \qquad \boldsymbol{B}(\boldsymbol{u}) \text{ continue}, \tag{9.12a}$$

$$\exists\, \gamma > 0 : \ \forall \boldsymbol{u} \in \mathbb{R}, \qquad \min \Lambda\left[\boldsymbol{B}(\boldsymbol{u})\right] \geq \gamma; \tag{9.12b}$$

avec $\Lambda[\boldsymbol{B}]$ l'ensemble fini des valeurs propres de $\boldsymbol{B} \in \mathbb{R}^{N\times N}$. Sous certaines adaptations mineures, ce même résultat de convergence peut être établi pour la version *relaxée* de l'algorithme à la condition que θ soit pris dans $]0;2[$. Cette dernière condition est nécessaire pour garantir le caractère monotone non croissant de la suite $\mathcal{J}$; c'est ce que montre la proposition suivante :

Proposition 4 *Soit $J : \mathbb{R}^N \to \mathbb{R}$ une fonction C^1, et $(\mathcal{X}, \Xi, \theta)$ un schéma itératif à pas fixe vérifiant (9.8) et (9.2) avec $\boldsymbol{B}(\cdot)$ un opérateur tel que l'hypothèse (9.10) est vérifiée. Si $\theta \in]0;2[$, alors on a :*

$$J(\boldsymbol{x}^{(n)}) \geq J(\boldsymbol{x}^{(n+1)}).$$

Preuve La preuve est donnée en annexe II, page 368. ■

*Lien entre l'approximation supérieure et la règle d'*ARMIJO

Dans le cadre d'un schéma itératif à pas fixe $(\mathcal{X}, \Xi, \theta)$ vérifiant (9.8) et (9.2), la proposition suivante montre que la condition d'approximation supérieure (9.10) entraîne *de facto* la condition (9.9) prise en $\omega = 1 - \theta/2$.

Proposition 5 *Soit $J : \mathbb{R}^N \to \mathbb{R}$ une fonction C^1 et $\tilde{J}$ une approximation qua-*

dratique de J (cf. définition 8.40) et vérifiant l'hypothèse (9.10). Si on a

$$\forall \boldsymbol{u}, \quad \boldsymbol{B}(\boldsymbol{u})\,\boldsymbol{\xi}(\boldsymbol{u}) = -\boldsymbol{\nabla} J(\boldsymbol{u}),$$

*alors l'inégalité d'*ARMIJO *(9.4) est vraie pour tout $\boldsymbol{u}$ pour la valeur $\omega = 1 - \theta/2$.*

Preuve La preuve est donnée en annexe II, page 369 ■

Conformément à la section 9.2, les conditions techniques (9.12) sont suffisantes pour garantir le caractère gradient-relié de la suite Ξ produite par l'algorithme de WEISZFELD généralisé. À la lumière du résultat précédent, on peut affirmer que *les CS de convergence obtenues par la condition d'approximation supérieure peuvent être plus restrictives que celles découlant de la règle d'*ARMIJO. Ainsi, les schémas itératifs $(\mathcal{X}, \Xi, \Theta)$ pour lesquels $\Theta = \theta$ vérifie l'hypothèse (9.9) seulement pour $\omega \in]0; \omega_{\max}[$ avec $\omega_{\max} < 1 - \frac{\theta}{2} < 1$ sont convergents, dans la mesure où Ξ est gradient-reliée, bien que ne découlant pas d'une approximation supérieure. On en déduit finalement que la classe des algorithmes de type « approximation majorante » est *incluse* dans celle définie par les algorithmes à directions gradient-reliées et à pas fixe admissible au sens d'ARMIJO.

Incidence pour cette étude

Les derniers développements indiquent qu'une étude de convergence basée sur la règle d'ARMIJO conduit potentiellement à des CS moins restrictives que celles obtenues en assurant le caractère « majorant » du schéma itératif. Pour l'algorithme SQ de GEMAN et YANG (abrég. GY), ce résultat s'avère fructueux : on a en effet montré à la fin du chapitre précédent que l'algorithme de GY définit un schéma itératif de type approximation majorante *sous réserve que les hypothèses propres à la construction SQ considérée soient vérifiées.* Or, nous verrons en section 9.4-9.4.1 que la règle d'ARMIJO permet d'assurer la convergence de l'algorithme même si ces hypothèses ne sont pas vérifiées.

En revanche, concernant l'algorithme de GEMAN et REYNOLDS (abrég. GR), les CS de convergences obtenues par la règle d'ARMIJO restent sensiblement identiques.

3 – Autre CS de convergence pour le pas fixe

On indique qu'une autre CS de convergence apparaît dans [Bertsekas, 1995, Prop. 1.2.3] pour un schéma itératif $(\mathcal{X}, \Xi, \theta)$. Cette CS s'appuie sur la possibilité de construire une approximation majorante monovariée de la fonction de θ

$$J(\boldsymbol{u} + \theta \boldsymbol{\xi}(\boldsymbol{u}))$$

et de montrer que le pas minimisant l'approximation quadratique vérifie la règle d'ARMIJO. Cette démarche s'est avérée moins facilement manipulable sur nos problèmes et a conduit à des résultats plus restrictifs que ceux déjà obtenus dans le cadre de cette étude.

9.4 Admissibilité du pas fixe pour les algorithmes SQ

Nous avons montré au chapitre précédent (section 8.4) que les algorithmes SQ étaient des schémas itératifs à pas fixe. En adoptant le vocabulaire et les notations introduites dans ce chapitre, ces algorithmes définissent un schéma itératif $(\mathcal{X}, \Xi, \theta)$ pour un opérateur $\boldsymbol{B}(\cdot)$ distinct suivant la formulation adoptée. C'est sous cet angle que nous réexaminons les propriétés de convergence de ces algorithmes ; rappelons avant de débuter que J est un critère C^1, *strictement convexe et coercif* qu'on écrira[4] :

$$J(\boldsymbol{x}) = Q(\boldsymbol{x}) + \Phi(\boldsymbol{x}) \qquad \text{avec} \qquad \Phi(\boldsymbol{x}) = \sum_{i=1}^{I} \phi(\delta_i) \tag{9.13}$$

avec $\phi : \mathbb{R} \to \mathbb{R}$, $\delta_i = \langle \boldsymbol{v}_i, \boldsymbol{x} \rangle$, et $\boldsymbol{v}_i \in \mathbb{R}^N$. Q est une forme quadratique *symétrique convexe* qu'on écrira :

$$Q(\boldsymbol{x}) = \langle \boldsymbol{Q}\boldsymbol{x}, \boldsymbol{x} \rangle - 2 \langle \boldsymbol{q}, \boldsymbol{x} \rangle + \mu,$$

[4]Au chapitre précédent, nous avions posé $\Phi(\boldsymbol{x}) = \sum_i \phi(\delta_i - w_i)$; afin de limiter la taille des expressions apparaissant dans ce chapitre (notamment celles apparaissant dans les preuves) nous avons choisi de poser $w_i = 0$. On se convaincra néanmoins aisément que les résultats de *convergence* énoncés dans ce chapitre restent valides.

c.à.d. que $\boldsymbol{Q}$ est une matrice définie non négative (abrég. DNN) *symétrique.* Le critère J retenu étant strictement convexe et coercif, la convergence va être établie dans le cadre de la proposition 3.

Admissibilité du pas fixe pour les critères pénalisés

Une formulation équivalente à l'inégalité d'ARMIJO pour la famille des critères pénalisés nous sera utile par la suite pour étudier la convergence des algorithmes considérés.

Proposition 6 *Soit $J : \mathbb{R}^N \to \mathbb{R}$ un critère C^1 défini par (9.13). Pour $\boldsymbol{\xi}(\boldsymbol{u})$ défini par*

$$\boldsymbol{B}(\boldsymbol{u})\,\boldsymbol{\xi}(\boldsymbol{u}) = -\boldsymbol{\nabla} J(\boldsymbol{u}),$$

*l'inégalité d'*ARMIJO *(9.4) s'écrit :*

$$\theta \langle [(1-\omega)\boldsymbol{B}(\boldsymbol{u}) - \theta \boldsymbol{Q}]\,\boldsymbol{\xi}(\boldsymbol{u}), \boldsymbol{\xi}(\boldsymbol{u})\rangle \geq \sum_i \left[\phi(\delta_i) - \phi\left(\delta_i^+\right) + \Delta_i \phi'(\delta_i)\right], \qquad (9.14)$$

où on a posé $\delta_i = \langle \boldsymbol{v}_i, \boldsymbol{u}\rangle$, $\delta_i^+ = \langle \boldsymbol{v}_i, \boldsymbol{u} + \theta\boldsymbol{\xi}(\boldsymbol{u})\rangle$, et $\Delta_i = \delta_i^+ - \delta_i$.

Preuve La preuve est donnée en annexe II, page 370. ■

Partant de cette proposition et des résultats de la section 9.2 sur le caractère gradient-relié de Ξ, nous sommes capable de réexaminer la convergence globale des

algorithmes SQ de GEMAN et YANG et de GEMAN et REYNOLDS.

9.4.1 Convergence globale des algorithmes SQ de GY

Nous commençons par une étude des algorithmes dérivés des deux formulations SQ de GY : soit la formulation avec mise à l'échelle et la formulation découlant de la dualité multivariée. Le principal résultat de cette section est un affaiblissement significatif des CS de convergence pour ces deux variantes de l'algorithme de GY.

1 – Schéma itératif de GY (dualité monovariée)

Dans le cas où la relaxation SQ est mise en œuvre sans *sur* ou *sous* relaxation des variables duales, nous avons montré au chapitre précédent que l'algorithme SQ de GY avec mise à l'échelle est un schéma itératif $(\mathcal{X}, \Xi, \theta)$ vérifiant (9.8) et (9.2) avec un opérateur $\boldsymbol{B}(\boldsymbol{u}) = \boldsymbol{B}^{a}_{\mathrm{GY}}$ défini par

$$\boldsymbol{B}^{a}_{\mathrm{GY}} = 2\boldsymbol{Q} + \frac{1}{a}\boldsymbol{V}\boldsymbol{V}^{T}, \tag{9.15}$$

où $a > 0$ est un paramètre dit d' « échelle ». Nous renvoyons respectivement en section 8.3.2 et 8.4 pour une présentation de cet algorithme dans son cadre primal/dual et de sa forme primale à pas fixe.

Conformément à la démarche exposée en début de section, nous analysons la convergence globale du schéma itératif de GY. Le résultat suivant établit dans

quelle mesure ce schéma itératif produit des suites de directions gradient-reliées :

Proposition 7 *Soit $\mathcal{X}$ et Ξ les suites d'itérées et de directions générées respectivement par les relations (9.2) et (9.8). Si $\boldsymbol{B}^{(n)}$ est donnée par (9.15) avec*

$$\ker\{\boldsymbol{Q}\} \cap \ker\{\boldsymbol{V}^T\} = \{\boldsymbol{0}\},$$

alors la suite de direction Ξ est gradient reliée.

Preuve La preuve est donnée en annexe II, page 372. ■

La proposition suivante donne les conditions d'admissibilité du pas fixe pour la règle d'ARMIJO pour un tel schéma itératif :

Proposition 8 *Soit $J : \mathbb{R}^N \to \mathbb{R}$ une fonction C^1 définie par (9.13) avec $\boldsymbol{Q}$ symétrique, ϕ strictement convexe et telle que*

$$\exists\ 0 < \widehat{a} < \infty : \quad \widehat{g}(u) = \frac{u^2}{2} - \widehat{a}\phi(u) \quad \textit{convexe.} \tag{9.16}$$

*Pour un schéma itératif $(\mathcal{X}, \Xi, \theta)$ défini par (9.8) et (9.2) avec $\boldsymbol{B}^{(n)}$ définie par (9.15), le pas fixe $0 < \theta < 2$ est admissible au sens d'*ARMIJO *pour tout $(\mathcal{X}, \Xi)$ si $a < 2\widehat{a}/\theta$.*

Preuve La preuve est donnée en annexe II, page 373. ■

Le lemme suivant nous apprend que la condition (9.16) est *équivalente* à la condition de Lipschitz de constante $L = 1/\widehat{a}$ pour ϕ'. D'autre part, si la condition (9.16) est vérifiée, alors on montre aisément que toute valeur de a prise dans l'intervalle (non vide) $[0;\widehat{a}]$ rend la fonction $g_a = (\cdot)^2/2 - a\phi(\cdot)$ convexe. Par la suite, $\widehat{a}$ s'entend comme la borne supérieure sur a qui permet la convexité de g_a.

Lemme 1 *Soit ϕ une fonction C^1 convexe. On pose $\widehat{g} = (\cdot)^2/2 - \widehat{a}\phi(\cdot)$; les deux assertions suivantes sont alors équivalentes :*

$$\begin{array}{lll} (a) & \widehat{g} & \textit{convexe,} \\ (b) & \phi' & \textit{L-Lipschitz pour } L = 1/\widehat{a}\,; \end{array} \tag{9.17}$$

d'autre part, pour $0 < a < \widehat{a}$, g est strictement convexe et l'inégalité de Lipschitz est strictement vérifiée.

Preuve La preuve est donnée en annexe II, page 363. ■

Les propositions 7 et 8 nous permettent de donner des CS de convergence de l'algorithme de GY :

Théorème 2 *Soit $J : \mathbb{R}^N \to \mathbb{R}$ une fonction C^1 définie par (9.13) avec $\boldsymbol{Q}$ symétrique, ϕ strictement convexe et telle que la condition (9.16) est réalisée. Pour*

$(\mathcal{X}, \Xi, \theta)$ un schéma itératif défini par (9.8) et (9.2), si $\boldsymbol{B}^{(n)}$ définie par (9.15) vérifie les hypothèses de la proposition 7, alors $\mathcal{X}$ converge vers $\boldsymbol{x}^*$, minimum global de J, pour $\theta \in]0;2[$ et $a < 2\widehat{a}/\theta$.

Preuve Montrons que la proposition 3 s'applique : les hypothèses concernant J sont bien remplies (J est C^1, coercive et strictement convexe puisque l'intersection des noyaux de $\boldsymbol{Q}$ et $\boldsymbol{V}^T$ se réduit au vecteur nul) ; la propositions 7 garantit le caractère gradient-relié de la suite Ξ, et la proposition 8 donne les conditions d'admissibilité au sens d'ARMIJO du pas fixe. ■

2 – Schéma itératif de GY (dualité multivariée)

On rappelle que dans une version sans sur relaxation des variables duales, l'algorithme SQ de GY *multivarié* est un schéma itératif $(\mathcal{X}, \Xi, \theta)$ vérifiant (9.8) et (9.2) avec un opérateur $\boldsymbol{B}(\boldsymbol{u}) = \boldsymbol{B}_\nu^a$ défini par

$$\boldsymbol{B}_\nu^a = 2\boldsymbol{Q} + \frac{1}{a}\boldsymbol{I}, \tag{9.18}$$

où $a > 0$ est un paramètre dit d' « échelle » et $\boldsymbol{I}$ la matrice identité de taille appropriée ; les sections 8.3.2 et 8.4 présentent, respectivement, la formulation de cet algorithme et son équivalence primale à pas fixe. On notera que $\boldsymbol{B}_\nu^a$ est nécessairement uniformément DP, ce qui assure le caractère gradient-relié de Ξ. La proposition suivante donne une CS assurant l'admissibilité du pas fixe pour ce

schéma itératif :

Proposition 9 *Soit* $J : \mathbb{R}^N \to \mathbb{R}$ *une fonction* C^1 *définie par (9.13) avec* $\boldsymbol{Q}$ *symétrique,* ϕ *strictement convexe et telle que*

$$\exists\ 0 < \widehat{a} < \infty : \quad \widehat{h}(\boldsymbol{u}) = \frac{||\boldsymbol{u}||^2}{2} - \widehat{a}\Phi(\boldsymbol{u}) \text{ convexe.} \tag{9.19}$$

Pour un schéma itératif $(\mathcal{X}, \Xi, \theta)$ *défini par (9.8) et (9.2) avec* $\boldsymbol{B}^{(n)}$ *définie par (9.18), le pas fixe* $0 < \theta < 2$ *est admissible au sens d'*ARMIJO *pour tout* $(\mathcal{X}, \Xi)$ *si* $a < 2\widehat{a}/\theta$.

Preuve La preuve est donnée en annexe II, page 375. ■

Pour ϕ de classe C^2, on montre facilement qu'il est possible de relier explicitement la valeur de $\widehat{a}$ avec le rayon spectral ρ de la matrice $\boldsymbol{V}\boldsymbol{V}^T$: $\widehat{a} = 1/\rho$. D'autre part, comme pour la formulation monovariée, si il existe $\widehat{a}$ tel que $\widehat{h}$ convexe, alors on montre facilement que toute valeur de a prise dans $]0; \widehat{a}[$ rend la fonction $h_a = ||\boldsymbol{u}||^2/2 - a\Phi(\boldsymbol{u})$ convexe. Finalement, moyennant une adaptation directe du théorème 2, la proposition ci-dessus fournit les CS de convergence globale pour l'algorithme de GY vectoriel.

3 – Algorithmes de GY : comparaison avec les CS existantes

Les résultats de convergence obtenus dans ce chapitre sont maintenant comparés à ceux produits par des outils différents. La construction de GY ayant été principalement développée dans le cadre de l'analyse convexe, nous commençons par donner les CS découlant de l'étude de convexité du critère primal/dual.

Via l'étude du critère primal/dual

On rappelle que l'algorithme SQ de GY étudié dérive d'un schéma de relaxation sur un critère primal/dual J^*_{GY} minimisant alternativement suivant toutes les variables primales, puis toutes les variables duales — *i.e.* l'algorithme de mise à jour « groupée » défini par (8.20) et (8.22) au chapitre précédent. Une étude de convergence à base d'analyse convexe est menée dans [Idier, 2001*b*, Sec. III] pour cet algorithme de relaxation ; on débute en donnant certains résultats saillants issus de cette référence. Introduisons la fonction scalaire g suivante pour l'étude de la forme monovariée

$$g \triangleq (\cdot)^2/2 - a\phi,$$

$a \geq 0$; en adoptant nos notations, la convergence de l'algorithme de GY vers le minimum *global* de J est assurée pour :

$$\begin{aligned} &(a) \quad \phi \textit{ strictement convexe,} \\ &(b) \quad g_a \textit{ strictement convexe,} \\ &(c) \quad \lim_{|u|\to\infty} \phi(u)/u^2 < 1/2a. \end{aligned} \tag{9.20}$$

Pour ϕ convexe et telle que (9.20b) puisse tenir, les valeurs de a qui rendent g_a convexe sont uniquement celles appartenant à un intervalle non vide du type $a \in [0;\widehat{a}]$; la stricte convexité tenant pour $0 \leq a < \widehat{a}$, cf. [Idier, 2001*b*, Sec. III.B].

En résumé, pour ϕ strictement convexe et a choisi[5] dans $]0;\widehat{a}[$, l'hypothèse technique (9.20c) suffit à assurer la convexité stricte du critère primal/dual, ce qui permet de garantir la convergence de l'algorithme de relaxation avec mise à jour « groupée ». Mentionnons finalement que la convergence des versions relaxées est assurée, sous réserve que le facteur de relaxation θ (respectivement γ) des variables primales (resp. duales) soit pris dans $]0;2[$ — resp. $]0;1[$.

L'objectif est maintenant de comparer les différentes conditions techniques suffisantes pour assurer la convergence, suivant qu'elles découlent de notre étude ou de l'étude de convexité du critère primal/dual effectué dans [Idier, 2001*b*]. Comme

[5]Bien que $a = 0$ permet de construire un g_a strictement convexe, la construction primale/duale avec « mise à l'échelle » n'est définie que pour $a \neq 0$.

hypothèse de travail initiale, on pose la stricte convexité de ϕ tout comme celle de J, ce qui permet d'assurer que l'opérateur

$$2\boldsymbol{Q} + \frac{1}{a}\boldsymbol{V}\boldsymbol{V}^T$$

est inversible (car DP), et ainsi que l'algorithme est bien défini. Le lemme 2 nous apprend que le caractère C^1 de ϕ n'a pas a être posé *a priori* pour garantir la convergence de l'algorithme de GY[6].

