

UNIVERSITÉ DE MONTRÉAL

TRAITEMENT DE LA POSTLUMINESCENCE EN TOMOGRAPHIE À RAYONS X

SYLVAIN ARRECKX
INSTITUT DE GÉNIE BIOMÉDICAL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE BIOMÉDICAL)

JUIN 2009



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-53892-0
Our file Notre référence
ISBN: 978-0-494-53892-0

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

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

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

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

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé:

TRAITEMENT DE LA POSTLUMINESCENCE EN TOMOGRAPHIE À RAYONS X

présenté par: ARRECKX Sylvain

en vue de l'obtention du diplôme de: Maîtrise ès sciences appliquées

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

M. BERTRAND Michel, Ph.D., président

M. GOUSSARD Yves, Ph.D., membre et directeur de recherche

M. ORBAN Dominique, Doct. Sc., membre

Qui joue perd. C'est mathématique. Les statistiques ne trompent pas. Et pourtant, il faut jouer pour gagner, même si les chances sont minces. [Normand Reid]

À l'école, en algèbre, j'étais du genre Einstein. Mais plutôt Franck qu'Albert. [Philippe Geluck]

REMERCIEMENTS

Je tiens à remercier les différentes personnes qui ont permis de réaliser ce projet.

Ces remerciements s'adressent particulièrement à mon directeur de recherche, M. Yves Goussard, qui a corrigé et apporté des conseils afin d'enrichir ce travail.

Mes collègues de laboratoire, pour avoir partagé des idées sur mes travaux et pour leur humour divertissant durant les moments de fatigue.

Ainsi qu'à Johanne Raymond qui a passé de longues heures pour embellir mon mémoire.

Je remercie vivement les membres de mon jury, les Professeurs Michel Bertrand et Dominique Orban, pour s'être rendus disponibles, pour avoir accepté de s'investir dans ce travail et pour leur contribution à une lecture critique de ce mémoire.

La réalisation de ce travail a aussi été rendue possible par l'accueil toujours très agréable que m'a réservé le personnel de radiologie de l'hôpital Notre-Dame afin de faciliter les acquisitions.

Enfin, un grand merci à Claire, qui a su faire preuve de beaucoup de patience tout au long de ce travail.

RÉSUMÉ

Ce travail est incorporé dans un projet de plus grande envergure. Celui-ci vise à réduire les artefacts métalliques dans les reconstructions tomographiques à rayons X afin d'améliorer le diagnostic de resténose chez les patients possédant des stents. Des études antérieures se sont déjà intéressées à ce problème, notamment en adoptant une modélisation polychromatique de la source des rayons X. Cependant, ces travaux ne prennent pas en compte les imperfections du système de détection dans la formation des données. La postluminescence des détecteurs est le défaut majeur du système de détection qui dégrade la qualité des images reconstruites. La postluminescence correspond à la persistance momentanée des détecteurs après la mesure.

Dans le travail présenté ici, nous visons tout d'abord à comprendre l'influence de la postluminescence sur les données. Ensuite, nous modélisons ce phénomène par une convolution avec une réponse impulsionnelle. Puis, nous caractérisons la réponse impulsionnelle des détecteurs par un estimateur MAP, à partir de données médicales.

Enfin, à la lumière de travaux antérieurs effectués sur la postluminescence, nous proposons un traitement, robuste au bruit et facilement implémentable, afin de pallier ce phénomène. Ce traitement consiste à estimer avec un estimateur MAP les données affranchies de postluminescence en utilisant la réponse impulsionnelle estimée précédemment.

La caractérisation de la réponse impulsionnelle et le traitement de la postluminescence fonctionnent correctement en simulation avec des rapports signal à bruit conventionnels en tomographie (25-35 dB). Cependant, l'application de ces techniques sur données médicales engendre des résultats peu convaincants.

ABSTRACT

This work is part of a wider project that aims at reducing metal artifacts in X-ray computed tomography reconstructed images in order to improve in-stent restenosis assessment. Previous studies already treated this topic, especially in incorporating a polychromatic model of the X-ray source into the reconstruction problem. However, these studies did not take into account imperfections coming from the detection system. Detector afterglow is the major imperfection of the detection system that affects reconstructed image quality. Afterglow corresponds to a temporary detectors persistence after measurements are taken.

Our work aims first at understanding the influence of afterglow on the data. Then we model this phenomenon by a convolution with the detector impulse response. Next, we use a MAP estimator to estimate the detector impulse response from medical raw data. Finally from what was developed on afterglow treatments we propose a treatment against afterglow which is more robust with respect to noise and easy to compute. This method consists in estimating afterglow-free data with a MAP estimator using the previously obtained impulse response.

Impulse response characterization and afterglow treatment work well on simulated data with conventional tomographic signal to noise ratios (25-35 dB). However, application of these techniques on real medical data reveals less convincing results.

TABLE DES MATIÈRES

DÉDICACE	iv
REMERCIEMENTS	v
RÉSUMÉ	vi
ABSTRACT	vii
TABLE DES MATIÈRES	viii
LISTE DES FIGURES	xiii
LISTE DES NOTATIONS ET DES SYMBOLES	xvi
LISTE DES TABLEAUX	.xviii
LISTE DES ANNEXES	xx
CHAPITRE 1 INTRODUCTION	1
1.1 Contexte	1
1.2 Objectifs	4
CHAPITRE 2 TOMOGRAPHIE À RAYONS X	6
2.1 Présentation générale	7
2.2 Principales évolutions technologiques	8
2.2.1 Première génération : un détecteur	8
2.2.2 Deuxième génération : barrette de détecteurs	8
2.2.3 Troisième génération : faisceau large de rayons X	10
2.2.4 Quatrième et cinquième générations	10
2.3 Systèmes de détection	12

2.4	Formation des mesures	14
2.4.1	Interactions photon-matière	14
2.4.2	Modèle déterministe	16
2.4.3	Modèle statistique	17
2.5	Reconstruction	18
2.5.1	Méthodes analytiques	19
2.5.1.1	Transformée de Radon et sinogramme	19
2.5.1.2	Théorème de la tranche centrale de Fourier et inversion directe	21
2.5.1.3	La rétroprojection filtrée	22
2.5.1.4	Extensions et limites	24
2.5.2	Méthodes algébriques	25
2.5.2.1	Discrétisation du problème	25
2.5.2.1.1	Cadre gaussien et critère des moindres carrés	28
2.5.2.1.2	Cadre poissonnien	28
2.5.2.1.3	Régularisation	29
2.5.2.1.4	Choix du critère	30
2.5.3	Conclusion	31
2.6	Artefacts	32
2.6.1	Quelques artefacts en tomographie	32
2.6.2	Postluminescence des détecteurs	35
2.7	Conclusion	37
CHAPITRE 3	ESTIMATION D'UN SYSTÈME LINÉAIRE	39
3.1	Introduction	39
3.2	Position du problème	40
3.3	Approches « naïves »	41
3.4	Régularisation	42

3.5	Approche bayésienne	46
3.5.1	Cadre du problème	46
3.5.2	Estimateurs	48
3.5.2.1	Estimateur du maximum de vraisemblance	48
3.5.2.1.1	Propriétés	49
3.5.2.2	Estimateur du maximum <i>a posteriori</i>	50
3.5.2.2.1	Propriétés	52
3.5.3	Choix de l'estimateur	52
3.5.4	Lien avec les méthodes déterministes	53
3.6	Estimation par optimisation itérative du critère	54
3.6.1	Conditions d'optimalité	55
3.6.1.1	Condition du premier ordre	56
3.6.1.2	Conditions du second ordre	56
3.6.2	Méthodes de direction de descente	57
3.6.2.1	Conditions assurant la convergence d'un algorithme	58
3.6.2.2	Algorithme de la plus forte pente	60
3.6.2.3	Algorithme du gradient conjugué linéaire et non linéaire	60
3.6.2.3.1	Le gradient conjugué linéaire	61
3.6.2.3.2	Le gradient conjugué non linéaire	62
3.6.2.4	Algorithme de Newton	63
3.6.2.5	Algorithme quasi-Newton	66
3.7	Conclusion	68
CHAPITRE 4	POSTLUMINESCENCE	70
4.1	Introduction	70
4.2	Modélisation de la postluminescence	70
4.3	Traitement de la postluminescence	71
4.3.1	Traitement proposé par Hsieh	72

4.3.2	Méthode proposée	75
4.3.2.1	Discrétisation du système linéaire	75
4.3.2.2	Déconvolution	77
4.4	Identification des paramètres du modèle	78
4.4.1	Méthodes proposées	79
4.4.1.1	Méthode 1 : vecteur h	81
4.4.1.2	Méthode 2 : ajustement d'un modèle multiexponentiel	82
4.4.1.2.1	Comment obtenir le jacobien $J(x)$	85
4.4.1.3	Méthode 3 : estimation directe des α_n et τ_n	86
4.5	Conclusion	87
CHAPITRE 5	RÉSULTATS ET DISCUSSION	89
5.1	Introduction	89
5.2	Estimation de la réponse impulsionnelle	90
5.2.1	Méthodologie des simulations	90
5.2.2	Estimation sur données simulées	93
5.2.2.1	Méthode 1	93
5.2.2.2	Méthode 2	96
5.2.2.3	Méthode 3	98
5.2.3	Estimation sur données réelles	98
5.2.4	Conclusions	102
5.3	Traitement de la postluminescence	103
5.3.1	Méthodologie des simulations	104
5.3.2	Estimation sur données simulées	105
5.3.3	Estimation sur données réelles	109
5.4	Conclusion	111
CHAPITRE 6	CONCLUSION	113
6.1	Revue des objectifs	113

6.2 Perspectives	114
RÉFÉRENCES	116
ANNEXES	119

LISTE DES FIGURES

FIG. 1.1	Pose d'un stent dans une artère coronarienne (Heart and Institute, 2009).	2
FIG. 2.1	Principe de fonctionnement de la tomographie d'après Macovski (1983).	7
FIG. 2.2	Différentes générations de scanners d'après Kalender (2005). .	9
FIG. 2.3	Détecteur au xénon (Kalender, 2005).	12
FIG. 2.4	Détecteur de scintillation (Kalender, 2005).	13
FIG. 2.5	Spectre de la source du scanner <i>Siemens SOMATOM Sensation 16</i> . .	15
FIG. 2.6	Faisceau de rayons parallèles et son système de coordonnées. .	20
FIG. 2.7	Fantôme Shepp Logan et son sinogramme associé.	21
FIG. 2.8	Filtres utilisés pour la rétroprojection.	24
FIG. 2.9	Exemples typiques d'artefacts sur des images CT provenant a) de problèmes électroniques aux détecteurs ; b) du mouvement du patient ; c) du durcissement de rayons ; d) des effets de volume partiel ; e) d'implants métalliques ou f) du patient excédant les limites du champ de mesures d'après Kalender (2005).	33
FIG. 2.10	Réponses de plusieurs détecteurs à un échelon : caractéristiques de décroissance (Hsieh et al., 2000).	36
FIG. 2.11	Mise en évidence de la postluminescence des détecteurs en fonction de la vitesse de rotation d'après Hsieh et al. (2000).	37
FIG. 2.12	MTF pour un câble en pointillé gris pour une vitesse de rotation de 0,5 s de rotation (sans correction) ; en pointillé noir, de 2,0 s ; et la courbe noire, de 0,5 s (avec correction) d'après Hsieh (2000).	38
FIG. 3.1	Représentation des fonctions de pénalisation : L_1 , L_2 et L_2L_1 . .	45
FIG. 4.1	Courbe de postluminescence la plus récemment publiée.	79

FIG. 4.2	Méthodologie employée pour estimer la réponse impulsionnelle des détecteurs.	80
FIG. 5.1	Démarche suivie pour la validation de la restauration d'entrée dans le cas de la postluminescence.	90
FIG. 5.2	Fantôme Catphan 600 module CTP404.	91
FIG. 5.3	Réponses impulsionnelles pour différentes vitesses de rotation. .	93
FIG. 5.4	Validation de la technique d'estimation de réponse impulsionnelle sur des données sans bruit ($v = 0,5$ s/rotation).	94
FIG. 5.5	Influence du paramètre de régularisation sur l'erreur quadratique en fonction du SNR.	95
FIG. 5.6	Validation de la technique d'estimation de réponse impulsionnelle sur des données avec bruit.	96
FIG. 5.7	Modèle multiexponentiel de la réponse impulsionnelle des détecteurs ($v = 0,5$ s/rotation), ...	97
FIG. 5.8	Estimation directe des paramètres du modèle multiexponentiel de la réponse impulsionnelle des détecteurs en fonction du SNR ($v = 0,5$ s/rotation).	99
FIG. 5.9	Estimation de la réponse impulsionnelle sur des données Siemens provenant du protocole <i>InnerEarSeq</i>	100
FIG. 5.10	Estimation de la RI sur le protocole <i>InnerEarSeq</i>	101
FIG. 5.11	Estimation de la réponse impulsionnelle sur des données Siemens provenant du protocole <i>Trauma</i>	102
FIG. 5.12	Effet de la postluminescence en fonction des vitesses de rotations : zoom sur un câble de la section CTP404 du Catphan 600.	105
FIG. 5.13	Déconvolution de Hsieh sur données simulées avec un SNR de 25dB ($v = 0,5$ s/rotation).	107

FIG. 5.14	Reconstruction d'images associées à un sinogramme dégradé par la postluminescence (SNR = 25 dB) puis corrigé par un estimateur MAP ($v = 0,5$ s/rotation).	109
FIG. 5.15	Déconvolution sur données réelles : protocole <i>InnerEarSeq</i> . . .	110
FIG. 5.16	Déconvolution sur données réelles : protocole <i>Trauma</i>	112
FIG. I.1	Plan du fantôme <i>Catphan 600</i>	119
FIG. I.2	Composition du fantôme <i>Catphan 600</i> section 404.	120

LISTE DES NOTATIONS ET DES SYMBOLES

Notations

FFS :	Flying Focal Spot
UFC :	Ultra Fast Ceramics
MTF :	Modulation transfert function
$\mathcal{F}$:	Opérateur de la transformée de Fourier
$\mathcal{N}$:	Distribution suivant une loi normale
$\mathcal{P}$:	Distribution suivant une loi de Poisson
$\mathcal{R}$:	Opérateur de la transformée de Radon
$\mathcal{Z}[]$:	Opérateur de la transformée en Z
$\hat{x}$:	Estimée d'une variable
A :	Matrice
a_{ij} :	Éléments de la matrice A

Symboles physiques

N_0 :	Nombre de photons émis par la source
μ :	Coefficient d'atténuation linéique

Symboles liés à la reconstruction

θ :	Angle de projection
$p_\theta(r)$:	Projection selon l'angle θ
$p_\theta^*(r)$:	Projection filtrée selon l'angle θ
N_p :	Nombre de projections
N_d :	Nombre de détecteurs
P :	Nombre total de projections

Symboles liés à l'estimation

X :	Variable aléatoire
σ_B :	Ecart-type du bruit
σ_X :	Ecart-type des mesures
R_B :	Matrice de covariance du bruit
R_X :	Matrice de covariance des mesures
MAP :	Maximum <i>a posteriori</i>
MV :	Maximum de vraisemblance
$f_{X Z=z}$:	Loi conditionnelle de X sachant que $Z = z$
PE_{MAP} :	Propagation des erreurs de l'estimateur MAP
PE_{MV} :	Propagation des erreurs de l'estimateur MV
EQM :	Erreur quadratique moyenne

Symboles liés à l'optimisation

λ :	Coefficient de régularisation
R :	Fonction de régularisation
C :	Critère de minimisation
∇f :	Gradient de la fonction f
$\nabla^2 f$:	Hessien de la fonction f
t_k :	Pas de descente
d_k :	Direction de descente
α :	Longueur du pas
L-BFGS :	Algorithme de Broyden, Fletcher, Goldfarb et Shanno à mémoire limitée

LISTE DES TABLEAUX

TAB. 5.1	Caractéristiques géométriques des simulations pour l'estimation de la réponse impulsionnelle	92
TAB. 5.2	Erreurs quadratiques moyennes pour l'estimation des RI en fonction du SNR pour une vitesse de rotation de 0,5 s/rotation . . .	95
TAB. 5.3	Erreurs commises lors de l'ajustement de la RI par un modèle multiexponentiel (méthode 2)	98
TAB. 5.4	Erreurs quadratiques entre la réponse impulsionnelle idéale et la réponse impulsionnelle estimée par la méthode 3 en fonction du SNR	98
TAB. 5.5	Valeurs des paramètres α_n et τ_n obtenus par la méthode 2 pour le protocole <i>InnerEarSeq</i>	100
TAB. 5.6	Valeurs des paramètres α_n et τ_n obtenus par la méthode 2 pour le protocole <i>Trauma</i>	102
TAB. 5.7	Erreurs quadratiques moyennes entre le sinogramme exact et le sinogramme dégradé (EQM_{conv}); entre le sinogramme exact et celui corrigé par l'estimateur MAP (EQM_{MAP}); entre le sinogramme exact et celui corrigé par la méthode de Hsieh (EQM_{Hsieh}) pour une vitesse de rotation de 0,5 s/rotation en fonction du SNR	106
TAB. 5.8	Erreurs quadratiques moyennes entre l'image réelle et l'image reconstruite avec un sinogramme dégradé (EQM_{conv}); entre l'image réelle et celle reconstruite avec un sinogramme corrigé par l'estimateur MAP (EQM_{MAP}); entre l'image réelle et celle reconstruite avec un sinogramme corrigé par la technique de Hsieh (EQM_{Hsieh}) pour une vitesse de rotation de 0,5 s/rotation en fonction du SNR	108

TAB. I.1	Paramètres géométriques des ellipses constituant le module CTP404 du fantôme Catphan 600©	121
----------	--	-----

LISTE DES ANNEXES

ANNEXE I	DESCRIPTION DU FANTÔME CATPHAN 600	119
I.1	Plan du fantôme	119
I.2	Composition du fantôme	120
I.3	Paramètres géométriques du fantôme	121

CHAPITRE 1

INTRODUCTION

1.1 Contexte

Les maladies cardio-vasculaires sont la première cause de mortalité au Canada. Elles représentaient 32 % des décès en 2004 selon Statistiques Canada (2009). L'athérosclérose est la pathologie qui cause le plus de maladies cardio-vasculaires (54 %). Elle se caractérise par un rétrécissement du diamètre d'artères ou de veines, empêchant une circulation adéquate du sang dans ces vaisseaux. L'obturation de ceux-ci est due à l'accumulation de graisses ou de cellules endothéliales.

Lorsque de telles obturations surviennent, la mise en place d'endoprothèses, encore appelées *stents*, est parfois nécessaire afin de maintenir l'artère ouverte et de rétablir la circulation sanguine. Le stent est une structure maillée composée majoritairement de métal. Afin de l'introduire dans ce vaisseau, on insère un cathéter par lequel un ballonnet gonflant est guidé jusqu'à la lésion puis gonflé pour déployer le stent (Fig. 1.1).

Cependant, il arrive parfois que l'artère se rebouche à l'intérieur du stent. On parle alors de *resténose*. Il est donc nécessaire de suivre médicalement les patients possédant des stents pour s'assurer que l'artère conserve un diamètre suffisant pour la circulation sanguine.

Actuellement, deux techniques principales permettent d'assurer ce type de suivi : l'**angiographie** et l'**échographie endovasculaire**.

L'angiographie est un examen médical qui consiste à rendre visible les artères et les

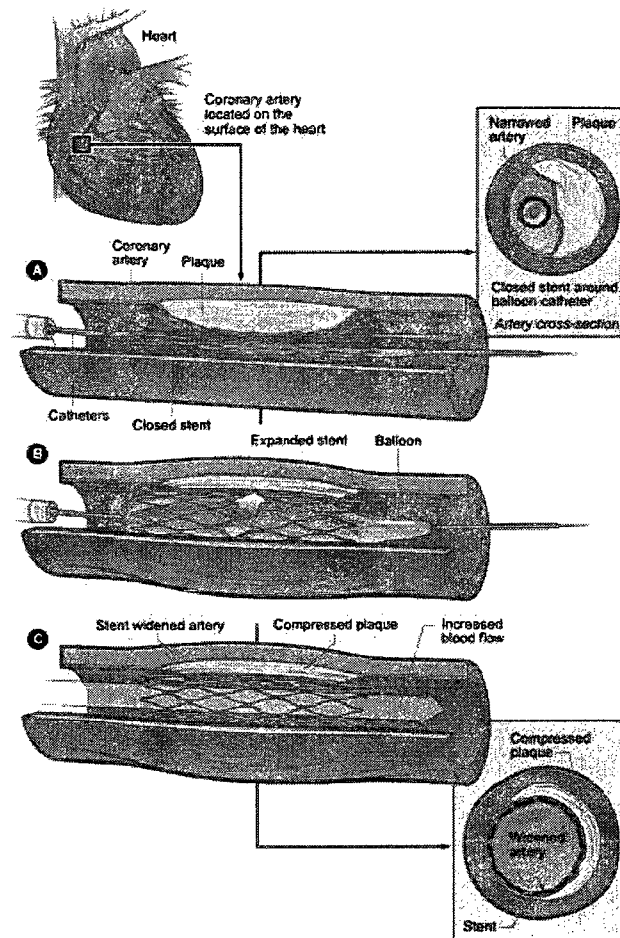


FIG. 1.1: Pose d'un stent dans une artère coronarienne (Heart and Institute, 2009).

veines lors d'une radiographie. Un cathéter sert à injecter un agent de contraste iodé, qui, par ses propriétés radio-opaques, met en évidence le système vasculaire et facilite le diagnostic des resténoses.

Cependant, cette technique est invasive. En effet, la mise en place d'un cathéter se fait par incision et l'agent de contraste est toxique pour l'organisme, notamment pour les reins.

La deuxième technique est l'IVUS (*intravascular ultrasound*), c.-à-d. l'échographie endovasculaire. Par celle-ci, le système vasculaire est observé grâce à une sonde échogra-

phique guidée par cathéter. Cette technique est moins invasive que la technique précédente, car elle ne requiert plus d'agent de contraste contrairement à la première technique. Mais cet examen est long (environ 1 h 30) et demande une grande dextérité de la part des techniciens en échographie.

La tomographie à rayons X peut aussi être utilisée pour le suivi des resténoses. En effet, cette technique présente une résolution suffisamment fine pour observer des lumières artérielles de quelques millimètres. L'utilisation d'un agent de contraste est toutefois nécessaire pour relever le contraste du système vasculaire, mais l'examen présente l'avantage d'être de courte durée.

Mes travaux s'inscrivent dans un projet de plus grande envergure dont le but est de diagnostiquer des resténoses en utilisant la tomographie à rayons X. Ils ont été réalisés dans le *Laboratoire d'imagerie par optimisation numérique* (LION) de l'École Polytechnique de Montréal, sous la direction de Yves GOUSSARD, en collaboration avec les chercheurs du *Centre de recherche du Centre hospitalier de l'Université de Montréal* (CRCHUM) pour les acquisitions de données médicales sur les scanners *Siemens* de l'hôpital Notre-Dame.

Les images reconstruites en tomographie sont entachées d'artefacts causés par la présence d'objets métalliques. De ce fait, le diagnostic d'une resténose devient difficile à réaliser. L'approche de reconstruction qui a été choisie pour s'affranchir de tels artefacts est l'utilisation de méthodes de reconstruction algébriques polychromatiques. Cependant, ces travaux se basent sur l'hypothèse que les données recueillies par le scanner (sinogramme) ne sont pas corrompues par le système de détections. Cette hypothèse est fautive dans la réalité. En effet, le système de détection permettant de capter les rayons X n'est ni parfait ni idéal. Plusieurs défauts viennent entâcher les mesures, comme par exemple les radiations afocales, les diaphonies entre détecteurs et la postluminescence des détecteurs.

La postluminescence des détecteurs, aussi appelée *postluminescence*, correspond à la persistance momentanée des détecteurs après la mesure. Ce phénomène a lieu lors de la conversion de la radiation en lumière visible. Il influence grandement la qualité des sinogrammes. Mon projet consiste à traiter les sinogrammes issus des scanners afin de diminuer l'effet de la postluminescence.

1.2 Objectifs

L'objectif principal des travaux liés à ce projet est le développement de techniques permettant d'affranchir les sinogrammes de la postluminescence du système de détection.

Afin d'atteindre cet objectif principal, nous proposons trois sous-objectifs :

1. **Compréhension et modélisation de la postluminescence des détecteurs.** En supposant que la postluminescence des détecteurs est un phénomène linéaire, nous modéliserons celle-ci par un modèle par réponse impulsionnelle.
2. **Estimation des paramètres de ce modèle.** Dans la littérature, peu de données sont fournies pour caractériser la postluminescence des détecteurs. Seules des courbes approximatives ou des données anciennes, sur des détecteurs qui ne sont plus présents dans les tomographes, sont disponibles. Ainsi, il est nécessaire d'obtenir une estimation des paramètres de notre modèle propres aux scanners utilisés, à partir de données recueillies sur ceux-ci.
3. **Restauration des sinogrammes.** Une fois les paramètres du modèle obtenus, il s'agit de corriger les sinogrammes de l'effet de postluminescence des détecteurs. On attendra de cette correction une meilleure résolution spatiale sur les reconstructions.

Afin d'atteindre ces objectifs, il est nécessaire de faire une revue de littérature en deux parties. Premièrement, nous exposons le contexte général dans lequel s'inscrit le projet :

la tomographie à rayons X. Deuxièmement, nous présentons les principaux résultats d'estimation sur lesquels nous nous sommes appuyé pour mener le projet. i

CHAPITRE 2

TOMOGRAPHIE À RAYONS X

Les idées et les concepts de la tomographie par rayons X ont été développés au début du $XX^{\text{ème}}$ siècle avec les travaux du physicien H.A. Lorentz en 1905 et ensuite du mathématicien J.H. Radon en 1917. Cependant, ce n'est qu'en 1972 qu'une mise en œuvre pratique de ces théories a été réalisée par l'ingénieur anglais G.N. Hounsfield, qui est maintenant reconnu comme l'inventeur de la tomographie. Cette nouvelle technique fut accueillie par un milieu médical enthousiaste et est considérée comme la plus importante invention dans le diagnostic radiologique depuis la découverte des rayons X. Hounsfield et A.M. Cormack (sur lequel Hounsfield a basé ses travaux) ont reçu le prix *Nobel de médecine* en 1979 pour leur invention.

Dans un premier temps, nous développons brièvement les principes généraux de la tomographie à rayons X, puis les principales évolutions technologiques qui ont permis d'améliorer la qualité des images, depuis la création de la tomographie jusqu'aux tomodiagraphes les plus récents. Cet exposé est basé sur le livre de Kalender (Kalender, 2005). Ensuite, nous examinons le système de détection et plus particulièrement les détecteurs et leur postluminescence.

Dans un deuxième temps, nous nous intéressons aux processus physiques régissant la formation des données, c'est-à-dire l'interaction entre les rayons X et la matière traversée.

Dans un troisième temps, nous détaillons les deux méthodes usuelles de reconstruction d'images (analytique et algébrique) afin de justifier notre choix de méthode de reconstruction.

Enfin, nous positionnons notre projet par rapport aux différentes étapes de l'acquisition tomographique.

2.1 Présentation générale

De manière générale, le principe de la tomographie à rayons X consiste à insérer un objet tridimensionnel entre un système produisant des rayons X (source) en direction de l'objet et un système de détection qui collecte l'énergie du faisceau ayant traversé l'objet (Fig. 2.1).

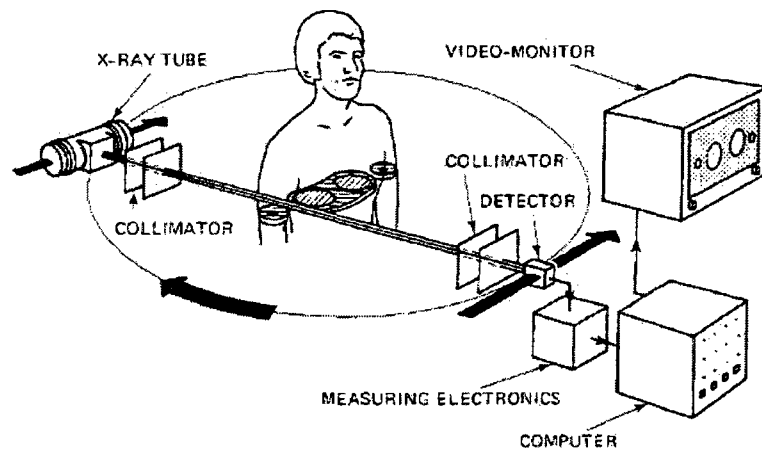


FIG. 2.1: Principe de fonctionnement de la tomographie d'après Macovski (1983).

En fonction de la composition de l'objet, les rayons X émis sont plus ou moins atténués. Sur le plan de détection, nous recueillons ainsi une projection de l'objet.

L'obtention de projections suivant plusieurs directions nous permet de reconstruire une carte spatiale de la distribution des coefficients d'atténuation, qui constitue l'objet imagé.

2.2 Principales évolutions technologiques

Depuis la construction du premier tomographe au début des années 1972, de nombreuses améliorations ont été apportées afin notamment d'augmenter la qualité des images produites, la vitesse d'acquisition des images, mais aussi de réduire la dose de rayons X envoyés aux patients, puisque ce sont des rayonnements ionisants et donc dangereux pour la santé.

2.2.1 Première génération : un détecteur

Le premier tomographe, construit en 1970 à des fins d'expérimentations, est composé d'une source émettant un faisceau étroit de rayons X et d'un unique détecteur. Tout d'abord, l'ensemble source-détecteurs se déplace en translation, constituant ainsi un ensemble de rayons parallèles pour une projection. Ensuite, il tourne d'un angle de projection autour de l'objet pour mesurer une deuxième projection (Fig. 2.2 (a)). Cet ensemble de translations puis de rotations rend le temps d'acquisition d'une tranche très long (supérieur à 24 heures pour 180 rotations de 160 translations) pour des images de 80 sur 80 pixels. Il faut attendre l'apparition des premiers tomographes commerciaux (la deuxième génération) pour diminuer de façon radicale le temps d'acquisition.

2.2.2 Deuxième génération : barrette de détecteurs

Les tomographes de deuxième génération, apparus pendant l'année 1972, fonctionnent sur le même principe de translations-rotations que la première génération (Fig. 2.2(b)). Cependant, afin de diminuer le temps d'acquisition, des détecteurs supplémentaires ont été ajoutés. Cela a pour effet de diviser le nombre de translations d'un facteur 5.

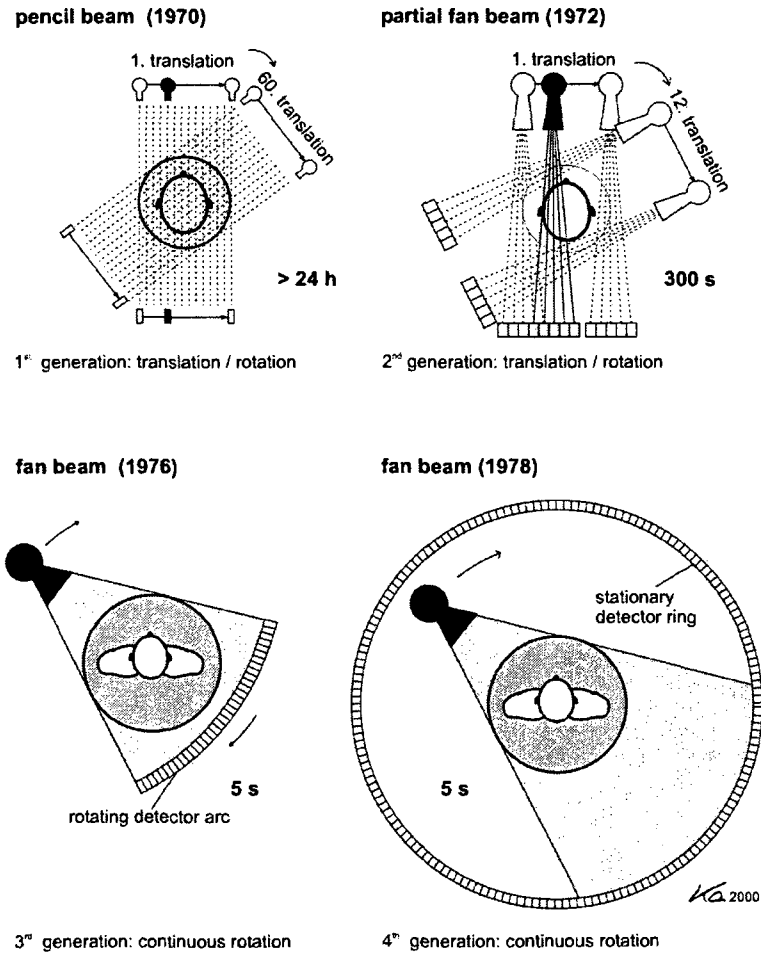


FIG. 2.2: Différentes générations de scanners d'après Kalender (2005).

Une tranche de résolution semblable à celle des premières générations est alors parcourue en quelques minutes (entre 3 et 5 min). Ces scanners étaient uniquement utilisés pour imager la tête. Des essais portant sur l'imagerie du corps entier furent réalisés. L'objectif est alors de réduire le temps d'acquisition d'une tranche en dessous des 20 secondes pour réaliser des images du corps entier pendant une seule respiration. L'introduction de la troisième génération répond à cet objectif.

2.2.3 Troisième génération : faisceau large de rayons X

En 1976, les tomographes de troisième génération sont mis au point. Leur principe de fonctionnement est complètement différent de celui des générations précédentes. Les translations du système source-détecteurs sont supprimées. L'introduction d'un faisceau de rayons X en éventail ainsi qu'un arc de détecteurs plus large ont autorisé la mesure d'une projection complète simultanément (Fig. 2.2 (c)). Cette nouveauté simplifie grandement le mécanisme de l'appareil. Le portique source-détecteurs se déplace alors uniquement selon un mouvement de rotation continu autour de l'objet.

2.2.4 Quatrième et cinquième générations

Avec les tomographes de quatrième génération (1978), la mécanique est encore simplifiée. Le système de détection est constitué d'un anneau de détecteurs entourant entièrement l'objet (Fig. 2.2 (d)). De ce fait, seule la source tourne autour de l'objet. Cependant, ces scanners sont peu utilisés dans le milieu clinique car ils sont plus coûteux que ceux de troisième génération et ne présentent qu'une amélioration non significative de la qualité des images.

La source des scanners de cinquième génération ne fait plus partie du portique rotatif. Elle est retirée de ce portique et le faisceau de rayons X se reflète sur des anodes régulièrement espacées sur le portique. Pour les mêmes raisons que la quatrième génération, ces scanners sont peu utilisés.

Pour tous nos travaux, nous mettons l'accent sur les scanners de troisième génération, puisque l'hôpital Notre-Dame où nous avons réalisé toutes nos acquisitions est équipé principalement de scanners de cette génération.

Cette génération a cependant beaucoup évolué dans les années 1990 avec la technologie

« slip ring ». En effet, en 1976 les scanners ne pouvaient effectuer qu'une rotation à la fois, car les câbles alimentant la source devaient être rembobinés avant toute nouvelle rotation. Dans la technologie « slip ring », des balais glissent sur des anneaux métalliques, fournissant ainsi de l'énergie électrique au tube producteur de rayons X, ce qui permet une rotation continue sur plus d'un tour à la fois. Cette technologie est à l'origine des scanners hélicoïdaux dans lesquels la table se déplace de manière continue pendant que le portique source-détecteurs décrit plusieurs rotations autour de lui. Ainsi lorsqu'on se place dans un référentiel lié à la table, le portique suit une trajectoire hélicoïdale.

Depuis 1990, nous assistons à une augmentation du nombre de rangées de détecteurs (barrettes) réduisant davantage le temps d'acquisition de grands volumes anatomiques. Actuellement les derniers scanners commercialisés comportent 256 barrettes de détecteurs.

De plus, pour accroître la précision des images reconstruites, Siemens a développé, en 2004, une technologie de source volante appelée FFS (*Flying Focal Spot*, (Kachelriess et al., 2006)). Cela consiste à créer à partir d'un seul tube radiogène plusieurs points de source virtuels en faisant dévier le faisceau d'électrons touchant l'anode. Cette technologie permet d'acquérir des images de résolution spatiale équivalente à 512 barrettes mais avec uniquement 256 barrettes.

Toutes ces évolutions technologiques et/ou techniques ont eu pour objectifs de réduire les temps d'acquisition de façon considérable et d'améliorer la qualité des images reconstruites afin de réaliser de meilleurs diagnostics médicaux.

D'autres évolutions portant sur le système de détection ont permis d'affiner la qualité des images reconstruites, notamment le choix de détecteurs en fonction de leur temps de réponse et leur efficacité de détection.

2.3 Systèmes de détection

Le système de détection est l'une des parties les plus critiques, technologiquement parlant, de tout le scanner CT. Il est chargé de transformer l'intensité photonique incidente en un signal électrique correspondant puis de l'amplifier. Ses composants principaux sont les détecteurs sensibles à l'intensité photonique, leur configuration géométrique et les pré-amplificateurs. Dans cette partie, nous nous concentrerons uniquement sur les détecteurs et leur configuration.

En tomographie, deux méthodes de conversion de l'intensité photonique en signaux électriques sont prédominantes (Kalender, 2005) :

- conversion par une chambre d'ionisation, remplie par un gaz noble (le xénon) sous haute pression (Fig.2.3) ;
- par des détecteurs de scintillation sous forme de cristaux, comme l'iodure de césium ou le tungstate de cadmium ou encore des matériaux céramiques tels que l'oxysulfure de gadolinium (Fig.2.4).

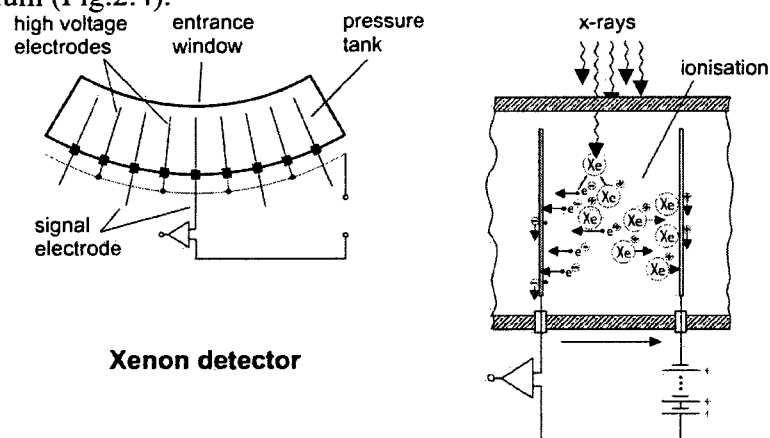


FIG. 2.3: Détecteur au xénon (Kalender, 2005).

Les détecteurs de scintillation sont composés de petits blocs de matériaux de scintillation ($CdWO_4$, Gd_2O_2S , ...) recouverts d'un matériau réfléchissant et couplés à des photodiodes. Un schéma du fonctionnement se trouve dans la figure 2.4. Un photon incident

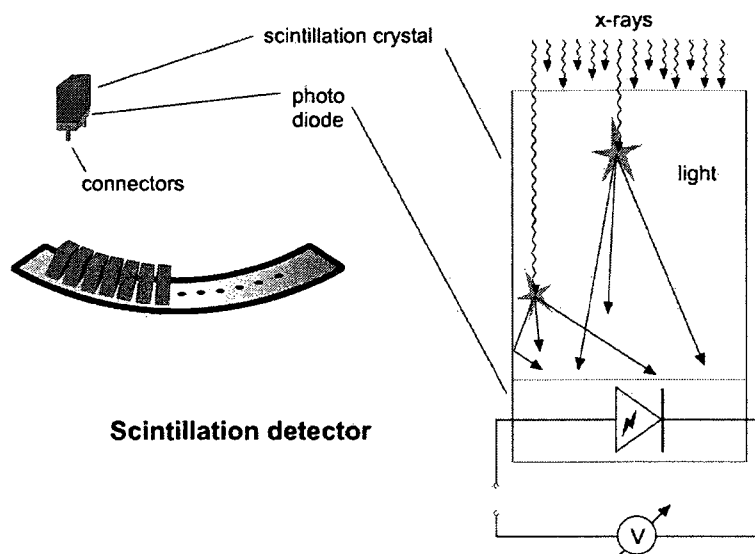


FIG. 2.4: Détecteur de scintillation (Kalender, 2005).

provoque une interaction photoélectrique avec le scintillateur. Le photoélectron relâché par cette interaction traverse le scintillateur. Les photons de lumière produits par la scintillation se dispersent dans toutes les directions. Ainsi, un matériau réflecteur permet de diriger les photons lumineux vers les photodiodes, ce qui produit un signal électrique mesurable.

Les détecteurs à chambre d'ionisation fonctionnent aussi sur un principe d'interaction photoélectrique. En effet, lorsqu'un photon de rayons X frappe une chambre d'ionisation, on assiste à l'ionisation du xénon par une interaction photoélectrique. Cette interaction relâche des photo-électrons très énergétiques, qui ionisent à leur tour d'autres molécules de gaz. La chambre d'ionisation est composée de deux plaques métalliques (Fig. 2.3) sur lesquelles deux potentiels électriques (0 V et 500 V) sont appliqués. La plaque à 0 V collecte les noyaux de xénon ionisés et les électrons libres sont collectés sur l'autre plaque afin de créer un courant électrique mesurable.

Les détecteurs avec chambre d'ionisation au xénon offrent plusieurs avantages dont l'un est d'avoir une réponse temporelle avec une rapide décroissance et peu de postlumines-

cence. Nous expliquerons cet aspect plus tard dans la section 2.6.2.

De plus, ils présentent une haute efficacité par rapport à la dose de rayons X envoyée. C'est pourquoi, ils furent considérés comme le meilleur choix de détecteurs dans les années 80.

La demande pour des temps de décroissance très faibles a été motivée par l'apparition des scanners à acquisition sub-seconde (c.-à-d. avec un temps de rotation inférieur à 1 s). Avec ces nouvelles générations de scanners sont apparus les détecteurs en céramique (UFC : Ultra Fast Ceramics) qui ont supplanté les détecteurs au xénon. En effet, ceux-ci possèdent un désavantage majeur : leur construction. Le principe de construction d'une barrette unique de détecteurs est relativement facile à réaliser. Cependant un système de détection multi-barrettes est extrêmement difficile. De ce fait, tous les scanners multi-barrettes ont été construits avec des cristaux de scintillation ou des matériaux céramiques.

2.4 Formation des mesures

Dans la partie précédente, nous avons exposé les principes de fonctionnement du système de détection des tomographes à rayons X. Ici, nous détaillons les modèles de formation des mesures à partir desquels les images tomographiques sont reconstruites. Les méthodes de reconstruction sont basées sur des approximations de ces modèles, comme nous le verrons dans la section suivante.

2.4.1 Interactions photon-matière

Les rayons X sont des ondes constituées de photons. Ces photons possèdent une énergie caractéristique inversement proportionnelle à sa longueur d'onde. Dans les tomographes

commerciaux, cette énergie est comprise entre 25 et 125 KeV (Fig. 2.5)

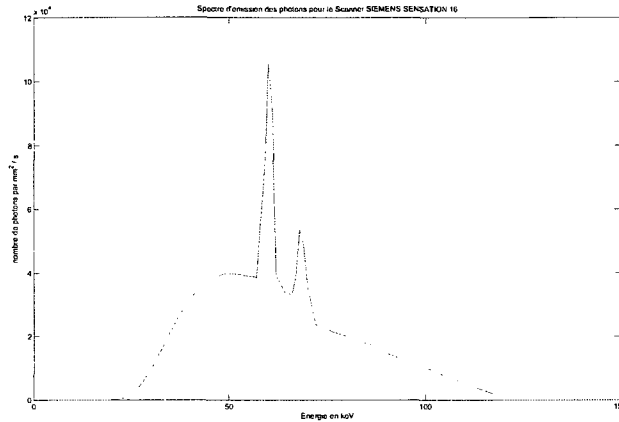


FIG. 2.5: Spectre de la source du scanner *Siemens SOMATOM Sensation 16*.

Les interactions entre ces photons et la matière sont à l'origine des mesures tomographiques. Certains photons X traversent la matière sans interaction avec celle-ci. Ils sont responsables de parties sombres de l'image tomographique. Les interactions entre la matière et le photon peuvent se classer selon trois catégories :

- *L'effet photoélectrique* : phénomène le plus important pour la formation d'images radiologiques. Il consiste en l'émission d'un électron par un atome lorsque celui-ci est frappé par un photon d'énergie supérieure à celle de liaison. Il se produit principalement pour des énergies faibles (50-80 KeV). Il forme ainsi les parties claires de l'image.
- *L'effet Compton* : dans ce cas, le photon interagit avec un électron. Lors de cette collision, l'électron gagne de l'énergie cinétique et le photon perd de l'énergie et est dévié. Il se produit sous haute énergie.
- *La production de paires* : un photon de haute énergie entre en interaction avec le noyau d'un atome. Ce photon est entièrement absorbé et le noyau émet un électron et un positron (anti-électron, c.-à-d. un électron chargé positivement).

Les scanners *Siemens SOMATOM 16* et *64* émettent des photons à des énergies com-

prises entre 25 et 125 keV comme le montre la figure 2.5. À ces énergies, les effets photoélectriques et Compton sont prédominants.

Dans les trois interactions que nous venons de voir, un photon incident selon une direction a peu de probabilités d'émerger de la matière dans la même direction. Lorsque le photon incident est dévié, il est généralement écarté du système de détection grâce à un collimateur situé à l'avant des détecteurs. Ces photons déviés seront modélisés par du bruit de diffusion puisqu'il est impossible d'assurer leur traçabilité.

Tous ces phénomènes sont à l'origine de deux types de modèles du processus d'émission et de production des rayons X : le modèle déterministe et le modèle statistique. Dans les deux sous-sections suivantes, nous nous attacherons à comprendre les différentes hypothèses que sous-tend chacun de ces modèles.

2.4.2 Modèle déterministe

Ce modèle est basé sur deux approximations :

- Le faisceau de rayons X est infiniment fin et se propage en ligne droite ;
- L'émission des photons est déterministe.

Considérons un faisceau de rayons X comportant N_0 photons et traversant un matériau homogène mince d'une épaisseur w . Le nombre moyen de photons absorbés (dN) par le matériau est donné par la relation suivante :

$$dN = \mu N_0 w \quad (2.1)$$

dans laquelle la variable μ représente le coefficient d'atténuation linéique du matériau en question, que l'on appellera simplement par la suite coefficient d'atténuation, exprimé en cm^{-1} . Ce coefficient est propre à chaque matériau et varie aussi en fonction de l'énergie ; il permet de modéliser les trois phénomènes vus dans la section précédente.

Ainsi, La loi de *Beer-Lambert* permet de quantifier le nombre de photons mesurés aux détecteurs :

$$N = N_0 e^{-w\mu} \quad (2.2)$$

Cependant, le corps humain n'étant pas constitué d'un seul et unique matériau, on aboutit alors à une expression intégrale plus complexe du nombre de photons mesurés :

$$N = N_0 e^{-\int_0^w \mu(w) dw} \quad (2.3)$$

Cette relation, à la base de nombreux algorithmes de reconstruction, signifie que le nombre de photons émergents suit une loi exponentielle. Par cette approche, nous avons considéré que le nombre de photons incidents est connu (déterministe), de valeur N_0 . Dans la section suivante, nous allons relaxer cette hypothèse et modéliser de manière statistique la production de rayons X.

2.4.3 Modèle statistique

Considérons les mêmes hypothèses de rayons fins et de monochromaticité des photons. Supposons maintenant que la production de rayons X est une variable aléatoire qui suit une loi de Poisson de paramètre N_0 :

$$N \sim \mathcal{P}(N_0) \quad (2.4)$$

Selon Macovski (1983), la combinaison de la génération des rayons X (modélisée par une loi de Poisson) et de leur atténuation (modélisée de la même manière que dans la section précédente) peut être aussi décrite par une loi de Poisson.

Ainsi le nombre de photons moyens détectés est une variable aléatoire de Poisson :

$$N \sim \mathcal{P}(N_0 e^{-\int_0^W \mu(w) dw}) \quad (2.5)$$

Ce modèle présente l'avantage d'être plus flexible que le précédent. En effet, dans celui-ci, il est possible d'inclure le caractère polychromatique des rayons X ainsi que l'épaisseur des rayons. Ainsi, ce modèle offre la possibilité d'être plus réaliste que le précédent. Dans la section suivante, nous abordons la reconstruction des données à partir d'un des modèles développés ci-dessus.

2.5 Reconstruction

Notre projet porte sur le traitement des sinogrammes avant la reconstruction. Une façon d'observer une amélioration quelconque de ce traitement est de la quantifier sur les images reconstruites. C'est pourquoi nous présentons dans cette section les différentes approches de reconstruction.

La reconstruction permet d'obtenir, à partir des sinogrammes, des images tomographiques représentant la distribution spatiale des coefficients d'atténuation $\mu(x, y)$.

Deux familles d'algorithme se distinguent :

- Les reconstructions *analytiques* (ex : la rétroprojection filtrée) : elles utilisent des techniques d'inversion exactes qui doivent ensuite être discrétisées et elles sont basées sur un modèle déterministe de formation de données. Cette technique est principalement utilisée dans les tomographes médicaux actuels.
- Les reconstructions *algébriques* : dans celles-ci, on considère les pixels de l'image à reconstruire comme des inconnues, et un système d'équations, possiblement non linéaire, est établi entre ces inconnues et les projections mesurées. Ce système est ca-

ractérisé par le modèle choisi de formation des données (déterministe ou statistique). Il sera ensuite inversé itérativement afin de déterminer la valeur des inconnues. Nous présenterons les hypothèses et les fondements mathématiques de chacune de ces méthodes dans les sections suivantes. Pour de plus amples détails sur les calculs mathématiques, le lecteur peut se référer au livre de Kak et Slanley (1987).

2.5.1 Méthodes analytiques

Les méthodes analytiques sont la base des algorithmes de reconstruction des toмоgraphes commerciaux. Elles s'appuient sur les hypothèses théoriques suivantes :

- Les mesures sont continues. Cette hypothèse n'est pas respectée à la lettre puisque nous possédons uniquement un nombre limité de mesures. Cependant, si l'échantillonnage des mesures est dense, alors on peut la considérer comme vérifiée.
- Les hypothèses de la loi de Beer-Lambert sont valides, c.-à-d. rayons X monochromatiques, infiniment minces et émission des rayons X *déterministe* (c.f. 2.4.2).

Ces hypothèses nous permettent de bâtir un cadre théorique pour la reconstruction analytique. Commençons par introduire les notations et les outils mathématiques.

2.5.1.1 Transformée de Radon et sinogramme

Plaçons-nous dans le cadre théorique de la figure 2.6. Chaque rayon X fait un angle θ avec l'axe des ordonnées et est à une distance r de l'origine. L'objet à imager est caractérisé complètement par la répartition spatiale des coefficients d'atténuation $\mu(x, y)$.

Définissons la projection $p_\theta(r)$ de notre objet suivant l'angle θ par :

$$p_\theta(r) = \int_{L_{\theta,r}} \mu(x, y) ds \quad (2.6)$$

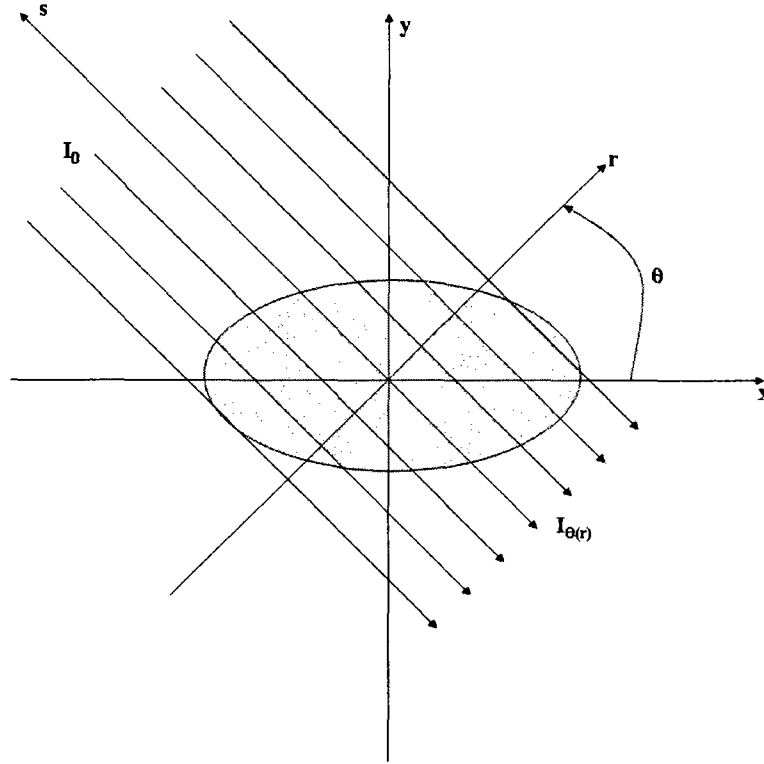


FIG. 2.6: Faisceau de rayons parallèles et son système de coordonnées.

où $L_{\theta,r}$ désigne la droite orientée parallèlement à l'axe s et à une distance r de celui-ci.

L'ensemble des projections $p_{\theta}(r)$ est appelé *sinogramme*. Un exemple de sinogramme pour un fantôme classique en tomographie (le fantôme de Shepp Logan) est tracé sur la figure 2.7. Ce fantôme est composé de plusieurs ellipses de matériaux différents permettant de simuler les différents tissus du cerveau.

En mathématiques, on appelle *transformée de Radon*, la transformation qui à toute distribution $f(x, y)$ associe son sinogramme :

$$p_{\theta}(r) = \mathcal{R}(f(x, y)) \quad (2.7)$$

$$= \int_{-\infty}^{+\infty} f(r \cos \theta - s \sin \theta, r \sin \theta + s \cos \theta) ds \quad (2.8)$$

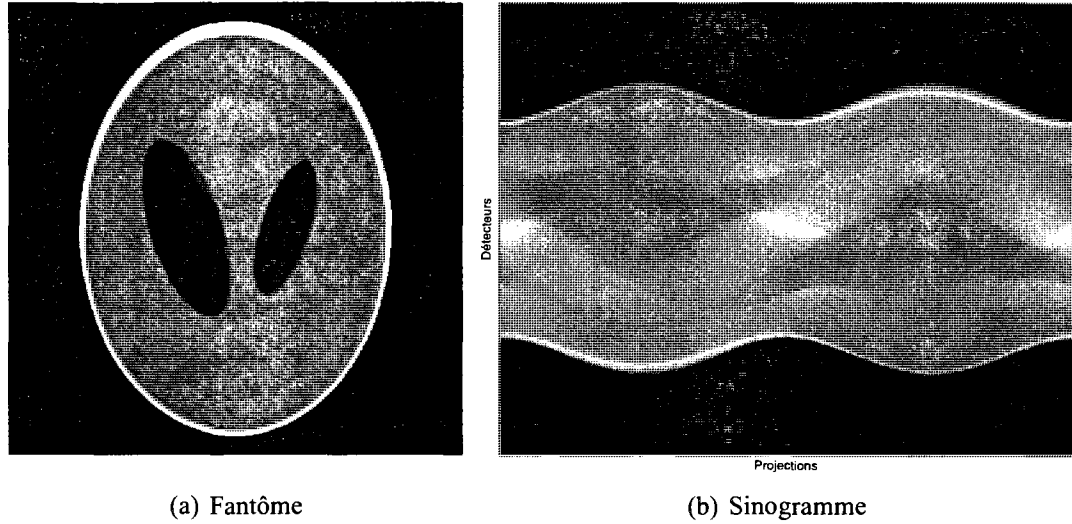


FIG. 2.7: Fantôme Shepp Logan et son sinogramme associé.

Dans les années 1920, Radon formule cette transformée de manière analytique valable uniquement pour les géométries à rayons parallèles. Une extension de cette transformée de Radon existe pour les géométries en éventail (Kak and Slaney, 1987), mais nous ne la présenterons pas par souci de simplicité.

Les méthodes analytiques reposent sur l'inversion de cette transformée de Radon. C'est ce que nous verrons dans la section suivante.

2.5.1.2 Théorème de la tranche centrale de Fourier et inversion directe

Le théorème de la tranche centrale de Fourier permet d'établir une relation entre les projections $p_{\theta}(r)$ et l'objet $f(x, y)$ quel que soit θ .

Théorème 2.5.1 (Tranche centrale de Fourier). *Soit $F(u, v)$, la transformée de Fourier de $f(x, y)$:*

$$F(u, v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) e^{-2\pi i(ux+vy)} dx dy$$

et $P_\theta(\rho)$, la transformée de Fourier de $p_{\theta(r)}$:

$$P_\theta(\rho) = \int_{-\infty}^{+\infty} p_{\theta(r)} e^{-2\pi i \rho r} dr, \text{ où } \rho = \sqrt{u^2 + v^2}$$

alors

$$P_\theta(\rho) = F(\rho \cos \theta, \rho \sin \theta) \quad (2.9)$$

Ainsi, d'après ce théorème, la transformée de Fourier d'une projection d'angle θ forme une ligne du plan de Fourier de notre objet à cet angle-là. Pour θ variant entre 0 et π , tout le plan de Fourier de f est bien couvert par les transformées de Fourier des p_θ . Donc la transformée de Fourier de f peut être connue en tout point (x, y) d'un repère cartésien.

Cette méthode est très peu utilisée dans la pratique, car elle nécessite une transformée de Fourier inverse 2D. On utilise plus couramment la rétroprojection filtrée.

2.5.1.3 La rétroprojection filtrée

La méthode de rétroprojection dite filtrée a été développée à partir du même théorème. Il est possible de réécrire l'équation 2.9 de façon à obtenir f qu'à partir de transformées de Fourier inverses 1D moins coûteuses.

Elle repose sur l'équation suivante :

$$f(x, y) = \int_0^\pi p^*(x \cos \theta + y \sin \theta, \theta) d\theta \quad (2.10)$$

dans laquelle

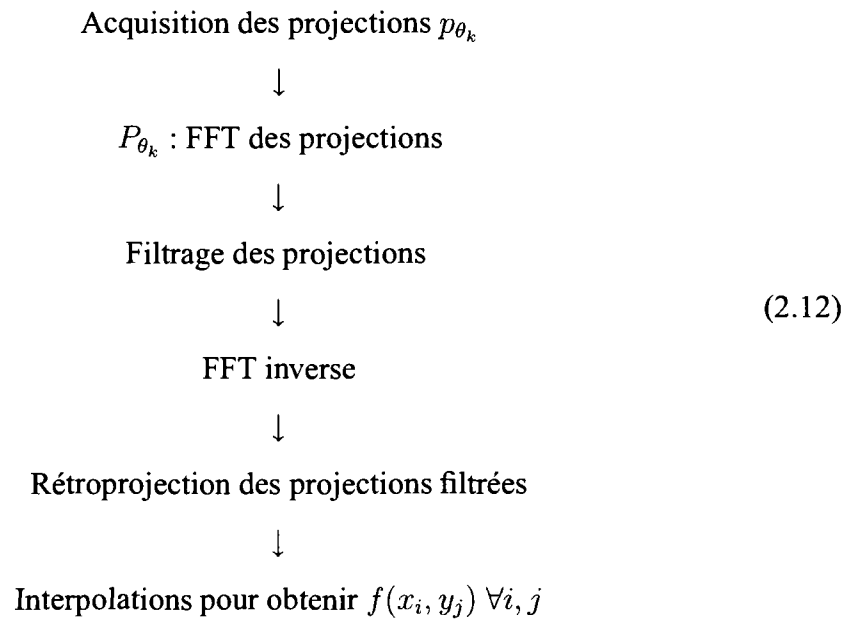
$$p^*(r, \theta) = \int_{-\infty}^{+\infty} |\rho| P(\rho, \theta) e^{2\pi i \rho r} d\rho, \text{ avec } r = x \cos \theta + y \sin \theta \quad (2.11)$$

Donc notre objet $f(x, y)$ peut être obtenu en *rétroprojetant* (c.-à-d. en épandant) les

projections filtrées $p^*(r, \theta)$ par un filtre rampe suivant tous les angles θ .

La rétroprojection filtrée, telle que présentée ci-dessus, est très pratique puisque l'inversion est analytique et peu coûteuse. Cependant, ces formules ne sont exactes que pour des fonctions dépendants de variables p_θ continues dans l'intervalle $[0; 2\pi[$. Les tomographes permettent d'acquérir uniquement des projections régulièrement espacées pour un nombre fini d'angles. De même, le système de détection contient un nombre fini de détecteurs. Ainsi, toutes les données acquises sont discrètes ; il serait donc illusoire de vouloir reconstruire la fonction f de manière continue. On ne reconstruit f que sur une grille discrète de taille finie. Cette grille nous amène à réaliser des interpolations afin d'attribuer un coefficient d'atténuation à chaque couple (x_i, y_i) . Ces interpolations sont responsables d'une détérioration de la qualité des reconstructions.