Lemme 2 *Soit ϕ une fonction convexe. Si $g_a = (\cdot)^2/2 - a\phi(\cdot)$ est convexe, alors ϕ et g sont des fonctions C^1.*

Preuve La preuve est donnée en annexe II, page 361. ■

À ce niveau de l'exposé, il importe de souligner que la convergence de l'algorithme est garantie *par l'étude de convexité du critère primal/dual* pour :

$$(\theta, a) \in]0; 2[\times]0; \widehat{a}[;$$

cette condition doit être comparée à celle obtenue par l'admissibilité du pas fixe

[6]Le caractère C^1 de ϕ aurait pu donc être omis dans le cadre de l'étude de convergence de la section 9.4.1, page 289. Bien que redondante avec la convexité de ϕ et le caractère Lipschitz de sa dérivée, nous la conservons cependant comme préliminaire de manière à conserver à cette étude son degré de généralité.

(cf. théorème 2) :

$$(\theta, a) \in]0; 2[\times]0; 2\widehat{a}/\theta[\,,$$

qui, bien que fonction du facteur de relaxation θ, est *moins restrictive* que la précédente. La figure 9.2 montre les domaines de convergence sur un graphe (θ, a). *Outre l'intérêt purement formel d'un tel résultat, cette extension du domaine de convergence présente un intérêt pratique majeur : des expériences préliminaires menées en simulations semblent indiquer qu'une vitesse de convergence plus élevée se produit pour* $a \approx 2\widehat{a}/\theta$ *;* nous reviendrons sur ce point lors de l'étude de vitesse asymptotique menée en section 9.5.

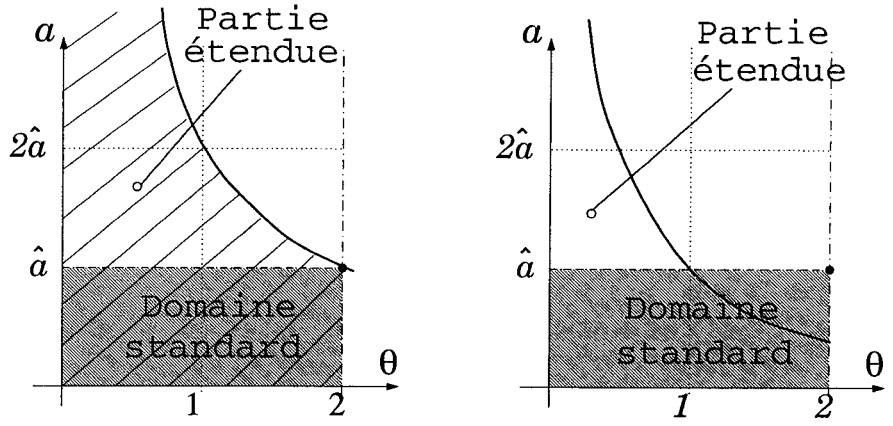


Figure 9.2: Domaines de convergence découlant : de l'étude de convexité du critère primal/dual (domaine standard) ; de l'admissibilité du pas fixe pour la règle d'ARMIJO (←) ou de la condition d'approximation supérieure (→).

Enfin, deux points supplémentaires méritent d'être soulignés avant d'examiner le cas multivarié : tout d'abord, l'étude de convergence par analyse convexe permet d'examiner l'effet d'une relaxation des variables duales, alors que la démarche adop-

tée dans notre étude se révèle inadéquate pour le faire ; d'autre part, la condition technique (9.20c) apparaissant dans l'étude de convexité du critère primal/dual disparaît dans les conditions techniques de notre étude.

Concernant la forme multivariée, la convergence vers le minimiseur global de J se déduit sans encombre des résultats scalaires [Idier *et al.*, 2002]. Introduisons la fonction $h_a : \mathbb{R}^N \to \mathbb{R}$ définie par

$$h_a(\cdot) \triangleq ||\cdot||^2/2 - a\Phi(\cdot).$$

Des conditions techniques similaires à celles données par (9.20) peuvent alors être énoncées :

$$\begin{array}{lll} (a) & \phi & \textit{strictement convexe,} \\ (b) & h_a & \textit{strictement convexe,} \\ (c) & \lim\limits_{||u||\to\infty} \Phi(u)/||u||^2 < 1/2a. & \end{array} \tag{9.21}$$

Ainsi, sous réserve que la condition technique (9.21c) tienne, la convergence globale est assurée si a est choisi dans la plage qui rend h_a convexe, i.e. $a \in]0;\widehat{a}[$, et pour les paramètres de relaxation primale θ et duale ν : $(\theta,\nu) \in]0;2[\times]0;1[$.

Une comparaison avec les domaines de convergence produits par cette étude montre que le domaine de convergence suivant le paramètre a peut, là encore, être

étendu puisque la proposition 9 garantit l'admissibilité du pas fixe pour :

$$a \in]0; 2\widehat{a}/\theta[\,.$$

Nous n'avons pas encore établi de résultat équivalent au lemme 1 pour la formulation multivariée. D'autre part, on indique que la condition technique (9.21c) n'a pas d'équivalent dans notre analyse et que le formalisme adopté dans cette étude ne permet pas l'étude de la relaxation des variables duales.

Via la condition d'approximation supérieure

Conformément à l'énoncé de la proposition 5, la condition d'approximation supérieure pour le schéma itératif à pas fixe de GY conduit à vérifier que l'inégalité ci-dessous tient pour tout $\boldsymbol{u}$:

$$\theta^2 \left\langle \left[\frac{1}{2}\boldsymbol{B}(\boldsymbol{u}) - \boldsymbol{Q}\right] \boldsymbol{\xi}(\boldsymbol{u}), \boldsymbol{\xi}(\boldsymbol{u}) \right\rangle - \sum_i \left[\phi\left(\delta_i\right) + \Delta_i \phi'\left(\delta_i\right) - \phi\left(\delta_i^+\right)\right] \geq 0; \quad (9.22)$$

outre les conditions techniques assurant le caractère gradient relié de la suite Ξ, on en déduit naturellement que les CS assurant la convergence sont, *à un facteur deux près*, celles données par la proposition 8 :

$$(\theta, a) \in]0; 2[\times]0; \widehat{a}/\theta[\,.$$

Finalement, les contraintes imposées sur le couple (θ, a) par cette approche sont les moins intéressantes de toutes celles obtenues jusqu'alors : elles sont plus restrictives qu'avec une admissibilité du pas fixe au sens d'ARMIJO, et elle ne permettent pas de garantir la convergence pour les couples (θ, a) qui sont actuellement largement reconnues comme les plus intéressants[7] : $a \approx \widehat{a}$ et $2 > \theta > 1$.

9.4.2 Convergence globale de l'algorithme SQ de GR

Dans le cas où la relaxation SQ est mise en œuvre sans *sur* ou *sous* relaxation des variables duales, nous avons montré au chapitre précédent que l'algorithme SQ de GR est un schéma itératif $(\mathcal{X}, \Xi, \theta)$ vérifiant (9.8) et (9.2) pour un opérateur $\boldsymbol{B}(\boldsymbol{u}) = \boldsymbol{B}_{\mathrm{GR}}(\boldsymbol{u})$ défini par

$$\boldsymbol{B}_{\mathrm{GR}}(\boldsymbol{u}) = 2\boldsymbol{Q} + \boldsymbol{V}\boldsymbol{L}(\boldsymbol{u})\boldsymbol{V}^T, \quad \text{avec} \quad \boldsymbol{L}(\boldsymbol{u}) = \mathrm{diag}(\phi'(\delta_i)/\delta_i) \tag{9.23}$$

avec $\delta_i = \langle \boldsymbol{v}_i, \boldsymbol{u} \rangle$. Nous renvoyons respectivement en section 8.3.2 et 8.4 pour une présentation de cet algorithme dans son cadre primal/dual et de sa forme primale à pas fixe.

[7] En ce sens qu'elle permettent la vitesse de convergence la plus intéressante pour l'algorithme ; nous reviendrons sur ce point lors de l'étude de vitesse asymptotique des algorithmes SQ développé en section 9.5.

1 – Convergence du schéma itératif de GR

Notons que cet algorithme est uniquement défini si l'inverse de $\boldsymbol{B}_{\mathrm{GR}}^{(n)}$ existe pour tout n. Cette condition est automatiquement vérifiée si les conditions de la proposition ci-dessous sont assurées.

Proposition 10 *Soit $\mathcal{X}$ et Ξ les suites d'itérées et de directions générées respectivement par les relations (9.2) et (9.8). Pour une fonction ϕ de classe C^1 et telle que*

$$\phi \quad \textit{paire}, \tag{9.24a}$$

$$\exists\, \infty > B > 0 : \forall u \in \mathbb{R}, \qquad \phi'(u)/u \leq B, \tag{9.24b}$$

si $\boldsymbol{B}^{(n)}$ découle de $\boldsymbol{B}(\boldsymbol{u})$ définie par (9.23) et vérifie

$$\ker\{\boldsymbol{Q}\} \cap \ker\{\boldsymbol{V}^T\} = \{\boldsymbol{0}\}, \tag{9.25}$$

alors l'itération de GR est bien définie et la suite de direction Ξ est gradient reliée.

Preuve La preuve est donnée en annexe II, page 377. ■

Reste à fournir les conditions d'admissibilité du pas fixe pour la règle d'ARMIJO. C'est le propos de la proposition suivante que nous faisons suivre du théorème de

convergence pour l'algorithme de GR.

Proposition 11 *Soit* $J : \mathbb{R}^N \to \mathbb{R}$ *une fonction* C^1 *définie par (9.13) avec* $\boldsymbol{Q}$ *symétrique,* ϕ *strictement convexe et telle que*

$$\phi(\sqrt{u}) \text{ concave sur } \mathbb{R}^+. \tag{9.26}$$

Pour un schéma itératif $(\mathcal{X}, \Xi, \theta)$ *défini par (9.8) et (9.2) avec* $\boldsymbol{B}^{(n)}$ *déduit de (9.23), le pas fixe* $0 < \theta < 2$ *est admissible au sens d'*ARMIJO *pour tout* $(\mathcal{X}, \Xi)$.

Preuve La preuve est donnée en annexe II, page 378. ■

Théorème 3 *Soit* $J : \mathbb{R}^N \to \mathbb{R}$ *une fonction* C^1 *définie par (9.13) avec* $\boldsymbol{Q}$ *symétrique et* ϕ *strictement convexe et telle que les conditions (9.24) et (9.26) sont réalisées. Pour* $(\mathcal{X}, \Xi, \theta)$ *un schéma itératif défini par (9.8) et (9.2), si* $\boldsymbol{B}^{(n)}$ *déduit de (9.23) vérifie la condition (9.25), alors* $\mathcal{X}$ *converge vers* $\boldsymbol{x}^*$, *minimum global de* J, *pour* $\theta \in]0; 2[$.

Preuve En tirant partie des propositions 10 et 11, la démonstration suit le même cheminement que la démonstration de convergence de l'algorithme de GY. ■

2 – Algorithmes de GR : comparaison avec les CS existantes

Comme pour l'algorithme de GY, les résultats de convergence que nous venons d'obtenir pour l'algorithme de GR sont comparés à ceux obtenus par une étude de

convexité du critère primal/dual d'une part, et par la condition d'approximation supérieure d'autre part.

Via l'étude du critère primal/dual

On rappelle que l'algorithme SQ de GR étudié dérive d'un schéma de relaxation sur un critère primal/dual J^*_{GR} minimisant alternativement suivant toutes les variables primales, puis toutes les variables duales — *i.e.* l'algorithme de mise à jour « groupée » défini par (8.20) et (8.21) au chapitre précédent. Une présentation de la construction de GR ainsi qu'une étude de convergence à base d'analyse convexe est menée dans [Idier, 2001*b*, Sec. IV]. On donne dans ce paragraphe les points marquant pour notre étude issus de cette référence. Rappelons en premier lieu les conditions permettant la construction du critère primal/dual de GR :

$$\begin{array}{lll} (a) & \phi & \textit{paire,} \\ (b) & \phi(\sqrt{\cdot}) & \textit{concave sur } \mathbb{R}^+, \\ (c) & \phi & \textit{continue à l'origine et } C^1 \textit{ sur } \mathbb{R} \setminus \{0\}. \end{array} \tag{9.27}$$

La convergence de l'algorithme de GR vers le minimum *global* de J est alors assurée si la matrice $(\boldsymbol{Q} + \boldsymbol{V}\boldsymbol{V}^T)$ est inversible et si les conditions techniques

suivantes sont réalisées [Idier, 2001*b*, Théorème 3] :

$$\begin{aligned} &(a) \quad \phi \textit{ strictement convexe,} \\ &(b) \quad \lim_{|u|\to\infty} \phi'(u)/u = 0, \\ &(c) \quad \lim_{|u|\to 0} \phi'(u)/u < \infty, \end{aligned} \tag{9.28}$$

et que la condition (9.28b) tient de manière *stricte*, i.e.

$$\phi(\sqrt{\cdot}) \textit{ strictement concave sur } \mathbb{R}^+. \tag{9.29}$$

Comparons maintenant ces CS de convergence avec celles obtenues dans ce document. Pour les deux études, on pose comme hypothèse initiale la stricte convexité de ϕ et le caractère disjoint des noyaux de $\boldsymbol{Q}$ et de $\boldsymbol{V}^T$ — cette dernière condition permettant d'assurer la stricte convexité du critère J. On notera que l'hypothèse (9.27c) est légèrement plus faible que l'hypothèse de continue différentiabilité de ϕ faite dans notre étude.

Intéressons nous aux conditions (9.24) qui permettent d'assurer dans notre étude le caractère gradient-reliée de la suite Ξ : on notera que pour ϕ de classe C^1, le seul point de discontinuité de $\phi'(u)/u$ est éventuellement à l'origine, ce qui implique l'équivalence des conditions (9.28c) et (9.24b) ; la parité de ϕ est supposée par les deux études alors que la condition (9.28b) ne semble pas avoir de contrepartie dans

notre cadre d'analyse[8]. On notera enfin que la concavité stricte est nécessaire pour montrer la convergence dans [Idier, 2001*b*] alors que notre étude se satisfait d'une concavité au sens large pour que le pas fixe soit admissible. Ceci peut être intéressant dans la mesure où la pénalisation de HUBER ne vérifie pas cette condition strictement : cette étude semble donc élargir le résultat de convergence pour cette pénalisation.

Hormis l'extension des CS à la fonction de HUBER, les conditions techniques produites par notre étude sont très similaires à celles obtenues dans le cadre de l'analyse convexe.

Via la condition d'approximation supérieure

Les CS obtenues par la condition d'approximation supérieure sont en tous points identiques à ceux de notre étude. Le lecteur pourra d'ailleurs constater que l'admissibilité du pas fixe au sens d'ARMIJO est démontrée avec la valeur particulière $\omega = 1 - \theta/2$ (cf. preuve de la proposition 11, ce qui équivaut effectivement à considérer l'inégalité (9.22) découlant de la condition d'approximation supérieure. Soulignons ici que cette démarche n'est pas réductrice : considérer $\omega = 1-\theta/2$ nous mène assez directement à la condition technique (9.26) ; ce résultat tient toujours pour une valeur quelconque de $\omega \in]0;1[$ mais la preuve est plus longue...

[8]On indique néanmoins que cette condition est équivalente à $\lim_{u\to\infty} \phi(u)/u^2$ qui est souvent vérifiée pour les pénalisations adoptées en traitement de signal et d'image.

9.5 Convergence locale et vitesse asymptotique

Nous présentons maintenant certains résultats de régime asymptotique pour les algorithmes de GY et de GR. Dans cette section, nous tentons de répondre à certaines questions importantes : (1) *quel est l'ordre et le taux de convergence de ces deux algorithmes*, (2) *avons nous un critère quantitatif fiable permettant de choisir entre ces deux familles d'algorithmes*, (3) *parmi les formes paramétrées de GY, quelle est la plus rapide.* En dépit de nos efforts, nous ne sommes pas parvenu à donner une réponse complète à chacune de ces questions. Ceci étant, certain éléments intéressants sont présentés dans le cadre de ce chapitre.

Hypothèses préliminaires et notations

Dans ce qui suit, on note $\text{int}(D)$ l'intérieur de l'ensemble D, $\Lambda[\boldsymbol{A}] = \{\lambda_i\}_{i=1}^N$ l'ensemble des N valeurs propres (possiblement complexes) de $\boldsymbol{A} \in \mathbb{R}^{N\times N}$. On rappelle que les éléments de $\Lambda[\boldsymbol{A}]$ sont nécessairement réels si $\boldsymbol{A}$ est une matrice symétrique ; dans ce cas, il sera parfois commode de noter la plus petite et la plus grande de ces valeurs propres respectivement par $\lambda_{\min}[\boldsymbol{A}]$ et $\lambda_{\max}[\boldsymbol{A}]$. Enfin, on rappelle que *le rayon spectral* de $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ est défini par

$$\rho(\boldsymbol{A}) = \max_{i=1,\cdots,N} \{|\lambda_i|\}$$

où $|\cdot|$ désigne le module. On considère pour cette section J deux fois continûment différentiable (abrég. C^2) de manière à simplifier l'analyse de convergence asymptotique. La base de cette étude s'appuie largement sur [Ortega et Rheinboldt, 1970] auquel nous empruntons une partie du formalisme.

Ordre et taux de convergence : résultats généraux

Dans un cadre général, on introduit la classe des *algorithmes itératifs stationnaires du premier ordre* défini par

$$\boldsymbol{x}^{(n+1)} = \mathcal{M}(\boldsymbol{x}^{(n)}) \quad n = 1, \cdots, \tag{9.30}$$

avec $\mathcal{M} : D \subset \mathbb{R}^N \to \mathbb{R}^N$. Le théorème d'OSTROWSKI [Ortega et Rheinboldt, 1970, 10.1.3] énoncé ci-dessous permet d'assurer que l'application (9.30) est contractante ; dans ce qui suit, $\mathcal{M}'(\boldsymbol{x})$ est la différentielle de $\mathcal{M}$ en $\boldsymbol{x} \in \mathbb{R}^N$.

Théorème 4 *Supposons que $\mathcal{M} : D \subset \mathbb{R}^N \to \mathbb{R}^N$ ait un point fixe $\boldsymbol{x}^\star \in \mathrm{int}(D)$ et qu'elle soit différentiable en $\boldsymbol{x}^\star$. Si le rayon spectral $\rho(\mathcal{M}'(\boldsymbol{x}^\star)) < 1$, alors $\boldsymbol{x}^\star$ est un point d'attraction de l'itération (9.30).*

Sous réserve que les conditions du précédent théorème soient vérifiées, on peut établir le *taux* et l'*ordre* de convergence en racine de l'itération (9.30).

Théorème 5 *Supposons que les hypothèses du théorème 4 soient vérifiées. Alors*

le taux de convergence en racine de l'itération (9.30) est $\rho(\mathcal{M}'(\boldsymbol{x}^\star))$. *De plus, si* $\rho(\mathcal{M}'(\boldsymbol{x}^\star)) > 0$, *alors l'ordre de convergence est linéaire.*

Le théorème précédent est issu de [Ortega et Rheinboldt, 1970, 10.1.4] ; une définition précise de l'ordre et du taux de convergence en racine peut être trouvée à la section 9.2 du même ouvrage. On rappelle que le taux $\rho(\mathcal{M}'(\boldsymbol{x}^\star))$ est d'autant plus intéressant qu'il est proche de zéro ; pour $\rho(\mathcal{M}'(\boldsymbol{x}^\star)) = 0$, la convergence est *super linéaire.* Avant de poursuivre, insistons néanmoins sur le caractère *asymptotique* de ces indices : en particulier, l'ordre et le taux de convergence ne permettent pas de conclure sur la vitesse de convergence « loin » de la solution.