Les étapes de la méthode de rétroprojection filtrée sont résumées de la façon suivante :



De ce fait, la rétroprojection filtrée ne traite que des quantités 1D, rendant les calculs moins laborieux que pour l'inversion directe par le théorème de la tranche centrale de

Fourier. De plus, il est possible de traiter les projections en temps réel. Directement après l'acquisition d'une projection, il est possible de la filtrer puis de la rétroprojeter.

Dans l'équation 2.11, le filtre $\rho = |\rho|$ peut être remplacé par d'autres filtres qui permettent, par exemple, de réduire le bruit sur les reconstructions, au prix d'une détérioration de la résolution spatiale de l'image. Parmi ceux-ci, citons les plus couramment utilisés : les filtres de Ramachandran et Lakshminarayanan (RAMLAK), de Shepp Logan, de Hamming. Tous ces filtres sont représentés sur la figure 2.8.

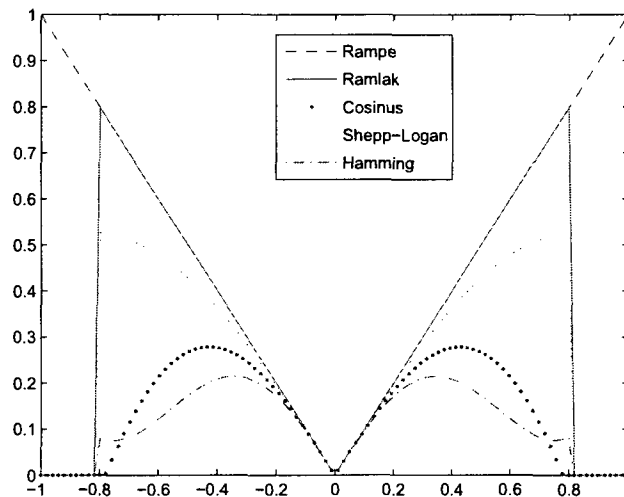


FIG. 2.8: Filtres utilisés pour la rétroprojection.

2.5.1.4 Extensions et limites

Les formules développées précédemment ne sont valides que dans le cas 2D et pour des rayons parallèles. Certains auteurs ont développé des algorithmes pour le cas 3D (Pan et al., 2004; Stierstorfer et al., 2004) et pour des géométries avec un faisceau en cône (Feldkamp et al., 1984; Tuy, 1983) et une trajectoire de la source hélicoïdale (Katsevich, 2002; Pan et al., 2004).

Ces techniques sont souvent très rapides et très efficaces. C'est pour cela qu'elles sont actuellement employées dans l'industrie. Cependant, leur modèle de formation des données est déterministe et rigide. Il s'adapte difficilement, en particulier lorsqu'on veut s'affranchir de la monochromaticité des rayons X, notamment pour la visualisation d'objets métalliques de petite taille.

Les manufacturiers s'affranchissent de ces effets en appliquant des corrections *a posteriori* sur les données et sur les reconstructions. Mais ces corrections ne résolvent qu'en partie les erreurs dues à la modélisation de la formation de données.

Les algorithmes dits *algébriques* ont été développés dans le but d'adapter le modèle de formation des données à des réalités pratiques. Ils permettent d'intégrer, entre autres, la polychromaticité des rayons X, un modèle statistique d'émission des rayons X et des imperfections du portique source-détecteurs (rotation non continue, ...).

Nous expliquerons dans la section suivante les principes de ces méthodes algébriques.

2.5.2 Méthodes algébriques

Les méthodes algébriques sont apparues dans les années 1980. Elles présentent l'avantage d'être souples par rapport au modèle de formation des données.

Dans cette section, nous nous plaçons dans un cas 2D, plus propice au développement des équations.

2.5.2.1 Discrétisation du problème

Dans les méthodes analytiques, nous supposons l'objet continu et le nombre de projections infini. Ce n'est évidemment pas le cas dans la réalité. Nous allons introduire dans

cette section les notations concernant la discrétisation de l'objet, des projections et des détecteurs.

Supposons que notre objet $f(x, y)$ est une combinaison linéaire de fonctions de base. Ici, nous choisirons comme fonctions de base le pixel carré. Il existe d'autres bases de décomposition (par exemple, le pixel circulaire (Allain et al., 2002)), mais nous nous limiterons au pixel carré, qui est le plus couramment utilisé. Ainsi,

$$f(x, y) = \sum_{j=1}^J f_j b_j(x, y) \quad (2.13)$$

où f_j est la valeur du pixel j de l'image.

L'image que nous cherchons à reconstruire dans sa forme conventionnelle peut être assimilée à une matrice 2D des coefficients d'atténuation de taille $N \times N$. Cependant, afin de simplifier les équations et les calculs suivants, nous noterons notre image : f , vecteur de taille $(N^2, 1)$.

$$f = \begin{pmatrix} f_1 \\ \vdots \\ f_{N^2} \end{pmatrix}$$

Avec ces notations, la valeur d'une projection devient alors :

$$p_j = \sum_{i=1}^{N^2} A_{i,j} f_i \text{ pour } j = 1, 2, \dots, P \quad (2.14)$$

avec P , le nombre total de projections. Pour un scanner comportant N_p projections de chacune N_d détecteurs, $P = N_p \times N_d$.

En prenant en compte toutes ces équations pour chacune des projections p_j , nous pou-

vons écrire l'équation précédente sous une forme matricielle :

$$p = A\mu \quad (2.15)$$

où

- p est le vecteur mesure de taille $P = N_p \times N_d$
- μ est le vecteur image recherché de taille N^2
- la matrice $A = (a_{i,j})$, appelée *matrice de projection*, constitue la discrétisation de l'opération de projection de la transformée de Radon. Elle dépend uniquement du processus d'acquisition et donc peut être calculée une fois pour toute. Ces coefficients $a_{i,j}$ sont caractéristiques de l'intersection du rayon i avec le pixel j . Des études menées sur l'influence d'un rayon mince ou épais sur la reconstruction n'ont pas montré l'avantage de l'une ou l'autre de ces méthodes. Par conséquent, nous retiendrons l'approche rayons minces. Par cette approche, les coefficients $a_{i,j}$ correspondent à la longueur de l'intersection du rayon i avec le pixel j .

Ainsi, la reconstruction algébrique consiste à résoudre ce système d'équations. L'un des enjeux majeurs des méthodes algébriques est la taille que prend ce système. En effet, en 2D, pour une géométrie type de $N_p = 2320$ projections, $N_d = 672$ détecteurs et une taille d'image de 512×512 , nous avons à résoudre un système de 1 559 040 équations à 262 144 inconnues. Bien que ce système soit surdéterminé d'un facteur 6, il reste difficile à résoudre.

Il s'agit alors d'estimer μ de la meilleure manière possible afin qu'il décrive de façon rigoureuse les données y dont nous disposons. Le chapitre 3 est consacré uniquement aux problèmes d'estimation. Nous n'entrerons donc pas dans les détails de la théorie de l'estimation dans cette partie.

Considérons les projections P et les coefficients d'atténuation U comme des variables aléatoires. L'estimation consiste alors à extraire μ , connaissant les mesures P , c.-à-d. à

chercher la loi conditionnelle de U , sachant que les mesures sont réalisées ($f_{U|P=p}(\mu)$) :

$$\hat{\mu} = \arg \max_{\mu} f_{U|P=p}(\mu) \quad (2.16)$$

$$= \arg \min_{\mu} -f_{U|P=p}(\mu) \quad (2.17)$$

$$= \arg \min_{\mu} f(\mu) \quad (2.18)$$

On appelle critère la fonction $f(\mu)$.

2.5.2.1.1 Cadre gaussien et critère des moindres carrés La production de rayons X est un phénomène qui suit la loi de Bernoulli. Pour un grand nombre de photons émis, on peut assimiler le nombre de photons émergents à une loi normale.

Chaque projection p_j devient ainsi une réalisation d'un processus gaussien de moyenne $[A\mu]_j$ et de variance σ_B^2 :

$$p_j \sim \mathcal{N}\{[A\mu]_j, \sigma_B^2\} \quad (2.19)$$

Considérant qu'aucune information *a priori* n'est apportée sur la solution, on obtient un critère de minimisation qui prend la forme suivante :

$$f(\mu) = \frac{1}{2}(p - A\mu)(p - A\mu)^T \quad (2.20)$$

Ce critère correspond aux moindres carrés linéaires et est donc quadratique.

2.5.2.1.2 Cadre poissonnien Un autre choix de critère $f(\mu)$ est le critère poissonnien. Il est basé sur une modélisation plus fidèle de la réalité de production des rayons X.

Comme nous l'avons vu dans la section 2.4.3 sur la formation des données :

$$N_j \sim \mathcal{P}(N_0 e^{-\int_0^w \mu(w) dw}) \quad (2.21)$$

Le critère de minimisation qui lui est associé prend alors la forme suivante, pour l'ensemble des projections N_j :

$$f(\mu) = \sum_j N_0 e^{-(A\mu)_j} + [A\mu]_j N_j \quad (2.22)$$

Ce critère correspond au maximum de la fonction de vraisemblance entre le modèle poissonnien de formation des données et les données réelles.

En comparant ce critère à celui des moindres carrés (eq. 2.20), on remarque qu'il est beaucoup plus complexe car il n'est plus quadratique ! Cependant, le modèle de formation de données étant plus juste, il donne une meilleure adéquation aux données.

2.5.2.1.3 Régularisation Le problème de reconstruction tomographique est connu pour être *mal conditionné* (voir section 3.4). Un effet visuel de ce mauvais conditionnement est la qualité des images reconstruites. Un mauvais conditionnement est aussi responsable de problèmes calculatoires lors de la minimisation. La vitesse de convergence des algorithmes de minimisation est beaucoup plus lente, car ceux-ci doivent réaliser plus d'itérations.

Il est possible d'améliorer le conditionnement de notre problème de reconstruction en ajoutant une information *a priori* à notre critère de minimisation (Tikhonov and Arsenin, 1976). C'est ce qu'on appelle la régularisation (cf. 3.4). Le critère s'écrit alors

$$\hat{\mu} = \arg \min_{\mu} f(\mu) + \lambda R(\mu) \quad (2.23)$$

où λ est un scalaire positif qui permet de fixer l'importance de l'information *a priori* par rapport à la fidélité aux mesures. Si λ est nul, on revient au critère initial où l'on n'apporte aucune information. S'il est grand, la solution de notre minimisation dépendra fortement de nos *a priori*.

L'information *a priori* peut, par exemple, porter sur la régularité de l'image reconstruite, l'amplitude des coefficients d'atténuation ou encore la positivité des coefficients d'atténuation. Plusieurs fonctions de régularisation seront développées dans le chapitre 3.

2.5.2.1.4 Choix du critère Le choix du critère dépend du compromis entre qualité et temps de calcul. En effet, minimiser un critère qui n'est pas quadratique est plus coûteux en temps de calcul, puisque chaque itération de minimisation de ce critère demande le calcul supplémentaire de l'exponentiel d'une matrice. Cependant, la qualité des images reconstruites est meilleure.

Pour les reconstructions algébriques futures, nous utiliserons le critère des moindres carrés régularisé qui nous donne une solution acceptable pour des temps de calculs substantiellement plus faibles que pour le critère poissonien :

$$\hat{\mu} = \arg \min_{\mu} \frac{1}{2}(p - A\mu)(p - A\mu)^T + \lambda R(\mu) \quad (2.24)$$

Une fois le modèle de formation de données choisi, et donc le critère à minimiser établi, il ne reste plus qu'à réaliser la minimisation proprement dite. La présentation des différents algorithmes de minimisation est reportée dans la partie 3.6.2.

2.5.3 Conclusion

Dans cette section, nous avons décrit deux approches de la reconstruction d'images tomographiques.

L'approche analytique présente l'avantage d'être rapide et peu chère en espace mémoire. En effet, dans cette approche, il n'est pas nécessaire de stocker de grosses matrices, comme la matrice de projection. Cependant, cette technique est basée sur des hypothèses fortes non (ou seulement en partie) vérifiées en pratique. De plus, cette approche est basée sur une formulation continue du problème, contrairement à l'approche algébrique qui prend en compte directement le caractère discret de notre problème dans son modèle. L'approche algébrique permet aussi de prendre en considération, par une interprétation probabiliste du problème, la nature des fluctuations statistiques affectant les données. Ensuite, il est possible d'introduire des connaissances *a priori* sur notre modèle au moyen de techniques de régularisation. Cependant, les temps de calculs et l'espace mémoire se voient augmentés dans une forte proportion.

Ainsi, l'approche algébrique possède plus d'avantages théoriques que l'approche analytique au prix d'un temps de calcul significativement plus élevé.

Dans le projet général dans lequel mes travaux s'intègrent, notre équipe s'intéresse à reconstruire des images dans lesquelles il y a présence d'objets métalliques. Dans ce cas, les méthodes analytiques montrent leurs limites. Le cadre théorique de la reconstruction algébrique permettrait de limiter ces artefacts.

Dans la section suivante, nous allons aborder de manière générale les artefacts de reconstruction sur les images tomographiques.

2.6 Artefacts

Les images obtenues par les différents algorithmes de reconstruction doivent être observées afin de vérifier leur fidélité par rapport à l'objet réel. En effet, le système d'imagerie peut produire des structures artificielles, que nous appelons *artefacts*. Lors de l'analyse d'acquisitions sur des données de patients, il n'est pas toujours facile de discerner ce qui est correct de ce qui ne l'est pas. Cet exercice requiert une expérience énorme de la part du praticien ainsi qu'une bonne connaissance du comportement des systèmes d'imagerie.

Nous développons, dans le paragraphe suivant, quelques causes d'artefacts dans les images tomographiques, comme le mouvement du patient, le durcissement de rayons, les radiations afocales, les effets de volume partiel, les implants métalliques, les erreurs de discrétisation et de champ de reconstruction (c.-à-d. que le patient excède les limites du champ de mesures). Le livre de Kalender (2005) détaille plus amplement ces différents artefacts.

Nous nous intéressons ensuite à une dégradation de l'image causée par les caractéristiques du système de détection, soit la postluminescence, artefact sur lequel porte l'essentiel de mes travaux de maîtrise.

2.6.1 Quelques artefacts en tomographie

Prenons le cas des artefacts causés par le mouvement du patient (Fig. 2.9 b)). En tomographie, les mouvements du patient ne causent pas uniquement un flou local des contours (contrairement à la radiographie), mais aussi des perturbations dans toute l'image. Compenser les contributions indésirables hors du pixel en question est uniquement possible si l'objet ne bouge pas ou ne change pas pendant l'acquisition, c.-à-d. lorsque les données

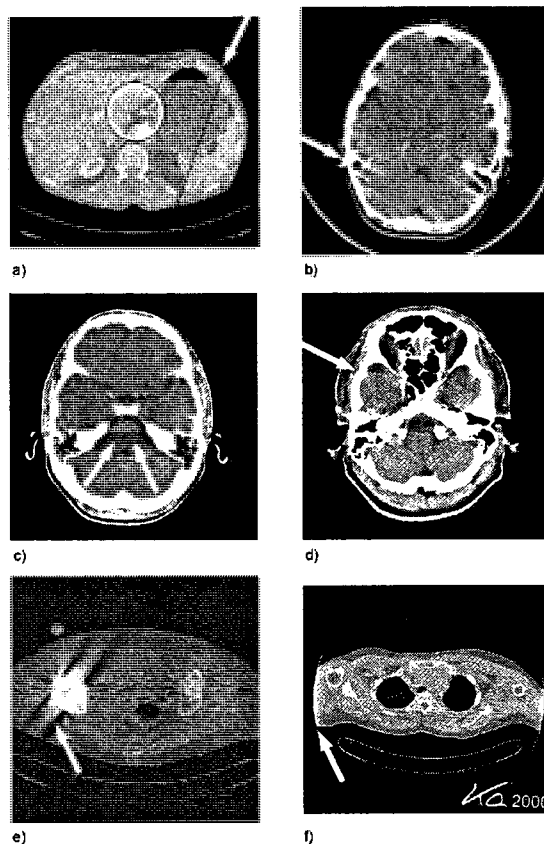


FIG. 2.9: Exemples typiques d'artefacts sur des images CT provenant a) de problèmes électroniques aux détecteurs ; b) du mouvement du patient ; c) du durcissement de rayons ; d) des effets de volume partiel ; e) d'implants métalliques ou f) du patient excédant les limites du champ de mesures d'après Kalender (2005).

des projections selon les différentes directions décrivent de manière consistante le même objet.

Cet artefact est une conséquence du système d'imagerie : les incohérences dans les données acquises causent toujours des perturbations dans toute l'image. Mais ces perturbations peuvent ne pas être immédiatement apparentes. Elles s'expriment parfois sous forme de valeurs reconstruites erronées, comme c'est le cas pour les artefacts dus au durcissement de rayons (Fig. 2.9 c)).

En général, on peut dire que les perturbations sur les images sont plus prononcées à proximité de leur site d'origine, mais elles peuvent encore affecter l'image à des plus grandes distances de ce site.

Le durcissement des rayons apparaît sous forme de zones ou de stries sombres dans les structures denses. Puisque la source émet un large spectre d'énergie, les rayons X sont atténués différemment en fonction de l'énergie de radiation, du type d'objet et de la direction de projection. Cela cause une augmentation variable de l'énergie moyenne du spectre quand ils rencontrent des objets durs, spécialement des structures osseuses. Pour les tissus mous, qui ressemblent à l'eau par leurs caractéristiques énergétiques, le durcissement des rayons peut être corrigé, mais une correction simultanée pour les os, l'eau et les contrastes de moyenne amplitude est difficile.

Les artefacts de volume partiel se présentent lorsque des structures de haut contraste s'étendent seulement partiellement dans la tranche examinée (Fig. 2.9 d)). La seule correction adéquate contre ces artefacts est l'acquisition de tranches minces. Cependant, ce type d'acquisition augmente le bruit dans chacune des fines tranches. Ainsi, il est préférable d'en additionner plusieurs, ce qui produira une tranche plus épaisse avec un niveau de bruit inférieur, mais cette fois-ci sans les artefacts.

Avec les implants métalliques, les artefacts de durcissement de rayon et de volume partiel sont intensifiés et peuvent complètement masquer le contenu de l'image dans le voisinage des objets métalliques (Fig. 2.9 e)). Certaines méthodes sont en cours de développement pour réduire le bruit et les artefacts métalliques (Hamelin et al., 2008)

Des artefacts peuvent aussi apparaître si une partie du patient est dans le portique source-détecteurs, mais positionné hors du champ de mesure. Ce phénomène mène à des zones d'hyperdensité au voisinage des régions périphériques (Fig. 2.9 f)). Ces artefacts peuvent être atténués par des corrections sur les mesures (Ohnesorge et al., 2000; Sourbelle et al.,

2005).

2.6.2 Postluminescence des détecteurs

Les images tomographiques sont aussi touchées par des artefacts dus aux propriétés des détecteurs captant les rayons X. Pour l'observation des sténoses artérielles, deux sources d'artefacts influencent la qualité des images : les artefacts métalliques et les artefacts causés par le système de détection. Les premiers ont fait l'objet d'un doctorat (Hamelin et al., 2008) dans lequel des solutions ont été apportées. Dans cette section, nous étudions l'artefact majeur causé par le système de détection : la postluminescence (Hsieh et al., 2000).

Depuis l'apparition des tomographes à acquisition sub-seconde, les détecteurs doivent capturer les rayons X de plus en plus rapidement. On définit la *postluminescence* des détecteurs par leur capacité à retourner à leur état stable après avoir été excités par un rayonnement X. Ce processus est exponentiel-décroissant avec une constante de temps caractéristique de la configuration de l'électronique sous-jacente. On appelle généralement cette constante de temps *vitesse primaire du scintillateur*.

À cause de la présence d'impuretés dans les cristaux de scintillation des détecteurs, un faible pourcentage des électrons excités est bloqué dans le scintillateur pendant une période relativement longue avant de retrouver leur état de stabilité. Ce phénomène provoque une décroissance plus lente, qui est modélisable par des multi-exponentielles décroissantes, conjointement avec la vitesse primaire du scintillateur.

Ainsi, une postluminescence des détecteurs et un temps de décroissance long indiquent une contamination importante du signal courant par les signaux précédents. Ce comportement est similaire au phénomène observé après avoir fermé la télévision dans une pièce obscure : l'écran de télévision ne revient pas au noir directement !

La figure 2.10 représente les caractéristiques de décroissance de plusieurs types de détecteurs. On comprend alors pourquoi les fabricants de tomographes ont adopté la technologie *UFC* (Ultra Fast Ceramics) après l'arrivée des scanners subseconde, puisqu'ils sont caractérisés par une décroissance plus rapide.

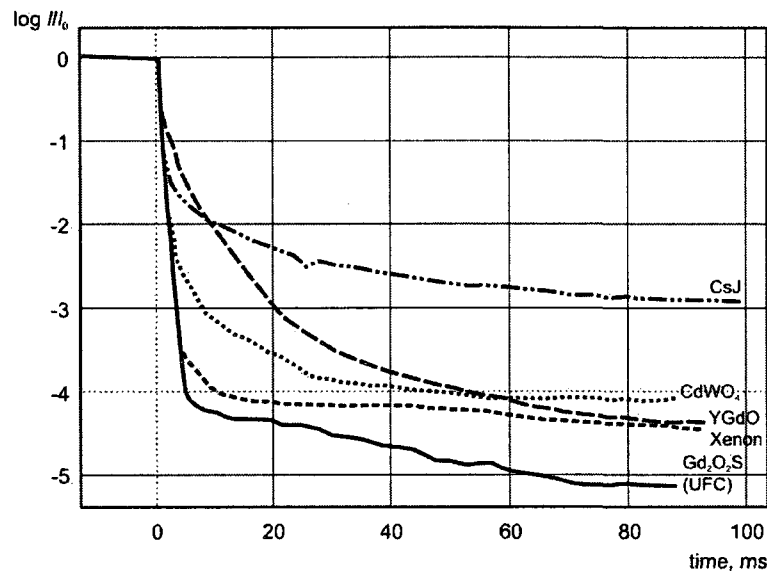


FIG. 2.10: Réponses de plusieurs détecteurs à un échelon : caractéristiques de décroissance (Hsieh et al., 2000).

Leurs effets sur la résolution spatiale et sur la qualité d'image sont importants, comme montré sur la figure 2.11. Cette figure présente le fantôme de performance de GE parcouru selon deux vitesses de rotation (0,5 s et 2,0 s) avec le même flux de rayons X, afin de réduire les autres facteurs qui pourraient biaiser les mesures. Ce fantôme est constitué d'une série de barrettes métalliques de largeur décroissante et d'un câble métallique au centre d'un disque de tissu mou. Il est clair sur cette figure qu'il y a une forte dégradation de la résolution spatiale de l'image reconstruite pour une vitesse de rotation de 0,5 s. En effet, les barres rectangulaires sont plus floues et ne sont plus tout à fait rectangulaires.

Afin de quantifier l'impact de la postluminescence sur la résolution spatiale, nous pouvons calculer la *MTF* (modulation transfer function) du câble. La MTF est la réponse fréquentielle d'un système d'imagerie. Cette fréquence est typiquement mesurée en *paire*

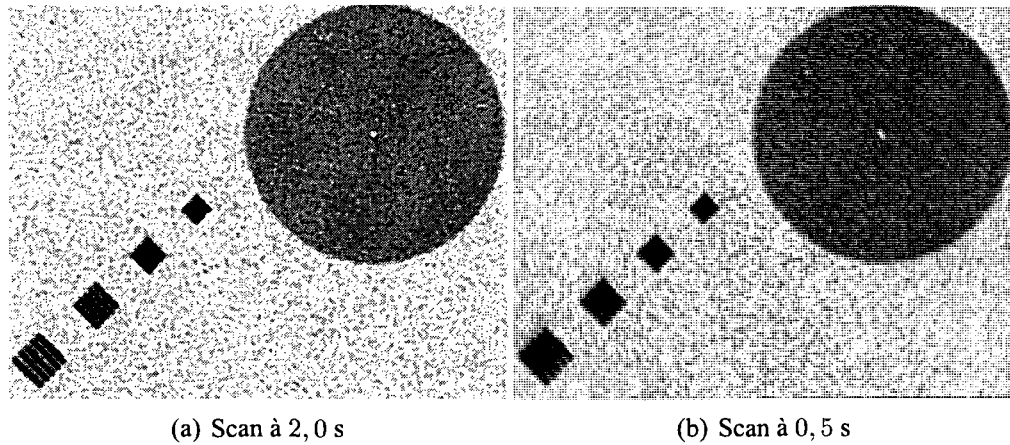


FIG. 2.11: Mise en évidence de la postluminescence des détecteurs en fonction de la vitesse de rotation d'après Hsieh et al. (2000).

de lignes par millimètre (lp/mm). Les hautes fréquences représentent les détails les plus fins d'une image. On cherche à obtenir une réponse instrumentale la plus proche possible de 1, même lorsque la fréquence spatiale est élevée. Pour le calcul de la MTF, la transformée de Fourier 1D de la projection du câble selon l'axe vertical a été calculée et son amplitude tracée en fonction de la fréquence spatiale.

Les résultats de la MTF du câble sont présentés sur la figure 2.12. La perte de résolution est manifeste puisque la courbe avec une vitesse de rotation de 0,5 s est en dessous de celle à 2,0 s.

2.7 Conclusion

Dans ce chapitre, nous avons décrit les différentes composantes qui permettent d'obtenir une image tomographique : des interactions photon-matière aux techniques de reconstruction des données.

Bien que la reconstruction par rétroprojection filtrée soit la méthode la plus utilisée en

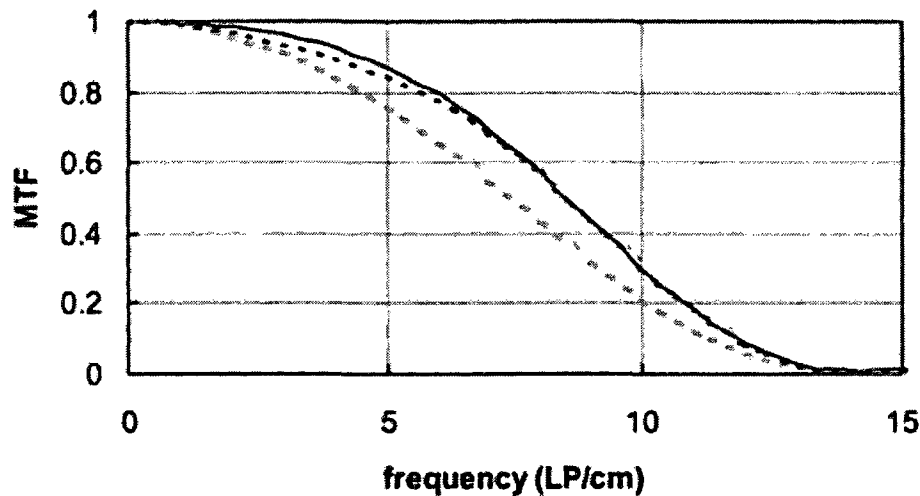


FIG. 2.12: MTF pour un câble en pointillé gris pour une vitesse de rotation de 0, 5 s de rotation (sans correction) ; en pointillé noir, de 2, 0 s ; et la courbe noire, de 0, 5 s (avec correction) d'après Hsieh (2000).

milieu clinique en raison de sa rapidité, les méthodes statistiques possèdent de nombreux avantages du point de vue de la flexibilité du modèle. De ce fait, ces dernières années, de nombreux travaux ont été réalisés sur ces techniques et sur les modélisations de la production de données. Cependant, peu de personnes se sont intéressées aux traitements des sinogrammes pour estomper l'effet de la postluminescence, malgré l'importance de ce phénomène dans les tomographes actuels.

Mes travaux de maîtrise se concentrent donc sur cet aspect du traitement des données, puisque la postluminescence est la dégradation majeure qui s'applique sur les données mesurées par les détecteurs. Nous verrons dans le chapitre *Postluminescence*, que cette dernière peut être modélisée par un système linéaire dont nous devons estimer les paramètres et restaurer l'entrée.

Le chapitre suivant traitera de la théorie de l'estimation tout d'abord dans un cadre général, puis dans le cas particulier où le système est linéaire et les inconnues gaussiennes.

CHAPITRE 3

ESTIMATION D'UN SYSTÈME LINÉAIRE

3.1 Introduction

Dans la plupart des systèmes de mesures d'un phénomène physique, l'accès direct à la grandeur souhaitée n'est pas possible. Il se fait par des observations indirectes, comme c'est le cas pour la postluminescence des détecteurs ainsi que pour certains problèmes de reconstruction tomographique.

Les techniques d'estimation ont pour but, à partir des observations, d'extraire l'information utile concernant la grandeur recherchée. Dans ce chapitre, les deux approches les plus utilisées en estimation sont décrites. Bien que développées par deux communautés de scientifiques, ces approches se rejoignent sur certains points.

Nous nous intéresserons uniquement aux systèmes linéaires. Les phénomènes physiques étudiés peuvent, en effet, être modélisés sous hypothèse linéaire avec de bons résultats. Puis, nous étudierons les principales caractéristiques des estimateurs de *maximum du vraisemblance* et du *maximum a posteriori*.

Certains problèmes d'estimation ne peuvent pas être résolus par des estimateurs sous forme explicite. Ainsi, nous exposerons les techniques itératives d'estimation permettant d'y parvenir.

3.2 Position du problème

Soit un système linéaire continu idéal défini par :

$$y = A(x) \tag{3.1}$$

où x et y sont des éléments d'espaces fonctionnels de dimensions infinies respectivement $\mathcal{X}$ et $\mathcal{Y}$ et $A : \mathcal{X} \mapsto \mathcal{Y}$ est un opérateur linéaire.

Dans le cas discret, x et y appartiennent à des espaces de dimension finie de taille N et M respectivement, et l'opérateur A devient une matrice de taille $M \times N$.

Ce système linéaire peut, par exemple, être une convolution entre un signal d'entrée et une réponse instrumentale, comme pour la postluminescence. L'équation 3.1 peut s'écrire sous forme de convolution, comme suit :

$$y(t) = (x * h)(t) \tag{3.2}$$

où h est la réponse impulsionnelle du dispositif de mesure. Sous cette formulation, on distingue deux types de problèmes : la restauration d'entrée (comme le traitement de la postluminescence et la reconstruction tomographique) et l'identification de paramètres.

Le but de la restauration d'entrée $x(t)$ est alors d'estimer x à partir des données y en supposant h connue.

Cependant, puisque le produit de convolution est commutatif, on peut aussi l'écrire de la manière suivante : $y(t) = (h * x)(t)$. Ainsi, cette équation nous permet d'estimer la réponse impulsionnelle h à partir des données y en supposant x connu. On parle alors d'identification des paramètres d'un système linéaire.

Formellement, ces deux problèmes sont de la même forme. Mais en pratique, ils diffèrent puisque le rapport nombre d'inconnues sur nombre de mesures est différent. Dans le cas de la restauration d'entrée, ce rapport est généralement petit, ce qui facilite l'estimation. Par contre, dans les problèmes d'identification de paramètres, ce rapport est plus grand que 1. De ce fait, le problème d'estimation devient plus complexe.

3.3 Approches « naïves »

Une solution évidente et naturelle de l'équation 3.1 est :

$$\hat{x}^{IG} = A^\dagger(y) \quad (3.3)$$

$A^\dagger$ est appelée *inverse généralisée* de A . Elle généralise l'inversion à des opérateurs qui ne sont pas carrés ou qui sont singuliers. Par définition, l'inverse généralisée de A doit satisfaire les quatre conditions suivantes (conditions de Moore-Penrose) :

1. $AA^\dagger A = A$;
2. $A^\dagger AA^\dagger = A^\dagger$;
3. $(AA^\dagger)^T = AA^\dagger$ et
4. $(A^\dagger A)^T = A^\dagger A$.

Sous ces conditions, l'inverse généralisée est unique. Pour une matrice A carrée et inversible, $A^\dagger = A^{-1}$. L'inverse généralisée est facilement calculable par décomposition en valeurs singulières ou par factorisation QR (Israel and Greville, 1973).

$\hat{x}^{IG}$ correspond à la solution de norme minimale du problème aux moindres carrés :

$$\hat{x}^{MC} = \arg \min_x \|y - A(x)\|^2 \quad (3.4)$$

Même si cette solution paraît de prime abord acceptable, elle n'est pas satisfaisante. En effet, pour l'obtention de cette solution, on néglige le bruit de mesures. Mais ce bruit se trouve souvent amplifié lors de l'inversion, même lorsque le rapport signal à bruit est relativement faible et les résultats sont dégradés.

Tous les problèmes classiques de la physique mathématique sont bien posés au sens d'Hadamard. On définit un problème bien posé au sens d'Hadamard lorsqu'il vérifie les conditions suivantes :

- Une solution existe ;
- La solution est unique ;
- La solution dépend de façon continue des données.

Bien que les problèmes directs de la physique soient bien posés, leurs problèmes inverses associés ne le sont généralement pas.

Pour un problème inverse mal posé, au sens d'Hadamard, une petite erreur δy sur les données ne se propage pas par une petite erreur δx sur la solution

$$|\delta y| \rightarrow 0 \nRightarrow |\delta x| \rightarrow 0. \quad (3.5)$$

En dimension finie, le problème est toujours bien posé. Cependant, la discrétisation d'un problème mal posé conduit généralement à un problème mal conditionné dont la solution est sensible aux petites variations présentes sur les observations.

Le nombre de conditionnement ($Cond(A)$) s'évalue par le rapport entre la plus grande valeur propre de la matrice A et la plus petite. Un problème « bien » conditionné a un nombre de conditionnement égal à 1.

Les problèmes mal posés ou les problèmes bien posés au sens d'Hadamard mais mal conditionnés doivent être traités par des méthodes de régularisation.

3.4 Régularisation

Obtenir la vraie solution à partir de données bruitées ou corrompues est impossible. En régularisant un problème, on cherche une solution acceptable parmi un ensemble de solutions possibles défini par

$$\{\hat{x} : \|y - Ax\| \leq \|b\|\} \quad (3.6)$$

où b représente le bruit.

Idier (2001) classe les méthodes de régularisation en deux grandes familles :

- celles à contrôles de dimension ;
- celles qui opèrent par minimisation d'un critère composite.

Les méthodes de la première classe contournent le problème mal posé

- soit en minimisant un critère dans un sous-espace de dimension réduite. Par exemple, grâce à une décomposition en valeurs singulières ou un changement de discrétisation ;
- soit en minimisant un critère dans l'espace initialement choisi mais par une méthode itérative dont on limite le nombre d'itérations (Demoment, 1989).

Dans ce chapitre, nous ne traitons pas les méthodes de première classe. Nous nous penchons essentiellement sur les méthodes de régularisation de la deuxième classe car elles nécessitent moins de réglages empiriques et sont de fait plus autonomes.

Les méthodes de régularisation par minimisation d'un critère composite demandent à la solution de réaliser un compromis entre la fidélité aux mesures et une fidélité à une information *a priori* (Tikhonov and Arsenin, 1976).

On considère comme solution admissible une solution non exacte, mais qui rendrait $A\hat{x}$ peu distante des données y et qui, dans l'ensemble des solutions admissibles, serait considérée comme physiquement la plus raisonnable.

Il s'agit alors de rechercher une solution $\hat{x}$ qui minimise un critère de la forme

$$C(x) = d(Ax - y) + \lambda R(x) \quad (3.7)$$

où d représente une distance quelconque et λ un scalaire positif qui permet de fixer l'importance de l'information *a priori* (R) par rapport à la fidélité aux mesures. Si λ est nul, on revient à un critère qui n'apporte aucune information. S'il est grand, la solution de notre minimisation dépendra fortement de nos *a priori*.

Il existe des méthodes de réglage du coefficient de régularisation λ . Ces méthodes sortent du cadre de ce document. Thompson y consacre un chapitre dans (Thompson et al., 1991).

En imagerie, le choix de la fonction de pénalisation $R(x)$ est primordial pour la réduction du bruit et la netteté des frontières de notre objet. Une synthèse détaillée sur les fonctions de pénalisation peut être trouvée dans le livre d'Idier (2001).

Une manière simple de mesurer la rugosité d'une image consiste à lui appliquer un opérateur de différentiation approprié puis d'en calculer la norme :

$$R(x) = \|D_k x\|^2$$

où la matrice D est la matrice des différences d'ordre k . Cette mesure est quadratique et Herman (1976) montre que l'utilisation d'une fonction quadratique de régularisation conduit à une réduction du bruit au détriment du lissage des contours.

Pour préserver davantage les discontinuités de l'objet, il est possible d'employer une pénalisation non quadratique. Ce genre de pénalisation croît moins vite qu'une parabole afin de moins pénaliser les variations importantes d'une image.

En reconstruction tomographique, on privilégie, généralement, les fonctions $L_2 L_1$, car

elles sont quadratiques à l'origine et asymptotiquement linéaires, ce qui permet de réduire le bruit dans les zones homogènes de l'objet tout en conservant des frontières franches (Fig. 3.1). De plus, ce sont des fonctions convexes, si bien que les algorithmes de minimisation sont assurés de converger.

Il existe aussi des fonctions L_2L_0 qui, elles, sont asymptotiquement constantes et donc non convexes.

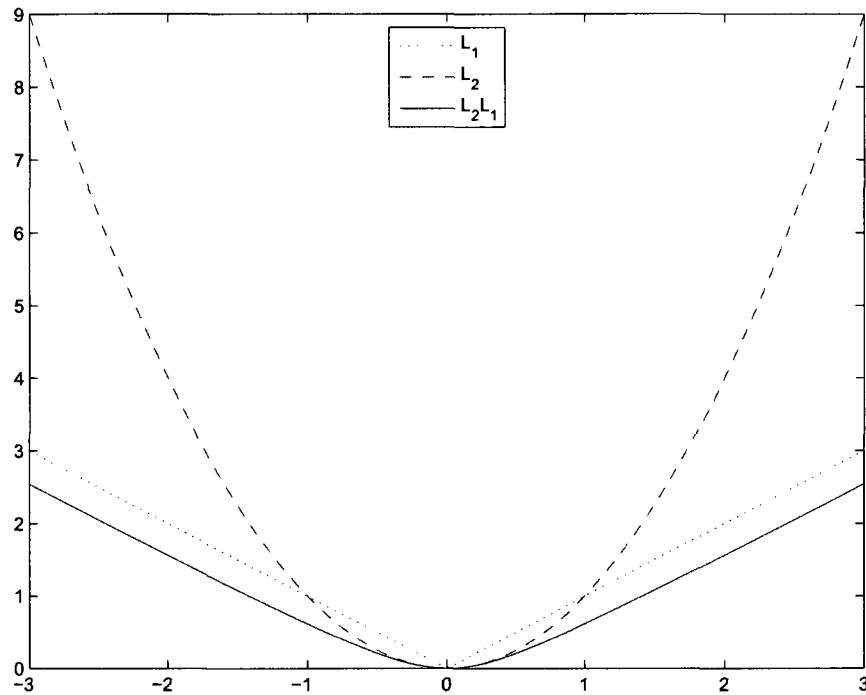


FIG. 3.1: Représentation des fonctions de pénalisation : L_1 , L_2 et L_2L_1 .

Le critère composite ainsi obtenu doit être minimisé. Déterminer l'estimateur qui minimise l'équation 3.7 est simplifié lorsque la distance d et la fonction R sont quadratiques :

$$\hat{x} = \arg \min_x \{ \|y - Ax\|_W^2 + \lambda \|x\|_Q^2 \} \quad (3.8)$$

La solution de cette équation est unique et linéaire par rapport aux données et peut être

calculée explicitement :

$$\hat{x} = (A^T W^{-1} A + \lambda Q)^{-1} A^T W^{-1} y \quad (3.9)$$

où W et Q sont des matrices symétriques définies positives choisies pour traduire certaines caractéristiques de la mesure de proximité.

Cependant, lorsque les quantités de données à traiter prennent des tailles extrêmement grandes, les matrices mises en jeu dans les estimateurs deviennent difficiles à manipuler. Ainsi, l'estimée ne peut plus être obtenue à partir des formules analytiques des estimateurs. Dans ce cas, des méthodes approximatives sont utilisées. Elles sont généralement itératives et demandent la minimisation d'une fonction objectif réalisée grâce à des algorithmes d'optimisation, comme ceux de la section 3.6.

3.5 Approche bayésienne

Dans la section précédente, nous avons vu que les problèmes inverses sont souvent des problèmes mal posés soit parce que l'opérateur A est singulier, soit parce que le dispositif expérimental, qui nous permet d'acquérir les données, n'est pas complètement affranchi d'une incertitude.

L'approche probabiliste, et notamment l'approche bayésienne, cherche à prendre en compte toute l'information disponible et à éviter de supposer disponible une information qui ne l'est pas. L'emploi d'un cadre probabiliste s'avère donc commode pour décrire une situation d'informations incomplètes.

3.5.1 Cadre du problème

Soit x , une quantité scalaire ou vectorielle à estimer, et z les observations (ou mesures). x et z sont deux quantités incertaines ; de ce fait, on associe à chacune d'elles une variable aléatoire (VA), respectivement X et Z .

En prenant en compte les incertitudes, nous pouvons écrire la relation matricielle suivante :

$$z = Ax + b \quad (3.10)$$

dans laquelle A représente la relation linéaire entre z et x , et b désigne le vecteur incertitudes. Cette relation représente l'hypothèse linéaire de la structure du problème.

L'estimation consiste alors à extraire l'information sur x en connaissant les mesures. En termes probabilistes, il s'agit de trouver la loi conditionnelle de X sachant que $Z = z$ est réalisé, c.-à-d. $f_{X|Z=z}(x)$.

Supposons que les incertitudes b soient gaussiennes. Ces incertitudes sont de moyenne nulle et on note sa matrice de covariance $\sigma_B^2 R_B$, où le plus grand élément de R_B est normalisé à 1. Cette notation de la matrice de covariance permet de séparer l'importance des incertitudes (variance σ_B^2) de la corrélation entre les différents éléments du vecteur b (représentée par R_B).

Soit $f_{Z|X=x}(z)$, la probabilité d'obtenir les mesures sachant que notre objet est connu ($X = x$). Cette probabilité correspond aux incertitudes faites sur le modèle de formation des mesures. Supposons qu'elle soit gaussienne de moyenne Ax et de covariance $\sigma_B^2 R_B$:

$$f_{Z|X=x}(z) = \frac{1}{(2\pi\sigma_B^2)^{N/2} \sqrt{|R_B|}} \exp \left(-\frac{(z - Ax)^T R_B^{-1} (z - Ax)}{2\sigma_B^2} \right) \quad (3.11)$$

où N représente la taille du vecteur x .

Maintenant que la formulation du problème est établie, nous pouvons nous intéresser à la restauration de l'entrée. Dans la prochaine section, nous développons des estimateurs qui vont permettre la restauration de l'entrée à partir des observations.

3.5.2 Estimateurs

Dans la littérature, il existe plusieurs estimateurs. Dans cette section, nous en verrons deux : l'estimateur du maximum de vraisemblance, grâce auquel on cherche à obtenir les inconnues qui rendent les mesures les plus vraisemblables, et l'estimateur bayésien du maximum *a posteriori*.

3.5.2.1 Estimateur du maximum de vraisemblance

L'estimateur par maximum de vraisemblance (**MV**) est basé sur la maximisation de la probabilité d'observer les données z . Ainsi, on définit l'estimateur MV par :

$$\hat{x}_{MV} = \arg \max_x \{f_{Z|X=x}(z)\} \quad (3.12)$$

Si l'on remplace dans cette expression $f_{Z|X=x}$ par son expression obtenue au paragraphe précédent, on trouve :

$$\hat{x}_{MV} = \arg \max_x \left\{ \frac{1}{(2\pi\sigma_B^2)^{N/2} \sqrt{|R_B|}} \exp \left(-\frac{(z - Ax)^T R_B^{-1} (z - Ax)}{2\sigma_B^2} \right) \right\} \quad (3.13)$$

La maximisation d'une exponentielle revient à minimiser l'opposé de son argument, puisque la fonction exponentielle est strictement croissante. De plus, lors d'une maximisation, les termes constants de la fonction à maximiser peuvent être enlevés. Ainsi,

$$\hat{x}_{MV} = \arg \min_x \left\{ \frac{(z - Ax)^T R_B^{-1} (z - Ax)}{2\sigma_B^2} \right\} \quad (3.14)$$

La fonction à minimiser est une forme quadratique et donc convexe. De ce fait, une condition suffisante pour trouver son minimum est l'annulation de son gradient, c.-à-d.

$$A^T R_B^{-1}(z - Ax) = 0 \quad (3.15)$$

Ainsi, en extrayant x de l'équation précédente, on obtient :

$$\hat{x}_{MV} = (A^T R_B^{-1} A)^{-1} A^T R_B^{-1} z \quad (3.16)$$

Cette expression est explicite par rapport aux données. Par conséquent, il n'est pas nécessaire d'avoir recours à des méthodes d'optimisation numériques pour la calculer. La mise en œuvre d'un tel estimateur est donc facilitée.

3.5.2.1.1 Propriétés Les techniques d'estimation nous donnent rarement la valeur exacte de la grandeur recherchée. En effet, les estimées sont entachées d'erreurs causées par les incertitudes sur les mesures. Il est important de quantifier ces incertitudes sur l'estimée afin de se positionner quant à la vraie grandeur. Une manière simple de calculer l'écart de l'estimée par rapport à la vraie grandeur est de calculer les caractéristiques d'ordre 2, par exemple la moyenne de la loi de probabilité de la quantité à estimer.

L'estimateur du maximum de vraisemblance peut se réécrire de la manière suivante :

$$\hat{x}_{MV} = (A^T R_B^{-1} A)^{-1} A^T R_B^{-1} z \quad (3.17)$$

$$= x + S_{MV} b, \text{ avec } S_{MV} = (A^T R_B^{-1} A)^{-1} A^T R_B^{-1} \quad (3.18)$$

Un inconvénient de cet estimateur est lié au bruit : il s'amplifie. En effet, dans l'équation 3.18, le bruit est multiplié par $S_{MV} = (A^T R_B^{-1} A)^{-1} A^T R_B^{-1}$.

Thiebaut (2006) définit la propagation des erreurs PE_{MV} par la relation suivante :

$$PE_{MV} = E[(S_{MV}b)(S_{MV}b)^T] \quad (3.19)$$

$$= (R_X^{-1} + A^T R_B^{-1} A)^{-1} \quad (3.20)$$

Nous allons nous servir de ce calcul de la propagation des erreurs à des fins de comparaison entre cet estimateur et l'estimateur du maximum *a posteriori*.

3.5.2.2 Estimateur du maximum *a posteriori*

L'estimateur du maximum *a posteriori* (MAP) est un estimateur bayésien qui prend en compte de l'information *a priori* sur les valeurs recherchées.

Avec cet estimateur MAP, nous cherchons à maximiser la probabilité du modèle étant donné les mesures.

$$\hat{x}_{MAP} = \arg \max_x f_{X|Z=z}(x) \quad (3.21)$$

Par le théorème de Bayes, il est possible de calculer cette quantité :

$$f_{X|Z=z}(x) = \frac{f_{Z|X=x}(z)f_X(x)}{f_Z(z)} \quad (3.22)$$

dans laquelle $f_Z(z)$ représente la probabilité de réalisation des mesures. Cependant, ce terme est souvent ignoré dans les procédures d'estimation puisqu'il est constante.

Cette méthode est appelée *approche bayésienne* puisqu'elle est fournie par le théorème

de Bayes.

$$\hat{x}_{MAP} = \arg \max_x f_{X|Z=z}(x) \quad (3.23)$$

$$= \arg \max_x f_{Z|X=x}(z) f_X(x) \quad (3.24)$$

Afin de réaliser cet estimateur, nous devons choisir la densité de probabilité $f_X(x)$. La quantité x est de moyenne x_0 et de matrice de covariance $\sigma_X^2 R_X$, où R_X est la matrice de covariance normalisée de X . Soit $f_X(x)$, l'incertitude sur les connaissances *a priori* sur x :

$$f_X(x) = \propto \exp(-\|D_k x\|^2) \quad (3.25)$$

où la matrice D_k représente la matrice de différence d'ordre k . Cette formulation permet d'englober une information de rugosité quadratique ou semi-quadratique.

Pour la suite des calculs, nous nous plaçons dans le cas où $D_k = D_0$, ce qui correspond à un cas gaussien :

$$f_X(x) = \frac{1}{(2\pi\sigma_X^2)^{P/2} \sqrt{|R_X|}} \exp\left(-\frac{(x-x_0)^T R_X^{-1} (x-x_0)}{2\sigma_X^2}\right) \quad (3.26)$$

De la même façon que pour l'estimateur MV, remplaçons les densités de probabilités dans cette expression et faisons les simplifications avec les termes constants. Annuler le gradient de $f_X(x)$ permet d'obtenir son minimum :

$$-\frac{A^T R_B^{-1} (z - Ax)}{\sigma_B^2} + \frac{R_X^{-1} (x - x_0)}{\sigma_X^2} = 0 \quad (3.27)$$

Extraire x de cette expression conduit à l'expression de l'estimateur MAP :

$$\hat{x}_{MAP} = x_0 + \left(A^T R_B^{-1} A + \frac{\sigma_B^2}{\sigma_X^2} R_X^{-1} \right)^{-1} A^T R_B^{-1} (z - Ax_0) \quad (3.28)$$

De la même manière que pour l'estimateur MV, l'estimateur MAP est aussi explicite par rapport aux données. De plus, on constate que l'estimateur MV peut être vu comme un cas particulier de l'estimateur MAP. En effet, si l'on prend $\sigma_B^2/\sigma_X^2 = 0$ dans l'équation (3.28), on retrouve l'équation de l'estimateur MV (3.16). Physiquement, on peut interpréter la mise à zéro de σ_B^2/σ_X^2 par le fait que les valeurs probables de x couvrent un domaine infini ; c'est à dire qu'elles peuvent prendre n'importe quelles valeurs. Ainsi, en faisant tendre $\sigma_B^2\sigma_X^2$ vers 0, on n'apporte aucune information *a priori* sur les valeurs de x .

3.5.2.2.1 Propriétés Nous pouvons réécrire les équations 3.28 et 3.28 sous la forme suivante :

$$\hat{x}_{MAP} = x_0 + \left(A^T R_B^{-1} A + \frac{\sigma_B^2}{\sigma_X^2} R_X^{-1} \right)^{-1} A^T R_B^{-1} (z - Ax_0) \quad (3.29)$$

$$= x_0 + S_{MAP} A (x - x_0) + S_{MAP} b \quad (3.30)$$

où $S_{MAP} = \left(A^T R_B^{-1} A + \frac{\sigma_B^2}{\sigma_X^2} R_X^{-1} \right)^{-1} A^T R_B^{-1}$.

Pour cet estimateur, la propagation des erreurs s'écrit :

$$PE_{MAP} = E[(S_{MAP} b)(S_{MAP} b)^T] \quad (3.31)$$

$$= (R_X^{-1} + PE_{MV}^{-1})^{-1} PE_{MV}^{-1} (R_X^{-1} + PE_{MV}^{-1})^{-1} \quad (3.32)$$

$$= (PE_{MV} R_X^{-1} + I)^{-2} PE_{MV} \quad (3.33)$$

$$PE_{MAP} \leq PE_{MV} \quad (3.34)$$

Ainsi, le calcul de la propagation d'erreurs de cet estimateur nous montre que l'erreur propagée est moins importante pour l'estimateur MAP que pour l'estimateur MV.

3.5.3 Choix de l'estimateur

Les question de mises en œuvre des estimateurs ne sont pas en général déterminantes dans le choix de l'estimateur. Le choix dépend essentiellement de la difficulté du problème d'estimation. Celle-ci s'évalue en fonction de trois paramètres :

1. La taille de x et de z . Plus le rapport de ces tailles se rapproche de 1, plus l'estimation est difficile ;
2. La nature de H . L'application linéaire H peut modéliser une transformation qui supprime certaines composantes du signal x dans les observations z . L'estimation de x sera donc difficile à réaliser sans l'apport d'*a priori* sur la solution ;
3. L'importance des incertitudes b dans les observations.

Lors de la résolution d'un problème d'estimation, il est conseillé d'introduire le minimum d'information *a priori* sur la solution afin de ne pas restreindre l'espace de solutions vraisemblables. Il est recommandé de tester un estimateur des moindres carrés et, en fonction des résultats, d'introduire, ou non, une information *a priori*.

3.5.4 Lien avec les méthodes déterministes

Les méthodes déterministes de critère régularisé (cf. 3.4) et les méthodes bayésiennes sont fortement liées. En effet, dans le cadre linéaire et gaussien qui nous intéresse, la minimisation d'un critère régularisé est équivalente à choisir la solution qui maximise la loi *a posteriori* suivante :

$$f_{X|Z=z}(x) \propto e^{-\frac{1}{2\sigma^2}(d(z-Ax)+\lambda R(x))} \quad (3.35)$$

où σ^2 représente la variance du bruit.

Si la distance d est quadratique, par exemple une norme euclidienne, on peut décomposer

$f_{X|Z=z}$ selon :

$$f_{Z|X=x}(x) \propto e^{\frac{1}{2\sigma^2}d(y-A(x))} \quad (3.36)$$

et la loi des *a priori* sur les variables recherchées prend alors la forme :

$$f_X(x) \propto e^{\frac{\lambda}{2\sigma^2}R(x)} \quad (3.37)$$

Ces relations ne sont valides que si les lois *a priori* et conditionnelle précédentes sont propres, c.-à-d. que leur intégrale converge.

De ce fait, l'estimateur MAP se confond avec le minimiseur du critère pénalisé de l'équation 3.7 :

$$\hat{x}_{MAP} = \arg \max_x f_{X|Z=z}(x) \quad (3.38)$$

$$= \arg \max_x \{d(z - Ax) + \lambda R(x)\} \quad (3.39)$$

Ainsi, le cadre bayésien donne un sens statistique à la minimisation des critères pénalisés. De plus, il permet de déterminer des propriétés statistiques des estimateurs, par exemple l'erreur quadratique moyenne (EQM).

3.6 Estimation par optimisation itérative du critère

Dans la section précédente, nous avons déterminé des expressions d'estimateurs sous forme analytique et compacte. Cependant, dans certains cas, le problème d'estimation ne peut pas être résolu par ces formes compactes. C'est le cas en reconstruction tomographique, puisque les matrices mises en jeu sont de trop grosse taille pour pouvoir en calculer l'inverse (cf. 2.5.2.1). Ainsi une solution explicite est impossible à obtenir.

On fait alors appel à des techniques itératives de résolution. La minimisation du critère

est réalisée par un algorithme d'optimisation. Dans ce chapitre, les discussions sur les méthodes d'optimisation sont valables pour des critères convexes et quadratiques (ou semi-quadratiques). De ce fait, nous limiterons l'exposé à un petit nombre d'algorithmes qui se sont révélés efficaces pour la résolution de ces problèmes. Pour de plus amples détails sur la théorie de l'optimisation, le livre de Nocedal et Wright (2000) sera une référence.

Nous commencerons par exposer les conditions sous lesquelles il est assuré d'avoir convergence des algorithmes de minimisation. Ensuite, nous discuterons de la méthode à direction de descente, qui permet d'obtenir le minimum recherché si les conditions précédentes sont respectées. Dans cette partie, nous évoquerons les algorithmes de la plus forte pente, de Newton et de Newton modifiés. Nous en décrirons les propriétés et les avantages pour l'estimation.

3.6.1 Conditions d'optimalité

Plaçons-nous dans un cadre général d'optimisation sans contrainte dans lequel il faut trouver :

$$x^* \in \arg \min_{x \in \mathbb{R}^n} f(x) \quad (3.40)$$

Une approche itérative génère une suite $\{x_k\}_{k=0,\dots,K}$ de solutions qui converge vers x^* si les conditions d'optimalité sont respectées.

Soit $f(x)$, une fonction de $\mathbb{R}^n \mapsto \mathbb{R}$, différentiable deux fois.

Précisons la notion de minimum. En optimisation, on définit généralement quatre types de minimum. Soit $x^* \in \mathbb{R}^n$, on dit que x^* est

- un minimum *global* de f si et seulement si $\forall x \in \mathbb{R}^n, f(x) \geq f(x^*)$,
- un minimum *local* de f si et seulement si $\exists \epsilon > 0$ tel que $\forall x \in \mathcal{B}(x^*, \epsilon), f(x) >$

- $f(x^*)$,
- un minimum *local strict* de f si et seulement si $\exists \epsilon > 0$ tel que $\forall x \in \mathcal{B}(x^*, \epsilon), x \neq x^*, f(x) > f(x^*)$,
 - un minimum *local isolé* de f si et seulement si $\exists \epsilon > 0$ tel que x^* est le seul minimum local de f dans $\mathcal{B}(x^*, \epsilon)$.

Afin d'identifier les minimums candidats, nous devons poser les conditions que tout minimum doit satisfaire. Ce sont les conditions nécessaires d'optimisation.

3.6.1.1 Condition du premier ordre

Théorème 3.6.1 (Condition nécessaire du premier ordre). *Si $x^* \in \mathbb{R}^n$ est un minimum local de f dans la relation 3.40 et f est une fonction de classe C^1 sur une boule ouverte centrée en x^* , alors*

$$\nabla f(x^*) = 0 \quad (3.41)$$

Soit un point $x \in \mathbb{R}^n$, il est dit *stationnaire* lorsque $\nabla f(x) = 0$.

Cette condition est nécessaire mais pas suffisante ; ainsi, elle est surtout utile pour montrer la contraposée de ce théorème : si $\nabla f(x^*) \neq 0$, alors x^* n'est pas un minimum local de f . Ce n'est donc pas un point stationnaire pour f .

Toutefois, la réciproque de ce théorème est fausse : un point stationnaire n'est pas forcément un minimum local. Il peut être un minimum local, un maximum local, ou ni l'un ni l'autre. Pour préciser la nature de ces points, une condition sur les dérivées secondes est nécessaire.

3.6.1.2 Conditions du second ordre

Théorème 3.6.2 (Conditions nécessaires du second ordre). *Si $x^* \in \mathbb{R}^n$ est un minimum local de l'équation 3.40 et f est une fonction de classe C^2 sur une boule ouverte centrée en x^* , alors*

$$\nabla^2 f(x^*) \text{ est semi définie positive.} \quad (3.42)$$

De la même façon que la condition nécessaire du premier ordre, ce théorème sert à montrer que x^* n'est pas un minimum local de f .

Théorème 3.6.3 (Conditions suffisantes du second ordre). *Soit x^* un point stationnaire de f et f une fonction de classe C^2 sur une boule ouverte centrée en x^* . Si $\nabla f(x^*) = 0$ et $\nabla^2 f(x^*)$ est définie positive, alors x^* est minimum local isolé de f .*

3.6.2 Méthodes de direction de descente

Les méthodes de directions de descente ont pour but de trouver x^* en se servant d'informations sur les dérivées de f . Ces informations, très importantes, représentent les variations locales de la fonction f .

Algorithme 1 Algorithme de descente

Initialisation : on choisit un point de départ $x_0 \in \mathbb{R}^n$, une tolérance d'arrêt $\epsilon_0 > 0$ et on met le compteur d'itérations à 0

tantque $\|\nabla f(x_k)\| > \epsilon_0$ **faire**

À partir de x_k , on calcule une direction de descente d_k et un pas t_k (recherche linéaire) tels que :

$$f(x_k + t_k d_k) < f(x_k). \quad (3.43)$$

Mise à jour de l'itéré $x_{k+1} = x_k + t_k d_k$ et du compteur d'itérations.

fin tantque

Soit $f : \mathbb{R}^n \mapsto \mathbb{R}$; on définit une *direction de descente* $d \in \mathbb{R}^n$ de f en x si $\nabla f(x)^T d < 0$.

Il existe toujours une direction de descente pour f en x tant que $\nabla f(x) \neq 0$. Cependant, lorsque $\nabla f(x) = 0$, il n'y en a pas, mais ce point est un point stationnaire ! Dans la pratique, plusieurs directions de descente existent et font référence à des catégories distinctes d'algorithmes que nous allons aborder plus loin.

La recherche linéaire d'un pas de descente permet d'obtenir, à partir d'un point courant x_k et d'une direction de descente choisie d_k , une valeur de la fonction objectif inférieure à celle de l'objectif courant.

3.6.2.1 Conditions assurant la convergence d'un algorithme

Sous certaines conditions touchant la direction de descente et la recherche linéaire, les algorithmes sont assurés de converger vers un minimum local de f . La recherche d'un pas optimal est complexe et elle est un problème d'optimisation en elle-même. Ainsi, nous nous contenterons de trouver une valeur de pas t_k telle que $f(x_k + t_k d_k) < f(x_k)$.