Algorithmes à pas fixe

Intéressons nous maintenant au cas particulier des algorithmes à pas fixe : on introduit l'application $\mathcal{M}_\theta : \mathbb{R}^N \to \mathbb{R}^N$ de la forme

$$\mathcal{M}_\theta(\boldsymbol{u}) = \boldsymbol{u} - \theta \boldsymbol{B}(\boldsymbol{u})^{-1} \boldsymbol{\nabla} J(\boldsymbol{u}) ; \tag{9.31}$$

qui permet de redéfinir les schémas itératifs à pas fixe $(\mathcal{X}, \Xi, \theta)$ vérifiant (9.8) et (9.2) au travers de la relation de récurrence

$$\boldsymbol{x}^{(n+1)} = \mathcal{M}_\theta(\boldsymbol{x}^{(n)}); \tag{9.32}$$

clairement, ces formes algorithmiques sont des cas particuliers d'algorithmes itératifs stationnaires du premier ordre. Pour J strictement convexe et coercif, on garantit l'existence et l'unicité d'un point fixe $\boldsymbol{x}^\star$. D'autre part, sous réserve que $\mathcal{M}'_\theta$ existe effectivement, on a [Ortega et Rheinboldt, 1970, 10.2.1] :

$$\mathcal{M}'_\theta(\boldsymbol{u}) = \boldsymbol{I} - \theta \boldsymbol{B}(\boldsymbol{u})^{-1} \boldsymbol{H}(\boldsymbol{u}) \tag{9.33}$$

où $\boldsymbol{I}$ et $\boldsymbol{H}$ désignent respectivement l'opérateur identité et le Hessien de J. On indique qu'il suffit que $\boldsymbol{B}(\boldsymbol{u})$ soit continue sur $\mathbb{R}^N$ et J de classe C^2 pour garantir que $\mathcal{M}$ est différentiable pour tout $\boldsymbol{u} \in \mathbb{R}^N$ (donc y compris pour $\boldsymbol{u} = \boldsymbol{x}^\star$, le minimiseur de J). Sous les hypothèses que nous venons d'énoncer, le taux de convergence en racine de l'algorithme à pas fixe $\rho(\mathcal{M}'_\theta(\boldsymbol{x}^\star))$ s'écrit :

$$\rho\left(\boldsymbol{I} - \theta \boldsymbol{B}(\boldsymbol{x}^\star)^{-1} \boldsymbol{H}(\boldsymbol{x}^\star)\right) \equiv \sigma(\theta, \boldsymbol{x}^\star); \tag{9.34}$$

ce taux est alors d'autant plus intéressant (*i.e.* $\ll 1$) que $\boldsymbol{B}(\boldsymbol{x}^\star)$ est « proche » de $\boldsymbol{H}(\boldsymbol{x}^\star)$. Dans le cas particulier $\boldsymbol{B} \equiv \boldsymbol{H}$ et $\theta = 1$, les itérations sont celles d'un algorithme de NEWTON : le taux de convergence est alors égal à zéro et l'ordre de convergence est *quadratique* pour J suffisamment régulière — cf. par ex. [Gilbert, 1999, Th 7.1]. Néanmoins, contrairement aux schémas itératifs à pas fixe de GR et GY, la convergence globale de l'algorithme de NEWTON ne semble pas garantie

pour notre problème.

Propriétés asymptotiques des algorithmes SQ

Les algorithmes de GY et de GR sont des schémas itératifs découlant d'une application $\mathcal{M}_\theta$ définie par (9.31) et pour lesquels $\boldsymbol{B}(\boldsymbol{u})$ prend une forme dépendant de la construction adoptée. À partir des résultats précédents, il n'est pas difficile de montrer que l'ordre de convergence de ces algorithmes est au moins linéaire, et que leur taux en racine s'écrit (9.34) pour :

$$\forall \boldsymbol{u} \in \mathbb{R}^N, \quad \boldsymbol{B}(\boldsymbol{u}) \equiv \begin{cases} \boldsymbol{B}_{\mathrm{GR}}(\boldsymbol{u}) & \text{définie par (9.23),} \\ \boldsymbol{B}^a_{\mathrm{GY}} & \text{définie par (9.15),} \\ \boldsymbol{B}^a_{\mathcal{V}} & \text{définie par (9.18);} \end{cases} \tag{9.35}$$

ceci est établi précisément pour l'algorithme de GR par la proposition suivante.

Proposition 12 *Soit $J : \mathbb{R}^N \to \mathbb{R}$ une fonction de la forme (9.13) avec ϕ une fonction strictement convexe au moins C^2 et vérifiant les hypothèses du théorème 3. Supposons que J est minimisée en $\boldsymbol{x}^\star$; sous réserve que $\theta \in]0;2[$, alors l'algorithme de GR a un ordre de convergence au moins linéaire et son taux de convergence en racine s'écrit $\rho\left(\boldsymbol{I} - \theta \boldsymbol{B}_{\mathrm{GR}}(\boldsymbol{x}^\star)^{-1}\boldsymbol{H}(\boldsymbol{x}^\star)\right) \equiv \sigma_{\mathrm{GR}}(\theta)$.*

Preuve La preuve est donnée en annexe II, page 379. ■

Ce résultat de vitesse asymptotique a été précédemment établis par [Chan et Mulet, 1999, Sec. 6] en tirant partie du fait que l'algorithme de GR est de type WEISZFELD généralisé ; pour notre part, nous préférons nous appuyer sur les résultats plus généraux que constituent les théorèmes 4 et 5.

La preuve de la proposition 12 s'adapte aisément de manière à obtenir un résultat similaire pour l'algorithme de GY monovarié et multivarié : ainsi pour $\theta \in]0;2[$ et $a \in]0;2\widehat{a}/\theta[$, les ordres de convergence sont linéaires et le taux s'écrit respectivement :

$$\rho\left(\boldsymbol{I} - \theta \boldsymbol{B}^{a}_{\mathrm{GY}}(\boldsymbol{x}^{\star})^{-1}\boldsymbol{H}(\boldsymbol{x}^{\star})\right) \equiv \sigma_{\mathrm{GY}}(\theta, a) \qquad \text{et} \qquad \rho\left(\boldsymbol{I} - \theta \boldsymbol{B}^{a}_{\mathcal{V}}(\boldsymbol{x}^{\star})^{-1}\boldsymbol{H}(\boldsymbol{x}^{\star})\right) \equiv \sigma_{\mathcal{V}}(\theta, a).$$

Contrairement à la forme itérative de GR, on note que les formes itératives de GY conduisent à un taux de convergence dépendant de *deux* paramètre, soit le pas fixe θ et le « paramètre d'échelle » a.

Si on écarte la pénalisation quadratique[9], alors on constate invariablement qu'en pratique, les trois algorithmes SQ ont un ordre de convergence *au mieux linéaire. Dans ce cas, la seule comparaison des taux asymptotique ne permet pas de déterminer lequel de ces algorithmes est le plus intéressant* : un algorithme de moindre coût peut en particulier s'avérer plus intéressant même si son taux n'est pas le

[9]Si $\phi(u) = u^2$, alors on montre facilement que pour $\theta = 1$ et $a = \widehat{a}$, les itérations de GR et GY sont identiques à un algorithme de NEWTON.

meilleur[10]. Il s'agit donc ici de nuancer un point de vue largement répandu qui est de privilégier *systématiquement* l'algorithme de GR puisque son taux de convergence est semble t-il meilleur — cf. ci-dessous. Il est alors certainement bon de rappeler qu'au contraire des algorithmes de GY, cet algorithme nécessite la résolution d'une équation normale par itération...

Comparaison des vitesses asymptotiques des constructions SQ

Expérimentalement, les taux de convergence semblent vérifier l'inégalité suivante

$$\forall a > 0, \qquad \sigma_{\mathcal{V}}(\theta^\star_{\mathcal{V}}, a) \geq \sigma_{\mathrm{GY}}(\theta^\star_{\mathrm{GY}}, a) \geq \sigma_{\mathrm{GR}}(\theta^\star_{\mathrm{GR}}) > 0. \tag{9.36}$$

où $\theta^\star_\bullet$ correspond à la valeur de paramètre θ permettant la meilleur convergence asymptotique pour chaque algorithme. Dans un cadre plus rigoureux, la comparaison *explicite* des taux reste formellement difficile si on excepte des cas particuliers trop simplifiés pour être intéressants en pratique. *Dans ce qui suit, on choisit alors de privilégier une analyse qualitative basée sur une simple comparaison de la forme du Hessien et des opérateurs $\boldsymbol{B}$ associés aux algorithmes SQ.*

De manière claire, le taux de convergence des algorithmes SQ dépend fortement

[10]Le taux de convergence en *quotient* [Ortega et Rheinboldt, 1970, Sec. 9.1] donne une idée claire de l'impact de l'implantation. Considérons deux algorithmes $\mathcal{M}_1$ et $\mathcal{M}_2$ de coût par itération respectif égale à C_1 et C_2, d'un ordre de convergence linéaire et d'un taux respectif égal à γ_1 et γ_2 ; pour tenir compte du coût par itération, on constate simplement à partir de la définition du taux de convergence en quotient que les quantités γ_1 et $\gamma_2^{C_1/C_2}$ doivent être comparés.

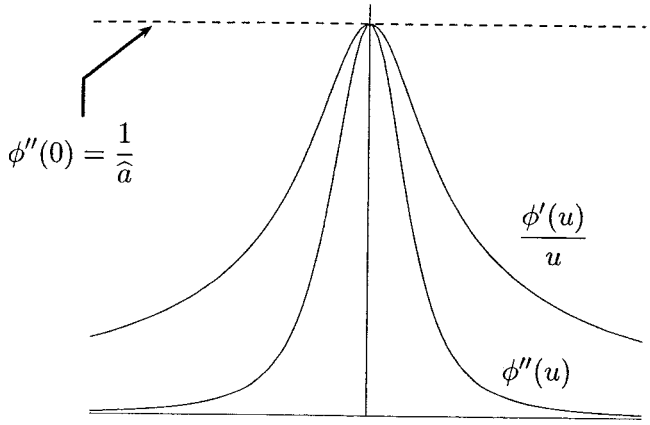


Figure 9.3: Tracé de $\phi''(u)$ et de $\phi'(u)/u$ pour $\phi(u) = \sqrt{u^2 + s^2}$; on remarquera que le maximum de $\phi'(u)/u$ est atteinte en zéro et qu'elle correspond à la courbure maximale de ϕ.

du produit $\boldsymbol{B}(\boldsymbol{u})^{-1}\boldsymbol{H}(\boldsymbol{u})$ où le Hessien s'écrit à partir de (9.13)

$$\boldsymbol{H}(\boldsymbol{u}) = 2\boldsymbol{Q} + \boldsymbol{V}\,\mathrm{diag}(\phi''(\delta_i))\boldsymbol{V}^T; \tag{9.37}$$

et avec $\boldsymbol{B}$ choisie parmi les trois opérateurs suivant :

$$\boldsymbol{B}_{\mathrm{GR}}(\boldsymbol{u}) = 2\boldsymbol{Q}+\boldsymbol{V}\,\mathrm{diag}(\phi'(\delta_i)/\delta_i)\boldsymbol{V}^T, \qquad \boldsymbol{B}^a_{\mathrm{GY}} = 2\boldsymbol{Q}+\frac{1}{a}\boldsymbol{V}\boldsymbol{V}^T, \qquad \boldsymbol{B}^a_{\nu} = 2\boldsymbol{Q}+\frac{1}{a}\boldsymbol{I};$$

on constate donc finalement que $\boldsymbol{H}(\boldsymbol{u})$ et $\boldsymbol{B}(\boldsymbol{u})$ diffèrent par le second terme. Pour les fonctions ϕ adoptées en traitement d'image et de signal, $\boldsymbol{B}_{\mathrm{GR}}(\boldsymbol{u})$ constitue souvent une meilleure approximation de $\boldsymbol{H}(\boldsymbol{u})$ que ne le sont $\boldsymbol{B}^a_{\mathrm{GY}}$ et $\boldsymbol{B}^a_{\nu}$; c'est en particulier vrai pour les fonctions ϕ convexes « *préservant les contours* » [11]

[11] Le terme de fonction « edge preserving » est largement adoptée par la littérature du traitement d'image : elle correspond aux fonctions ϕ dont le pouvoir de pénalisation pour les fortes variations

largement répandues qui vérifient très souvent

$$\lim_{||u||\to\infty} \frac{\phi(u)}{u^2} = 0. \tag{9.38}$$

Pour ces fonctions, on montre que $\phi''(u)$ tend vers zéro à l'infini et que $\phi'(u)/u$ vérifie l'inégalité suivante — cf. lemmes 4 et 5 en annexe :

$$\forall u \in \mathbb{R}, \qquad \phi''(0) \geq \frac{\phi'(u)}{u} \geq \phi''(u). \tag{9.39}$$

Comme la figure 9.3 l'illustre, les quantités $\phi''(\delta_i)$ apparaissant dans le Hessien $\boldsymbol{H}(\boldsymbol{u})$ sont qualitativement bien approchées par les $\phi'(\delta_i)/\delta_i$ pour ce type de fonctions ϕ. En guise de comparaison, la matrice à coefficient constant $\boldsymbol{B}^a_{\text{GY}}$ approche la diagonale $\text{diag}(\phi''(\delta_i))$ par un *unique* paramètre $1/a$ dont l'ajustement revient à l'utilisateur ; enfin, la matrice $\boldsymbol{B}^a_{\mathcal{V}}$ approche le second terme $\boldsymbol{V}\,\text{diag}(\phi''(\delta_i))\boldsymbol{V}^T$ par une matrice diagonale constante $\boldsymbol{I}/a$. Ces constations simples expliquent sur une base heuristique la différence de comportement asymptotique des trois algorithmes SQ et la plus grande vitesse de convergence de l'algorithme de GR. Ce résultat n'a évidemment pas valeur de démonstration : $\boldsymbol{Q}$ et $\boldsymbol{V}^T\boldsymbol{V}$ ne se diagonalisant pas dans une même base en général, nous ne sommes pas parvenu à montrer que l'inégalité (9.39) implique la relation (9.36).

---

locales est moins accentuée que la pénalisation quadratique $\phi(u) = u^2$; c'est ce que formalise la limite ci-dessus.

Avant de poursuivre, on indique que l'approximation quadratique locale obtenue par la fonction $\phi'(u)/u$ est la plus « ajustée » [12] des approximation quadratiques majorante [Huber, 1981, Sec. 7.8]. Certain auteurs concluent de ce résultat que l'algorithme de GR est l'algorithme *le plus rapide* parmi les algorithmes de type approximation quadratique majorants ; voir par exemple [Erdogan et Fessler, 1999, p. 805].

Réglage des paramètres θ et a

L'ajustement du paramètre θ — et du paramètre a pour les algorithmes de GY — influe en général sensiblement sur le taux de convergence : dès lors, dégager des règles claires permettant un réglage *a priori* de ce(s) paramètre(s) serait d'un intérêt majeur. Le lemme suivant fournit des premiers éléments intéressants sur cette question.

Lemme 3 *Soit $\mathcal{D}_\theta : \mathbb{R}^N \to \mathbb{R}^{N\times N}$ une application de la forme $\mathcal{D}_\theta(\boldsymbol{u}) = \boldsymbol{I} - \theta \boldsymbol{P}(\boldsymbol{u})$ avec $\boldsymbol{P} : \mathbb{R}^N \to \mathbb{R}^N$ un opérateur continu dont le spectre est réel et strictement positif. Alors $\rho(\mathcal{D}'_\theta(\boldsymbol{u})) \equiv \sigma(\theta, \boldsymbol{u})$ s'écrit :*

$$\sigma(\theta) = \begin{cases} 1 - \theta\lambda_{min}[\boldsymbol{P}(\boldsymbol{u})] & si \qquad \theta \leq \widehat{\theta} \\ \theta\lambda_{max}[\boldsymbol{P}(\boldsymbol{u})] - 1 & sinon. \end{cases} \tag{9.40}$$

[12] C'est à dire qu'elle conduit à l'approximation quadratiques majorante de courbure la plus faible.

avec $\widehat{\theta} \equiv 2/(\lambda_{min}[\boldsymbol{P}(\boldsymbol{u})] + \lambda_{max}[\boldsymbol{P}(\boldsymbol{u})])$, *et on a :*

$$\min_{\theta>0} \sigma(\theta) = \frac{\lambda_{max}[\boldsymbol{P}(\boldsymbol{u})] - \lambda_{min}[\boldsymbol{P}(\boldsymbol{u})]}{\lambda_{max}[\boldsymbol{P}(\boldsymbol{u})] + \lambda_{min}[\boldsymbol{P}(\boldsymbol{u})]} \equiv \widehat{\sigma} < 1. \tag{9.41}$$

Preuve La preuve est donnée en annexe II, page 363. ■

La figure 9.4 rassemble les divers éléments géométriques associés à ce lemme. On notera également qu'on déduit facilement de (9.40) que l'ensemble $\Gamma = \{\theta > 0 \ : \ \sigma(\theta) < 1\}$ est un intervalle ouvert $]0; \bar{\theta}[$ avec $\bar{\theta} \equiv 2/\lambda_{\max}[\boldsymbol{P}(\boldsymbol{u})]$.

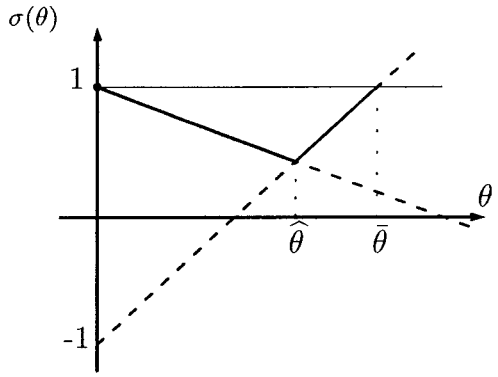


Figure 9.4: Évolution du rayon spectral $\sigma(\theta, \boldsymbol{u})$ en fonction du pas fixe θ.

Pour les algorithmes de GY et GR, l'opérateur $\boldsymbol{B}(\boldsymbol{u})$ choisi parmi (9.35) est symétrique et DP, et il n'est pas difficile de montrer que le produit $\boldsymbol{B}(\boldsymbol{u})^{-1}\boldsymbol{H}(\boldsymbol{u})$ est également DP et à spectre réelle. Ainsi, le lemme 3 permet de caractériser

l'évolution du taux de convergence en fonction de θ de ces algorithmes en posant[13] $\boldsymbol{P} = \boldsymbol{H}^{-1}\boldsymbol{B}$. Le calcul du $\widehat{\theta}$ associé au taux optimal $\widehat{\sigma}$ reste néanmoins délicat dans la mesure où son expression fait intervenir les valeurs propres extrêmes du produit $\boldsymbol{H}^{-1}\boldsymbol{B}$ qui dépendent de la solution $\boldsymbol{x}^\star$ du problème. Il semble néanmoins qu'adopter un pas fixe $\theta \in]1;2[$ permet d'accélérer la convergence des algorithmes SQ avec des fonctions ϕ de type $\ell_2\ell_1$; cf. par exemple [Ciuciu et Idier, 2002, Sec. 5] ou encore l'exemple en tomographie de la figure 9.5.

Pour les algorithmes de GY, le taux de convergence dépend également du paramètre a qui doit être choisi dans $]0; 2\widehat{a}/\theta[$. Partant de (9.41), la valeur $a^\star$ qui permet d'aboutir au meilleur taux de convergence minimise en a la fonction suivante :

$$\widehat{\sigma}(a) = \frac{\lambda_{\max}[\boldsymbol{P}(a;\boldsymbol{x}^\star)] - \lambda_{\min}[\boldsymbol{P}(a;\boldsymbol{x}^\star)]}{\lambda_{\max}[\boldsymbol{P}(a;\boldsymbol{x}^\star)] + \lambda_{\min}[\boldsymbol{P}(a;\boldsymbol{x}^\star)]} \qquad \text{où} \quad \boldsymbol{P}(a;\boldsymbol{x}^\star) = \boldsymbol{H}(\boldsymbol{x}^\star)^{-1}\boldsymbol{B}_{\text{GY}}^{a} \quad (9.42)$$

si on considère par exemple la forme monovariée de GY. Le minimiseur $\widehat{\sigma}(a)$ n'a pas de forme explicite et on note que cette fonction est non convexe. Formellement, la valeur de $a^\star$ dépend là encore du problème, pourtant choisir a à la frontière de son *ancien* domaine de convergence (c.à.d. si $a \approx \widehat{a}$) semble systématiquement conduire à une convergence accélérée de l'algorithme ; cf. [Ciuciu et Idier, 2002, Sec. 5.3]. Avec l'extension du domaine de convergence qui a découlé de notre étude, on

[13] On note que le caractère DP de $\boldsymbol{P}(\boldsymbol{u})$ permet d'écrire $\lambda_{\max}[\boldsymbol{P}(\boldsymbol{u})] = 1/\lambda_{\min}[\boldsymbol{P}(\boldsymbol{u})^{-1}]$, ce qui conduit à une expression équivalente de $\widehat{\sigma}(\boldsymbol{u})$ où $\boldsymbol{P}(\boldsymbol{u})$ est remplacé par $\boldsymbol{P}(\boldsymbol{u})^{-1}$ dans (9.41).

peut s'interroger légitimement sur l'intérêt de choisir un paramètre $a > \widehat{a}$ pris dans $]0; 2\widehat{a}/\theta[$. Expérimentalement, le constat qui s'est dégagé est qu'il est effectivement possible d'obtenir une vitesse de convergence augmentée en choisissant a sur la frontière de son *nouveau* domaine ; c.à.d. proche de l'hyperbole de la figure 9.2.

Convergence accélérée dans de domaine étendu

Contrairement au réglage précédent $a \approx \widehat{a}$, le placement de a sur l'hyperbole dépend de la valeur de $\theta^\star$ qui dépend elle du problème. En guise de support pour la discussion, nous allons considérer la figure 9.5 qui illustre les courbes de niveau du taux de convergence en racine $\sigma_{\text{GY}}(\theta, a)$ calculé à partir de la relation suivante :

$$\sigma_{\text{GY}}(\theta, a) = \rho\left(\boldsymbol{I} - \theta\big(\boldsymbol{B}^a_{\text{GY}}\big)^{-1}\boldsymbol{H}(\boldsymbol{x}^\star)\right); \tag{9.43}$$

le calcul a été effectué sur une grille de 60×50 valeurs du couple (θ, a) couvrant l'intérieur du domaine de convergence étendu. On note que la valeur de $\widehat{a}$ pour ce problème est $\widehat{a} = 0{,}1$ et que $\boldsymbol{x}^{\text{num}}$ a été utilisé à la place de $\boldsymbol{x}^\star$ dans (9.43), $\boldsymbol{x}^{\text{num}}$ étant la solution du problème à la précision numérique près (10^{-12}). À titre indicatif, le critère J considéré pour cette simulation est de type « moindre carrés pénalisés » appliqué à la tomographie axiale :

$$J(\boldsymbol{x}) = ||\boldsymbol{y} - \boldsymbol{R}\boldsymbol{x}||^2 + \alpha \sum_c \sqrt{\langle \boldsymbol{v}_c, \boldsymbol{x}\rangle^2 + s^2}$$

où nous avons adopté les notations du chapitre 6. Le problème a été choisi de petite taille (15×15 pixels) de manière à permettre le calcul rapide du rayon spectral.

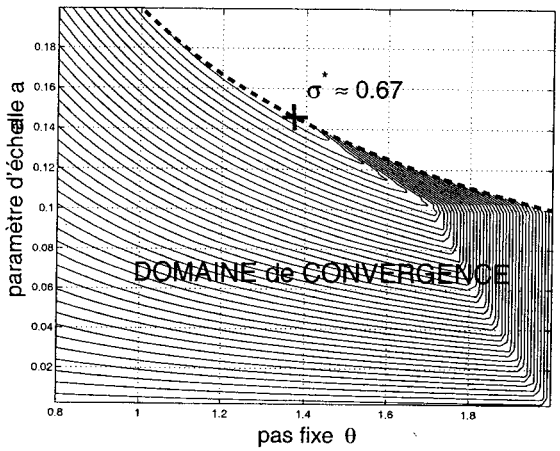


Figure 9.5: Courbes de niveaux pour $\sigma_{\mathrm{GY}}(\theta, a)$ à l'intérieur du nouveau domaine de convergence défini par $\theta \in]0;2[$ et $a \in]2\widehat{a}/\theta[$ pour $\widehat{a} = 0,1$; le meilleur taux de convergence noté $\sigma^\star$ est atteint sur la frontière.

Un certain nombre d'informations intéressantes peuvent être extraites de cette figure :

(1) les lignes de niveaux suivent globalement l'hyperbole délimitant la frontière suggérant ainsi que le taux de convergence reste identique pour un nombre important de couples (θ, a) tels que $a\theta =$ constante ;

(2) cette simulation montre clairement que, pour θ quelconque, le meilleur taux n'est pas systématiquement atteint sur la frontière du nouveau domaine : le taux remonte clairement pour $a > \widehat{a}$ si θ est pris dans $]1{,}7;2[$, et il remonte également sur une partie du domaine si $a > \widehat{a}$ pour θ est pris dans $]1{,}45;1{,}7[$;

(3) il est néanmoins remarquable sur cet exemple que le meilleur taux de conver-

gence $\sigma^\star$ est effectivement atteint sur la frontière du nouveau domaine en $a^\star = 2\widehat{a}/\theta^\star$ avec $\theta^\star \approx 1{,}4$; ce même phénomène a également été souligné pour une application en contrôle non destructif par [Fiani, 2001, Sec. IV.5.3].

Établir formellement que le taux optimal $a^\star$ est atteint sur la frontière pour $2\widehat{a}/\theta^\star$ paraît particulièrement ardu. Une première étape serait néanmoins de montrer que $\sigma(a)$ décroît sur son ancien domaine de convergence $]0;\widehat{a}[$ de manière à assurer que l'optimum est atteint à l'intérieur de la partie étendue. Malheureusement, on peut construire des contres exemples à cette conjecture pour un simple problème en dimension $N = 2$ même en tenant compte du caractère $\ell_2\ell_1$ de ϕ. Pour le moment, une étude empirique plus large est donc nécessaire pour confirmer ou infirmer le phénomène identifié au point (3).

9.6 En résumé

Dans ce chapitre, la forme primale des algorithmes SQ nous a permis d'examiner les propriétés de convergence par les outils de l'optimisation non contrainte. Une étude basée sur la règle d'ARMIJO a notamment permis une extension sensible du domaine de convergence des algorithmes de GY. On indique que nous nous sommes consacrés dans ce chapitre aux formes SQ dites à « mise à jour groupée » ; les principaux résultats obtenus (extension du domaine de convergence pour les formes de GY et étude du taux de convergence asymptotique) ne s'appliquent pas

directement aux formes SSU et « par bloc » même si leur extension paraît réalisable en utilisant les résultats de convergence des schémas *relaxation*, ex. [Bertsekas, 1995, Prop. 2.7.1] ou [Ortega et Rheinboldt, 1970, 10.3].

En termes pratiques, l'extension du domaine de convergence pour la forme de GY prend tout son sens si elle conduit effectivement à une convergence accélérée, ce que nous avons constaté en pratique. Cette constatation mérite néanmoins d'être étayée par des résultats théoriques, ou à défaut par des simulations sur différents problèmes distincts et représentatifs des applications traitées. Indéniablement, un travaille important reste à effectuer sur cet aspect.

CONCLUSION

Dans le cadre de la conception d'implants personnalisés du genou, nous nous sommes intéressés principalement dans ce document au problème de reconstruction d'images 3D de *précision* en tomographie hélicoïdale. Afin d'apporter une réponse efficace à cette problématique, il est nécessaire de développer deux composantes complémentaires relevant, d'une part des éléments *méthodologiques* adoptés pour améliorer la précision des reconstructions, et d'autre part des aspects de *mise en œuvre* très délicats en imagerie tridimensionnelle (abrég. 3D). D'autre part, les considérations algorithmiques abordées au cours de cette thèse ont conduit à investir une part importante de nos efforts de recherche dans l'étude des algorithmes semi quadratiques (abrég. SQ) et de leurs liens forts avec des schémas itératifs à pas fixe sur le critère pénalisé.

Nous concluons maintenant brièvement sur ces deux aspects et soulignons quelques perspectives qui nous paraissent intéressantes à court ou moyen terme.

Méthodologie de reconstruction

Du point de vue méthodologique, nous pensons que l'approche standard basée conjointement sur l'*interpolation* des projections et la *rétroprojection convoluée* (abrég. RPC) conduit à dégrader sensiblement la précision de l'image 3D. Cette thèse propose plutôt d'adopter *une formulation 3D algébrique* du problème de re-

construction afin de s'affranchir simultanément de l'interpolation et de la RPC. Celle-ci repose sur la construction d'un modèle d'observation $\boldsymbol{H}$ décrivant le procédé de projection en géométrie hélicoïdale. De plus, nous ajoutons à la formulation du problème un modèle d'image *a priori* soigneusement choisi de manière à stabiliser la reconstruction.

L'image 3D reconstruite est alors obtenue en minimisant un critère des *moindres carrés pénalisés* de très grande taille, éventuellement sous contraintes séparables. L'emploi de modèles d'image convexes à base de champs de Gibbs 3D favorise la localisation des interfaces franches dans le volume sans hypothéquer pour autant les aspects liés au coût d'implantation. Testée sur des données synthétiques, cette approche permet un gain significatif de précision dans le volume reconstruit. Des tests sur données réelles sont néanmoins nécessaires pour confirmer ces résultats, et la signature d'un accord de confidentialité avec GENERAL ELECTRIC a été initiée en ce sens. Nous espérons donc avoir la possibilité à moyen terme de quantifier expérimentalement l'apport de notre méthode de reconstruction.

Indiquons enfin que la méthodologie adoptée semble pouvoir s'adapter sans difficulté à la nouvelle génération de tomographes à hélice « *multicoupes* ». Des obstacles d'implantation sont néanmoins à prévoir dans la mesure où le nombre de données à traiter augmente de manière sensible.

Problématiques d'implantation

La mise en œuvre de ces approches pénalisées soulève des difficultés d'implantation liées à la taille très importante du problème de reconstruction. La *construction* et le *stockage* de $\boldsymbol{H}$ sont à ce titre particulièrement délicat ; ces obstacles peuvent néanmoins être surmontés si l'on tire parti des éléments suivant :

(a) le modèle d'observation est construit à partir d'opérateurs de projection en *géométrie axiale* de structure bien connue ;

(b) une invariance spatiale du modèle est exploitée afin de limiter l'explosion du coût mémoire associé à son stockage.

Par conséquent, la construction et le stockage d'un modèle paramétrique $\boldsymbol{H}_r$ de taille réduite est suffisant pour décrire entièrement le modèle d'observation. Indiquons néanmoins que la taille de $\boldsymbol{H}_r$ reste importante : à l'avenir, il est donc souhaitable de réduire encore le nombre d'éléments réellement stockés en tirant parti de certaines symétries du problème (cf. page 170).

Minimisation itérative

La seconde difficulté de mise en œuvre se situe à l'étape de *minimisation itérative* du critère pénalisé. En particulier, la taille du problème d'optimisation est une entrave sévère à l'implantation de la majorité des algorithmes standard. Deux approches sont néanmoins proposées dans ce mémoire pour reconstruire un volume

de grande taille en un temps raisonnable.

(c) La première approche consiste à implanter un schéma de *relaxation SOR* de manière à garder la maîtrise du volume de calcul à chaque itération (cf. page 200). Le schéma itératif SOR choisi correspond à l'algorithme semi quadratique (abrég. SQ) *single site update* (abrég. SSU) introduit initialement dans [Brette et Idier, 1996]. Des tests effectués sur données synthétiques montrent par ailleurs que la convergence de cet algorithme reste intéressante pour l'application en tomographie hélicoïdale.

(d) Dans un registre différent, une démarche simple consiste à adopter une approximation du problème pénalisé initial de manière à le rendre « séparable » (cf. page 173). Cette seconde approche permet de considérer la reconstruction de chaque plan du volume comme un sous problème de dimension réduite ; des tests sur données synthétiques ont montrés que cette démarche ne conduit qu'à une légère dégradation de précision sur l'image finale.

Ces résultats encourageant doivent, là encore, être étayés par des tests effectués avec des données réelles.

Les approches (c) et (d) ne constituent pas les uniques solutions envisageables pour reconstruire rapidement le volume d'intérêt. La recherche d'un algorithme de reconstruction efficace reste un problème largement ouvert en inversion pénalisé. Ceci est notamment vrai pour les problèmes de grande taille comme le nôtre, et

nous présentons maintenant les voies qui nous semblent les plus prometteuses.

Variations algorithmiques sur la relaxation SQ

Dans un premier temps, une démarche légitime est de tenter d'améliorer les performances du schéma itératif (c), c.à.d. à rester dans la famille des algorithmes de relaxation. Or, il semble que l'algorithme de relaxation que nous mettons en œuvre constitue actuellement un des choix les plus intéressants pour la reconstruction tomographique axiale [Erdogan et Fessler, 1999] : il est donc raisonnable de penser qu'il y a finalement « peu à gagner » à investir dans la recherche d'une version rapide dans cette famille.

Dans un deuxième temps, une démarche alternative peut consister à adopter un schéma de relaxation « par blocs de coordonnées » tel que l'algorithme SQ « par blocs » de GEMAN et YANG (abrég. GY) ou de GEMAN et REYNOLDS (abrég. GR) — cf. page 249. L'intérêt de cette démarche reste néanmoins hypothétique puisque le volume de calcul par itération augmente rapidement à mesure que le nombre d'éléments mis à jour par relaxation croît. Une relaxation *plan par plan* nécessite par exemple de résoudre une équation normale de grande taille pour mettre à jour l'image bidimensionnelle (abrég. 2D) courante. Une mise en œuvre *rapide* de ce type d'algorithme reste finalement sujette à l'une des deux conditions suivantes :

(e) le choix d'une taille de bloc de relaxation suffisamment faible pour permettre

un coût par itération raisonnable ;

(f) la possibilité de résoudre efficacement (ex. dans le domaine de FOURIER) les équations normales successives pour la mise à jour d'un grand nombre de voxels.

Si on adopte la stratégie (e), il reste encore à démontrer que ce choix permet un gain substantiel de vitesse de reconstruction. Par ailleurs, la stratégie (f) reste difficile à utiliser dans notre situation : la possibilité de résoudre rapidement l'équation normale associée à la géométrie hélicoïdale reste en particulier à démontrer ; [Delaney et Bresler, 1996] constitue néanmoins une référence intéressante bien que relative à la tomographie axiale.

D'autres alternatives algorithmiques

Dans le cadre de la tomographie hélicoïdale, certaines approches algorithmiques distinctes méritent certainement d'être approfondies. On pense par exemple à la mise en œuvre d'un algorithme du *gradient conjugué préconditionné* sur la succession de critères pénalisés réduits dans l'inversion approximative (d). Soulignons cependant que ce problème reste de taille importante (pour des plans de 512×512 voxels, le problème « réduit » reconstruit près de 800 000 variables) et que, si des préconditionneurs efficaces existent en tomographie axiale [Fessler et Booth, 1999], leurs analogues en géométrie hélicoïdale restent néanmoins à concevoir.

Une autre alternative intéressante serait de substituer au gradient conjugué

préconditionné un algorithme de type l-BFGS. Cet algorithme nécessite un volume de calcul restreint à chaque itération et ses besoins en termes de stockage restent faibles [Nocedal et Wright, 2000, Sec. 91.] ; reste néanmoins que la mise en œuvre nécessite le recours à une *recherche linéaire* qui peut sensiblement augmenter la charge de calcul. Enfin, il semble important de s'intéresser aux approches spécialement introduites pour la minimisation des très grands systèmes ; à ce titre, les approches les plus prometteuses nous semble être les *techniques de décomposition* ainsi que les méthodes tirant partie de la *séparabilité de la fonction objectif* (cf. page 217).

Autour des algorithmes SQ...

Finalement, cette thèse apporte une contribution originale sur certains *points méthodologiques* ayant trait à la minimisation des critères pénalisés. Plus précisément, les algorithmes « semi quadratiques » largement employés en restauration et reconstruction d'image sont réexaminés au travers du lien fort qui les unis à des algorithmes standard « à pas fixe »— formes newtoniennes à pas fixe, relaxation SOR par blocs à pas fixe. Dans un deuxième temps, ces algorithmes SQ sont reliés aux algorithmes de WEISZFELD *généralisés* [Weiszfeld, 1937] dont les représentants les plus connus sont l'algorithme *Expectation-Maximization* et ses nombreuses variantes.

La forme « purement primale » des itérations SQ permet de réexaminer ces algo-

rithmes par les outils standard de la programmation mathématique. Une étude des propriétés de convergence globale et de vitesse asymptotique a donc été conduite, et un certain nombre de résultats intéressants ont pu être établis.

Convergence globale des algorithmes SQ

Pour étudier les propriétés de convergence globale de ces algorithmes à pas fixe, la notion d'*admissibilité du pas fixe au sens d'*ARMIJO a été introduite. Cette démarche conduit en particulier à des conditions suffisantes de convergence significativement *moins restrictives* pour les algorithmes de GY. Ce résultat est important en pratique dans la mesure où on constate expérimentalement une *convergence accélérée* de ces algorithmes sur la frontière de leur nouveau domaine de convergence.

Indiquons que l'outil que nous avons introduit permet d'analyser la convergence d'algorithmes SQ *effectivement implantés.* La possibilité de conduire une telle analyse est à souligner dans la mesure où les preuves de convergence apparaissant notamment dans [Charbonnier *et al.*, 1997] ou [Delaney et Bresler, 1998] ne permettent pas d'assurer la convergence des mises en œuvre « tronquées » qui y sont présentées. À titre d'exemple, considérons un algorithme SQ pour lequel l'équation normale est résolue *approximativement* par p itérations d'un algorithme de relaxation SOR. Dans ce cas, il est facile de montrer que l'itération SQ suivante s'écrit

— cf. [Ortega et Rheinboldt, 1970, Sec. 7.4] :

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \theta \boldsymbol{S}^{(k)} \boldsymbol{\nabla} J(\boldsymbol{x}^{(k)}) \tag{9.44}$$

où $\boldsymbol{\nabla} J$ est le gradient du critère pénalisé J, et $\boldsymbol{S}^{(k)}$ est une matrice qui s'écrit à partir de l'opérateur $\boldsymbol{B}^{(k)}$ associé indifféremment à la forme SQ de GY ou GR. La convergence de ce schéma itératif peut alors s'analyser directement dans le cadre de l'admissibilité du pas fixe développé au chapitre 9.

Un certain nombre de travaux restent à effectuer dans le cadre d'une étude de convergence globale. Tout d'abord, on indique que les résultats de convergence établis dans le cadre du chapitre 9 n'ont pas encore été étendues aux autres formes de mise à jour SQ — c.à.d. la mise à jour SSU ou « par bloc ». D'autre part, nous envisageons d'étudier la convergence pour une forme générale d'algorithme à pas fixe et à matrice *constante*

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \theta \boldsymbol{B}^{-1} \boldsymbol{\nabla} J(\boldsymbol{x}^{(k)}); \tag{9.45}$$

cette itération constituant, en quelque sorte, une forme « généralisée » d'algorithme de GY.