La condition d'Armijo établit une condition sur le pas t_k afin que celui-ci soit choisi selon un critère strict de décroissance de l'objectif.

Condition d'Armijo. Soit $\alpha \in]0, 1[$, le pas t_k doit satisfaire

$$f(x_k + t_k d_k) < f(x_k) + \alpha \nabla f(x_k)^T d_k \quad (3.44)$$

Le coefficient α fixe l'importance de la décroissance. Pour de petits pas, une faible décroissance est autorisée à condition de satisfaire la condition d'Armijo.

Ces deux conditions réunies forment les *conditions de Wolfe* : soit $\alpha \in]0, 1[$ et $\gamma \in]\alpha, 1[$,

le pas t_k doit satisfaire les deux conditions suivantes :

$$f(x_k + t_k d_k) < f(x_k) + \alpha \nabla f(x_k)^T d_k \quad (3.45)$$

$$\nabla f(x_k + t_k d_k)^T d_k \geq \gamma \nabla f(x_k)^T d_k \quad (3.46)$$

La deuxième condition, qui porte sur la courbure de l'objectif, assure que le pas fera descendre rapidement l'objectif, écartant ainsi les pas trop petits. Cette condition peut être renforcée par une condition plus stricte. On parle alors des conditions de Wolfe fortes. Il s'agit de borner la pente de décroissance, contraignant le pas à n'être ni trop petit, ni trop grand.

Conditions de Wolfe fortes. Soit $\alpha \in]0, 1[$ et $\gamma \in]\alpha, 1[$, le pas t_k doit satisfaire les deux conditions suivantes :

$$f(x_k + t_k d_k) < f(x_k) + \alpha \nabla f(x_k)^T d_k \quad (3.47)$$

$$|\nabla f(x_k + t_k d_k)^T d_k| \leq \gamma |\nabla f(x_k)^T d_k| \quad (3.48)$$

On définit l'angle θ_k , entre la direction de descente d_k et l'opposé du gradient de f en x_k , par

$$\cos \theta_k = \frac{-\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\| \|d_k\|} \quad (3.49)$$

Théorème 3.6.4 (Théorème de Zoutendijk). *Soit un algorithme quelconque de descente imposant soit la condition d'Armijo, soit les conditions de Wolfe pour le choix d'une longueur de pas. Soit $\mathcal{L} = \{x \in \mathbb{R}^n | f(x) \leq f(x_0)\}$.*

Si on suppose que f est $\mathcal{C}^1$ sur un ensemble ouvert contenant $\mathcal{L}$ et que ∇f est Lipschitz-continue sur $\mathcal{L}$, alors soit $\{f(x_k)\} \rightarrow -\infty$ ou $\sum_{k=0}^{+\infty} \cos^2 \theta_k \|\nabla f(x_k)\|^2 < +\infty$.

On remarque que les hypothèses de ce théorème sont peu contraignantes et donc souvent vérifiées en pratique.

Le théorème 3.6.4 implique que :

$$\lim_{k \rightarrow +\infty} \cos^2 \theta_k \|\nabla f(x_k)\|^2 = 0 \quad (3.50)$$

Utilisons cette limite pour déterminer la convergence des algorithmes de descente. Si cette méthode sélectionne une direction de descente suffisamment éloignée de l'orthogonalité au gradient (c.-à-d. que $|\cos \theta_k| \geq \epsilon > 0$), alors

$$\lim_{k \rightarrow +\infty} \|\nabla f(x_k)\| = 0 \quad (3.51)$$

Cette limite signifie que le gradient de f tendra vers 0 si les directions de descente sélectionnées sont suffisamment éloignées de l'orthogonalité avec le gradient. Sous cette condition, l'algorithme de descente est assuré de converger.

Dans les sections suivantes, nous allons aborder plusieurs algorithmes d'optimisation qui diffèrent selon leur direction de descente et le choix de recherche linéaire.

3.6.2.2 Algorithme de la plus forte pente

Cet algorithme est un algorithme de descente dans lequel la direction de descente est intuitive et correspond à l'opposé du gradient de la fonction objectif :

$$d_k = -\nabla f(x_k) \quad (3.52)$$

Le théorème de Zoutendijk assure une convergence globale de la méthode de la plus forte pente, puisque $\theta_k = 0$ pour tout k .

À cause des propriétés du gradient, cette méthode possède de faibles performances et oscille fortement, même proche du minimum. Ainsi, il est nécessaire de réaliser beaucoup

d'itérations de cet algorithme (non souhaitable lorsque le gradient coûte cher en temps d'évaluation).

3.6.2.3 Algorithme du gradient conjugué linéaire et non linéaire

Les algorithmes du gradient conjugué permettent de minimiser une fonction objectif quadratique. Cependant, ils ont été étendus aux objectifs non quadratiques. Nous expliquerons, dans un premier temps, les résultats pour les objectifs quadratiques et, dans un deuxième temps, pour ceux non quadratiques.

3.6.2.3.1 Le gradient conjugué linéaire Plaçons-nous dans le cas de minimisation d'une fonction objectif quadratique comme suit :

$$f(x) = \frac{1}{2} \|y - Cx\|^2 \text{ où } C \text{ est une matrice carrée de taille } n \times n \quad (3.53)$$

Posons $H = C^T C$, le Hessien de f et $r(x) = y - Cx$, le résidu de f et $\nabla f(x) = -C^T r(x)$, son gradient.

La direction de descente dans cette méthode est donnée par :

$$d_{k+1} = -r_{k+1} + \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k} d_k \quad (3.54)$$

Cette nouvelle direction est peu coûteuse en calcul et en mémoire, car elle s'obtient à partir de la direction précédente et des résidus courant et précédent.

Soit u et v , deux vecteurs, ils sont dits *H-conjugués* si et seulement si $u^T H v = 0$. Puisque $d_{k+1} H d_k = 0$ pour tous les k , les directions sont dites *conjuguées*.

De plus, la recherche linéaire pour trouver une longueur de pas optimale peut s'obtenir

analytiquement par $t_k = \frac{r_k^T r_k}{d_k^T H d_k}$.

Cet algorithme converge plus rapidement que la méthode de la plus forte pente vue dans le paragraphe précédent. Néanmoins, la vitesse de convergence dépend fortement de la distribution des valeurs propres du Hessien H . On peut démontrer que la méthode du gradient conjugué sur un problème quadratique converge en au plus m itérations, m étant le nombre de valeurs propres distinctes de H . Ainsi, pour les problèmes dans lesquels les valeurs propres ne sont pas proches les unes des autres, il faut adapter la méthode du gradient conjugué en le préconditionnant. Ce préconditionnement permet de rassembler les valeurs propres en paquets de valeurs semblables afin de réduire le nombre d'itérations de l'algorithme du gradient conjugué. Nous ne présenterons pas cette version du gradient conjugué.

L'algorithme du gradient conjugué n'est valide que dans le cas d'une minimisation strictement convexe. Lorsque l'objectif ne l'est pas, une adaptation de cet algorithme doit être faite. C'est l'objet de la section suivante.

3.6.2.3.2 Le gradient conjugué non linéaire Supposons maintenant que la fonction objective ne soit pas quadratique. Nous ne pouvons plus obtenir une formule analytique exacte pour la détermination du pas de descente. La structure générale de cet algorithme reste la même à l'exception du calcul du pas t_k , qui se fait maintenant par recherche linéaire classique et du terme β_{k+1} .

Fletcher et Reeves (1964) proposent une généralisation de l'algorithme du gradient conjugué linéaire aux fonctions non quadratiques. Ainsi, ils définissent le coefficient β_{k+1} :

$$\beta_{k+1}^{FR} = \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k^T \nabla f_k} \quad (3.55)$$

Cette expression permet de rejoindre la formulation du β_{k+1} pour des fonctions quadra-

tiques. En 1997, Powell montra que cette méthode peut conduire à de petits déplacements et donc à une lente convergence.

Afin d'accélérer la convergence du gradient conjugué non linéaire, Polak et Ribière expriment β_{k+1} de façon légèrement différente (Nocedal and Wright, 2000) :

$$\beta_{k+1}^{PR} = \frac{\nabla f_{k+1}^T (\nabla f_{k+1} - \nabla f_k)}{\nabla f_k^T \nabla f_k} \quad (3.56)$$

Powell en 1984 a montré que cette expression associée avec une recherche linéaire exacte peut tourner indéfiniment autour de l'objectif sans jamais l'atteindre.

Pour allier les propriétés de convergence assurée du coefficient de Fletcher et Reeves aux propriétés numériques de Polak et Ribière, Nocedal (1992) propose la formulation suivante :

$$\beta_{k+1}^{PR+} = \max(0, \beta_{k+1}) \quad (3.57)$$

L'algorithme du gradient conjugué non linéaire avec β_{k+1}^{PR+} , couplé avec une recherche linéaire satisfaisant les conditions de Wolfe fortes, possède une convergence prouvée pour des fonctions de classe $\mathcal{C}^1$.

Les algorithmes précédents utilisent des informations d'ordre 1 (le gradient) afin de minimiser le critère. Cependant, lorsque le critère est complexe à minimiser, une information plus riche peut être nécessaire. Elle est obtenue par la courbure de l'objectif, c.-à-d. l'information d'ordre 2. Les algorithmes, que nous allons détailler dans la suite de ce chapitre, sont basés sur une information d'ordre 2 sur la fonction objectif.

3.6.2.4 Algorithme de Newton

Ces algorithmes prennent en compte une information plus riche sur l'objectif : sa courbure, c.-à-d. le Hessien de la fonction objectif ($\nabla^2 f$).

Cette direction de descente est obtenue par développement de Taylor d'ordre 2 de la fonction objectif :

$$f(x + d) \approx f(x) + \nabla f(x)^T d + \frac{1}{2} d^T \nabla^2 f(x) d \quad (3.58)$$

La différentiation de $f(x + d)$ par rapport à d , nous permet d'extraire la direction d :

$$\nabla f(x) + \nabla^2 f(x) d = 0 \quad (3.59)$$

L'utilisation d'information sur la courbure de l'objectif augmente la vitesse de convergence. Cependant, pour des problèmes de grosses tailles, dans lesquels le Hessien est coûteux en calcul, cette méthode est à proscrire.

Quand le calcul du Hessien de f est impossible, la direction de Newton n'est pas calculable. Il est alors d'usage de faire appel à des méthodes, dites *quasi-Newton*, qui calculent une approximation du Hessien de f .

Enfin, lorsque le Hessien de f n'est pas défini positif, cette direction n'est plus une direction de descente. On fait alors appel à des algorithmes dits de *Newton modifié*.

Ainsi, au lieu de la direction de Newton, nous cherchons à obtenir une direction de descente qui vérifie :

$$(\nabla^2 f(x) + M_k) d = -\nabla f(x) \quad (3.60)$$

où la matrice M_k est symétrique définie positive. Elle est choisie telle que :

- M_k est nulle si $\nabla^2 f(x)$ est définie positive ;
- $\nabla^2 f(x) + M_k$ est définie positive si $\nabla^2 f(x)$ n'est pas définie positive ;
- M_k est la plus petite possible.

Un exemple de cette méthode de Newton modifié est l'algorithme de *Levenberg*. Il propose comme matrice M_k :

$$M_k = \lambda I_n \quad (3.61)$$

Lorsque λ est petit, on retrouve la direction de Newton. Tandis que si λ est grand, la direction obtenue est équivalente à une direction de la plus forte pente. Pour des valeurs intermédiaires de λ , on se trouve entre ces deux méthodes.

Dans cet algorithme, le coefficient λ est modifié à chaque itération. Si d_k est une direction de descente, alors on le diminue (en le divisant par 10, par exemple) pour se rapprocher d'une méthode de Newton. Par contre, lorsque d_k n'est pas une direction de descente, on augmente λ (en le multipliant par 10, par exemple) pour se rapprocher de la méthode de la plus forte pente.

Cet algorithme a ensuite été amélioré par Marquardt (Roweis, 2009). Le pas de l'itération est défini cette fois par :

$$M_k = \lambda \text{diag}(H) \quad (3.62)$$

L'introduction de la diagonale de H au lieu de la matrice identité permet de modifier le comportement de l'algorithme lorsque λ est grand. Par cette modification, on se déplace plus vite dans les directions vers lesquelles le gradient est plus fiable, pour ne pas rester bloqué sur un plateau.

Cet algorithme est appelé algorithme de *Levenberg-Marquardt* synthétisé dans l'algorithme 2.

Algorithme 2 Algorithme de Levenberg Marquardt

Initialisation : on choisit un point de départ $x_0 \in \mathbb{R}^n$, une tolérance d'arrêt $\epsilon_0 > 0$, un paramètre $\lambda = \lambda_0$ et on met le compteur d'itérations à 0

tantque $\|\nabla f(x_k)\| > \epsilon_0$ **faire**

$$H_k = \nabla^2 f(x_k) + \lambda \text{diag}(H)$$

Obtention de d_k en solutionnant $H_k d_k = -\nabla f(x)$

si d_k est une direction de descente, c.-à-d. $d_k^T \nabla f_k < 0$ **alors**

$$\lambda = 10 * \lambda \text{ et } x_{k+1} = x_k$$

sinon

$$\lambda = \frac{\lambda}{10}$$

$$x_{k+1} = x_k - H_k^{-1} d_k.$$

fin si

Mise à jour du compteur d'itération $k = k + 1$.

fin tantque

3.6.2.5 Algorithme quasi-Newton

L'algorithme quasi-Newton est une adaptation de l'algorithme de Newton lorsque l'information sur les dérivées secondes n'est pas disponible. Il s'agit de construire une approximation du Hessien B_k de f symétrique définie positive dans le but d'éviter son calcul à chaque itération.

La direction de descente vérifie alors :

$$B(x_k)d_k = -\nabla f(x_k) \tag{3.63}$$

La construction de la matrice $B(x_k)$ est déterminante dans l'efficacité de l'algorithme. Les méthodes quasi-Newton peuvent être très efficaces si une quelconque information

Algorithme 3 Algorithme BFGS

Initialisation : on choisit un point de départ $x_0 \in \mathbb{R}^n$, une tolérance d'arrêt $\epsilon_0 > 0$ et on met le compteur d'itérations à 0

tantque $\|\nabla f(x_k)\| > \epsilon_0$ **faire**

Obtention de la direction de descente d_k vérifiant

$$B_k d_k = -\nabla f(x_k) \quad (3.64)$$

Calcul du pas de descente t_k par une recherche linéaire.

Mise à jour de l'itérée : $x_{k+1} = x_k + t_k d_k$ et du compteur d'itération $k = k + 1$.

Mise à jour de B_k^{-1} par :

$$B_{k+1}^{-1} = B_k^{-1} + (s_k s_k^T) \frac{s_k^T y_k + y_k^T B_k^{-1} y_k}{(s_k^T y_k)^2} - \frac{B_k^{-1} y_k s_k^T + s_k y_k^T B_k^{-1}}{s_k^T y_k} \quad (3.65)$$

avec $s_k = x_{k+1} - x_k$ et $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$

fin tantque

sur la courbure de l'itérée courant est introduite dans la matrice $B(x_k)$.

Un type d'algorithme quasi-Newton est celui développé par Broyden, Fletcher, Goldfarb et Shanno (1970). Dans cet algorithme (algo 3), l'approximation du Hessien suit la formule de récurrence suivante :

$$B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k (B_k s_k)^T}{s_k^T B_k s_k} \quad (3.66)$$

où $s_k = x_{k+1} - x_k$ et $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$. L'obtention de la direction de descente demande la résolution d'un système linéaire impliquant B_{k+1} , qui peut être difficile à réaliser.

Grâce au théorème de Sherman-Morrisson sur l'inversion de matrice, nous obtenons directement l'inverse de B_{k+1} par une formule de récurrence :

$$B_{k+1}^{-1} = B_k^{-1} + (s_k s_k^T) \frac{s_k^T y_k + y_k^T B_k^{-1} y_k}{(s_k^T y_k)^2} - \frac{B_k^{-1} y_k s_k^T + s_k y_k^T B_k^{-1}}{s_k^T y_k} \quad (3.67)$$

Pour cette approximation du Hessien, de nombreux produits matriciels ainsi que le stockage de la matrice B_k^{-1} de l'itération précédente sont nécessaires. Pour des problèmes de grosses tailles, comme en tomographie, le stockage de cette matrice est impossible. Ainsi, une version à mémoire limitée (L-BFGS) a été développée par Liu et al. (1989). Cette version allège la mémoire par l'utilisation d'une formule de récurrence sur le calcul de B_k^{-1} .

Pour cet algorithme, la convergence n'a pas été prouvée formellement. Cependant, en pratique, on constate qu'il converge.

3.7 Conclusion

Ce chapitre a été consacré à la théorie de l'estimation afin de l'appliquer à la reconstruction tomographique par approche algébrique ainsi qu'au traitement de la postluminescence.

Nous avons développé la méthodologie de l'estimation qui permet de résoudre la majorité des problèmes d'estimation.

On a pu remarquer que l'inversion du problème direct sans tenir compte des erreurs sur les mesures donne de très mauvais résultats, parce que le problème est mal posé. Il apparaît que l'estimateur MAP est meilleur au sens de la proximité par rapport à la solution et en termes de propagation d'erreur. Cependant l'estimateur explicite MAP demande plus de ressources de calculs qu'un simple estimateur MV.

Pour le traitement de la postluminescence, nous utiliserons un estimateur MAP, parce qu'il est plus précis et que la taille des données et des variables recherchées permettent le calcul de l'estimateur de manière explicite.

Contrairement à la postluminescence, la reconstruction tomographique par l'approche algébrique ne peut se faire avec un estimateur explicite. Par conséquent, nous utiliserons un estimateur MAP minimisé par un algorithme itératif de type BFGS à mémoire limitée. Des travaux sont en cours au laboratoire pour déterminer le type d'algorithme à utiliser pour obtenir un compromis temps/résolution d'image acceptable. Tous les résultats de reconstruction seront obtenus par L-BFGS, car cet algorithme nous fournit une solution acceptable en un temps plus faible que l'algorithme du gradient conjugué.

La cadre de travail de notre projet est maintenant défini. Nous allons développer, dans le chapitre suivant, le traitement de la postluminescence.

CHAPITRE 4

POSTLUMINESCENCE

4.1 Introduction

Ce chapitre est consacré aux traitements de la postluminescence. Ce phénomène se produit lors de la conversion de la radiation en lumière visible (cf. section 2.6.2). Les cristaux de scintillation commencent à briller puis retournent à leur état d'origine. Ainsi, plus le scintillateur est rapide pour retourner à son état d'équilibre, plus les contours des objets seront précis.

Tout d'abord, nous modélisons mathématiquement ce phénomène par un modèle de convolution. Puis, nous étudions le traitement le plus efficace présenté dans la littérature scientifique (Hsieh et al., 2000). Ensuite, nous remplaçons ce traitement dans le cadre de la théorie de l'estimation et nous formulons des traitements alternatifs. Enfin, chacune des méthodes qui sont présentées demande des informations spécifiques sur les paramètres du modèle. Nous développons trois techniques qui permettent de les obtenir.

4.2 Modélisation de la postluminescence

Ici, nous supposons que la postluminescence est un phénomène linéaire et que le système est invariant dans le temps. Ces hypothèses sont largement utilisées dans la littérature sur la postluminescence (Hsieh et al., 2000; LaRivière et al., 2006). D'après la théorie du signal, la réponse de détecteurs à un signal d'entrée est la convolution de la réponse impulsionnelle de ces détecteurs avec le signal d'entrée. Ainsi, la réponse du détecteur

$y(t)$ au signal $x(t)$ s'écrit :

$$y(t) = h(t) * x(t) \quad (4.1)$$

$$= \int_{-\infty}^{+\infty} h(t - \tau)x(\tau)d\tau \quad (4.2)$$

où

- $y(t)$ sont les données observées ;
- $h(t)$ représente la réponse impulsionnelle des détecteurs, également appelée noyau de convolution. La figure 2.10 montre quelques réponses à un échelon pour plusieurs détecteurs ;
- $x(t)$ est l'entrée à restaurer, soit, dans notre cas, le sinogramme.

On appelle ce type d'équation linéaire une *équation intégrale de Fredholm du premier type*. Elle correspond au cas où les observations ne sont pas entachées de bruit ni d'incertitudes. Ceci n'est pas vrai en réalité, puisqu'au moins un bruit numérique s'ajoute sur les observations.

Une représentation plus fidèle de la réalité s'écrit sous la forme :

$$z(t) = \int_{-\infty}^{+\infty} h(t - \tau)x(\tau)d\tau + b(t) \quad (4.3)$$

dans laquelle $b(t)$ représente un bruit additif, et $z(t)$ les données observées bruitées.

4.3 Traitement de la postluminescence

Dans la section précédente, le modèle de la postluminescence par convolution a été introduit. Ici, nous nous intéressons à la déconvolution de la postluminescence. Nous cherchons à obtenir $x(t)$, connaissant les observations $z(t)$ et en supposant la fonction h connue. Nous verrons dans la section 4.4 comment obtenir cette réponse impulsionnelle.

Premièrement, nous étudions la méthode proposée par Hsieh (2000). Cette méthode est considérée comme le traitement de référence de la postluminescence en tomographie. Nous replaçons cette méthode dans le contexte de l'estimation vue au chapitre 3.

Ensuite, nous proposons des méthodes plus robustes face au bruit présent dans les données et basées sur la théorie de l'estimation.

4.3.1 Traitement proposé par Hsieh

La littérature sur la correction de la postluminescence est plutôt rare. LaRivière et al. (2006) recensent les différentes techniques de compensation de la postluminescence et concluent que la méthode la plus efficace est celle de Hsieh (Hsieh et al., 2000). En effet, cette méthode est citée plusieurs fois dans la littérature et a fait l'objet d'un brevet pour son implantation dans les scanners commerciaux de Siemens.

Présentons cette méthode qui nous sert de point de référence pour nos développements futurs.

Dans l'article de Hsieh, la réponse impulsionnelle des détecteurs est modélisée par des exponentielles décroissantes :

$$h(t) = \sum_{n=1}^N \frac{\alpha_n}{\tau_n} e^{-\frac{t}{\tau_n}} U(t) \quad (4.4)$$

où

$$U(t) = \begin{cases} 1 & t \geq 0 \\ 0 & \text{sinon} \end{cases}$$

α_n représente le poids accordé à l'exponentielle de constante de temps τ_n et N le nombre d'exponentielles.

Puisque les fonctions x et h n'existent que pour $t > 0$, la convolution de postluminescence de l'équation 4.2 s'écrit :

$$y(t) = h(t) * x(t) \quad (4.5)$$

$$= \sum_{n=1}^N \frac{\alpha_n}{\tau_n} \int_0^t x(v) e^{-\frac{(t-v)}{\tau_n}} dv \quad (4.6)$$

En discrétisant la relation précédente et en supposant que sur un intervalle Δt , x est constant, on obtient une équation permettant de déterminer x à l'instant $k\Delta t$ en fonction de y au même instant et des x aux instants précédents :

$$x(k\Delta t) = \frac{y(k\Delta t) - \sum_{n=1}^N \alpha_n (1 - e^{-\Delta t/\tau_n}) \sum_{j=1}^{k-1} x(j\Delta t) e^{-(k-j)\Delta t/\tau_n}}{\sum_{n=1}^N \alpha_n (1 - e^{-\Delta t/\tau_n})} \quad (4.7)$$

On pose $\beta_n = \alpha_n (1 - e^{-\frac{\Delta t}{\tau_n}})$. Afin de réaliser un traitement de la postluminescence simultanément à l'acquisition de données, Hsieh réécrit cette équation sous la forme suivante :

$$x(k\Delta t) = \frac{y(k\Delta t) - \sum_{n=1}^N \beta_n e^{-\frac{\Delta t}{\tau_n}} S_{nk}}{\sum_{n=1}^N \beta_n} \quad (4.8)$$

où

$$S_{nk} = x((k-1)\Delta t) + e^{-\frac{\Delta t}{\tau_n}} S_{n(k-1)} \quad (4.9)$$

Lorsqu'une nouvelle donnée $y(k\Delta t)$ est acquise, S_{nk} est mise à jour par l'équation 4.9, et uniquement le dernier S_{nk} est utilisé pour la correction de la prochaine donnée. Lors d'exams tomographiques cliniques, le temps entre deux exams est suffisamment long. De ce fait, l'effet de la postluminescence de l'examen précédent peut être ignoré et on initialise S_{n0} avec des valeurs nulles.

Cette méthode de résolution peut être assimilée à du filtrage inverse. En effet, $x(k\Delta t)$

est fonction des données y et des estimations de x précédentes. L'équation 4.7 de filtrage obtenue par approximation de Hsieh peut être dérivée de manière exacte grâce à l'utilisation de la transformée en $\mathcal{Z}$.

La convolution de l'équation 4.6 s'écrit alors :

$$\mathcal{Z}[y(t)] = \mathcal{Z}[h(t) * x(t)] \quad (4.10)$$

$$= \mathcal{Z}[h(t)] \times \mathcal{Z}[x(t)] \quad (4.11)$$

$$= H(z)X(z) \quad (4.12)$$

Pour simplifier les calculs, nous écrivons l'équation précédente sous forme de fonction de transfert :

$$H(z) = \frac{Y(z)}{X(z)} \quad (4.13)$$

$$= \sum_{k=1}^N c_k \frac{1}{1 - e^{\frac{-\Delta t}{\tau_k}} z^{-1}} \quad (4.14)$$

D'une manière générale, nous pouvons écrire cette équation sous forme de quotient des polynômes $B(z)$ et $A(z)$:

$$H(z) = \frac{B(z)}{A(z)} = \frac{\sum_{i=0}^{N-1} b_i z^{-i}}{1 + \sum_{j=1}^N a_j z^{-j}}$$

Celle-ci équivaut à l'équation discrète suivante :

$$y_n = \sum_{i=0}^{N-1} b_i x_{n-i} - \sum_{j=1}^N a_j y_{n-j} \quad (4.15)$$

c.-à-d.

$$x_n = \frac{y_n + \sum_{j=1}^N (a_j y_{n-j} - b_j x_{n-j})}{b_0} \quad (4.16)$$

Cette équation est l'équation d'un filtrage inverse (cf. chapitre 3), puisqu'elle fait intervenir les données et les estimées précédentes pour obtenir l'estimée courante, sans prendre en compte le bruit. On constate alors que les équations 4.16 et 4.7 sont équivalentes.

Hsieh ne prend pas en compte les incertitudes sur les mesures. Cependant, aucun dispositif expérimental n'est complètement affranchi d'incertitudes. Or d'après le chapitre 3, on sait que réaliser un filtrage inverse sur des données bruitées conduit à des solutions instables.

Dans la section suivante, nous proposons une méthode d'estimation permettant le traitement de la postluminescence en tenant compte des incertitudes sur les données.

4.3.2 Méthode proposée

Afin d'utiliser les techniques d'estimation que nous avons vues dans le chapitre 3, il est nécessaire de discrétiser la modélisation de la postluminescence. C'est ce que nous verrons dans le paragraphe suivant.

4.3.2.1 Discrétisation du système linéaire

En tomographie, le signal à déconvoluer $z(t)$ se présente déjà sous forme discrète. Il est donc connu en un nombre fini de points $t_1, t_2, \dots, t_N$ espacés régulièrement.

Pour discrétiser l'équation 4.3, on considère x_n , l'entrée à restaurer correspondante aux échantillons de $x(t)$ avec le même pas d'échantillonnage que $z(t)$ et h_n , l'échantillonnage de $h(t)$. On peut donc écrire l'équation 4.3 :

$$z_n = z(t_n) = \sum_{m=1}^M h_{n-m} x_m + b_n, \quad n = 1, 2, \dots, N \quad (4.17)$$

Cette équation est appelée *convolution discrète* entre les échantillons de $x(t)$ et de la réponse impulsionnelle $h(t)$.

En concaténant les N équations 4.17, on obtient le système linéaire suivant :

$$z = Hx + b \quad (4.18)$$

Étudions maintenant la structure de la matrice H . On suppose que la réponse impulsionnelle des capteurs a un support fini, plus petit ou égal à l'intervalle de temps pendant lequel $x(t)$ est observé. Le support correspond au domaine pour lequel les valeurs des coefficients de la réponse impulsionnelle h_{n-m} sont significatives. Dans notre cas, la réponse impulsionnelle est causale de taille $P \leq N$, c.-à-d. $h = [h_1, h_1, \dots, h_P]^T$.

L'équation 4.18 s'écrit alors :

$$\begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix} = \begin{bmatrix} h_1 & 0 & \dots & 0 \\ h_2 & h_1 & 0 & & \\ \ddots & \ddots & \ddots & & \vdots \\ 0 & h_P & \dots & h_1 & \ddots \\ \vdots & \ddots & h_P & \dots & h_1 & 0 \\ 0 & \dots & 0 & h_P & \dots & h_1 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix} \quad (4.19)$$

La matrice H est donc de forme *Toeplitz* avec une structure de bande. Cette structure facilitera les calculs futurs.

4.3.2.2 Déconvolution

Dans la partie précédente, nous avons vu que la méthode de Hsieh est une technique de filtrage inverse basée sur un modèle parfait. En effet, le bruit n'est pas pris en compte dans son modèle. Cette méthode correspond à la restauration de x_k par l'inverse généralisée de H selon l'équation suivante $\hat{x}_{Hsieh} = H^\dagger z$, c'est ce qu'on a appelé, dans le chapitre 3, l'inversion directe caractérisée par une amplification du bruit. Ici, nous proposons des méthodes, basées sur la théorie de l'estimation, tenant compte du bruit.

La restauration d'entrée, ou déconvolution, est un problème particulièrement difficile puisque nous cherchons à estimer un vecteur de taille N à partir de données d'observations, elles aussi de taille N . Dans un tel cas, une approche MAP (ou régularisée) est requise, afin d'estimer la variable x .

Utilisons les résultats du chapitre 3 sur l'estimation pour développer une méthode de déconvolution. Supposons que le bruit soit blanc et gaussien. Il suit donc une distribution normale $b \sim \mathcal{N}(0, R_B)$.

Afin de simplifier les calculs, nous choisissons de conserver un cadre linéaire et gaussien. Pour cela, nous supposons que les incertitudes et l'information *a priori* sur x sont gaussiennes. Cette hypothèse sur les sinogrammes x est vérifiée en pratique, puisque ceux-ci sont doux et ne présentent pas de franches discontinuités d'une projection à l'autre.