Vitesse asymptotique des algorithmes SQ

La vitesse de convergence des algorithmes SQ a également motivé des recherches au cours de cette thèse. En particulier, nous avons cherché à valider certaines constatations expérimentales, la première étant :

(i) « La formulation de GR conduit-elle systématiquement au plus rapide des algorithme SQ ? »

Nous avons finalement répondu par la négative à cette question : en effet, si l'algorithme de GR bénéficie d'un taux de convergence plus intéressant en pratique — ce qu'un raisonnement de nature qualitative semble laisser entendre — son coût d'implantation plus élevé peu le rendre finalement moins intéressant en pratique qu'un algorithme de GY. En termes plus précis, il s'avère que dans tous les cas pratiques intéressant (c.à.d. pénalisation non quadratique) l'ordre de convergence des algorithmes SQ est *au mieux linéaire*, ce qui empêche d'établir une hiérarchie entre les algorithmes SQ *sur la seule base de leurs taux de convergence asymptotique.*

Une seconde question a été examinée avec un intérêt particulier dans la mesure où, à la lumière du résultat précédent, l'algorithme de GY peut être « accéléré » et devenir effectivement compétitif par rapport à un algorithme de GR :

(ii) « L'algorithme de GY converge t-il systématiquement plus vite sur le bord de son domaine de convergence ? ».

Une étude des propriétés asymptotiques de la l'algorithme de GY n'a cependant pas permis d'établir ce résultat ; un travail aussi bien expérimental que théorique reste donc certainement à mener en ce sens.

RÉFÉRENCES

ALLAIN, M., DORÉ, S., GOUSSARD, Y. et IDIER, J. (2000). Conception automatisée et fabrication rapide de prothèses personnalisées. Dans *ACFAS, Coopération France-Québec*, Montréal, Québec, Canada.

ALLAIN, M., GOUSSARD, Y. et IDIER, J. (2001). Approche régularisée en reconstruction tomographique 3D hélicoïdale. Dans *Actes 18e coll. GRETSI*, Toulouse.

ALLAIN, M., IDIER, J. et GOUSSARD, Y. (2002). On Global and Local Convergence of Half-Quadratic Algorithms. Dans *Proc. IEEE ICIP*, Rochester, USA.

ALLINEY, S. et RUZINSKY, S. A. (1994). An Algorithm for the Minimization of Mixed l_1 and l_2 Norms with Application to Bayesian Estimation. *IEEE Trans. Signal Processing*, *42*, 618–627.

ANDERSEN, A. H. et KAK, A. C. (1984). Simultaneaous algebraic reconstruction technique (SART) : A superior implementation of the ART algorith. *Ultrasonic Imaging*, *6*, 81–94.

ANDIA, B. I., SAUER, K. et BOUMAN, C. A. (2001). Nonlinear Backprojection for Tomographic Reconstruction. Soumis à IEEE Transactions on Nuclear Science numero du rapport, Université Notre-Dame, Notre Dame, Indiana, USA.

AUBERT, G. et VESE, L. (1997). A variational method in image recovery. *SIAM J. Num. Anal.*, *34*, 1948–1979.

AVRIEL, M. (1976). Nonlinear Programming : Analysis and Methods. Prentice-Hall, Englewood Cliffs, NJ.

BASS, J. (1974). Éléments de Calcul des Probabilités. Masson, Paris, 3e édition.

BERTSEKAS, D. P. (1995). Nonlinear programming. Athena Scientific, Belmont, MA, USA.

BERTSEKAS, D. P. (1999). Nonlinear programming. Athena Scientific, Belmont, MA, USA, 2e édition.

BLAKE, A. (1989). Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction. *IEEE Trans. Pattern Anal. Mach. Intell.*, *PAMI-11*, 2–12.

BLAKE, A. et ZISSERMAN, A. (1987). Visual reconstruction. The MIT Press, Cambridge, MA, USA.

BONNANS, J. F., GILBERT, J. C., LEMARÉCHAL, C. et SAGASTIZÁBAL, C. (1997). Optimization numérique. Mathématiques et Applications, Spinger, New York, NY, USA.

BOUMAN, C. A. et SAUER, K. D. (1993). A generalized Gaussian image model for edge-preserving MAP estimation. *IEEE Trans. Image Processing*, *2*, 296–310.

BOUMAN, C. A. et SAUER, K. D. (1996). A Unified Approach to Statistical Tomography Using Coordinate Descent Optimization. *IEEE Trans. Image Processing, 5*, 480–492.

BOWSHER, J., SMITH, M., PETER, J. et JASZCZAK, R. (1998). A Comparison of OSEM and ICD for Iterative Reconstruction of SPECT Brain Images. *Journal of Nuclear Medicine, 39*, 79.

BRÉMAUD, P. (1998). Markov chains. Gibbs fields and Monte Carlo. Cours ENSTA, Paris.

BRETTE, S. et IDIER, J. (1996). Optimized Single Site Update Algorithms for Image Deblurring. Dans *Proc. IEEE ICIP*, Lausanne, Suisse, 65–68.

BROWNE, J. A., HERMAN, G. T. et ODHNER, D. (1993). SNARK93 : A programming System for Image Reconstruction from Projections. Technical report no. mipg198, Medical Image Processing Group, University of Pennsylvania, Philadelphie.

BURRUS, C., BARRETO, J. et SELESNICK, I. (1994). Iterative reweighted least-squares design of FIR filters. *IEEE Trans. Signal Processing, 42*, 2926–2936.

CALVETTI, D., HANSEN, P. C. et REICHEL, L. (2001). L-curve curvature bounds via Lanczos bidiagonalization. rapport interne à paraître dans Electronic Transactions on Numerical Analysis IMM-TR-2001-5, Technical University of Denmark, Lyngby, Denmark.

CETIN, M. et KARL, W. (2001). Feature-Enhanced Synthetic Aperture Radar Image Formation Based on Nonquadratic Regularization. *IEEE Trans. Image Processing, 10*, 623–631.

CENSOR, Y., EGGERMONT, P. P. B. et GORDON, D. (1983). Strong Underrelaxation in Kaczmarz's Method for Inconsistent Systems. *Numerische Mathematik, 41*, 83–92.

CENSOR, Y. et ELFVING, T. (2001). Block-Iterative Algorithms with Diagonally Scaled Oblique Projections for the Linear Feasibility. Accepté dans SIAM Journal on Matrix Analysis and Applications.

CENSOR, Y., GORDON, D. et GORDON, R. (2001). BICAV : a block-iterative parallel algorithm for sparse systems with pixel-related weighting. *IEEE Trans. Medical Imaging, 20*, 1050–1060.

CHAN, T. F. et CHIU-KWONG, W. (1998). Total variation blind deconvolution. *IEEE Trans. Image Processing, 7*, 370–375.

CHAN, T. F. et MULET, P. (1999). On the convergence of the lagged diffusivity fixed point method in total variation image restoration. *SIAM Journal of Numerical Analysis, 36*, 354–367.

CHARBONNIER, P. (1994). Reconstruction d'image : régularisation avec prise en compte des discontinuités. Thèse de doctorat, Université de Nice-Sophia Antipolis, Nice.

CHARBONNIER, P., BLANC-FÉRAUD, L., AUBERT, G. et BARLAUD, M. (1994). Two Deterministic Half-Quadratic Regularization Algorithms for Computed Imaging. Dans *Proc. IEEE ICIP*, volume 2, Austin, TX, USA, 168–172.

CHARBONNIER, P., BLANC-FÉRAUD, L., AUBERT, G. et BARLAUD, M. (1997). Deterministic edge-preserving regularization in computed imaging. *IEEE Trans. Image Processing, 6*, 298–311.

CHENG, Q., CHEN, R. et LI, T.-H. (1996). Simultaneous Wavelet Estimation and Deconvolution of Reflection Seismic Signals. *IEEE Trans. Geosci. Remote Sensing, 34*, 377–384.

CHEUNG, W. K. et LEWITT, R. M. (1991). Modified Fourier Reconstruction Method Using Shifted Transform Sampled. *Physics in Medicine and Biology, 36*, 269–277.

CIARLET, P. G. (1988). Introduction à l'analyse numérique matricielle et à l'optimisation. Collection mathématiques appliquées pour la maîtrise, Masson, Paris.

CIUCIU, P. (2000). Méthodes markoviennes en estimation spectrale non paramétrique. Applications en imagerie radar Doppler. Thèse de doctorat, Université de Paris–Sud, Orsay.

CIUCIU, P. et IDIER, J. (2002). A Half-Quadratic block-coordinate descent method for spectral estimation. *Signal Processing, 82*, 941–959.

CIUCIU, P., IDIER, J. et GIOVANNELLI, J.-F. (2001). Regularized estimation of mixed spectra using a circular Gibbs-Markov model. *IEEE Trans. Signal Processing*, *49*, 2201–2213.

CRAWFORD, C. R. et KING, K. F. (1990). Computed tomography scanning with simultaneous patient translation. *Med. Phys.*, *17*, 967–982.

DAVID, K. (2002). Données relatives à la chirurgie et aux prothèses orthopédiques pour les procédures d'arthroplasties totales de la hanche et du genou pratiquées au Canada, de mai 2001 à mars 2002. Bulletin du RCRA, Institut canadien d'information sur la santé, Toronto (Ontario), Canada.

DEANS, S. R. (1983). The Radon transform and some of its applications. Wiley Interscience, New York.

DELANEY, A. H. et BRESLER, Y. (1995). Multiresolution Tomographic Reconstruction using Wavelets. *IEEE Trans. Image Processing*, *4*, 799–813.

DELANEY, A. H. et BRESLER, Y. (1996). A Fast and Accurate Fourier Algorthm for Iterative Parallel-Beam Tomography. *IEEE Trans. Image Processing*, *5*, 840–853.

DELANEY, A. H. et BRESLER, Y. (1998). Globally convergent edge-preserving regularized reconstruction : an application to limited-angle tomography. *IEEE Trans. Image Processing*, *7*, 204–221.

DEMOMENT, G. (1989). Image Reconstruction and Restoration : Overview of Common Estimation Structure and Problems. *IEEE Trans. Acoust. Speech, Signal Processing*, ASSP-*37*, 2024–2036.

Demoment, G. et Idier, J. (2001*a*). Problèmes inverses et problèmes mal-posés, chapitre 1. Dans Idier [2001*a*], 25–40.

Demoment, G. et Idier, J. (2001*b*). Régularisation d'un problème mal-posé : approches génériques, chapitre 2. Dans Idier [2001*a*], 41–58.

Descombes, X. et Goussard, Y. (2001). Problèmes non supervisés, chapitre 8. Dans Idier [2001*a*], 195–216.

DESCOMBES, X., MORRIS, R., ZERUBIA, J. et BERTHOD, M. (1999). Estimation of Markov Random Field prior parameters using Markov chain Monte Carlo Maximum Likelihood. *IEEE Trans. Image Processing*, *8*, 954–963.

DINTEN, J.-M. (1990). Tomographic reconstruction of axially symmetric objects : Regularization by a Markovian modelisation. Dans *Proc. of the Int. Conf. on Pattern Recog.*

DOBSON, D. C. et SANTOSA, F. (1996). Recovery of Blocky Images from Noisy and Blurred Data. *SIAM J. Appl. Mathematics*, *56*, 1181–1198.

DORé, S. et GOUSSARD, Y. (1997). Experimental Determination of CT Point Spread Function Anisotropy and Shift-Variance. Dans *19th Conference of the IEEE-EMBS*, Chicago, IL , USA, 788–791.

DUNMUR, A. P. et TITTERINGTON, D. M. (1997). Computational Bayesian Analysis of Hidden Markov Mesh Models. *IEEE Trans. Pattern Anal. Mach. Intell., PAMI-19*, 1296–1300.

EDHOLM, P. et HERMAN, G. T. (1988). Image Reconstruction from Linograms : Implementation and Evaluation. *IEEE Trans. Medical Imaging, 7*, 239–246.

EGGERMONT, P. et HERMAN, G. (1981). Iterative algorithms for large partitioned linear systems, with applications to image reconstruction. *Linear Algebra and Its Applications, 40*, 37–67.

ERDOGAN, H. et FESSLER, J. (1999). Monotonic Algorithms for Transmission Tomography. *IEEE Trans. Medical Imaging, 18*, 801–814.

FARIDANI, A. (1999). Mathematical problems in computed tomography. Dans *Mathematical Geophysics Summerschool : Geophysical Tomography and its Applications*, Édité par J. Berryman, G. Papanicolaou et W. Symes.

FESSLER, J., FICARO, E., CLINTHORNE, N. et LANGE, K. (1997). Grouped-Coordinate Ascent Algorithms for Penalized-Likelihood Transmission Image Reconstruction. *IEEE Trans. Medical Imaging, 16*, 166–175.

FESSLER, J. et HERO, A. (1995). Space-Alternating Generalized Expectation-Maximization Algorithm. *IEEE Trans. Signal Processing, 42*, 2664–2677.

FESSLER, J. A. (1994). Penalized weighted least-squares image reconstruction for positron emission tomography. *IEEE Trans. Medical Imaging, 13*, 290–300.

FESSLER, J. A. et BOOTH, S. D. (1999). Conjugate-Gradient Preconditionning Methods for Shift-Variant PET Image Reconstruction. *IEEE Trans. Image Processing*, *8*, 668–699.

FIANI, M. (2001). Reconstruction 3D à partir d'un nombre limité de radiographies. Application au contrôle non destructif des soudures. Thèse de doctorat, Université de Paris-Sud, Orsay.

FORTIER, N., DEMOMENT, G. et GOUSSARD, Y. (1993). GCV and ML methods of determining parameters in image restoration by regularization : Fast computation in the spatial domain and experimental comparison. *J. Visual Comm. Image Repres.*, *4*, 157–170.

FOURGEAUD, C. et FUCHS, A. (1972). Statistique. Dunod, Paris, 2e édition.

GAUTIER, S. (1996). Fusion de données gammagraphiques et ultrasonores. Application au contrôle non destructif. Thèse de doctorat, Université de Paris-Sud, Orsay.

Gautier, S., Champagnat, F. et Idier, J. (2001). Application de la déconvolution au contrôle non destructif par ultrasons, chapitre 9. Dans Idier [2001*a*], 219–240.

GEMAN, D. et REYNOLDS, G. (1992). Constrained Restoration and the Recovery of Discontinuities. *IEEE Trans. Pattern Anal. Mach. Intell.*, *14*, 367–383.

GEMAN, D. et YANG, C. (1995). Nonlinear Image Recovery with Half-Quadratic Regularization. *IEEE Trans. Image Processing*, *4*, 932–946.

GEMAN, S. et GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, *PAMI-6*, 721–741.

GEMAN, S. et MCCLURE, D. (1987*a*). Statistical Methods for Tomographic Image Reconstruction. Dans *Proceedings of the 46th Session of the* ICI, *Bulletin of the* ICI, volume 52, 5–21.

GEMAN, S. et MCCLURE, D. (1987*b*). Statistical Methods for Tomographic Image Reconstruction. *Bulletin of the Int. Stat. Inst.*, *52*, 5–21.

GILBERT, J. C. (1999). Optimisation Différentiable : Théorie et Algorithmes. Notes de cours, INRIA, Rocquencourt.

GILBERT, P. (1972). Iterative methods for the three-dimensional reconstruction of an object from projections. *J. Theor. Biol.*, *36*, 105–117.

GOLUB, G. H., HEATH, M. et WAHBA, G. (1979). Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics*, *21*, 215–223.

GOLUB, G. H. et VAN LOAN, C. F. (1996). Matrix computations. The Johns Hopkins University Press, Baltimore, Third édition.

GORDON, R., BENDER, R. et HERMAN, G. T. (1970). Algebraic Reconstruction Techniques (ART) for three-dimensional electron microscopy and X-ray photography. *Journal of Theoretical Biology*, *29*, 471–481.

GREEN, P. J. (1990). Bayesian Reconstructions From Emission Tomography Data Using a Modified EM Algorithm. *IEEE Trans. Medical Imaging, 9*, 84–93.

GUAN, H. et GORDON, R. (1994). A Projection Access Order for Speedy Convergence of ART (Algebraic Reconstruction Technique) : a Multilevel Scheme for Computed Tomography. *Physics in Medicine and Biology*, 2005–2022.

HANSEN, P. (1990). Truncated SVD Solutions to Discrete Ill-Posed Problems with Ill-Determined Numerical Rank. *SIAM J. Stat. Sci. Comp., 11*, 503–518.

HANSEN, P. (1992*a*). Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Rev., 34*, 561–580.

HANSEN, P. (1992*b*). Regularization Tools : a Matlab Package for Analysis and Solution of Discret Ill-Posed Problems. Rapport technique, Technical University of Denmark, Dept. Informatics and Mathematical Modelling, Lyngby, Danemark.

HANSEN, P., JACOBSEN, M., RASMUSSEN, J. et SøRENSEN, H. (2000). The PP-TSVD algorithm for image restoration problems. Dans *Methods and Applications of Inversion, Lecture Notes in Earth Sciences*, Édité par P. Hansen, B. Jacobsen et K. Mosegaard, volume 92, Springer, Berlin.

HANSON, K. M. et WECHSUNG, G. W. (1983). Bayesian Approach to Limited-Angle Reconstruction in Computed Tomography. *J. Opt. Soc. Amer., 73*, 1501–1509.

HANSON, K. M. et WECHSUNG, G. W. (1985). Local Basis-Function Approach to Computed Tomography. *Applied Optics*, *24*, 4028–4039.

HAUPT, R. L. (1995). An Introduction to Genetic Algorithm for Electromagnetics. *IEEE Ant. Propag. Mag.*, *37*, 7–15.

HEBERT, T. et LEAHY, R. (1989). A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors. *IEEE Trans. Medical Imaging*, *8*, 194–202.

HERMAN, G. T. (1980). Image reconstruction from projections. The fundamentals of computerized tomography. Academic Press, New York, NY, USA.

HERMAN, G. T., HURWITZ, H., LENT, A. et LUNG, H. P. (1979). On the Bayesian Approach to Image Reconstruction. *Inform. Contr.*, *42*, 60–71.

HERMAN, G. T. et LENT, A. (1976). Quadratic optimization for image reconstruction I. *Computer Graphics and Image Processing*, *5*, 319–332.

HERMAN, G. T. et MEYER, L. B. (1993). Algebraic Reconstruction Techniques Can Be Made Computationally Efficient. *ieeeMI*, *12*, 600–609.

HU, H. (1999). Multi-Slice Helical CT : Scan and Reconstruction. *Medical Physics*, *26*, 5–18.

HU, H. et SHEN, Y. (1998). Helical CT Reconstruction with Longitudinal Filtration. *Medical Physics*, *25*, 2130–2138.

HUBER, P. J. (1981). Robust Statistics. John Wiley, New York, NY, USA.

HUESMAN, R. H., GULLBERG, G., GREENBERG, W. L. et BUDINGER, T. F. (1977). RECLBL Library Users Manuals. Rapport technique, Lawrence Berkley Laboratory, University of California.

HUNT, B. R. (1973). The Application of Constrained Least Squares Estimation to Image Restoration by Digital Computer. *IEEE Trans. Communications, C-22*, 805–812.

IDIER, J. (2000). Problèmes inverses en restauration de signaux et d'images. Habilitation à diriger des recherches, Université de Paris-Sud, Orsay.

IDIER, J., éditeur (2001*a*). Approche bayésienne pour les problèmes inverses. Traité IC2, Série traitement du signal et de l'image, Hermès, Paris.

IDIER, J. (2001*b*). Convex half-quadratic criteria and interacting auxiliary variables for image restoration. *IEEE Trans. Image Processing, 10*, 1001–1009.

Idier, J. (2001). Modèles de Gibbs-Markov pour les images, chapitre 7. Dans Idier [2001*a*], 169–194.

Idier, J. et Blanc-Féraud, L. (2001). Déconvolution en imagerie. Dans *Approche bayésienne pour les problèmes inverses*, Édité par J. Idier, Traité IC2, Série traitement du signal et de l'image, Hermès, Paris, 139–165.

IDIER, J., CIUCIU, P., FIANI, M. et ALLAIN, M. (2002). Generalized Forms of Geman and Yang Half-quadratic Construction. Rapport technique, LSS-GPI.

IDIER, J., GOUSSARD, Y. et RIDOLFI, A. (2001). Unsupervised image segmentation using a telegraph parameterization of Pickard random fields. Dans *Spatial statistics. Methodological aspects and some applications*, Édité par M. Moore, volume 159 de *Lecture notes in Statistics*, Springer Verlag, New York, NY, USA, 115–140.

JAFFE, J. S. (1990). Limited Angle Reconstruction Using Stabilized Algorithms. *IEEE Trans. Medical Imaging, 9*, 338–344.

JEFFS, B. D. et PUN, W. H. (1996). Simple Shape Parameter Estimation from Blurred Observations for a Generalized Gaussian MRF Image Prior used in MAP Image Restoration. Dans *Proc. IEEE ICIP*, Lausanne, Suisse, 465–468.

JOHNSON, C. A. et SOFER, A. (2000). A Primal-Dual Method for Large-Scale Image Reconstruction in Emission Tomography. *SIAM J. Optimization, 11*, 691–715.

KAK, A. C. et SLANEY, M. (1987). Principles of Computerized Tomographic Imaging. IEEE Press, New York, NY, USA.

KALENDER, W., SEISSLER, W., KLOTZ, E. et VOCK, P. (1990). Spiral Volumetric CT with Single-Breath Technique Continuous Transport, and Continuous Scanner Rotation. *Radiology, 176*, 181–183.

KAUFMAN, L. (1987). Implementing and Accelerating the EM Algorithm for Positron Emission Tomography. *IEEE Trans. Medical Imaging, 6*, 37–51.

LAKSHMINARAYANAN, A. V. et LENT, A. (1979). Methods of Least Squares ans SIRT in Reconstruction. *J. Theor. Biol.*, *76*, 267–295.

LANGE, K. (1995). A Gradient Algorithm Locally Equivalent to the EM Algorithm. *J. R. Statist. Soc. B*, *57*, 425–437.

LANGE, K. et FESSLER, J. (1995). Globally Convergent Algorithms for Maximum a Posteriori Transmission Tomography. *IEEE Trans. Image Processing*, *4*, 1430–1438.

LEWITT, R. M. (1992). Alternative to Voxels for Image Representation in iterative reconstruction algorithms. *Physics in Medicine and Biology*, *37*, 705–716.

LI, S. Z. (1995). On Discontinuity-Adaptive Smoothness Priors in Computer Vision. *IEEE Trans. Pattern Anal. Mach. Intell.*, *PAMI-17*, 576–586.

LI, S. Z. (1998). Close-form solution and parameter selection for convex minimization-based edge-preserving smoothing. *IEEE Trans. Pattern Anal. Mach. Intell.*, *PAMI-20*, 916–932.

LI, S. Z., HUANG, Y. H. et FU, J. S. (1995). Convex MRF potential functions. Dans *Proc. IEEE ICIP*, volume 2, Washington DC, USA, 296–299.

LI, Y. et SANTOSA, F. (1996). A computational algorithm for minimizing total variation in image restoration. *IEEE Trans. Image Processing*, *5*, 987–995.

LUENBERGER, D. G. (1973). Introduction to Linear and Nonlinear Programming. Addison-Wesley, New York, NY, USA, 1er édition.

MACOVSKI (1983). Medical Imaging Systems. Prentice-Hall, Englewood Cliffs, NJ, USA.

MUMCUOGLU, E., LEAHY, R., CHERRY, S. et ZHOU, Z. (1994). Fast Gradient-Based Methods for Bayesian Reconstruction of Transmission and Emission PET Images. *IEEE Trans. Medical Imaging, 13*, 687–701.

NASHED, M. Z. (1981). Operator-theoretic and computational approaches to ill-posed problems with applications to antenna theory. *IEEE Trans. Ant. Propag., 29*, 220–231.

NATTERER, F. (1986). The mathematics of computerized tomography. John Wiley.

NATTERER, F. (1994). Resolution and Reconstruction for a Helical CT-Scanner. Rapport technique 20-96N, FachBereich Mathematik der Universitt Münster, Münster.

NATTERER, F. (1997). Algorithms in Tomography. Dans *The State of the Art in Numerical Analysis*, Clarendon Press, duff, i.s. and watson, g.a. édition.

NATTERER, F. (1999). Numerical Methods in Tomography. Dans *Acta Numerica*, volume 8, Cambridge University Press.

NATTERRER, F. (1993). Sampling in Fan Beam Tomography. *SIAM J. Appl. Mathematics, 53*, 358–380.

NIKOLOVA, M., IDIER, J. et MOHAMMAD-DJAFARI, A. (1998). Inversion of large-support ill-posed linear operators using a piecewise Gaussian MRF. *IEEE Trans. Image Processing*, *7*, 571–585.

NIKOLOVA, M. et NG, M. (2001). Fast image reconstruction algorithms combining half-quadratic regularization and preconditioning. Dans *Proc. IEEE ICIP*, Thessaloniki, Grèce, 277–280.

NOCEDAL, J. et WRIGHT, S. J. (2000). Numerical Optimization. Series in Operations Research, Springer Verlag, New York.

OGATA, I. (1999). Pitfalls in Image Reconstruction of Helical CT Angiography : an Experimental Stydy. *Computorized Medical Imaging and Graphics*, *23*, 143–154.

ORTEGA, J. et RHEINBOLDT, W. (1970). Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York, NY, USA.

O'SULLIVAN, J. A. (1995). Roughness Penalties on Finite Domains. *IEEE Trans. Image Processing*, *4*, 1258–1268.

PARODI, R. C., SARDANELLI, F., CASTALDI, A., CITTADINI, G. et RESCINITO, G. (1997). Helical versus Conventional CT in Detecting Meniscal Injuries. *La Radiologica Medica*, *94*, 591–594.

PENG, H. et STARK, H. (1989). One-Step Image Reconstruction from Incomplete Data in Computer Tomography. *IEEE Trans. Medical Imaging*, *8*, 16–30.

PICINBONO, B. (1993). Signaux aléatoires - Probabilités et variables aléatoires avec problèmes résolus, volume 1. Dunod Université, Paris.

RHEINBOLDT, W. C. (1998). Methods for solving systems of nonlinear equations, volume 70 de *CBMS-NSF Regional conference series in applied mathematics*. SIAM, Philadelphia, 2e édition.

ROCKAFELLAR, R. T. (1970). Convex Analysis. Princeton Univ. Press.

RUDIN, L., OSHER, S. et FATEMI, C. (1992). Nonlinear total variation based noise removal algorithm. *Physica D*, *60*, 259–268.

SAQUIB, S. S., BOUMAN, C. A. et SAUER, K. D. (1998). ML Parameter Estimation for Markov Random Fields with Applications to Bayesian Tomography. *IEEE Trans. Image Processing*, *7*, 1029–1044.

Sauer, K. et Thibault, J.-B. (2001). Imagerie à partir de données de faible intensité, chapitre 14. Dans Idier [2001*a*], 343–362.

SAUER, K. D. et BOUMAN, C. A. (1993). A Local Update Strategy for Iterative Reconstruction from Projections. *IEEE Trans. Signal Processing*, *41*, 534–548.

SHEPP, L. A. et LOGAN, B. F. (1974). The Fourier reconstruction of a head section. *IEEE Trans. Nuclear Sciences*, *21*, 21–43.

SMITH, T. K., C., S. D. et L., W. S. (1977). Practical and Mathematical Aspects of the Problem of Reconstructing Objects from Radiographs. *Bulletin of the American Mathematical Society*, *82*, 1227–1270.

SOTTHIVIRAT, S. et FESSLER, J. (2002). Image recovery using partitioned-separable paraboloidal surrogate coordinate ascent algorithms. *IEEE Trans. Image Processing, 11*, 306–317.

SOUSSEN, C. (2000). Reconstruction 3D d'un objet compact en tomographie. Thèse de doctorat, Université de Paris–Sud, Orsay.

TEBOUL, S., BLANC-FÉRAUD, L., AUBERT, G. et BARLAUD, M. (1998). Variational Approach for Edge-Preserving Regularization using Coupled PDE's. *IEEE Trans. Image Processing, special issue on partial differential equations and geometry driven diffusion in image processing and analysis, 7*, 387–397.

TIKHONOV, A. (1963). Regularization of incorrectly posed problems. *Soviet. Math. Dokl., 4*, 1624–1627.

TIKHONOV, A. et ARSÉNINE, V. (1976). Méthodes de résolution de problèmes mal posés. Éditions MIR, Moscou, Russie.

VILLAIN, N. (1997). Restauration d'images tomographiques par champ de Markov 3D. M.S. thesis, École Polytechnique de Montréal, Canada.

VILLAIN, N., GOUSSARD, Y., IDIER, J. et ALLAIN, M. (2001). 3D Edge-Preserving Image Enhancement for Computed Tomography. en révision dans *IEEE Trans. Medical Imaging.*

VOGEL, C. (1997). Nonsmooth Regularization. Dans *Inverse Problems in Geophysical Applications*, Édité par H. Engl, A. Louis et W. Rundell, SIAM, 1–11.

VOGEL, C. R. (2002). Computational Methods for Inverse Problems, volume 23 de *Frontiers in Applied Mathematics*. SIAM.

VOGEL, R. V. et OMAN, M. E. (1998). Fast, robust total variation-based reconstruction of noisy, blurred images. *IEEE Trans. Image Processing*, *7*, 813–823.

VOSS, H. et ECKHARDT (1980). Linear Convergence of Generalized Weiszfeld's Method. *Computing*, *25*, 243–251.

WANG, G. et VANNIER, M. W. (1993). Helical CT Image Noise — Analytical Results. *Med. Phys.*, *6*, 1635–1640.

WANG, G. et VANNIER, W. (1999). The Effetct of Pitch in Multislice Spiral/Helical CT. *Medical Physics*, *26*, 2648–2653.

WEISZFELD, E. (1937). Sur le point pour lequel la somme des distances de n points donnés est minimum. *Tôhoku Mathematical Journal*, *43*, 355–386.

WINKLER, G. (1995). Image Analysis, Random Fields and Dynamic Monte Carlo Methods. Springer Verlag, Berlin, Allemagne.

YARLAGADDA, R., BEDNAR, J. B. et WATT, T. L. (1985). Fast algorithms for l_p deconvolution. *IEEE Trans. Acoust. Speech, Signal Processing*, *ASSP-33*, 174–182.

YEN, S. Y., YAN, C. H., RUDIN, G. D. et NAPEL, S. (1999). Longitudinal Sampling and Aliasing in Spiral CT. *IEEE Trans. Medical Imaging*, *18*, 43–58.

ZHENG, J., SAQUIB, S. S., SAUER, K. et BOUMAN, C. A. (2000). Parallelizable Bayesian Tomography Algorithms with Rapid, Guaranteed Convergence. *IEEE Trans. Image Processing*, *9*, 1745–1759.

ZHOU, Z., LEAHY, R. et JINYI, Q. (1997). Approximate maximum likelihood hyperparameter estimation for Gibbs priors. *IEEE Trans. Image Processing*, *6*, 844–861.

ANNEXE I

Annexe au chapitre 8

Proposition 13 *Soit un algorithme défini par (8.41) et tel que* $\boldsymbol{B}_{\bullet}^{(k)}$ *est égale à* $2\boldsymbol{B}_{\mathrm{GR}}^{(k)}$, $\boldsymbol{B}_{\mathrm{GY}}^{a}$ *ou* $\boldsymbol{B}_{\nu}^{a}$. *Alors on a*

$$\forall \boldsymbol{x},\ k \in \mathbb{N} \qquad \tilde{J}(\boldsymbol{x}^{(k)}, \boldsymbol{x}) \geq J(\boldsymbol{x})$$

où $\tilde{J}$ *est défini par (8.43).*

Preuve

Dans un premier temps, notons que puisque $J = \boldsymbol{Q} + \Phi$ est partiellement quadratique, d'après l'expression (8.43) de $\tilde{J}$, on a

$$\forall \boldsymbol{u}, \boldsymbol{u}^{+} \in \mathbb{R}^{N},\quad \tilde{J}(\boldsymbol{u}, \boldsymbol{u}^{+}) - J(\boldsymbol{u}^{+}) = \tag{I.1}$$

$$\Phi(\boldsymbol{u}) - \Phi(\boldsymbol{u}^{+}) + \left\langle \boldsymbol{\nabla}\Phi(\boldsymbol{u}), (\boldsymbol{u}^{+} - \boldsymbol{u}) \right\rangle + \frac{1}{2}\left\langle \boldsymbol{P}(\boldsymbol{u})(\boldsymbol{u}^{+} - \boldsymbol{u}), (\boldsymbol{u}^{+} - \boldsymbol{u}) \right\rangle,$$

où l'expression de $P(\boldsymbol{u})$ diffère suivant la variante SQ adoptée :

$$P(\boldsymbol{u}) \equiv \boldsymbol{V}\,\mathrm{diag}(\phi'(\delta_i)/\delta_i)\boldsymbol{V}^T \qquad \text{(GR)}$$

$$P(\boldsymbol{u}) \equiv \boldsymbol{V}\boldsymbol{V}^T/a \qquad \text{(GY monovariée)}$$

$$P(\boldsymbol{u}) \equiv \boldsymbol{I}/a \qquad \text{(GY multivariée)}$$

avec δ_i défini par $\delta_i = \langle \boldsymbol{v}_i, \boldsymbol{u}\rangle - w_i$. On montre maintenant que pour les opérateurs $P(\boldsymbol{u})$ correspondant aux constructions de GR et GY, l'expression (I.1) est positive ou nulle sous les conditions respectivement imposées pour la construction du critère primale/duale correspondant.

① construction de GR [Chan et Mulet, 1999] :

Pour la construction de GR, la différence (I.1) s'écrit :

$$\begin{aligned}&\sum_i \phi(\delta_i) - \phi(\delta_i^+) + \Delta_i\phi'(\delta_i) + \Delta_i^2\frac{1}{2}\frac{\phi'(\delta_i)}{\delta_i}\\ &= \sum_i \phi(\delta_i) - \phi(\delta_i^+) + \frac{1}{2}\frac{\phi'(\delta_i)}{\delta_i}\left((\delta_i^+)^2 - \delta_i^2\right).\end{aligned} \qquad \text{(I.2)}$$

où on a posé $\delta_i^+ = \langle \boldsymbol{v}_i, \boldsymbol{u}^+\rangle - w_i$ et $\Delta_i = \delta_i^+ - \delta_i$. La positivité de la somme est évidemment démontrée si on peut montrer la positivité pour tout i de

$$\phi(\delta_i) - \phi(\delta_i^+) + \frac{1}{2}\frac{\phi'(\delta_i)}{\delta_i}\left((\delta_i^+)^2 - \delta_i^2\right).$$

On pose alors pour $u \in \mathbb{R}$, $\psi(u) = \phi(\sqrt{|u|})$ en soulignant que cette fonction est *concave* sur $\mathbb{R}^+$ par hypothèse de la construction de GR — cf. page 235 ; de plus, on a $\psi(u) = \phi(-\sqrt{|u|})$ car ϕ est également paire par hypothèse. Chaque terme de la somme ci-dessus se réécrit donc sous la forme,

$$\psi(\delta_i^2) - \psi((\delta_i^+)^2) + \psi'(\delta_i^2)\left((\delta_i^+)^2 - \delta_i^2\right),$$

cette dernière expression est positive ou nulle par la concavité de ψ ; la différence (I.1) est donc positive ou nulle ce qui montre le caractère majorant de l'approximation quadratique découlant de la construction de GR.

② construction de GY :

Pour la construction de GY, on pose $P(\boldsymbol{u}) = \boldsymbol{V}\boldsymbol{V}^T/a$ pour tout $\boldsymbol{u}$, et la différence (I.1) s'écrit :

$$\sum_i \phi(\delta_i) - \phi(\delta_i^+) + \Delta_i \phi'(\delta_i) + \frac{\Delta_i^2}{2a}; \tag{I.3}$$

supposons que ϕ' est de dérivée L-Lipschitzienne, alors le *lemme de descente* [Bertsekas, 1999, prop. A.24] permet d'écrire :

$$\phi(\delta_i) - \phi(\delta_i^+) + \Delta_i \phi'(\delta_i) + \frac{L\Delta_i^2}{2} \geq 0$$

ce qui permet d'assurer la positivité de la somme (I.3) si a est choisi tel que

$$L \leq 1/a. \tag{I.4}$$

Le point clé est alors de remarquer qu'une fonction ϕ vérifie l'hypothèse de construction primale/duale de GY (cf. page 236)

$$\exists \hat{a} \ : \ \forall a \in]0;\hat{a}] \qquad u^2/2 - a\phi(u) \stackrel{\triangle}{=} g_a(u) \quad \text{convexe sur } \mathbb{R},$$

si et seulement si sa dérivée ϕ' est Lipschitzienne de constante $L = 1/\hat{a}$. Ce résultat que nous établissons page 297 (lemme 2) permet d'affirmer que la somme (I.3) est positive si (I.4) est vraie avec $L = 1/\hat{a}$, et donc si a est choisi dans l'intervalle $]0, \hat{a}]$ pour lequel l'hypothèse de construction du critère de GY est valide. En d'autres termes, tant que a est choisi de manière à vérifier l'hypothèse de construction ci-dessus, alors l'algorithme primal découlant de la construction de GY définit une approximation supérieure.

Note : si ϕ est C^2, alors la constante de Lipstchitz de ϕ' est la valeur *maximale* de $\phi''(u)$.

③ construction de GY « multivariée » :

Dans le cas de la construction de GY multivariée, on pose $P(\boldsymbol{u}) = \boldsymbol{I}/a$ pour tout $\boldsymbol{u}$, alors la différence (I.1) s'écrit simplement :

$$\forall \boldsymbol{u}, \boldsymbol{u}^+ \in \mathbb{R}^N, \quad \tilde{J}(\boldsymbol{u}, \boldsymbol{u}^+) - J(\boldsymbol{u}^+) = \tag{I.5}$$
$$\Phi(\boldsymbol{u}) - \Phi(\boldsymbol{u}^+) + \left\langle \boldsymbol{\nabla}\Phi(\boldsymbol{u}), (\boldsymbol{u}^+ - \boldsymbol{u}) \right\rangle + \frac{1}{2a}||\boldsymbol{u}^+ - \boldsymbol{u}||^2.$$

En partant de l'hypothèse de construction (cf. page 239)

$$\exists \hat{a} \ : \ \forall a \in]0; \hat{a}] \qquad ||\boldsymbol{u}||^2/2 - a\Phi(\boldsymbol{u}) \stackrel{\triangle}{=} h_a(\boldsymbol{u}) \quad \text{convexe sur } \mathbb{R}^N,$$

et en écrivant l'inégalité de convexité associée

$$h\left(\boldsymbol{u}^+\right) - h(\boldsymbol{u}) \geq \left\langle \boldsymbol{\nabla} h(\boldsymbol{u}), \left(\boldsymbol{u}^+ - \boldsymbol{u}\right) \right\rangle ;$$

avec $\boldsymbol{\nabla} h$ le gradient de h ; on montre alors sans difficulté que la différence (I.5) est positive ou nulle, ce qui montre que la construction de GY multivariée définit effectivement un algorithme primal de type approximation supérieure. ■

ANNEXE II

Annexe au chapitre 9

II.1 Preuve du théorème 1

D'après la règle d'ARMIJO (qui constitue également l'une des deux conditions de Wolf; cf. [Gilbert, 1999, p.118]), on déduit que la suite $\mathcal{J}$ est monotone non croissante. Puisque J est bornée inférieurement, $\mathcal{J}$ converge vers une valeur finie, et en particulier on a $\lim_{n\to\infty} J(\boldsymbol{x}^{(n)}) - J(\boldsymbol{x}^{(n+1)}) = 0$. Étant donné le caractère coercif de J, la suite des $\boldsymbol{x}^{(n)}$ reste dans le compact $D_0 = \{\boldsymbol{x} \in \mathbb{R} \mid J(\boldsymbol{x}) \leq J\left(\boldsymbol{x}^{(0)}\right)\}$, et le théorème de Cauchy assure qu'il existe au moins un point d'accumulation de $\left\{\boldsymbol{x}^{(n)}\right\}$ dans D_0.