D'après le chapitre 3, ces choix sont équivalents à minimiser un critère régularisé composite :

$$\hat{x} = \arg \min_x \|z - Hx\|^2 + \lambda \|x - \bar{x}\|^2 \quad (4.20)$$

Il existe une solution explicite de ce minimiseur qui s'écrit :

$$\hat{x}_{MAP} = \bar{x} + (H^T R_B^{-1} H + \lambda R_X^{-1})^{-1} H^T R_B^{-1} (z - H\bar{x}) \quad \text{où } \lambda = \frac{\sigma_B^2}{\sigma_X^2}$$

Dans notre modèle, nous supposons que les incertitudes b sont indépendantes les unes des autres. Ainsi, $R_B = I$. L'estimateur MAP que nous utilisons pour traiter la postluminescence peut se résumer ainsi :

$$\hat{x}_{MAP} = \bar{x} + \left(H^T H + \frac{\sigma_B^2}{\sigma_X^2} R_X^{-1} \right)^{-1} H^T (z - H\bar{x}) \quad (4.21)$$

D'après le chapitre 3, cet estimateur est, d'un point de vue théorique, plus robuste face au bruit de mesure que celui développé par Hsieh, d'un point de vue théorique.

Dans cette section, nous avons développé un estimateur en supposant la réponse impulsionnelle h connue, formant la matrice H . Cette réponse impulsionnelle est propre à chaque scanner et n'est pas une donnée connue du problème. Dans la section suivante, nous mettons en place les outils qui nous permettront d'obtenir cette réponse impulsionnelle.

4.4 Identification des paramètres du modèle

Afin de réaliser la déconvolution des données observées, il est nécessaire d'identifier les paramètres du modèle. Il s'agit alors de déterminer la réponse impulsionnelle $h(t)$.

Il est très difficile de trouver de la littérature qui caractérise la réponse impulsionnelle du système de détection des scanners. Ces données restent des données propriétaires non accessibles. Les seules courbes de décroissance concernant la postluminescence sont

peu précises. La figure 4.1 montre un exemple de courbe de décroissance publiée par Siemens (2007). Cette courbe présente les caractéristiques de détecteurs d'anciennes

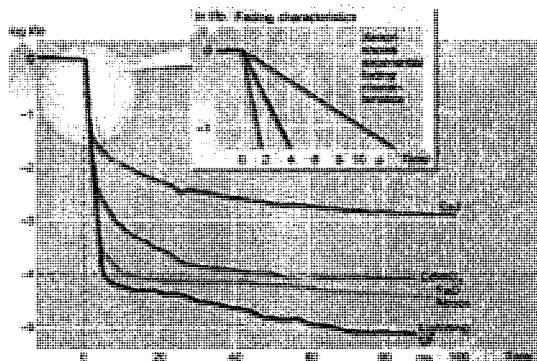


FIG. 4.1: Courbe de postluminescence la plus récemment publiée.

générations qui ne sont pas ceux utilisés sur les scanners Siemens SOMATOM dont nous nous servons à l'hôpital Notre-Dame. En effet, ces scanners sont équipés de la toute dernière technologie : les détecteurs UFC. Nous proposons donc d'estimer la réponse impulsionnelle à partir de données réelles.

4.4.1 Méthodes proposées

La caractérisation de la réponse impulsionnelle (RI) des détecteurs est réalisée par la mise en place d'estimateurs qui permettent de mesurer la RI directement à partir des sinogrammes bruts extraits du scanner. En effet, la mise en place de tels estimateurs ne nécessite pas l'introduction d'une quelconque connaissance sur la RI et ces estimateurs sont relativement simples à mettre en oeuvre.

La restauration de l'entrée qui nous intéresse dans la section 4.3 consiste à chercher x en connaissant z . La paramétrisation du système est un tout autre problème. Nous cherchons

à connaître h en supposant x connu dans l'équation :

$$z_n = \sum_{m=1}^M h_{n-m} x_m + b_n \quad n = 1, 2, \dots, N \quad (4.22)$$

Le vecteur z représente les données recueillies sur le scanner d'un fantôme dont nous connaissons parfaitement les caractéristiques géométriques ainsi que les matériaux utilisés. Le vecteur x correspond à un sinogramme idéal, également appelé analytique, obtenu par calcul analytique des projections du fantôme. Le schéma 4.2 résume la démarche employée pour l'estimation de la réponse impulsionnelle.

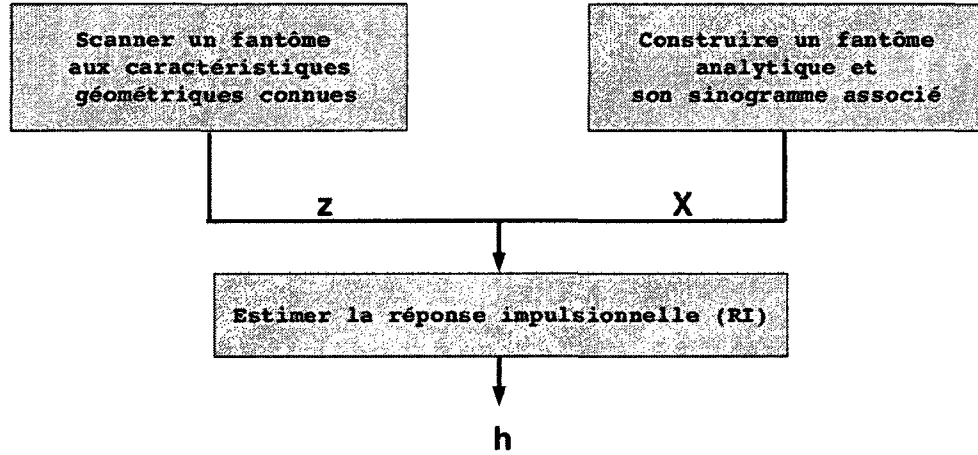


FIG. 4.2: Méthodologie employée pour estimer la réponse impulsionnelle des détecteurs.

Selon la technique de déconvolution choisie, nous n'avons pas besoin des mêmes informations concernant la réponse impulsionnelle. La déconvolution par l'estimateur MAP est basée sur la connaissance d'un vecteur $h = [h_1, h_1, \dots, h_P]^T$, représentant la fonction h en certains points. Nous développons donc un estimateur permettant d'obtenir ce vecteur directement à partir des données réelles (méthode 1).

La déconvolution développée par Hsieh se base, quant à elle, sur la connaissance des paramètres α_n et τ_n . Pour les obtenir, nous pouvons utiliser deux approches. La première

approche consiste à ajuster un modèle multiexponentiel sur le vecteur h obtenu par la méthode 1 et à en dégager les α_n et τ_n (méthode 2). Dans la deuxième approche, on estime directement les paramètres α_n et τ_n à partir des données réelles (méthode 3). Nous comparerons ces deux approches dans la section 5.

4.4.1.1 Méthode 1 : vecteur h

Cette méthode permet d'obtenir le vecteur h , nécessaire à la déconvolution par l'estimateur MAP.

L'équation 4.22 peut être réécrite sous la forme matricielle suivante :

$$z = Xh + b \quad (4.23)$$

où

$$X = \begin{bmatrix} x_1 & 0 & \dots & 0 \\ x_2 & x_1 & 0 & \dots & 0 \\ & \ddots & \ddots & \ddots & \\ x_N & \dots & & x_{N-p+1} \end{bmatrix} \quad (4.24)$$

La matrice X est une matrice contenant les valeurs exactes (sans postluminescence) du sinogramme analytique.

Cette équation est semblable à l'équation 3.10. Donc, à condition de supposer l'hypothèse gaussienne sur la distribution des incertitudes b et sur la loi *a priori* de h , on peut appliquer directement les résultats présentés dans la section 3.5.2. De plus, si on suppose que les incertitudes sont indépendantes les unes des autres (c.-à-d. $R_B = I$), les

estimateurs MV et MAP s'écrivent comme suit :

$$\hat{h}_{MAP} = (X^T X + \lambda R_H^{-1})^{-1} X^T z \quad (4.25)$$

$$\hat{h}_{MV} = (X^T X)^{-1} X^T z \quad (4.26)$$

Le coefficient de régularisation sera déterminé dans le chapitre suivant.

4.4.1.2 Méthode 2 : ajustement d'un modèle multiexponentiel

Comme nous l'avons vu dans la section 4.3.1, la technique de Hsieh se base sur un modèle multiexponentiel de la réponse impulsionnelle. Dans cette section et la suivante, nous cherchons à obtenir les α_n et τ_n qui caractérisent ce modèle. Cette méthode est un prolongement de la méthode précédente. En effet, on réalise un ajustement d'un modèle multiexponentiel sur le vecteur h obtenu par la méthode 1.

Soit les données (y_i, t_i) , $\forall i = 1, \dots, m$ représentant la réponse impulsionnelle h que nous cherchons à ajuster à un modèle multiexponentiel. Nous souhaitons obtenir une réponse impulsionnelle composée de P exponentielles décroissantes :

$$h(t) = \sum_{j=1}^P \frac{\alpha_j}{\tau_j} e^{-\frac{t}{\tau_j}} \quad \forall t \geq 0 \quad (4.27)$$

Nous cherchons ainsi les α_j et les τ_j qui caractérisent les différentes exponentielles.

Pour cela, nous nous plaçons dans un cas plus général dans lequel on cherche à ajuster un modèle composé de fonctions non linéaires f_j , dépendantes du temps et de x_j , et

pondérées par c_j :

$$y_i \triangleq \sum_{j=1}^P c_j f_j(x_j, t_i) \quad \forall i = 1, \dots, m \quad (4.28)$$

Si l'on concatène l'équation précédente pour tous les t_i , nous pouvons la réécrire en fonction des données y sous la forme matricielle suivante :

$$y = F(x)c \quad (4.29)$$

$$\text{où } x \in \mathbb{R}^P, c \in \mathbb{R}^P, F(x) = \begin{bmatrix} f_1(x_1, t_1) & \dots & f_p(x_p, t_1) \\ \vdots & & \vdots \\ f_1(x_1, t_m) & \dots & f_p(x_p, t_m) \end{bmatrix}$$

Dans notre modèle, la matrice $F(x)$ ne dépend que d'une seule inconnue x . De plus, la matrice c apparaît linéairement dans M comme étant le poids accordé aux composantes de la matrice $F(x)$. Ainsi, c est fonction du choix des variables x . Dans ce cas, nous pouvons réaliser une minimisation séparée sur $c(x)$ (solution des moindres carrés de $F(x)_k c(x)_k \simeq y$), puis sur $F(x)$.

Notre but est alors de chercher x qui minimise le critère suivant :

$$\hat{x} = \min_{x \in \mathbb{R}^P} \Phi(x) = \min_{x \in \mathbb{R}^P} \frac{1}{2} \|y - F(x)c(x)\|^2 \quad (4.30)$$

En définissant le résidu par $r(x) = y - F(x)c(x)$, on obtient $\Phi(x) = \frac{1}{2} r(x)^T r(x)$. Soit

x^* la solution de l'équation 4.30, x^* satisfait la condition d'optimalité d'ordre 1 :

$$\begin{aligned} \phi'(x^*) = 0 &\Leftrightarrow \phi'(x^*) = J(x^*)^T r(x^*) = 0 \\ \text{où } J(x^*) &= \begin{bmatrix} \nabla r_1(x^*)^T \\ \vdots \\ \nabla r_m(x^*)^T \end{bmatrix} \end{aligned} \quad (4.31)$$

Classiquement, nous pouvons utiliser une méthode de Gauss-Newton pour résoudre ce genre de problème. Par cette approche, le hessien H est approximé par $J^T(x)J(x)$. Ainsi, la direction de descente d^{GN} doit satisfaire l'équation :

$$J^T(x^*)J(x^*)d^{GN} = -J^T(x^*)r(x^*) \quad (4.32)$$

Cependant, pour des cas non idéaux, par exemple en présence de bruit, la stratégie de calcul de la direction de descente se trouve légèrement modifiée par l'ajout d'un terme de régularisation. On fait alors appel à l'équation 3.60 de l'algorithme 2 de Levenberg-Marquardt pour obtenir la direction de descente d^{LM} :

$$(J^T(x^*)J(x^*) + \lambda I)d^{LM} = -J^T(x^*)r(x^*) \quad \text{où } \lambda \geq 0.$$

Le récapitulatif de l'ajustement du modèle multiexponentiel sur la RI est exposé dans l'algorithme 4.

Algorithme 4 Récapitulatif de l'ajustement d'un modèle multiexponentiel de la RI

Initialisation : on choisit un point de départ $x_0 \in \mathbb{R}^p$, une tolérance d'arrêt $\epsilon_0 > 0$ et on met le compteur d'itérations à 0

tantque $\|G_k\| > \epsilon_0$ **faire**

Calcul d'une direction de descente par $d_k = (H_k + \lambda I)^{-1}G$ et du λ

Mise à jour du nouveau point : $x_{k+1} = x_k - H_k^{-1}d_k$

Évaluation de F au point x_{k+1}

Calcul de c_{k+1} , comme solution des moindres carrés de $F_{k+1}c_{k+1} \simeq y$

Calculs de J et H au point x_{k+1}

Mise à jour du compteur d'itération $k = k + 1$

fin tantque

4.4.1.2.1 Comment obtenir le jacobien $J(x)$ D'après l'équation 4.31, $J(x)$ s'écrit :

$$J(x^*) = \begin{bmatrix} \nabla r_1(x^*)^T \\ \vdots \\ \nabla r_m(x^*)^T \end{bmatrix}$$

c.-à-d. $(J(x))_{:,j} = (J_f(x))_{:,j} - F(x)c'_j(x)$, où $(J_f(x))_{:,j} = -F'_j(x)c$ et $(F'_j(x))_{ik} = \frac{\partial}{\partial x_j} f_k(x_j, t_i)$

Puisque $c(x)$ est solution de $F(x)c(x) \simeq y$, nous pouvons dire que :

$$F(x)^T y = F(x)^T F(x)c(x) \Rightarrow A(x)c(x) = b(x) \quad (4.33)$$

$$\Rightarrow Ac'_j = F_j'^T r + F^T (J_f(x))_{:,j}$$

En posant $F' = [(F'_1)_{:,1} \dots (F'_p)_{:,p}]$, $G = -J_f$ et $C' = [c'_1 \dots c'_p]$, cette dernière équation s'écrit : $AC' = [F'^T r] - F^T G$. Et dans ce cas, C' peut être calculé comme la solution des moindres carrés de $AC' \simeq [F'^T r] - F^T G$

Le jacobien de notre fonction objectif prend alors la forme matricielle suivante :

$$J = -G - FC'. \quad (4.34)$$

4.4.1.3 Méthode 3 : estimation directe des α_n et τ_n

Cette méthode sert à estimer les paramètres α_n et τ_n du modèle multiexponentiel proposé par Hsieh. Contrairement à la méthode précédente, nous cherchons à les estimer directement à partir des données réelles.

Dans la section 4.3.1, nous avons vu que la fonction de transfert du filtrage inverse de Hsieh peut s'écrire sous la forme :

$$H(z) = \sum_{k=1}^N c_k \frac{1}{1 - e^{\frac{-\Delta t}{\tau_k}} z^{-1}} = \frac{Z(z)}{X(z)}$$

D'une manière générale, cette fonction s'écrit sous la forme d'un quotient des polynômes $B(z)$ et $A(z)$:

$$H(z) = \frac{B(z)}{A(z)} = \frac{\sum_{i=0}^{N-1} b_i z^{-i}}{1 + \sum_{j=1}^N a_j z^{-j}}$$

Une telle fonction de transfert correspond à l'équation discrète suivante :

$$y_n = \sum_{i=0}^{N-1} b_i x_{n-i} - \sum_{j=1}^N a_j y_{n-j} \quad (4.35)$$

$$= X_n^T B - Y_n^T A \quad (4.36)$$

$$= [X_n | -Y_n]^T \begin{bmatrix} B \\ A \end{bmatrix} \quad (4.37)$$

où B et A sont, respectivement, les matrices des coefficients b_i et a_j .

En concaténant l'équation précédente, on obtient une équation de la même forme que l'équation 3.10. Donc, nous pouvons appliquer les résultats du chapitre 3. La matrice $\begin{bmatrix} B \\ A \end{bmatrix}$ peut être estimée par :

$$\begin{bmatrix} \hat{B} \\ \hat{A} \end{bmatrix}_{MAP} = ([X] - Y)[X] - Y]^T + \lambda I)^{-1} [X] - Y] y \quad (4.38)$$

Ainsi, à partir des données réelles y , nous pouvons obtenir les matrices des coefficients b_i et a_j . Par identification des polynômes et après résolution d'un système de $2 \times N$ équations à $2 \times N$ inconnues, nous sommes capables d'extraire les paramètres α_n et τ_n .

4.5 Conclusion

Ce chapitre a été consacré à la postluminescence. Dans un premier temps, nous avons modélisé celle-ci par un modèle de convolution avec une réponse impulsionnelle. Dans un deuxième temps, nous nous sommes intéressé à remettre dans le contexte de l'estimation la technique de traitement mise en place par Hsieh. Cette méthode apparaît être un filtrage inverse peu adapté dans les situations pratiques où les données d'observations sont toujours entachées de bruit. Suite à cela, nous avons proposé une technique basée sur l'estimateur MAP afin de prendre en compte le bruit des mesures et d'améliorer ce traitement de la postluminescence en supposant la réponse impulsionnelle connue.

Dans un troisième temps, nous avons développé des techniques permettant d'obtenir la réponse impulsionnelle des détecteurs à partir de données réelles. Si nous recherchons le vecteur h la représentant, alors nous pouvons utiliser un estimateur MAP, dont le paramètre de régularisation sera obtenu dans le chapitre suivant. Par contre, si nous recherchons les paramètres d'un modèle multiexponentiel caractérisant la réponse im-

pulsionnelle, alors nous pouvons soit ajuster un modèle multiexponentiel sur le vecteur h obtenu précédemment, soit les estimer par un estimateur MAP ou MV.

Le chapitre suivant est consacré aux résultats de ces développements sur données simulées puis sur données réelles.

CHAPITRE 5

RÉSULTATS ET DISCUSSION

5.1 Introduction

L'objet de ce chapitre est d'exposer la démarche méthodologique et les résultats des travaux de notre maîtrise. Il s'agit de tester les innovations mises en œuvre et de les comparer aux développements de la littérature. La démarche méthodologique sera identique pour les deux apports principaux que nous faisons dans ce mémoire :

1. Estimation de la réponse impulsionnelle des détecteurs propre au scanner utilisé à partir de données réelles ;
2. Traitement de la postluminescence des détecteurs sur sinogrammes.

Tout d'abord, nous validons les développements réalisés sur des données de simulations. Ces données exactes sont obtenues de manière analytique à partir de fantômes de tailles similaires à celles que nous rencontrons avec des données médicales. À partir de ces données, nous étudions le comportement de nos méthodes en fonction du rapport signal à bruit.

Ensuite, en nous plaçant dans des conditions expérimentales similaires à celles des acquisitions médicales, nous calibrons et ajustons les paramètres de nos méthodes.

Enfin, nous testons ces méthodes sur données réelles. Celles-ci ont été acquises en partenariat avec le Laboratoire de biorhéologie et d'ultrasonographie médicale (LBUM) de l'hôpital Notre-Dame sous la direction de Guy Cloutier. Nous concentrons nos travaux sur des fantômes scannés par des scanners Siemens SOMATOM 16, et nous comparons nos résultats à ceux de Hsieh.

Cette méthodologie est résumée dans la figure 5.1.

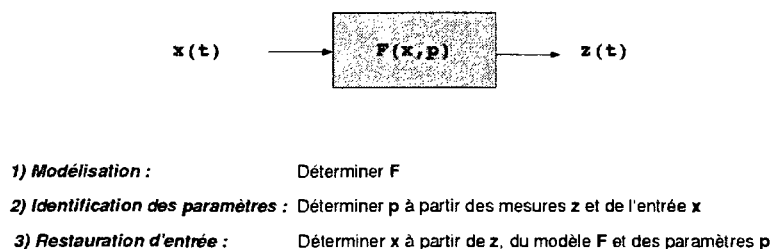


FIG. 5.1: Démarche suivie pour la validation de la restauration d'entrée dans le cas de la postluminescence.

5.2 Estimation de la réponse impulsionnelle

Cette section est consacrée à la validation de l'estimation de la réponse impulsionnelle de détecteurs. Puisque nous ne disposons pas d'une telle réponse, nous devons l'estimer.

Il s'agit de mettre en évidence qu'une technique d'estimation régularisée permet d'obtenir sur données réelles une estimation correcte de la réponse impulsionnelle compte tenu du niveau de bruit sur les données. Pour y parvenir, nous présentons plusieurs expériences sur des données de simulation et sur données réelles. La méthodologie est présentée dans le paragraphe suivant.

5.2.1 Méthodologie des simulations

Pour réaliser ces expériences, nous avons mis en œuvre un simulateur de sinogrammes. Ce simulateur calcule les projections d'un objet 3D constitué de plusieurs ellipsoïdes de matériaux homogènes pour des géométries semblables à celles des scanners Siemens.

Ce simulateur a été vérifié à partir d'un code développé par Jeffrey Fessler de l'Uni-

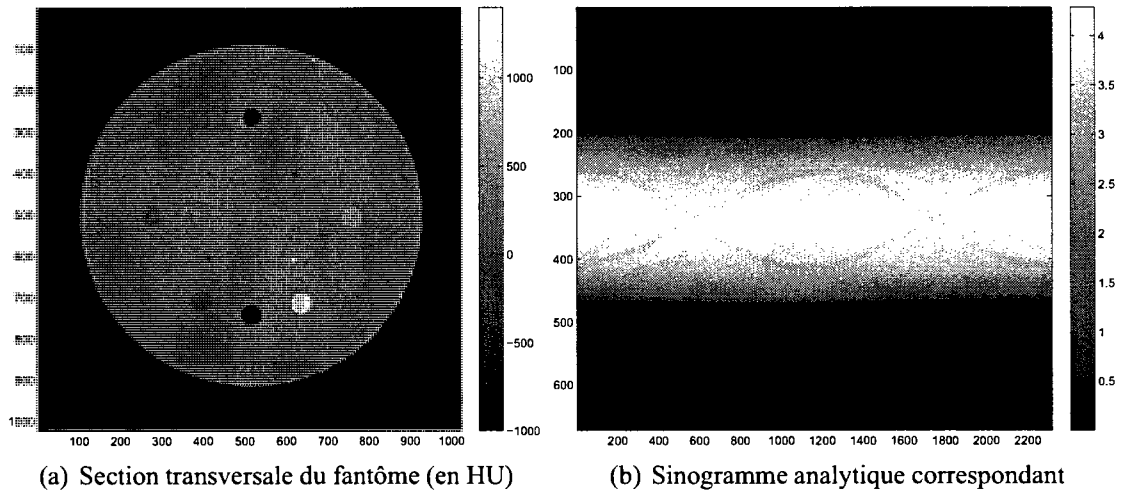


FIG. 5.2: Fantôme Catphan 600 module CTP404.

versité du Michigan (Fessler, 2008). Les comparaisons sont convaincantes et nous permettent de valider notre simulateur de tomographie. De plus, ce simulateur prend en compte les sources volantes, qu'elles soient angulaires ou longitudinales. Afin de tester ces deux fonctionnalités supplémentaires, nous nous sommes appuyé sur le code de David Gendron développé lors de sa maîtrise au Laboratoire d'imagerie par optimisation numérique et validé sur données réelles. Ainsi, notre simulateur semble affranchi d'erreurs.

Le simulateur de sinogrammes utilise les caractéristiques géométriques du scanner regroupées dans le tableau 5.2.1. Ces caractéristiques sont celles d'un tomographe Siemens (SOMATOM 16) de 16 barrettes de détecteurs. Chacune de ces barrettes contient 672 détecteurs UFC.

Le fantôme choisi pour les simulations est le fantôme Catphan 600©. Il s'agit d'un fantôme développé par The Phantom Laboratory aux États-Unis et utilisé en milieu clinique pour évaluer la performance de scanners axiaux, hélicoïdaux et multi-barrettes. Il est constitué de plusieurs modules qui permettent d'évaluer indépendamment la dose des rayons X, la sensibilité, la précision et le contraste du scanner.

TAB. 5.1: Caractéristiques géométriques des simulations pour l'estimation de la réponse impulsionnelle

ProjectionType : Fan	RayWidth : 0.077186
ProjectionMethod : Thin	nArrays : 1
nProjections : 1160	ArraySize : 0.7
nDetectors : 672	RefAngle : 0
RadiusFocus : 57	DetectorOffsetAngle : 0.0015234
RadiusDetector : 47	FFSMode : Angular
FanAngleIncrement : 0.0013541	aFFSAngle : 0.0007491

Le module de sensibilité (CTP404) est présenté à la figure 5.2 a). Nous utilisons ce module pour les simulations puisque les constructeurs du Catphan 600 nous donnent les caractéristiques géométriques précises de ce fantôme ainsi que la nature des matériaux le composant. Ces caractéristiques sont regroupées dans l'annexe I. Il nous est donc possible de construire des fantômes analytiques et des sinogrammes exacts, car les formes sont elliptiques. La figure 5.2 b) montre le sinogramme associé au module CTP404.

Une fois les données de simulation exactes créées, nous les convoluons avec une réponse impulsionnelle. Nous construisons une réponse impulsionnelle pour une vitesse de rotation de 1 s/rotation, semblable à celle présentée dans Siemens (2007) pour les détecteurs UFC (voir Fig. 5.3). La réponse impulsionnelle varie en fonction de la vitesse de rotation. Ainsi, nous interpolons cette courbe pour obtenir des réponses impulsionnelles pour des vitesses supérieures de 0,5, 0,375 et 0,1 s/rotation (Fig. 5.3). Puis, nous ajoutons un bruit gaussien généré par Matlab©. En tomographie, le bruit qui entache les données est compris entre 25 et 35 dB. Nous choisissons trois niveaux de bruit pour représenter cet intervalle : 25, 30 et 35 dB. Nous cherchons à estimer ces quatre réponses impulsionnelles bruitées. Enfin, nous comparons l'estimée avec la RI exacte en calculant l'erreur quadratique. Cette démarche permet de valider notre méthode avant de l'employer sur des données réelles.

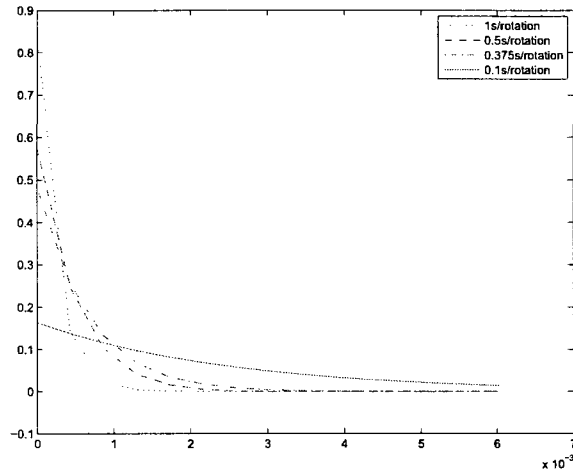


FIG. 5.3: Réponses impulsionnelles pour différentes vitesses de rotation.

5.2.2 Estimation sur données simulées

Tout d'abord, nous présentons les différents résultats pour les méthodes 1 à 3 d'estimation des réponses impulsionnelles (cf. 4.4). Puis, nous comparons les méthodes 2 et 3.

Les données de simulation correspondent au sinogramme du module CTP404 du Catphan 600 de la figure 5.2 b) convolué avec la réponse impulsionnelle de la figure 5.4 b) à la vitesse de 0,5 s/rotation pour former nos données d'observations $y(t)$ non bruitées. La ligne 400 de ce sinogramme convolué est présentée dans la figure 5.4 a).

5.2.2.1 Méthode 1

Dans cette méthode, nous cherchons à estimer directement le vecteur h représentant la réponse impulsionnelle.

Pour des données bruitées, une estimation de h utilisant la régularisation est préconisée.

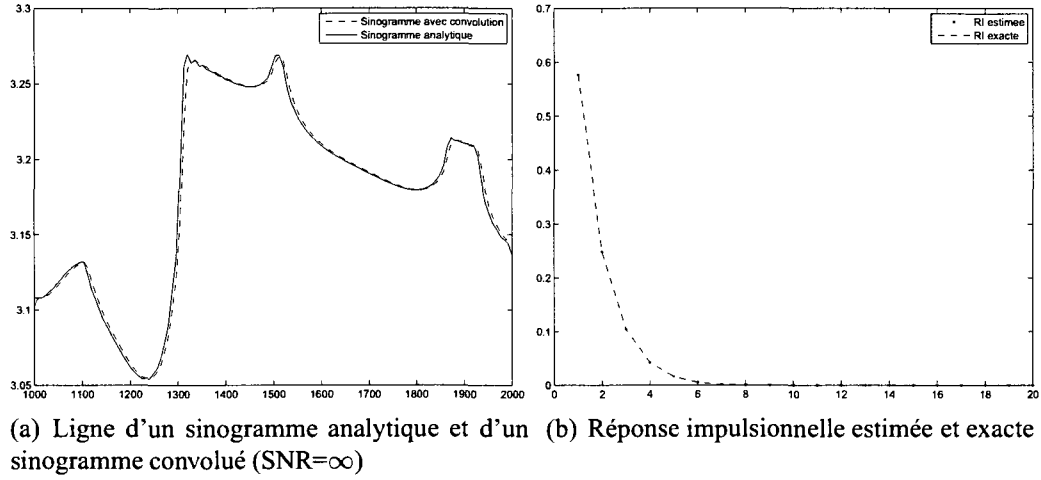


FIG. 5.4: Validation de la technique d'estimation de réponse impulsionnelle sur des données sans bruit ($v = 0, 5$ s/rotation).

Ainsi, dans cette méthode, l'estimée de h est obtenue par l'équation 4.26 :

$$\hat{h}_{MAP} = (X^T X + \lambda R_H^{-1})^{-1} X^T z \quad (5.1)$$

Grâce à la régularisation, la réponse impulsionnelle devient moins bruitée et donc plus lisse. Cependant, il faut accorder une grande importance au choix du paramètre λ de régularisation. En effet, s'il est trop petit, notre estimateur est équivalent à un estimateur non régularisé, et si λ est trop important, on accorde plus d'importance à l'information *a priori* et la solution est dégradée.