Si on considère d'une part que le caractère stationnaire des points d'accumulations est assuré par la proposition 2, et d'autre part que la stricte convexité de J garantit l'unicité des points stationnaires et que ce point stationnaire est un infimum, on en déduit la convergence de l'algorithme vers cet infimum de J qui est un minimum puisque D_0 est compact. ■

II.2 Preuve des lemmes

Preuve du lemme 1 :

Rappelons que $g = (\cdot)^2/2 - a\phi(\cdot)$ est convexe pour $0 \leq a \leq \widehat{a}$, avec $\widehat{a} > 0$ et que la stricte convexité de g tient si $a < \widehat{a}$, cf. [Idier, 2001*b*, Sec. III.B]. La condition de convexité de $\widehat{g} = (\cdot)^2/2 - \widehat{a}\phi(\cdot)$ équivaut à imposer $\widehat{g}'$ non décroissante ; sous l'hypothèse de convexité de ϕ, on déduit pour $v \geq u$:

$$v - \widehat{a}\phi'(v) \geq u - \widehat{a}\phi'(u) \;\Leftrightarrow\; \widehat{a}\phi'(v) - \widehat{a}\phi'(u) \leq v - u \;\Leftrightarrow\; |\phi'(v) - \phi'(u)| \leq L|v - u|,$$

où la dernière inégalité est bien celle de Lipschitz pour une constante $L = 1/\widehat{a}$. Notons que $\widehat{a}$ est différent de 0 puisque l'intervalle qui rend g convexe doit être non nul. Le lecteur pourra constater aisément que ce résultat est identique pour $v < u$: le sens des deux premières inégalités est inversé mais la convexité de ϕ permet d'aboutir dans les deux cas à l'inégalité de Lipschitz. D'autre part, on notera que le passage se fait par équivalence : imposer le caractère L-Lipschitz sur ϕ' équivaut donc à imposer $\widehat{g}'$ monotone non décroissante, ce qui est équivalent à $\widehat{g}$ convexe. Enfin, la convexité stricte de g tenant pour $0 < a < \widehat{a}$, on déduit : $\forall u, v,\ a \in]0; \widehat{a}[$,

$$|\phi'(v) - \phi'(u)| \leq \frac{1}{\widehat{a}}|v - u| < \frac{1}{a}|v - u|,$$

ce qui démontre que, pour tout a qui rend g strictement convexe, l'inégalité de Lipschitz est strictement vérifiée. ■

Preuve du lemme 2 :

La fonction ϕ (respectivement g) étant convexe, elle admet des demi dérivées directionnelles ϕ'_+ et ϕ'_- (g'_+ et g'_+) avec l'inégalité $\phi'_+ \geq \phi'_-$ ($g'_+ \geq g'_-$). On a donc :

$$\forall u, \qquad g'_+(u) = u - a\phi'_+(u) \;\geq\; g'_- = u - a\phi'_-(u), \tag{II.1}$$

dont on déduit $\phi'_- \geq \phi'_+$, donc $\phi'_- = \phi'_+$, c'est à dire que ϕ' est une fonction continue. D'autre part, puisqu'on a égalité des dérivées à gauche et à droite de ϕ, on en déduit que (II.1) est une égalité, donc que g est également C^1. ■

Preuve du lemme 3 :

Soit $\boldsymbol{v} \in \mathbb{S}$, avec $\mathbb{S}$ l'ensemble des vecteurs de $\mathbb{R}^N$ de norme unitaire, i.e. $\mathbb{S} = \{\boldsymbol{u} \in \mathbb{R}^N : ||\boldsymbol{u}|| = 1\}$, on pose

$$C(\theta, \boldsymbol{u}, \boldsymbol{v}) \equiv \boldsymbol{v}^t(\boldsymbol{I} - \theta \boldsymbol{P}(\boldsymbol{u}))\boldsymbol{v} = 1 - \theta \boldsymbol{v}^t \boldsymbol{P}(\boldsymbol{u})\boldsymbol{v}.$$

Le rayons spectral $\rho(\mathcal{D}_\theta(\boldsymbol{u}))$ sera noté $\sigma(\theta)$ et s'écrit

$$\begin{aligned}\sigma(\theta) &= \begin{cases} \max\limits_{\boldsymbol{v}\in\mathbb{S}} C(\theta,\boldsymbol{u},\boldsymbol{v}) & \text{si} \quad \max\limits_{\boldsymbol{v}\in\mathbb{S}} C(\theta,\boldsymbol{u},\boldsymbol{v}) \geq -\min\limits_{\boldsymbol{v}\in\mathbb{S}} C(\theta,\boldsymbol{u},\boldsymbol{v}), \\ -\min\limits_{\boldsymbol{v}\in\mathbb{S}} C(\theta,\boldsymbol{u},\boldsymbol{v}) & \text{sinon} \end{cases} \\ &= \begin{cases} 1-\theta\lambda_{\min}[\boldsymbol{P}(\boldsymbol{u})] & \text{si} \quad \theta \leq \dfrac{2}{\lambda_{\min}[\boldsymbol{P}(\boldsymbol{u})]+\lambda_{\max}[\boldsymbol{P}(\boldsymbol{u})]} \equiv \widehat{\theta} \\ \theta\lambda_{\max}[\boldsymbol{P}(\boldsymbol{u})]-1 & \text{sinon.} \end{cases}\end{aligned}$$

la seconde égalité utilise le caractère réel de $\Lambda[\boldsymbol{P}(\boldsymbol{u})]$. Par continuité de $\boldsymbol{P}(\boldsymbol{u})$, la fonction $\sigma(\theta)$ est continue et linéaire par morceaux. Le caractère positif des valeurs propres de $\boldsymbol{P}(\boldsymbol{u})$ assure la décroissance de $\sigma(\theta)$ sur $\theta \in]0;\widehat{\theta}]$ et sa croissance pour $\theta > \widehat{\theta}$ (cf. figure 9.4). Son minimum $\widehat{\sigma}$ est atteint en $\theta = \widehat{\theta}$ et vérifie (9.41).

Lemme 4 :

Lemme 4 *Soit ϕ une fonction C^2, paire et strictement convexe, alors la fonction,*

$$l(u) = \frac{\phi'(u)}{u},$$

est continue positive, paire, et prolongeable par continuité en $u = 0$ par $\phi''(0)$. D'autre part, si l'on suppose en plus que ϕ est sur $\mathbb{R}$ et que,

$$\lim_{|u|\to\infty} \frac{\phi(u)}{u^2} = 0,$$

alors $l(u)$ tend vers zéro pour $|u| \to \infty$.

Preuve Clairement, $l(u)$ est une fonction continue sur $\mathbb{R} \setminus \{0\}$ et on montre maintenant que $l(u)$ est prolongeable par continuité en $u = 0$: on a

$$\lim_{u \to 0} \frac{\phi'(u)}{u} = \lim_{u \to 0} \frac{\phi'(u) - \phi'(0)}{u} = \phi''(0)$$

la première égalité s'appuie sur le fait que $\phi'(0) = 0$ (une conséquence directe du caractère strictement convexe, coercive et paire de ϕ) et la seconde a effectivement un sens puisque ϕ est C^2. La parité et la positivité de $l(u)$ se déduisent immédiatement de la parité de ϕ (puisque ϕ' est alors une fonction impaire). Afin de montrer que $l(u)$ tend vers zéro à l'infini, on écrit l'inégalité de convexité pour ϕ prise pour un intervalle quelconque $]u; 2u[\in \mathbb{R}^+ \setminus \{0\}$:

$$\phi(2u) \geq \phi(u) + x\phi'(u),$$

ce qui, divisé par $(2u)^2 \neq 0$ donne,

$$\frac{\phi(2u)}{(2u)^2} \geq \frac{1}{4}\frac{\phi(u)}{u^2} + \frac{1}{4}\frac{\phi'(u)}{u}.$$

Quand $u \to +\infty$, le premier terme du membre de droite et le terme du membre de gauche tendent tous les deux vers 0, ce qui implique que $\phi'(u)/u$ tend également

vers cette limite à l'infini. Ce résultat tient également pour $u \to -\infty$ puisque $l(u)$ est paire. ■

Lemme 5 :

Lemme 5 *Soit ϕ une fonction C^2 vérifiant les hypothèses (9.27) propres à la construction de GR. Alors la fonction*

$$l(x) = \frac{\phi'(x)}{x},$$

est croissante sur $]-\infty; 0[$ et décroissante sur $]0; \infty[$. De plus, pour tout x dans $\mathbb{R}$, on a l'inégalité suivante :

$$\frac{\phi'(x)}{x} \geq \phi''(x).$$

Preuve Posons $h(.) = \sqrt{.}$ et écrivons le diagramme de composition suivant,

$$x \in \mathbb{R}^+ \overset{h}{\mapsto} u \in \mathbb{R}^+ \overset{\phi}{\mapsto} z \in \mathbb{R}^+ = \phi(u) = (\phi \circ h)(x).$$

La propriété de concavité de $(\phi \circ h)(.)$ sur $\mathbb{R}^+$ se traduit par,

$$\frac{\mathrm{d}^2(\phi \circ h)(x)}{\mathrm{d}x^2} = \frac{\mathrm{d}^2 z}{\mathrm{d}x^2} \leq 0 \quad \forall x \in \mathbb{R}^+. \tag{II.2}$$

La dérivation des fonctions composées nous amène classiquement à écrire pour $u = \sqrt{x} > 0$,

$$\frac{\mathrm{d}z}{\mathrm{d}x} = \frac{1}{2u}\frac{\mathrm{d}\phi(u)}{\mathrm{d}u}, \tag{II.3}$$

$$\frac{\mathrm{d}^2z}{\mathrm{d}x^2} = \frac{-1}{4u^3}\frac{\mathrm{d}\phi(u)}{\mathrm{d}u} + \frac{1}{4u^2}\frac{\mathrm{d}^2\phi(u)}{\mathrm{d}u^2}. \tag{II.4}$$

L'égalité (II.3 permet d'affirmer par (II.2) que $l(x)$ est décroissante sur $\mathbb{R}^+\setminus\{0\}$, et par parité de cette dernière, qu'elle est croissante sur $\mathbb{R}\setminus\{0\}$. D'autre part, (II.4) combinée avec (II.2) donne :

$$\frac{\phi'(u)}{u} \geq \phi''(u) \geq 0, \quad \forall u > 0;$$

la parité de l et de ϕ'' permet d'affirmer que cette inégalité tient également pour tout $u < 0$; finalement, puisque $l(x)$ peut être prolongé par continuité par $\phi''(0)$ en $x = 0$ (c.f. lemme 4), l'inégalité tient pour tout $x \in \mathbb{R}$. ■

II.3 Preuve des propositions

Preuve de la proposition 4 :

Soit $\tilde{J}$ une approximation quadratique de J au sens de la définition 5 et vérifiant l'hypothèse (9.10). Posons $\boldsymbol{\xi}(\boldsymbol{u}) \in \mathbb{R}^N$ la solution de l'équation normale[1] :

$$\boldsymbol{B}(\boldsymbol{u})\,\boldsymbol{\xi}(\boldsymbol{u}) = -\boldsymbol{\nabla} J(\boldsymbol{u}). \tag{II.5}$$

Posons $\boldsymbol{v}^+ = \boldsymbol{u} + \theta\boldsymbol{\xi}(\boldsymbol{u})$, la forme de $\tilde{J}$ nous permet d'écrire :

$$\begin{aligned} \tilde{J}(\boldsymbol{u}, \boldsymbol{v}^+) - J(\boldsymbol{u}) &= \theta \langle \boldsymbol{\nabla} J(\boldsymbol{u}), \boldsymbol{\xi}(\boldsymbol{u}) \rangle + \frac{\theta^2}{2} \langle \boldsymbol{B}(\boldsymbol{u})\,\boldsymbol{\xi}(\boldsymbol{u}), \boldsymbol{\xi}(\boldsymbol{u}) \rangle \\ &= \theta \left(\frac{\theta}{2} - 1 \right) \langle \boldsymbol{B}(\boldsymbol{u})\,\boldsymbol{\xi}(\boldsymbol{u}), \boldsymbol{\xi}(\boldsymbol{u}) \rangle ; \end{aligned}$$

la seconde égalité découle de (II.5) ; cette quantité est négative ou nulle si et seulement si $\theta \in]0;2[$ puisque la définition 5 assure que $\boldsymbol{B}(\cdot)$ est DP. Ceci montre que l'inégalité :

$$\tilde{J}(\boldsymbol{u}, \boldsymbol{v}^+) \leq J(\boldsymbol{u}), \tag{II.6}$$

tient si et seulement si $\theta \in]0;2[$. D'autre part, l'hypothèse (9.10) permet d'écrire :

$$\forall \boldsymbol{\xi} \in \mathbb{R}^N, \theta \in \mathbb{R} \qquad \tilde{J}(\boldsymbol{u}, \boldsymbol{u} + \theta\boldsymbol{\xi}) \geq J(\boldsymbol{u} + \theta\boldsymbol{\xi}) ;$$

[1]Notons que $\boldsymbol{\xi}(\boldsymbol{u})$ existe et est unique puisque $\boldsymbol{B}(\boldsymbol{u})$ est un opérateur DP par hypothèse.

cette dernière inégalité tient donc *a fortiori* pour $\boldsymbol{\xi} = \boldsymbol{\xi}(\boldsymbol{u})$ solution de (II.5) et $\theta \in]0;2[$; on a donc :

$$\forall \theta \in]0;2[\qquad \tilde{J}\left(\boldsymbol{u}, \boldsymbol{v}^{+}\right) \geq J\left(\boldsymbol{v}^{+}\right).$$

On en déduit donc par (II.6) :

$$\forall \theta \in]0;2[\qquad J\left(\boldsymbol{u}\right) \geq J\left(\boldsymbol{v}^{+}\right). \tag{II.7}$$

D'après les hypothèses faites au début de la preuve sur $\tilde{J}$, le schéma itératif à pas fixe $(\mathcal{X}, \Xi, \theta)$ défini par la correspondance $\boldsymbol{x}^{(n)} = \boldsymbol{u}$, $\boldsymbol{\xi}^{(n)} = \boldsymbol{\xi}(\boldsymbol{u})$ et $\boldsymbol{x}^{(n+1)} = \boldsymbol{v}^{+}$ est un algorithme relaxé de WEISZFELD généralisé, et l'inégalité (II.7) devient $J\left(\boldsymbol{x}^{(n)}\right) \geq J\left(\boldsymbol{x}^{(n+1)}\right)$ pour $\theta \in]0;2[$. ■

Preuve de la proposition 5 :

Soit $\tilde{J}$ une approximation quadratique de J au sens de la définition 5 et telle que l'hypothèse (9.10) soit vérifiée ; en posant $\boldsymbol{v} = \boldsymbol{u} + \theta\boldsymbol{\xi}$, on peut écrire :

$$\forall \boldsymbol{u}, \theta, \boldsymbol{\xi} \qquad J(\boldsymbol{u}) + \theta \left\langle \boldsymbol{\nabla} J\left(\boldsymbol{u}\right), \boldsymbol{\xi} \right\rangle + \frac{\theta^2}{2} \left\langle \boldsymbol{B}\left(\boldsymbol{u}\right)\boldsymbol{\xi}, \boldsymbol{\xi} \right\rangle \geq J\left(\boldsymbol{u} + \theta\boldsymbol{\xi}\right);$$

cette inégalité tient en particulier si $\boldsymbol{\xi} = \boldsymbol{\xi}(\boldsymbol{u})$, solution de l'équation normale (II.5) ; dans ce cas, $\boldsymbol{v}$ prend la valeur particulière $\boldsymbol{v} = \boldsymbol{u}^{+} = \boldsymbol{u} + \theta\boldsymbol{\xi}(\boldsymbol{u})$ et l'inégalité

précédente équivaut à :

$$\forall \boldsymbol{u}, \theta \qquad J(\boldsymbol{u}) - J\left(\boldsymbol{u}^{+}\right) + \theta\left(1 - \frac{\theta}{2}\right)\left\langle \boldsymbol{\nabla} J\left(\boldsymbol{u}\right), \boldsymbol{\xi}(\boldsymbol{u})\right\rangle \geq 0; \tag{II.8}$$

L'expression (II.8) est bien identique à l'inégalité d'ARMIJO (9.4) qui doit être vraie pour tout $\boldsymbol{u}$ et prise pour $\omega = 1 - \theta/2$. ■

Note : remarquons ici que la valeur $\omega = 1 - \theta/2$ appartient à $]0;1[$ pour tout pas fixe $\theta \in]0;2[$. Rappelons que la proposition 4 garantit le caractère non croissant de la suite $\mathcal{J}$ pour θ dans cet intervalle. On pourra alors préciser que, dans le cadre d'un schéma itératif à pas fixe vérifiant (9.8) et (9.2), l'hypothèse (9.10) est vérifiée *avec décroissance de l'objectif* si et seulement si l'hypothèse (9.9) est vraie pour $\omega = 1 - \theta/2$.

Preuve de la proposition 6 :

Réécrivons l'inégalité d'ARMIJO (9.4) en posant $\boldsymbol{u}^{+} = \boldsymbol{u} + \theta\boldsymbol{\xi}(\boldsymbol{u})$:

$$J\left(\boldsymbol{u}^{+}\right) - J\left(\boldsymbol{u}\right) - \omega\theta\left\langle \boldsymbol{\nabla} J\left(\boldsymbol{u}\right), \boldsymbol{\xi}(\boldsymbol{u})\right\rangle \leq 0 \tag{II.9}$$

En imposant à $\boldsymbol{\xi}(\boldsymbol{u})$ d'être la solution de (II.5), le membre de gauche de cette inégalité s'écrit :

$$J\left(\boldsymbol{u}^{+}\right)-J\left(\boldsymbol{u}\right)+\left(\omega-1\right)\left\langle \boldsymbol{B}(\boldsymbol{u})\boldsymbol{\xi}(\boldsymbol{u}),\boldsymbol{\xi}(\boldsymbol{u})\right\rangle -\left\langle \theta\boldsymbol{\nabla}J\left(\boldsymbol{u}\right),\boldsymbol{\xi}(\boldsymbol{u})\right\rangle ,$$

et en tenant compte de la forme du critère pénalisé, on en déduit :

$$\begin{aligned} &Q\left(\boldsymbol{u}^{+}\right)-Q\left(\boldsymbol{u}\right)-\left\langle \theta\boldsymbol{\nabla}Q\left(\boldsymbol{u}\right),\boldsymbol{\xi}(\boldsymbol{u})\right\rangle +\cdots \\ &\left(\omega-1\right)\left\langle \boldsymbol{B}(\boldsymbol{u})\boldsymbol{\xi}(\boldsymbol{u}),\boldsymbol{\xi}(\boldsymbol{u})\right\rangle +\left[\Phi\left(\boldsymbol{u}^{+}\right)-\Phi\left(\boldsymbol{u}\right)+\theta\left\langle \boldsymbol{\nabla}\Phi(\boldsymbol{u}),\boldsymbol{\xi}(\boldsymbol{u})\right\rangle \right]; \end{aligned} \quad \text{(II.10)}$$

on note alors que, puisque Q est une forme quadratique, les trois premiers termes du membre de gauche de (II.10) représentent la quantité $\theta^{2}\left\langle \boldsymbol{Q}\boldsymbol{\xi}(\boldsymbol{u}),\boldsymbol{\xi}(\boldsymbol{u})\right\rangle$. On en déduit que le membre de gauche de (II.9) s'écrit :

$$\begin{aligned} &J\left(\boldsymbol{u}^{+}\right)-J\left(\boldsymbol{u}\right)-\omega\theta\left\langle \boldsymbol{\nabla}J\left(\boldsymbol{u}\right),\boldsymbol{\xi}(\boldsymbol{u})\right\rangle = \\ &\theta\left\langle \left[\left(\omega-1\right)\boldsymbol{B}(\boldsymbol{u})+\theta\boldsymbol{Q}\right]\boldsymbol{\xi}(\boldsymbol{u}),\boldsymbol{\xi}(\boldsymbol{u})\right\rangle +\left[\Phi\left(\boldsymbol{u}^{+}\right)-\Phi\left(\boldsymbol{u}\right)-\theta\left\langle \boldsymbol{\nabla}\Phi(\boldsymbol{u}),\boldsymbol{\xi}(\boldsymbol{u})\right\rangle \right]. \end{aligned} \quad \text{(II.11)}$$

Enfin, d'après la forme de Φ donnée par (9.13), le gradient s'écrit :

$$\boldsymbol{\nabla}\Phi(\boldsymbol{u})=\boldsymbol{V}\boldsymbol{\phi}^{'}\left(\delta_{\bullet}\right)$$

où nous avons posé

$$\begin{aligned} \boldsymbol{V} &= [\boldsymbol{v}_1|\cdots|\boldsymbol{v}_I], \\ \delta_i &= \langle \boldsymbol{v}_i, \boldsymbol{u} \rangle, \\ \boldsymbol{\phi}'(\delta_\bullet) &= [\phi'(\delta_1), \ldots, \phi'(\delta_M)]^T; \end{aligned}$$

le membre de droite de (II.11) devient, en posant $\delta_c^+ = \langle \boldsymbol{v}_i, \boldsymbol{u}^+ \rangle$:

$$\left\langle \left[(\omega-1)\theta \boldsymbol{B}(\boldsymbol{u}) + \theta^2 \boldsymbol{Q}\right] \boldsymbol{\xi}(\boldsymbol{u}), \boldsymbol{\xi}(\boldsymbol{u}) \right\rangle + \sum_i \left[\phi\left(\delta_i^+\right) - \phi\left(\delta_i\right) - \left(\delta_i^+ - \delta_i\right) \phi'\left(\delta_i\right)\right].$$

■

Preuve de la proposition 7 :

Nous montrons que, pour tout compact $D_0 \subset \mathbb{R}$, le schéma itératif de GY produit une suite Ξ gradient-reliée à $\mathcal{X}$. Appuyons nous sur la proposition 1 ; la matrice

$$\boldsymbol{B}_{\mathrm{GY}}^a = 2\boldsymbol{Q} + \frac{1}{a}\boldsymbol{V}\boldsymbol{V}^T,$$

avec $a > 0$, définit le schéma itératif de GY. Cet opérateur étant à coefficients constants, ses valeurs propres sont invariantes et finies ; on note également que $\boldsymbol{Q}$ et $\boldsymbol{V}\boldsymbol{V}^T$ sont DNN, ce qui implique que $\boldsymbol{B}_{\mathrm{GY}}^a$ est DP sur $\mathbb{R}$ si et seulement si $\ker\{\boldsymbol{Q}\} \cap \ker\{\boldsymbol{V}^T\} = \{\boldsymbol{0}\}$. Les hypothèses de la proposition 1 sont donc réunies.■

Preuve de la proposition 8 :

On peut à présent donner la preuve de l'admissibilité du pas fixe pour l'algorithme de GY en s'appuyant sur la condition (9.9) donnée en section 9.3 et sur la proposition 6 ; ainsi en remplaçant $\boldsymbol{B}(\boldsymbol{u})$ par $\boldsymbol{B}^{a}_{\mathrm{GY}}$ dans (9.14) et en posant $\Delta_i = \delta_i^+ - \delta_i$, il faut montrer $\exists\, \omega \in]0;1[$ tel que, pour tout $\boldsymbol{u} \in \mathbb{R}^N$:

$$\begin{aligned}&(2-2\omega-\theta)\left\langle \boldsymbol{Q}\boldsymbol{\xi}(\boldsymbol{u}), \boldsymbol{\xi}(\boldsymbol{u})\right\rangle + \cdots \\ &\lambda \sum_i \left[\left(\frac{1-\omega}{a\theta}\right)\Delta_i^2 + \phi\left(\delta_i\right) + \delta_i \phi'\left(\delta_i\right) - \phi\left(\delta_i^+\right)\right] \geq 0. \end{aligned} \tag{II.12}$$

Une CS de positivité de (II.12) consiste à s'assurer que les deux éléments de l'expression sont positifs ou nuls pour tout $\boldsymbol{u}$; on souligne alors que :

i) le caractère DNN de $\boldsymbol{Q}$ assure que le premier terme est positif ou nul si $\omega \leq 1 - \frac{\theta}{2}$;

ii) Cherchons une condition permettant de garantir que la quantité

$$\left(\frac{1-\omega}{a\theta}\right)\Delta_i^2 + \phi\left(\delta_i\right) + \Delta_i \phi'\left(\delta_i\right) - \phi\left(\delta_i^+\right) \geq 0. \tag{II.13}$$

Faisons tout d'abord l'hypothèse que ϕ est de dérivée L-Lipschitz, i.e.

$$\exists\, 0 < L < \infty : \forall a, b \in \mathbb{R}, \qquad |\phi'(a) - \phi'(b)| \leq L|a-b|;$$

le *lemme de descente* [Bertsekas, 1999, prop. A.24] permet d'écrire :

$$\phi(\delta_i) - \phi(\delta_i^+) + \Delta_i \phi'(\delta_i) \geq -\frac{L}{2}\Delta_i^2,$$

ce qui permet de minorer (II.13) par :

$$\left[\left(\frac{1-\omega}{a\theta}\right) - \frac{L}{2}\right]\Delta_i^2,$$

dont la positivité est assurée si

$$\left(\frac{1-\omega}{a\theta}\right) \geq \frac{L}{2} \quad \Leftrightarrow \quad \omega \leq 1 - \frac{a\theta L}{2}.$$

Si on note maintenant que, d'après le lemme 1 (page 293), l'hypothèse L-lipschitzien sur ϕ' est identique à l'hypothèse (9.16) de la proposition 8, on déduit des points (i) et (ii) que si

$$\theta \in]0;2[\qquad \text{et} \qquad a \in]0; \frac{2}{\theta}\widehat{a}[,$$

avec $\widehat{a} = 1/L$, alors $\exists\, \omega \in]0;1[$ tel que, $\forall \boldsymbol{u}$, l'inégalité (II.12) est vérifiée. En accord avec la proposition 6, ce résultat permet d'affirmer que θ vérifie, pour tout $\boldsymbol{u}$, la règle d'ARMIJO (9.4). On en conclut que la condition (9.9) est vérifiée et le pas fixe θ est admissible pour la règle d'ARMIJO pour tout $(\mathcal{X}, \Xi)$ vérifiant (9.8) et (9.2).■

Preuve de la proposition 9 :

Suivons la même démarche que pour le cas GY monovariée : en remplaçant $\boldsymbol{B}(\boldsymbol{u})$ par $\boldsymbol{B}_{\nu}^{a}$ dans (9.14), il faut montrer $\exists\ \omega \in]0;1[$ tel que, pour tout $\boldsymbol{u} \in \mathbb{R}^N$:

$$(2-2\omega-\theta)\left\langle \boldsymbol{Q}\boldsymbol{\xi}(\boldsymbol{u}), \boldsymbol{\xi}(\boldsymbol{u})\right\rangle + \cdots$$
$$\left\{ \left(\frac{1-\omega}{a\theta}\right)\left\langle \boldsymbol{\xi}(\boldsymbol{u}), \boldsymbol{\xi}(\boldsymbol{u})\right\rangle + \sum_i \left[\phi\left(\delta_i\right) + \Delta_i\phi'\left(\delta_i\right) - \phi\left(\delta_i^{+}\right)\right] \right\} \geq 0. \quad \text{(II.14)}$$

Là encore, une condition suffisante consiste à garantir la positivité des deux parties du membre de gauche de (II.14). Le caractère DNN de $\boldsymbol{Q}$ assure que le premier terme est positif ou nul si $\omega \leq 1 - \theta/2$. Intéressons nous maintenant à la partie entre accolades. En se rappelant que $\Phi(\cdot) = \sum_i \phi\left(\langle \boldsymbol{v}_i, \cdot\rangle\right)$, on peut reformuler la somme par :

$$\Phi(\boldsymbol{u}) - \Phi\left(\boldsymbol{u}^{+}\right) + \left\langle \boldsymbol{\nabla}\Phi(\boldsymbol{u}), \left(\boldsymbol{u}^{+} - \boldsymbol{u}\right)\right\rangle, \quad \text{(II.15)}$$

où, là encore nous avons posé $\boldsymbol{u}^{+} = \boldsymbol{u} + \theta\boldsymbol{\xi}(\boldsymbol{u})$. Introduisons la fonction

$$\widehat{h}(\boldsymbol{u}) = \frac{||\boldsymbol{u}||^2}{2} - \widehat{a}\Phi(\boldsymbol{u}),$$

qu'on supposera convexe, i.e.

$$\widehat{h}\left(\boldsymbol{u}^{+}\right) - \widehat{h}(\boldsymbol{u}) \geq \left\langle \boldsymbol{\nabla}\widehat{h}(\boldsymbol{u}), \left(\boldsymbol{u}^{+} - \boldsymbol{u}\right)\right\rangle;$$

cette dernière inégalité nous permet d'obtenir aisément une minoration de la quantité (II.15) :

$$\Phi(\boldsymbol{u}) - \Phi\left(\boldsymbol{u}^{+}\right) + \theta \left\langle \boldsymbol{\nabla}\Phi(\boldsymbol{u}), \boldsymbol{\xi}(\boldsymbol{u}) \right\rangle \geq -\frac{\theta^2}{2\widehat{a}} \left\langle \boldsymbol{\xi}(\boldsymbol{u}), \boldsymbol{\xi}(\boldsymbol{u}) \right\rangle, \qquad \text{(II.16)}$$

ceci nous permet de minorer le second terme de (II.14) par l'expression suivante :

$$\theta^2 \left[\left(\frac{1-\omega}{a\theta} \right) - \frac{1}{2\widehat{a}} \right] \left\langle \boldsymbol{\xi}(\boldsymbol{u}), \boldsymbol{\xi}(\boldsymbol{u}) \right\rangle$$

qui est positive ou nulle si on a :

$$\omega \leq 1 - \frac{a\theta}{2\widehat{a}}.$$

Finalement, on garantit pour un couple (θ, a) tel que

$$\theta \in]0; 2[\qquad \text{et} \qquad a \in]0; 2\widehat{a}/\theta[,$$

qu'il existe $\omega \in]0; 1[$ tel que, $\forall \boldsymbol{u}$, l'inégalité (II.12) est vérifiée. En accord avec la proposition 6, ce résultat permet d'affirmer que θ vérifie, pour tout $\boldsymbol{u}$, la règle d'ARMIJO (9.4). On en conclut que, sous les hypothèses de la proposition 9, la condition (9.9) est vérifiée et le pas fixe θ est admissible pour cette règle pour tout $(\mathcal{X}, \Xi)$ vérifiant (9.8) et (9.2). ■

Preuve de la proposition 10 :

Pour l'algorithme de GR, le caractère gradient-relié de Ξ peut être montré par la proposition 1. Plus précisément, on montre que l'opérateur :

$$\boldsymbol{B}(\boldsymbol{u}) = 2\boldsymbol{Q} + \lambda \boldsymbol{V}\boldsymbol{L}(\boldsymbol{u})\boldsymbol{V}^T, \qquad \text{avec } \boldsymbol{L}(\boldsymbol{u}) = \operatorname{diag}\left(\frac{\phi'(\delta_i)}{\delta_i}\right)$$

est borné et uniformément DP sur $\mathbb{R}$ en invoquant que,

1. $\boldsymbol{Q}$ et $\boldsymbol{V}^T$ ont un noyau disjoint — condition (9.25) de la proposition 10,
2. ϕ étant convexe, ϕ' est monotone croissante et l'hypothèse (9.24a) assure que

$$\forall\, u \in \mathbb{R}^*, \quad \phi'(u)/u > 0$$

3. puisqu'on suppose que $\phi'(u)/u$ a une borne supérieure uniforme (condition (9.24b) de la proposition 10), alors $\lim_{u\to 0} \phi'(u)/u$ existe et est bornée et on écrit par (9.24a)

$$\lim_{u\to 0} \frac{\phi'(u)}{u} = \phi''(0)$$

qui est strictement positive puisque ϕ est strictement convexe.

Les deux derniers points assurent que, pour tout $\boldsymbol{u} \in \mathbb{R}^N$ l'opérateur diagonal $L(\boldsymbol{u})$ est uniformément positif et borné supérieurement ; en considérant ensuite le premier point, il est alors aisé de tirer une conclusion identique pour $\boldsymbol{B}(\boldsymbol{u})$. ■

Preuve de la proposition 11 (GR)

La preuve suit globalement la démarche présentée dans le cadre de l'algorithme de GY (cf. proposition 8) : en remplaçant $\boldsymbol{B}(\boldsymbol{u})$ par l'expression de $\boldsymbol{B}_{\text{GR}}(\boldsymbol{u})$ dans (9.14), faut montrer $\exists\, \omega \in]0;1[$ tel que, $\forall \boldsymbol{u} \in \mathbb{R}^N$

$$(2-2\omega-\theta)\left\langle \boldsymbol{Q}\boldsymbol{\xi}(\boldsymbol{u}), \boldsymbol{\xi}(\boldsymbol{u})\right\rangle + \cdots \qquad \text{(II.17)}$$
$$\lambda \sum_c \left[\left(\frac{1-\omega}{\theta}\right) \Delta_i^2 \frac{\phi'(\delta_i)}{\delta_i} + \phi\left(\delta_i\right) + \Delta_i \phi'\left(\delta_i\right) - \phi\left(\delta_i^+\right) \right].$$

Là encore, le caractère DP de $\boldsymbol{Q}$ assure que la première composante est positive si $\omega \leq 1 - \frac{\theta}{2}$.

D'autre part, si on impose $\omega = 1 - \frac{\theta}{2}$ (qui appartient à $]0;1[$ pour $0 < \theta < 2$), chaque terme de la somme s'écrit :

$$\begin{aligned} & \lambda \sum_c \left\{ \phi\left(\delta_i\right) - \phi\left(\delta_i^+\right) + \phi'(\delta_i)\, \Delta_i + \frac{1}{2} \frac{\phi'(\delta_i)}{\delta_i} \Delta_i^2 \right\} \\ = \; & \lambda \sum_c \left\{ \phi\left(\delta_i\right) - \phi\left(\delta_i^+\right) + \frac{1}{2} \frac{\phi'(\delta_i)}{\delta_i} \left(\left(\delta_i^+\right)^2 - \delta_i^2 \right) \right\}. \end{aligned}$$

On pose alors pour $u \in \mathbb{R}$, $\psi(u) = \phi(\sqrt{|u|})$ en soulignant que cette fonction est *concave* sur $\mathbb{R}^+$ par hypothèse de la construction de GR — cf. conditions (9.24) ; de plus, on a $\psi(u) = \phi(-\sqrt{|u|})$ car ϕ est également paire par hypothèse. Chaque

terme de la somme ci-dessus se récrit donc sous la forme,

$$\psi(\delta_i^2) - \psi((\delta_i^+)^2) + \psi'(\delta_i^2)\left((\delta_i^+)^2 - \delta_i^2\right),$$

cette dernière expression est positive ou nulle par la concavité de ψ.

Au final, on déduit que pour $0 < \theta < 2$, $\exists\, \omega = 1 - \frac{\theta}{2} \in]0;1[$ tel que la positivité de (II.17) est assurée pour tout $\boldsymbol{u}$. Ce résultat permet de garantir l'admissibilité du fixe θ au sens de la règle d'ARMIJO pour tout $(\mathcal{X}, \Xi)$ vérifiant (9.8) et (9.2) si les hypothèses de la proposition (11) sont vérifiées. ■

Preuve de la proposition 12

Soit

$$\mathcal{M}_{\text{GR}}(\boldsymbol{u}) = \boldsymbol{u} - \theta \boldsymbol{B}_{\text{GR}}(\boldsymbol{u})^{-1} \boldsymbol{\nabla} J(\boldsymbol{u})$$

l'application permettant de définir le schéma itératif à pas fixe de GR. Pour ϕ de classe C^2, on a simultanément J de classe C^2 et $\phi'(u)/u$ continu[2] sur $\mathbb{R}$, ce qui assure que $\boldsymbol{B}_{\text{GR}}(\boldsymbol{u})$ est continue sur $\mathbb{R}^N$. On garantit donc que la différentielle $\mathcal{M}'_{\text{GR}}$ de $\mathcal{M}_{\text{GR}}$ existe en $\boldsymbol{x}^\star$:

$$\mathcal{M}'_{\text{GR}}(\boldsymbol{x}^\star) = \boldsymbol{I} - \theta \boldsymbol{B}_{\text{GR}}(\boldsymbol{x}^\star)^{-1} \boldsymbol{H}(\boldsymbol{x}^\star)$$

[2]On montre en effet sans difficulté la continuité sur $\mathbb{R}^*$, et d'autre part, on peut prolonger par continuité $\phi'(u)/u$ en $u = 0$ par $\phi''(0)$; cf. lemme 4 dans cette annexe.

D'autre part, on vérifie aisément que $\boldsymbol{x}^\star$ est un point fixe de $\mathcal{M}_\theta$ — il annule le gradient. Par la suite, la preuve se déduit directement du théorème 5 si on peut montrer que pour $\theta \in]0;2[$, on a $\rho(\mathcal{M}'_{\text{GR}}(\boldsymbol{x}^\star)) < 1$.

Ce résultat intermédiaire peut être démontré sans difficulté en s'appuyant sur le lemme 3 pour la matrice $\boldsymbol{P} = \boldsymbol{B}_{\text{GR}}^{-1}\boldsymbol{H}$ (matrice dont on montre sans difficulté le caractère réel et positif du spectre) : on déduit de ce lemme que pour

$$\theta < \bar{\theta} \equiv \frac{2}{\lambda_{\max}[\boldsymbol{B}_{\text{GR}}^{-1}(\boldsymbol{x}^\star)\boldsymbol{H}(\boldsymbol{x}^\star)]},$$

on a $\rho(\mathcal{M}'_{\text{GR}}(\boldsymbol{x}^\star)) < 1$; donc, il suffit de montrer que $\bar{\theta} \leq 2$ pour garantir que $\rho(\mathcal{M}'_{\text{GR}}(\boldsymbol{x}^\star))$ est strictement inférieur à un pour tout θ pris dans $]0;2[$. Montrer que $\bar{\theta} \leq 2$ est équivalent à montrer la positivité de

$$\lambda_{\min}[\boldsymbol{H}(\boldsymbol{x}^\star)^{-1}\boldsymbol{B}_{\text{GR}}(\boldsymbol{x}^\star)] - 1. \tag{II.18}$$

On note alors que le Hessien de J s'écrit

$$\boldsymbol{H}(\boldsymbol{u}) = 2\boldsymbol{Q} + \boldsymbol{V}\,\text{diag}(\phi''(\delta_i))\boldsymbol{V}^T;$$

et que

$$\boldsymbol{B}_{\text{GR}}(\boldsymbol{u}) = 2\boldsymbol{Q} + \boldsymbol{V}\,\text{diag}(l_i)\boldsymbol{V}^T, \quad \text{avec} \quad l_i = \phi'(\delta_i)/\delta_i,$$

où, d'après le lemme 5 , on a $l_i \geq \text{diag}(\phi''(\delta_i)$. Il est donc possible d'écrire à partir de (II.18) :

$$\lambda_{\min}\left[\boldsymbol{H}(\boldsymbol{x}^\star)^{-1}\big(\boldsymbol{H}(\boldsymbol{x}^\star)+\boldsymbol{C}(\boldsymbol{x}^\star)\big)\right]-1,$$

avec $\boldsymbol{C}(\boldsymbol{x}^\star)$ une matrice DP. On a donc

$$\lambda_{\min}[\boldsymbol{H}(\boldsymbol{x}^\star)^{-1}\boldsymbol{B}_{\text{GR}}(\boldsymbol{x}^\star)]-1=\lambda_{\min}[\boldsymbol{H}(\boldsymbol{x}^\star)^{-1}\boldsymbol{C}(\boldsymbol{x}^\star)]\geq 0.$$

L'expression (II.18) est donc bien positive ou nulle ce qui implique que $\bar{\theta} \geq 2$, et finalement que $\rho(\mathcal{M}'_{\text{GR}}(\boldsymbol{x}^\star)) < 1$ pour $\theta \in]0;2[$.

■