Pour choisir le paramètre de régularisation adéquat, nous retenons celui qui minimise l'erreur quadratique entre le sinogramme analytique et le sinogramme déconvolué. La figure 5.5 illustre le comportement du paramètre de régularisation sur l'erreur quadratique en fonction des rapports signal à bruit. On constate sur cette figure que les approches non régularisées ($\lambda = 0$) donnent de moins bons résultats qu'une approche régularisée. En effet, l'erreur quadratique minimale est obtenue pour des valeurs de λ différentes de 0. Le tableau 5.2 résume la valeur minimale de l'erreur quadratique et le coefficient λ

TAB. 5.2: Erreurs quadratiques moyennes pour l'estimation des RI en fonction du SNR pour une vitesse de rotation de 0,5 s/rotation

SNR	∞	35 dB	25 dB
λ	0	0.7	3
EQM	1.2e-16	3.7e-5	1.5e-4

associé pour trois niveaux de bruit : 25, 30 et 35 dB.

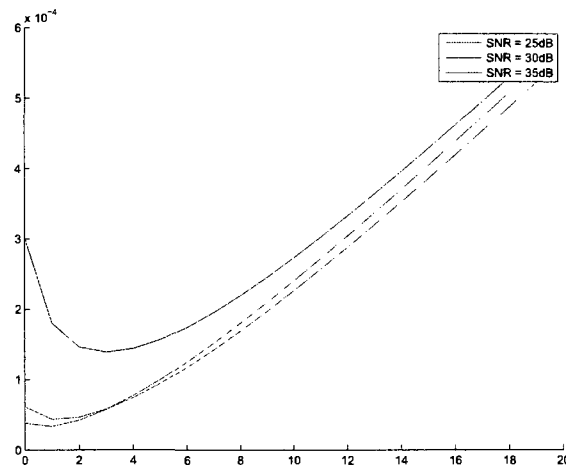


FIG. 5.5: Influence du paramètre de régularisation sur l'erreur quadratique en fonction du SNR.

Les résultats de l'estimateur MAP pour un niveau de bruit de 25 dB et un coefficient de régularisation de $\lambda = 3$ sont présentés dans la figure 5.6. Pour ce λ , l'erreur quadratique est minimale et est de l'ordre de $1,5 \times 10^{-4}$. L'erreur quadratique est suffisamment faible pour conclure que l'estimation du vecteur h représentant la réponse impulsionnelle est satisfaisante, même avec un niveau de bruit élevé.

Dans les deux méthodes qui suivent, nous allons chercher à obtenir les paramètres α_n et τ_n d'une réponse impulsionnelle sous forme multiexponentielle nécessaire à la déconvolution développée par Hsieh.

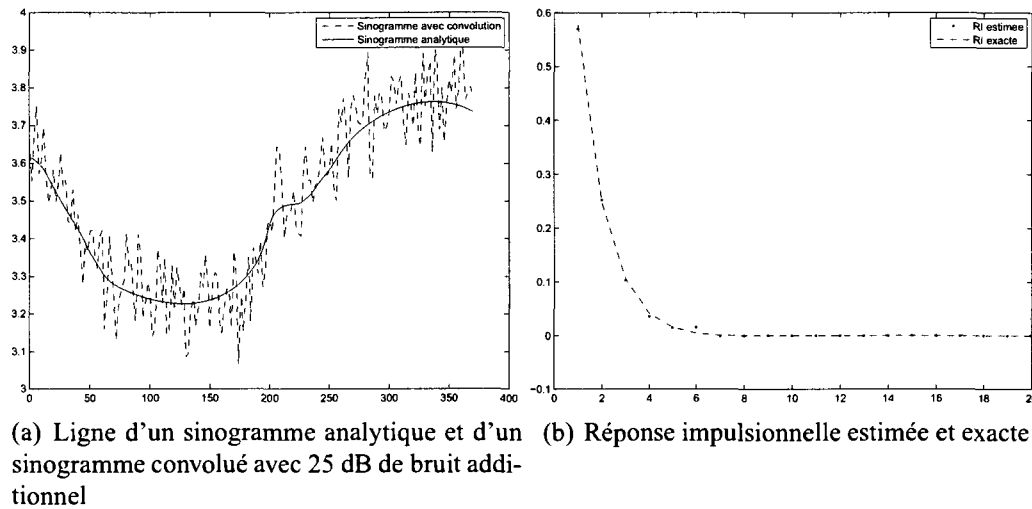


FIG. 5.6: Validation de la technique d'estimation de réponse impulsionnelle sur des données avec bruit.

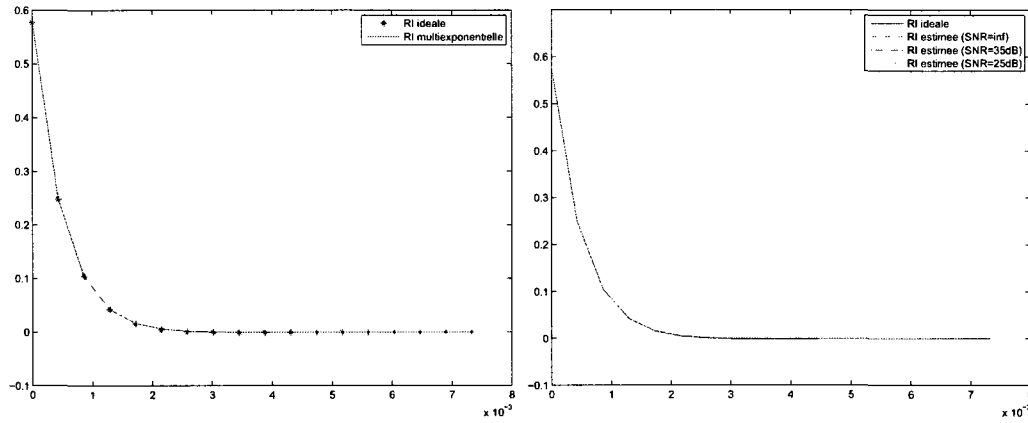
5.2.2.2 Méthode 2

Cette méthode est une continuité de la méthode précédente, puisque nous réalisons un ajustement d'un modèle multiexponentiel sur le vecteur h estimé.

Pour choisir le nombre d'exponentielles de la modélisation, nous prenons le nombre minimal d'exponentielles pour lequel l'erreur de modélisation est inférieure à 10^{-8} . Ainsi, nous avons choisi de nous limiter à deux exponentielles.

L'arrêt de l'algorithme de Levenberg-Marquardt est réalisé lorsque la norme du gradient de l'objectif est en dessous de 10^{-10} . Sous ces conditions, nous effectuons un ajustement du modèle multiexponentiel sur des réponses impulsionnelles de détecteurs selon les quatre vitesses de rotation. Les résultats pour une vitesse de rotation de 0,5 s/rotation sont présentés dans la figure 5.7. À cette vitesse, les constantes de temps valent $\tau_1 = 2,69e - 03$ s, $\tau_2 = 3,71e - 03$ s, et leur poids associés $c_1 = 2,04e - 01$ et $c_2 = -4,07e - 02$. On remarque dans la figure 5.7 a) que, pour un SNR de 25 dB, le modèle multiexponentiel de la réponse impulsionnelle estimée coïncide presque parfait-

tement avec la réponse impulsionnelle exacte. Des résultats identiques sont observables dans la figure 5.7 b) pour différents rapports signal à bruit. Sur cette figure, les réponses impulsionnelles estimées pour les différents rapports signal à bruit sont toutes confondues avec la réponse impulsionnelle idéale.



(a) Réponse impulsionnelle estimée et son modèle multiexponentiel ($SNR = 25$ dB) (b) Réponse impulsionnelle estimée et son modèle multiexponentiel, en fonction du SNR

FIG. 5.7: Modèle multiexponentiel de la réponse impulsionnelle des détecteurs ($v = 0, 5$ s/rotation),

Les erreurs quadratiques ont été calculées entre la réponse impulsionnelle exacte et celle correspondant à l'ajustement. Le tableau 5.3 résume ces erreurs. Il apparaît que les erreurs sur les constantes de temps τ_i et leurs poids c_i respectifs sont faibles (respectivement $< 10^{-8}$ et $< 10^{-6}$). Cependant, ce qui nous intéresse réellement, c'est l'erreur commise par rapport à la réponse impulsionnelle exacte. Ces erreurs quadratiques sont faibles, inférieures à $2 \cdot 10^{-20}$. Ainsi, la méthode 2 est performante. De plus, on remarque que ces erreurs sont plus faibles que dans la méthode 1, puisque l'ajustement de la RI à un modèle multiexponentiel tend à lisser les quelques fluctuations de la RI estimée par la méthode 1, c.-à-d. les dernières valeurs de la RI.

TAB. 5.3: Erreurs commises lors de l'ajustement de la RI par un modèle multiexponentiel (méthode 2)

SNR	∞		35 dB		25 dB	
Exponentielles	1	2	1	2	1	2
$Erreur_{\tau_i}$	2.78e-10	5.46e-10	9.57e-10	9.64e-10	1.57e-09	9.71e-09
$Erreur_{c_i}$	9.57e-07	4.85e-07	8.00e-07	4.21e-07	1.41e-07	7.92e-07
EQ	6.32e-21		9.75e-21		1.34e-20	

5.2.2.3 Méthode 3

La méthode 3 repose sur l'estimation directe des paramètres α_n et τ_n d'une réponse impulsionnelle. L'estimateur proposé semble de moins bonne qualité que celui de la méthode précédente. En effet, le tableau 5.4 présente les erreurs quadratiques entre la réponse impulsionnelle idéale et celle obtenue par l'estimation des paramètres α_n et τ_n . Ces erreurs sont beaucoup plus grandes que celles obtenues par la méthode précédente. Cet effet est remarquable sur la figure 5.8. On y observe un écart significatif.

TAB. 5.4: Erreurs quadratiques entre la réponse impulsionnelle idéale et la réponse impulsionnelle estimée par la méthode 3 en fonction du SNR

SNR	∞		35 dB		25 dB	
Exponentielles	1	2	1	2	1	2
$Erreur_{\tau_i}$	6.80e-05	5.46e-04	1.11e-04	7.22e-04	2.38e-04	8.17e-04
$Erreur_{\alpha_i}$	8.38e-04	8.33e-04	3.51e-04	3.35e-04	4.14e-04	3.67e-04
EQ	3.27e-07		5.19e-03		2.59e-02	

Pour l'obtention des α_n et τ_n , les résultats de simulation sont meilleurs pour la méthode 2. Ainsi, c'est cette méthode que nous retiendrons pour l'estimation sur données réelles.

5.2.3 Estimation sur données réelles

L'objectif de cette section est d'estimer la réponse impulsionnelle des détecteurs sur des données médicales. Rappelons que ces données ont été acquises à l'hôpital Notre-Dame

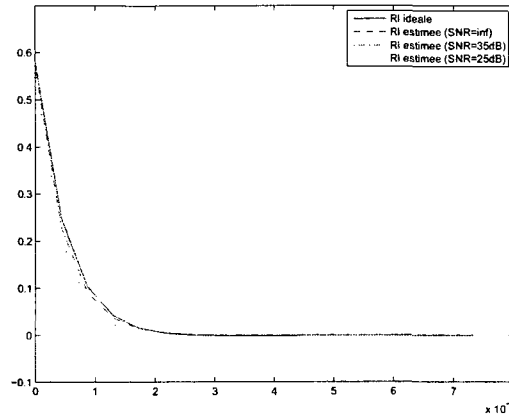


FIG. 5.8: Estimation directe des paramètres du modèle multiexponentiel de la réponse impulsionnelle des détecteurs en fonction du SNR ($v = 0,5$ s/rotation).

de Montréal.

Les simulations ont montré qu'en présence de bruit sur les données, il est nécessaire de régulariser pour obtenir une estimation convenable de la réponse impulsionnelle. Nous voulons mettre en évidence qu'une technique d'estimation régularisée de la réponse impulsionnelle est adaptée pour des données réelles forcément bruitées et que l'estimation de celle-ci est de bonne qualité.

Pour commencer l'estimation de la réponse impulsionnelle sur données réelles, plaçons-nous dans un cas favorable dans lequel le bruit est faible. Il s'agit de données provenant du protocole *InnerEarSeq* de Siemens. Ce protocole est utilisé en clinique pour obtenir des images de l'oreille interne de patients. On évalue le rapport signal à bruit de ce protocole à 35 dB. Cela correspond à un rapport signal à bruit favorable pour l'estimation. La vitesse de rotation (1 s/rotation) est la plus faible que nous ayons sur le scanner SOMATOM 16 de Siemens. Ainsi, la postluminescence sera peu marquée.

L'estimation de la réponse impulsionnelle sur ce protocole est réalisée en choisissant $\lambda = 1, 2$, qui correspond à un compromis juste entre le lissage de la réponse impulsionnelle et l'ajout trop important d'information sur la solution. En effet, pour un λ plus faible,

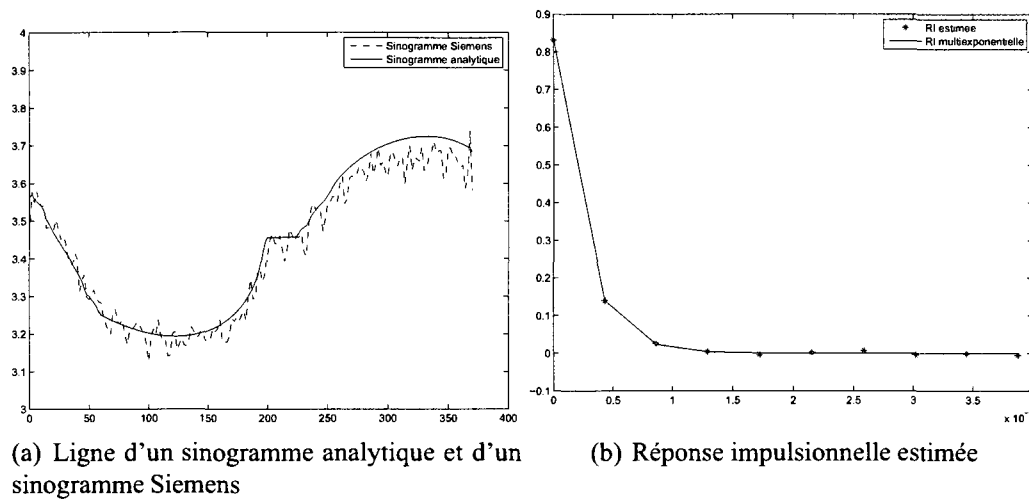


FIG. 5.9: Estimation de la réponse impulsionnelle sur des données Siemens provenant du protocole *InnerEarSeq*.

la solution oscille et pour un λ plus grand, l'amplitude de la réponse impulsionnelle estimée commence à diminuer.

Pour l'obtention des valeurs des coefficients α_n et τ_n , nous avons utilisé la méthode 2, puisqu'elle donne les meilleurs résultats en simulation. Les valeurs estimées des constantes de temps des exponentielles et des poids associés à ces exponentielles sont regroupés dans le tableau 5.5. Pour chaque ligne du sinogramme (c.-à-d. pour chaque dé-

TAB. 5.5: Valeurs des paramètres α_n et τ_n obtenus par la méthode 2 pour le protocole *InnerEarSeq*

Exponentielles	1	2
τ_i	4.51e-5	8.81e-3
α_i	3.87e-4	-1.06e-5

tecteur), nous estimons une réponse impulsionnelle. La figure 5.10 montre la réponse impulsionnelle moyenne obtenue et les écarts-types. On constate que ceux-ci sont faibles. De ce fait, nous pouvons utiliser une réponse impulsionnelle moyenne pour tous les détecteurs.

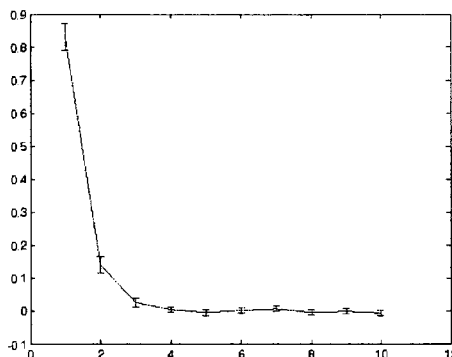


FIG. 5.10: Estimation de la RI sur le protocole *InnerEarSeq*.

La figure 5.9 nous montre que, même pour une faible vitesse de rotation, la postluminescence est déjà présente. En effet, dans le cas où nous n'aurions pas de postluminescence, la réponse impulsionnelle serait un dirac électronique.

Pour l'estimation des réponses impulsionnelles pour des vitesses plus grandes, les résultats sur données réelles sont surprenants. La figure 5.11 montre les résultats de l'estimation d'une RI sur des données provenant du protocole *Trauma* de Siemens. Pour réaliser l'estimation de la réponse impulsionnelle, nous choisissons $\lambda = 1, 7$, puisque le rapport signal à bruit est plus important que celui du protocole précédent (25 dB). Ce protocole a une vitesse de rotation de 0,375 s (vitesse la plus grande à notre disposition). Sur cette figure, on constate que la réponse impulsionnelle est pratiquement la même que celle estimée à partir du protocole *InnerEarSeq*, alors que nous nous attendions à obtenir une réponse impulsionnelle de la même forme que celle de la figure 5.3 ($v = 0,375$ s/rotation).

Les valeurs estimées des constantes de temps des exponentielles et des poids associés à ces exponentielles sont regroupées dans le tableau 5.6.

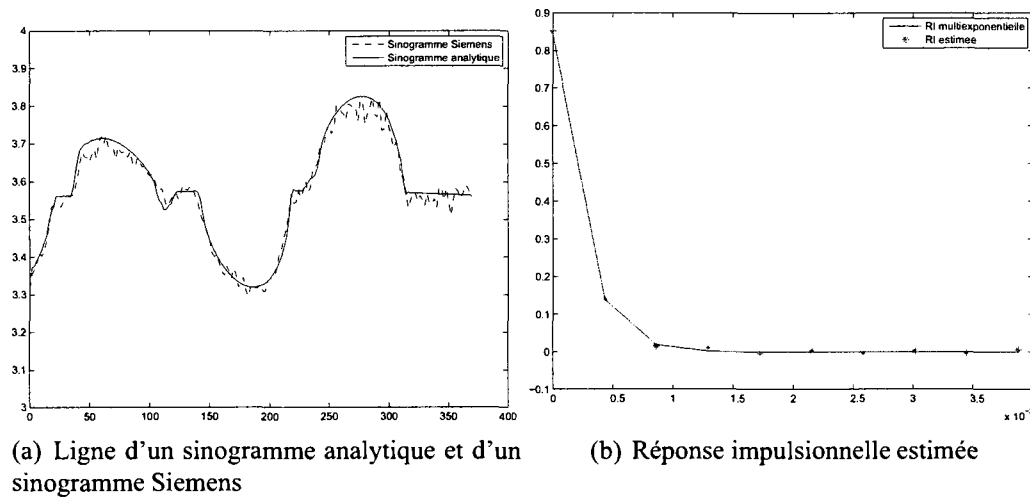


FIG. 5.11: Estimation de la réponse impulsionnelle sur des données Siemens provenant du protocole *Trauma*.

TAB. 5.6: Valeurs des paramètres α_n et τ_n obtenus par la méthode 2 pour le protocole *Trauma*

Exponentielles	1	2
τ_i	7.42e-4	9.63e-4
α_i	4.86e-4	-1.67e-4

5.2.4 Conclusions

Dans cette première partie, nous avons exposé les résultats concernant l'identification des paramètres de notre système.

La réponse impulsionnelle caractérisée par un vecteur h a été obtenue par une méthode d'estimation (méthode 1). Le choix d'une méthode régularisée telle que le maximum *a posteriori* est important puisque les observations sont entachées de bruit. Bien que le niveau de bruit soit important dans certains protocoles (25 dB), l'estimation de la réponse impulsionnelle reste bonne grâce à bon dosage de l'apport d'information *a priori* sur la solution. Le coefficient de régularisation a été choisi de manière empirique en minimisant l'erreur quadratique à la solution. Afin de simplifier les calculs lors du trai-

tement de la postluminescence, nous avons choisi de conserver une réponse impulsionnelle moyenne pour tous les détecteurs. Les variations entre les différentes réponses impulsionnelles sont minimales et influencent peu la qualité des reconstructions.

Pour ce qui est de l'estimation directe des coefficients α_n et τ_n , la méthode 2 est privilégiée, puisqu'elle est plus robuste face au bruit. Cette méthode permet l'ajustement d'un modèle multiexponentiel sur le vecteur h estimé par la méthode 1.

Les résultats sur données réelles sont mitigés par l'obtention d'une réponse impulsionnelle similaire pour deux protocoles utilisant des vitesses de rotation différentes. Ce résultat étrange ne l'est pas tout à fait. En effet, les données que nous possédons de Siemens ne sont pas complètement « brutes ». Siemens effectue sur ces données des prétraitements dont nous ignorons les détails et l'effet sur les données.

Maintenant que le modèle de la postluminescence est complètement caractérisé, nous pouvons aborder le traitement de la postluminescence proprement dit.

5.3 Traitement de la postluminescence

Dans la section précédente, nous avons complété notre modèle à réponse impulsionnelle en l'estimant sur les données brutes provenant d'un scanner Siemens. Ici, nous nous occupons exclusivement de la restauration de sinogramme en éliminant l'effet de la postluminescence.

L'objet de cette section est de montrer que la méthode que nous avons proposée permet de traiter la postluminescence correctement. Dans cette section, nous cherchons à mettre en évidence que :

1. La méthode proposée est efficace pour traiter la postluminescence ;
2. La méthode est robuste vis-à-vis des niveaux de bruit présents sur les données

d'observations.

5.3.1 Méthodologie des simulations

Afin d'atteindre les deux objectifs précédents, nous proposons de faire deux séries d'expériences. La première, sur données de simulation, consiste à prouver que la méthode développée est efficace pour plusieurs rapports signal à bruit. La deuxième série, sur données réelles, permet de conclure sur l'amélioration effective du traitement de la postluminescence sur des données médicales.

L'objectif final de ce projet est de montrer que le traitement de la postluminescence amène une amélioration sur la résolution spatiale des reconstructions de données médicales. Les reconstructions de cette section sont réalisées de manière algébrique en minimisant le critère suivant :

$$\hat{\mu} = \min_{\mu} \frac{1}{2} \|A\mu - y\|^2 + \gamma R(\mu) \quad (5.2)$$

où R est une fonction de pénalisation de type L_2L_1 . L'algorithme de minimisation employé est l'algorithme L-BFGS, du fait de la taille des matrices mises en jeu. Le critère d'arrêt de cet algorithme est une tolérance de 10^{-6} sur la norme du gradient de la fonction objectif. Le critère de régularisation est fixé à 0,01 pour un niveau de bruit de 25 dB.

La résolution spatiale des reconstructions est évaluée grâce à la fonction de transfert de modulation (MTF, cf. section 2.6.2). On rappelle que la MTF idéale est une ligne droite, ce qui signifie que même les plus petits détails d'une image sont parfaitement observables. La résolution se dégrade lorsque la MTF prend l'allure d'un filtre passe-bas. Ensuite, en simulation, on calcule l'erreur quadratique entre l'image réelle et l'image reconstruite.

5.3.2 Estimation sur données simulées

L'objectif de cette section est de valider notre méthode de déconvolution sur des données de simulation tout d'abord non bruitées, puis bruitées. Il s'agit aussi de la comparer à la méthode de Hsieh.

Pour les données simulées, avec bruit additionnel gaussien compris entre 25 et 35 dB, nous utilisons le fantôme Catphan 600 CTP404 avec la même géométrie que dans la section précédente (cf. tableau 5.2.1).

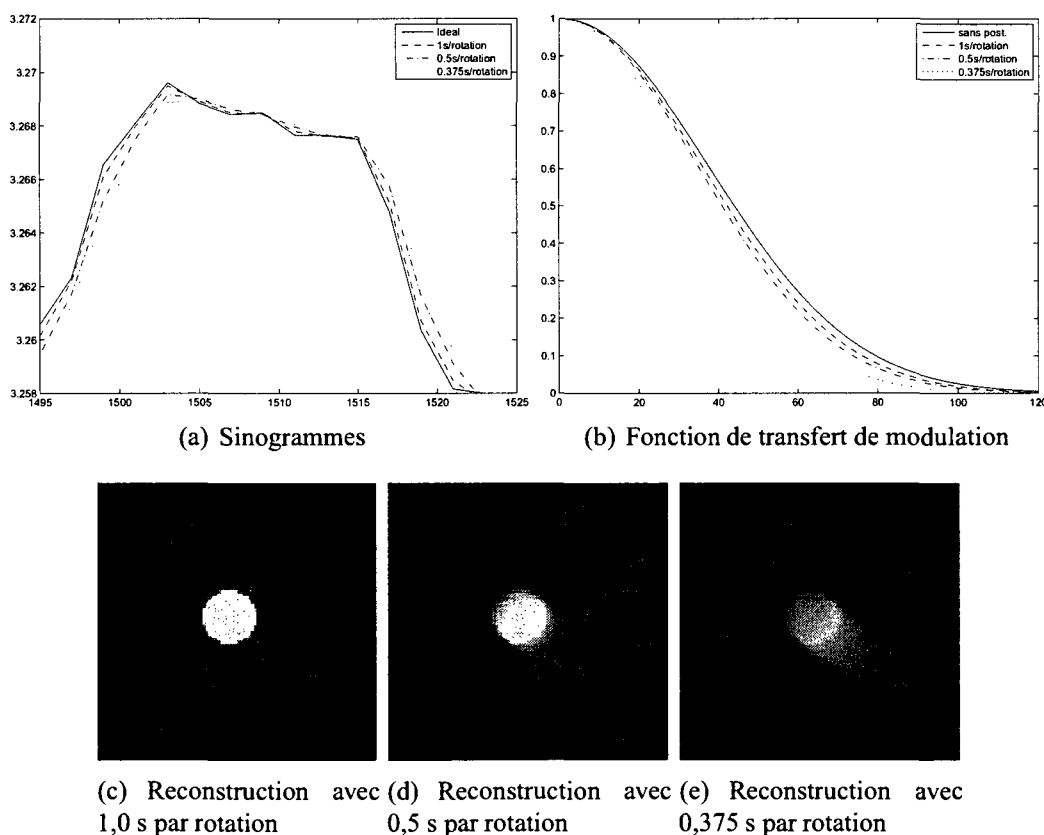


FIG. 5.12: Effet de la postluminescence en fonction des vitesses de rotations : zoom sur un câble de la section CTP404 du Catphan 600.

Premièrement, nous vérifions l'influence de la vitesse de rotation sur la qualité des images. Pour cela, nous convoluons le fantôme Catphan 600 CTP404 avec différentes

réponses impulsionnelles qui correspondent à des vitesses de rotations de 1 s à 0,375 s/rotation (Fig. 5.12 a)). Ces noyaux de convolutions, présentés dans la figure 5.7 a), sont obtenus par interpolation de la réponse impulsionnelle des détecteurs pour une vitesse de 1 s/rotation.

La figure 5.12 montre un zoom de reconstructions du Catphan 600 et indique que plus la vitesse augmente, plus la qualité de l'image est détériorée. Pour une vitesse faible de 1 s/rotation, l'effet de la postluminescence n'est pas marqué. Cependant, pour des vitesses plus élevées, sur les figures 5.12 d) et e), il apparaît des traînées blanches, rémanences de projections antérieures. De plus, on constate que les courbes de MTF sont de plus en plus basses, ayant pour effet de diminuer la résolution des reconstructions (Fig. 5.12 b)).

La déconvolution des sinogrammes sur données simulées sans bruit donne de très bons résultats avec notre méthode ainsi qu'avec celle de Hsieh. Ainsi, la méthode de restauration de sinogramme fonctionne correctement ($EQM_{MAP} \propto 10^{-15}$ et $EQM_{Hsieh} \propto 10^{-7}$).

Puisque la technique de Hsieh est une technique d'estimation non régularisée, elle a tendance à augmenter le bruit des données observées (entre 25 et 35 dB). En effet, on constate, dans le tableau 5.7, que l'erreur quadratique moyenne entre le sinogramme exact et le sinogramme déconvolué par la technique de Hsieh dégrade un petit peu le sinogramme.

TAB. 5.7: Erreurs quadratiques moyennes entre le sinogramme exact et le sinogramme dégradé (EQM_{conv}); entre le sinogramme exact et celui corrigé par l'estimateur MAP (EQM_{MAP}); entre le sinogramme exact et celui corrigé par la méthode de Hsieh (EQM_{Hsieh}) pour une vitesse de rotation de 0,5 s/rotation en fonction du SNR

SNR	∞	35 dB	25 dB
λ	0	0.31	1.37
EQM_{conv}	5.26e-05	6.08e-5	1.05e-4
EQM_{MAP}	4.37e-15	2.33e-5	7.39e-5
EQM_{Hsieh}	1.63e-7	1.71e-3	2.39e-3

De plus, sur la figure 5.13 a), on remarque que le niveau de bruit du sinogramme déconvolué est grandement augmenté par cette méthode. L'augmentation du bruit dans le

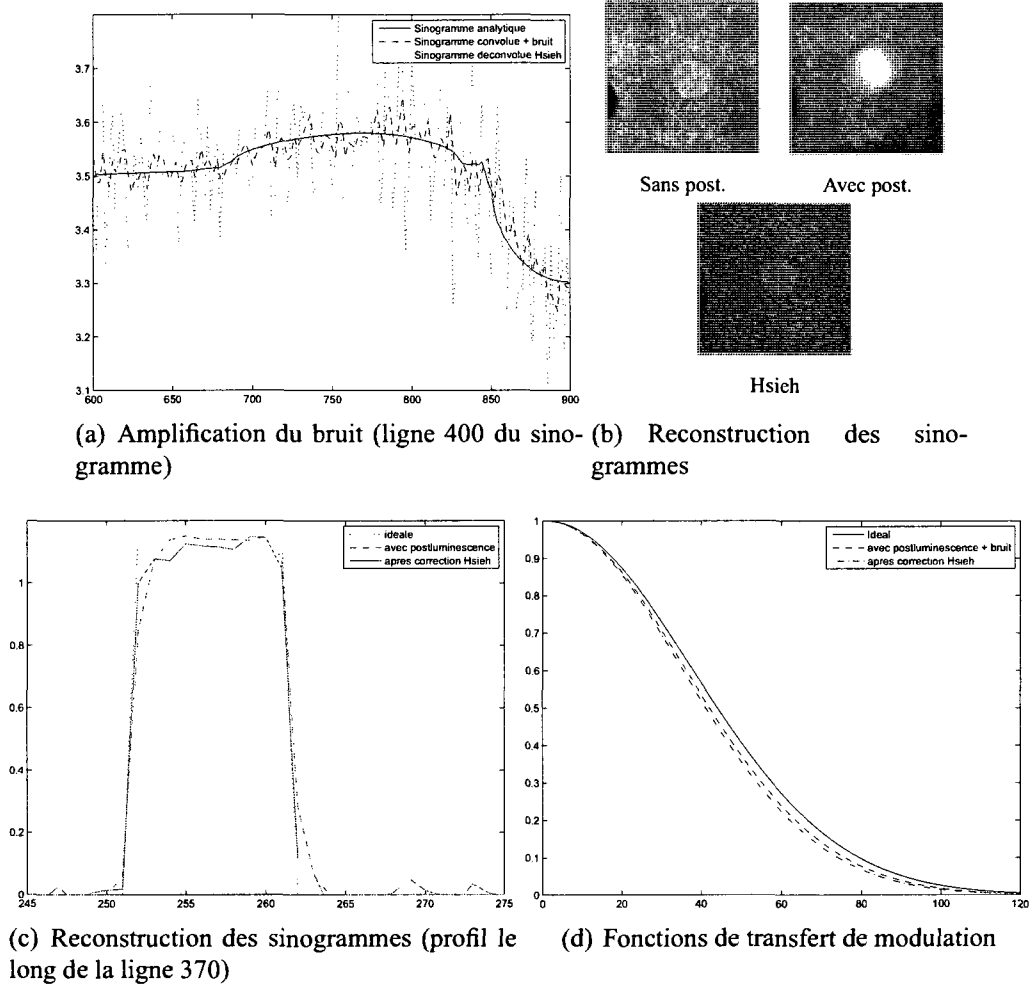


FIG. 5.13: Déconvolution de Hsieh sur données simulées avec un SNR de 25dB ($v = 0,5$ s/rotation).

sinogramme déconvolué par Hsieh amène l'apparition de bruit lors de la reconstruction (Fig. 5.13 b)). De plus, la déconvolution proposée par Hsieh entraîne une dégradation de la MTF (Fig. 5.13 c)).

La restauration de sinogramme par la méthode MAP améliore la qualité du sinogramme déconvolué. Dans le tableau 5.8, nous calculons l'erreur quadratique moyenne entre

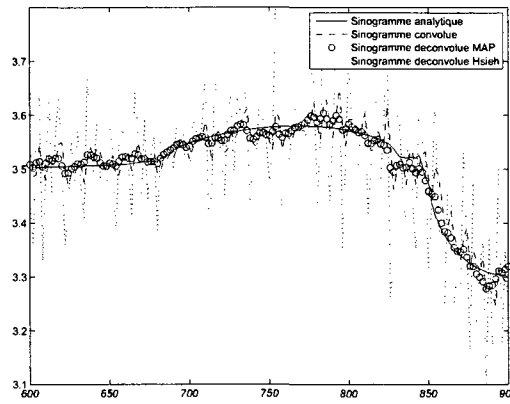
l'image réelle et l'image reconstruite, après convolution et ajout de bruit sur le sinogramme, puis lorsque le sinogramme a été corrigé par notre méthode. L'erreur quadratique entre la déconvolution MAP et le sinogramme exact est toujours inférieure ou égale à celle de la déconvolution proposée par Hsieh.

TAB. 5.8: Erreurs quadratiques moyennes entre l'image réelle et l'image reconstruite avec un sinogramme dégradé (EQM_{conv}); entre l'image réelle et celle reconstruite avec un sinogramme corrigé par l'estimateur MAP (EQM_{MAP}); entre l'image réelle et celle reconstruite avec un sinogramme corrigé par la technique de Hsieh (EQM_{Hsieh}) pour une vitesse de rotation de 0,5 s/rotation en fonction du SNR

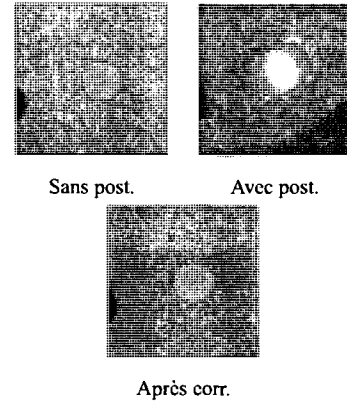
SNR	∞	35 dB	25 dB
λ	0	0.31	1.37
EQM_{conv}	3.52e-4	7.48e-4	2.25e-03
EQM_{MAP}	4.88e-4	5.23e-4	6.62e-04
EQM_{Hsieh}	8.62e-4	2.25e-3	3.95e-3

De plus, comme le montre la figure 5.14 a), le sinogramme déconvolué par l'estimateur MAP présente moins de bruit et la restauration du sinogramme semble correcte. Le choix du paramètre λ de l'estimateur MAP est obtenu par minimisation de l'erreur quadratique entre le sinogramme analytique et le sinogramme déconvolué, comme dans la section suivante. Bien évidemment, les paramètres de régularisation sont propres à chaque niveau de bruit. Les reconstructions des sinogrammes déconvolués par la méthode MAP sont proches de l'objet simulé comme le montre la figure 5.14 b), pour une vitesse de 0,5 s/rotation et un SNR de 25 dB.

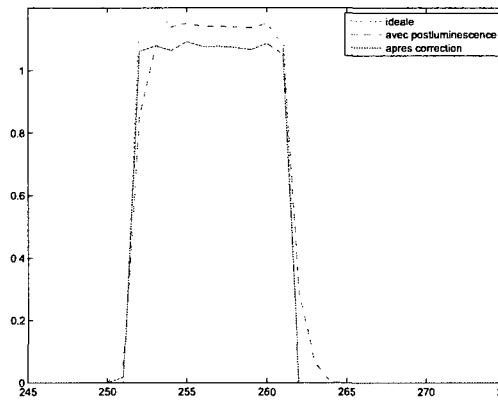
La méthode proposée fonctionne correctement et permet de restaurer le sinogramme dégradé. De plus, la restauration augmente la résolution des images reconstruites, qui est proche de celle idéale. Testons cette méthode sur données réelles.



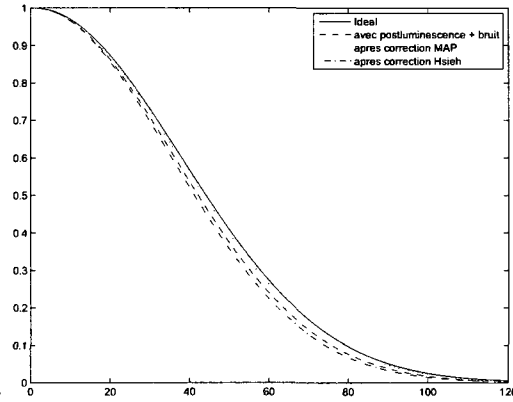
(a) Sinogrammes (ligne 400)



(b) Reconstruction des sinogrammes



(c) Reconstruction des sinogrammes (profil le long de la ligne 370)



(d) Fonctions de transfert de modulation

FIG. 5.14: Reconstruction d'images associées à un sinogramme dégradé par la postluminescence (SNR = 25 dB) puis corrigé par un estimateur MAP ($v = 0, 5$ s/rotation).

5.3.3 Estimation sur données réelles

Dans la section précédente, la méthode de traitement de la postluminescence a été validée et nous avons montré qu'elle donne de bons résultats même avec des données relativement bruitées. Ici, nous nous intéressons à employer notre méthode sur des données provenant du scanner Siemens CT16 SOMATOM de l'hôpital Notre-Dame, et à la comparer à la méthode proposée par Hsieh.

Le premier protocole sur lequel nous effectuons les tests est le protocole de Siemens *InnerEarSeq* dont les caractéristiques géométriques sont les mêmes que celles utilisées en simulation (cf. tableau 5.2.1). Il est utilisé en mode axial. Une estimation du rapport signal à bruit donne 35 dB et la vitesse de rotation est de 1 s/rotation, ce qui n'est pas favorable à l'apparition de rémanence dans les mesures. Le paramètre λ est fixé à 0,36, de la même façon que dans la section 5.2.3.

La figure 5.15 a) montre le sinogramme à déconvoluer.

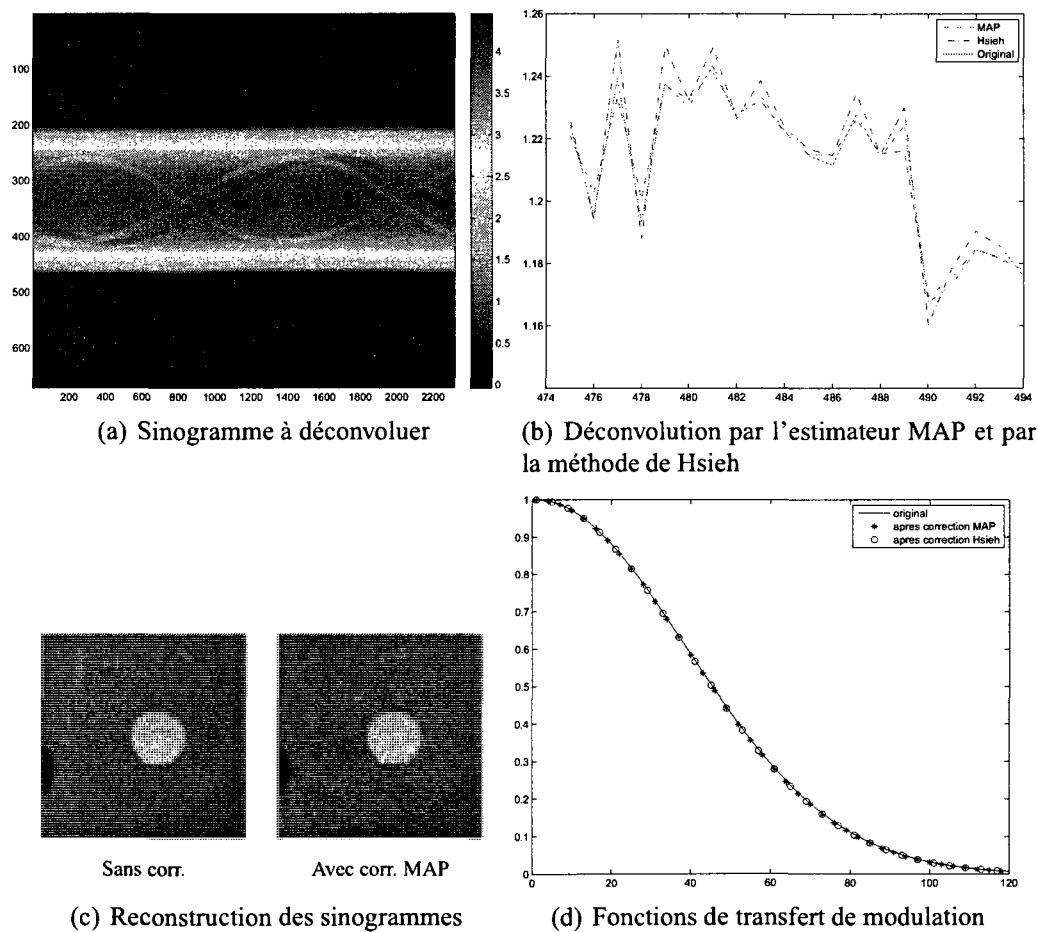


FIG. 5.15: Déconvolution sur données réelles : protocole *InnerEarSeq*.

La déconvolution MAP n'est pas un grand succès. En effet, lors de la reconstruction

de ce sinogramme déconvolué, la qualité de l'image reconstruite est similaire à celle de Siemens. Les MTF pour ces deux reconstructions sont indiscernables (Fig. 5.15 c)). Cependant, les résultats de la déconvolution de Hsieh nous donne les mêmes résultats. Ceci est rassurant pour nous. Nous nous attendions à ce résultat, puisque l'effet de la postluminescence n'est pas fortement marqué pour cette vitesse de rotation.

La déconvolution MAP a aussi été testée sur le protocole *Trauma* de Siemens. Ce protocole est utilisé en mode hélicoïdal pour imager le cerveau. La vitesse de rotation est supérieure à celle du protocole *InnerEarSeq* ($v = 0,375$ s/rotation). À cette vitesse, la postluminescence est censée être visible, puisque nous l'avons perçue en simulation (cf. Fig. 5.12 d)). Les déconvolutions MAP et celle proposée par Hsieh sont réalisées sur un sinogramme hélicoïdal correspondant au Catphan 600 CTP 404. Le rapport signal à bruit est évalué à 30 dB. Nous choisissons pour la déconvolution, $\lambda = 1, 2$, comme nous l'avons fait dans la section 5.2.3.

La figure 5.16 présente les résultats de cette déconvolution. Nous remarquons que la correction est minime : les MTF sont pratiquement confondues et les sinogrammes quasi identiques. D'après les résultats de la partie 5.2.3, nous nous attendions à cela.

5.4 Conclusion

Les résultats de la première partie de ce chapitre montrent que nous avons réussi à caractériser la postluminescence des détecteurs, en simulation. À partir du fantôme Catphan 600 CTP404, nous sommes capable d'estimer la réponse impulsionnelle de détecteurs, même avec des niveaux de bruit élevés. Cependant, pour les données réelles, les réponses impulsionnelles estimées des détecteurs sont identiques pour une vitesse de 1 s/rotation et pour une autre de 0,375 s/rotation.

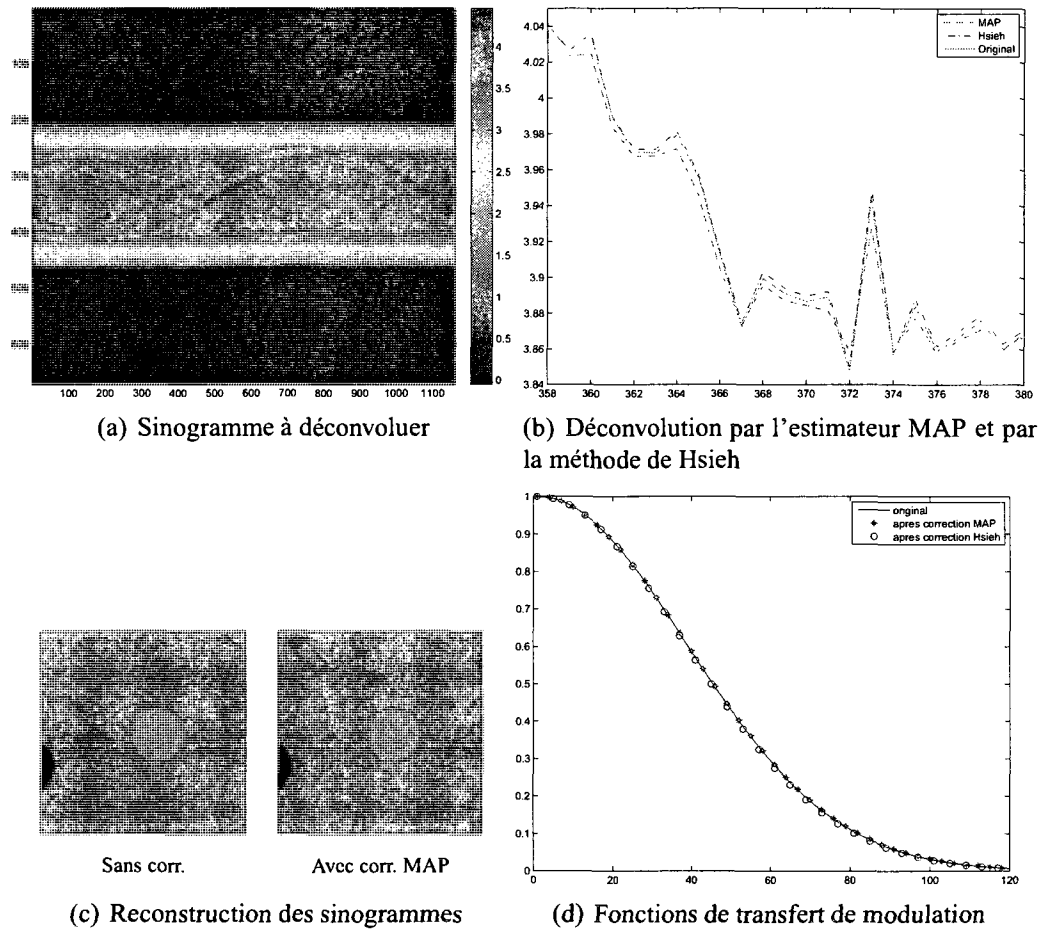


FIG. 5.16: Déconvolution sur données réelles : protocole *Trauma*.

Les résultats de la deuxième partie sont mitigés. Dans cette partie, nous avons vérifié que la vitesse de rotation influence la qualité des images reconstruites, car la rémanence des détecteurs est plus importante. Puis, nous avons montré qu'en simulations, les déconvolutions améliorent la qualité des images lors de la reconstruction. Cependant, pour la déconvolution de données réelles, l'amélioration est quasi nulle, que ce soit par notre technique ou celle de Hsieh.

Le constructeur Siemens ne nous fournissant pas les détails des prétraitements qu'il réalise sur les données « brutes », nous supposons qu'il réalise déjà une correction de la postluminescence.

CHAPITRE 6

CONCLUSION

Dans le cadre de notre maîtrise, l'objectif général était de mettre en place une procédure de correction de la postluminescence pour des sinogrammes de tailles médicales. Pour arriver à cet objectif, nous nous étions fixé :

1. **De comprendre et modéliser la postluminescence des détecteurs ;**
2. **D'estimer les paramètres de ce modèle ;**
3. **De restaurer les sinogrammes.**

Après une revue de l'état des objectifs atteints et de ceux non atteints, nous verrons les prolongements de ces travaux.

6.1 Revue des objectifs

La première partie de nos travaux a porté sur la description et la compréhension de la postluminescence. La postluminescence influence grandement la qualité des images de hautes résolutions lorsque la vitesse de rotation du portique source-détecteurs devient importante (< 1 s/rotation). Le modèle de convolution de réponse impulsionnelle permet de modéliser convenablement ce phénomène.

Ensuite, afin de caractériser les paramètres du modèle de convolution, nous avons mis en place une technique d'estimation de la réponse impulsionnelle des détecteurs, représentée soit par un vecteur, soit par les coefficients de multiexponentielles décroissantes. Ceci est une contribution originale puisque cette technique peut être employée sur n'importe quel tomographe dont on a scanné le fantôme Catphan 600 en usage dans le milieu

clinique pour la maintenance des scanners.

La deuxième partie du travail a consisté à replacer dans le contexte de l'estimation la méthode de déconvolution actuellement employée dans certains tomographes commerciaux. À partir de cette remise en contexte, nous avons tiré les limites de cette technique et proposé une technique plus robuste face au bruit présent sur les sinogrammes. La restauration du sinogramme se fait alors par une méthode d'estimation minimisant un critère composé d'un terme de fidélité aux données et d'un terme de régularisation. Cette minimisation peut être calculée de manière explicite.

Enfin, nous avons confronté notre méthode à celle de Hsieh. Nous avons ainsi observé qu'une approche régularisée améliore la qualité des sinogrammes. Cependant, les tests réalisés sur données réelles ne sont pas concluants. Nous supposons que Siemens effectue des pré-traitements sur les données « brutes ». Les sinogrammes « bruts » ne sont donc pas aussi bruts qu'on l'aurait souhaité. Des corrections sont déjà appliquées sur ces données, ce qui empêche la pleine restauration des sinogrammes.

6.2 Perspectives

Plusieurs perspectives peuvent être envisagées.

Dans nos travaux, nous avons obtenu des réponses impulsionnelles pour une vitesse de rotation de 1 s/rotation et 0,375 s/rotation. Les scanners actuels vont beaucoup plus vite que cela. Le dernier scanner Toshiba atteint des vitesses inférieures à 0,1 s/rotation. Par conséquent, obtenir des réponses impulsionnelles pour des vitesses inférieures à 0,1 s serait souhaitable afin de traiter des sinogrammes hélicoïdaux, qui sont maintenant les plus répandus.

Depuis janvier 2009, le LION reçoit des données de scanners Philips provenant de

l'Hôtel-Dieu de Montréal. Dans ces scanners, il est possible d'obtenir les données réellement brutes de pré-traitements du manufacturier. Ainsi, il serait souhaitable de réaliser des essais sur ces données afin d'obtenir des réponses impulsionnelles et de traiter les sinogrammes. Un travail de développement de code pourrait alors être réalisé afin de lire ces données brutes. Ceci permettrait d'effectuer aussi des tests plus poussés sur données médicales.

RÉFÉRENCES

- ALLAIN, M., GOUSSARD, Y., AND IDIER, J. (2002). Regularized approach in 3D helical computed tomography. *Proc. IEEE EMB Intern. Conf.*, **2**, 943–944.
- DEMOMENT, G. (1989). Image reconstruction and restoration : Overview of common estimation structure and problems. *IEEE Trans. Acoust. Speech, Signal Processing*, **ASSP-37**(12), 2024–2036.
- FELDKAMP, L. A., DAVIS, L., AND KRESS, J. W. (1984). Practical cone-beam algorithm. *J. Opt. Soc. Am. A*, **1**, 612–619.
- FESSLER, J. A. (2008). World Wide Web electronic publication. [http ://www.eecs.umich.edu/ fessler/irt/](http://www.eecs.umich.edu/fessler/irt/).
- HAMELIN, B., GOUSSARD, Y., GENDRON, D., DUSSAULT, J.-P., CLOUTIER, G., BEAUDOIN, G., AND SOULEZ, G. (2008). Iterative CT reconstruction of real data with metal artifact reduction. *Proc. IEEE ISBI* 1453–1456.
- HEART, N. AND INSTITUTE, L. B. (2009). Angioplasty. World Wide Web electronic publication. [http ://www.nhlbi.nih.gov/health/dci/Diseases/Angioplasty/-Angioplasty_All.html](http://www.nhlbi.nih.gov/health/dci/Diseases/Angioplasty/-Angioplasty_All.html).
- HSIEH, J., GURMEN, O. E., AND KING, K. F. (2000). Investigation of a solid-state detector for advanced computed tomography. *IEEE Trans. Medical Imaging*, **19**(9), 930–940.
- KACHELRIESS, M., KNAUP, M., PENSSSEL, C., AND KALENDER, W. A. (2006). Flying focal spot (ffs) in cone-beam ct. *IEEE Transactions on nuclear science*, **53**(3), 1238–1247.
- KAK, A. C. AND SLANEY, M. (1987). *Principles of Computerized Tomographic Imaging*. IEEE Press, New York, NY.
- KALENDER, W. A. (2005). *Computed Tomography*. Publicis.

- KATSEVICH, A. (2002). Theoretically exact FBP-type inversion algorithm for spiral CT. *SIAM Journal of Applied Mathematics*, **62**, 2012–2026.
- LARIVIÈRE, P. J., BIAN, J., AND VARGAS, P. A. (2006). Penalized-likelihood sinogram restoration for computed tomography. *IEEE Trans. Medical Imaging*, **25**(8), 1022–1036.
- LIU, D. C. AND NOCEDAL, J. (1989). On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, **45**, 503–528.
- MACOVSKI, A. (1983). *Medical Imaging Systems*.
- NOCEDAL, J. AND WRIGHT, S. J. (2000). *Numerical Optimization*. Operations Research. Springer Verlag, New York, NY.
- OHNESORGE, B., FLOHR, T., SCHWARZ, K., HEIKEN, J. P., AND BAE, K. T. (2000). Efficient correction for CT image artifacts caused by objects extending outside the scan field of view. *Medical Physics*, **27**(1), 39–46.
- PAN, X., XIAN, D., ZOU, Y., AND YU, L. (2004). A unified analysis of fbp-based algorithms in helical cone-beam and circular cone and fan-scans. *Physical medical biology*, **49**, 4349–4369.
- SIEMENS (2007). Ufc - ultra fast ceramics - for all your x-ray needs. World Wide Web electronic publication. http://health.siemens.com/ct_applications/ufc/index.html.
- SOURBELLE, K., KACHELRIESS, M., AND KALENDER, W. A. (2005). Reconstruction from truncated projections in ct using adaptive detruncation. *European Radiology*, **15**(5), 1008–1014.
- STATISTIQUES CANADA (2009). Santé. World Wide Web electronic publication. http://www41.statcan.gc.ca/2008/2966/ceb2966_000-fra.htm.
- STIERSTORFER, K., RAUSCHER, A., SCHALLER, S., AND FLOHR, T. (2004). Weighted fbp - a simple approximate 3d fbp algorithm for multislice spiral ct with good dose usage for arbitrary pitch. *Physics in Medecine and Biology*, **49**, 2209–2218.

- THOMPSON, A. M., BROWN, J. C., KAY, J. W., AND TITTERINGTON, D. M. (1991). A study of methods of choosing the smoothing parameter in image restoration by regularization. *IEEE Trans. Pattern Anal. Mach. Intell.*, **PAMI-13**(4), 326–339.
- TIKHONOV, A. N. AND ARSENIN, V. Y. (1976). *Méthodes de résolution de problèmes mal posés*. Éditions MIR, Moscou.
- TUY, H. K. (1983). An inversion formula for cone-beam reconstruction. *SIAM Journal Applied Mathematics*, **43**, 546–552.

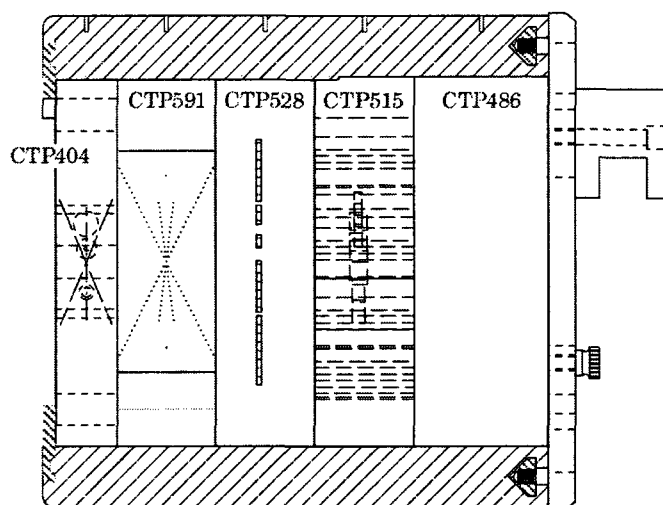
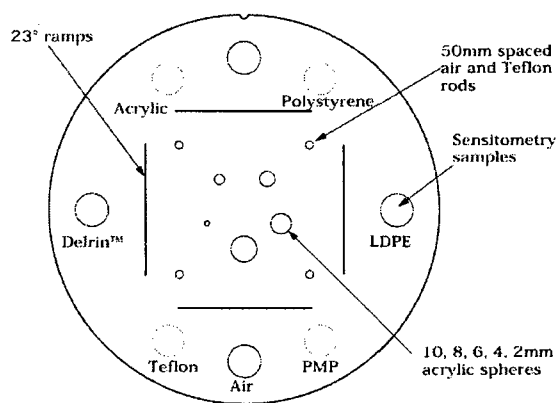
ANNEXE I**DESCRIPTION DU FANTÔME CATPHAN 600****I.1 Plan du fantôme**

FIG. I.1: Plan du fantôme *Catphan 600*.

I.2 Composition du fantôme



(a) Composition du fantôme.

Linear attenuation coefficient μ [units cm^{-1}]								
KEV	Teflon	Delrin	Acrylic	Polystyrene	Water	LDPE	PMP	Air
40	0.556	0.327	0.277	0.229	0.240	0.209	0.189	0
50	0.447	0.283	0.244	0.209	0.208	0.191	0.173	0
60	0.395	0.260	0.227	0.196	0.192	0.181	0.164	0
62	0.386	0.256	0.224	0.194	0.190	0.179	0.162	0
64	0.380	0.253	0.221	0.192	0.188	0.178	0.160	0
66	0.374	0.251	0.219	0.191	0.186	0.177	0.160	0
68	0.370	0.248	0.217	0.189	0.184	0.175	0.158	0
70	0.363	0.245	0.215	0.188	0.182	0.174	0.157	0
72	0.359	0.243	0.214	0.186	0.181	0.172	0.155	0
74	0.355	0.240	0.211	0.185	0.179	0.171	0.155	0
76	0.351	0.238	0.210	0.184	0.178	0.170	0.154	0
78	0.346	0.236	0.208	0.183	0.177	0.168	0.152	0
80	0.342	0.234	0.207	0.180	0.175	0.167	0.151	0
90	0.328	0.225	0.199	0.175	0.170	0.163	0.147	0
100	0.315	0.218	0.194	0.170	0.165	0.158	0.143	0

Nominal material formulation and specific gravity				
Material	Formula	Specific Gravity*	Electron Density (10^{20}e/g)	CT # est.
Air	75%N, 23.2%O, 1.3%A	0.00	3.007	-1000
PMP	$[\text{C}_8\text{H}_{12}(\text{CH}_2)]$	0.83	3.435	-200
LDPE	$[\text{C}_2\text{H}_4]$	0.92	3.429	-100
Water	$[\text{H}_2\text{O}]$	1.00	3.343	0
Polystyrene	$[\text{C}_8\text{H}_8]$	1.05	3.238	-35
Acrylic	$[\text{C}_5\text{H}_8\text{O}_2]$	1.18	3.248	120
Delrin™	Proprietary	1.41	3.209	340
Teflon	$[\text{CF}_2]$	2.16	2.889	990

(b) Coefficients d'atténuation des matériaux utilisés.

FIG. I.2: Composition du fantôme *Catphan 600* section 404.

I.3 Paramètres géométriques du fantôme

TAB. I.1: Paramètres géométriques des ellipses constituant le module CTP404 du fantôme Catphan 600©

	x centre en cm	y centre en cm	z centre en cm	rayon x en cm	rayon y en cm	rayon z en cm
1	0.085	0.183	0.000	10.059	10.059	10000
2	0.134	6.018	0.000	0.537	0.537	10000
3	0.037	-5.676	0.000	0.610	0.610	10000
4	2.917	5.261	0.000	0.610	0.610	10000
5	-2.795	5.261	0.000	0.610	0.610	10000
6	-5.750	0.183	0.000	0.610	0.610	10000
7	5.945	0.110	0.000	0.610	0.610	10000
8	-2.869	-4.871	0.000	0.610	0.610	10000
9	2.991	-4.919	0.000	0.610	0.610	10000
10	-2.380	2.673	0.000	0.146	0.146	10000
11	2.625	2.649	0.000	0.146	0.146	10000
12	-2.429	-2.332	0.000	0.146	0.146	10000
13	2.576	-2.356	0.000	0.146	0.146	10000
14	0.598	3.967	0.000	0.098	0.037	10000
15	-3.723	0.647	0.000	0.037	0.098	10000
16	-0.378	-3.650	0.000	0.098	0.037	10000
17	3.894	-0.330	0.000	0.037	0.098	10000
18	0.085	0.183	0.000	7.520	7.520	10000