

Titre: Optimisation d'un algorithme statistique itératif de reconstruction tomographique en utilisant les coordonnées cylindriques
Title: tomographique en utilisant les coordonnées cylindriques

Auteur: Adrien Wagner
Author:

Date: 2014

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Wagner, A. (2014). Optimisation d'un algorithme statistique itératif de reconstruction tomographique en utilisant les coordonnées cylindriques [Master's thesis, École Polytechnique de Montréal]. PolyPublie.
Citation: <https://publications.polymtl.ca/1511/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/1511/>
PolyPublie URL:

Directeurs de recherche: Yves Goussard
Advisors:

Programme: Génie biomédical
Program:

UNIVERSITÉ DE MONTRÉAL

OPTIMISATION D'UN ALGORITHME STATISTIQUE ITÉRATIF DE
RECONSTRUCTION TOMOGRAPHIQUE EN UTILISANT LES COORDONNÉES
CYLINDRIQUES

ADRIEN WAGNER
INSTITUT DE GÉNIE BIOMÉDICAL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE BIOMÉDICAL)
JUN 2014

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

OPTIMISATION D'UN ALGORITHME STATISTIQUE ITÉRATIF DE
RECONSTRUCTION TOMOGRAPHIQUE EN UTILISANT LES COORDONNÉES
CYLINDRIQUES

présenté par : WAGNER Adrien

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

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

M. SAVARD Pierre, Ph.D., président

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

M. LESAGE Frédéric, Ph.D., membre

À Marc, Sabine, Marjolaine et Mélusine.

REMERCIEMENTS

J'ai pu conduire cette maîtrise à terme grâce à l'aide de nombreuses personnes que j'aimerais remercier.

Tout d'abord, j'aimerais remercier mon directeur de recherche, le professeur Yves Gousard, pour sa passion, sa patience, son support, ses critiques et ses conseils avisés.

Merci aussi à mes collègues qui ont fait de cette maîtrise un parcours agréable, merci à Mahsa Golkar, Qingyi Han, Matthieu Voorons, Corentin Friedrich et Benjamin Seignot.

Je tiens aussi à remercier les membres de mon jury pour le temps qu'ils prendront pour lire et critiquer mon travail afin de le rendre meilleur.

Enfin, merci à mes amis rencontrés ici à Montréal pour tous les bons moments passés ensembles et à toute ma famille pour leurs encouragements.

RÉSUMÉ

La tomographie axiale calculée par ordinateur utilise ordinairement des techniques de analytiques pour la reconstruction d'image, mais les méthodes de reconstruction statistiques itératives peuvent être une bonne alternative à ces techniques conventionnelles. Le principal problème de ces méthodes est qu'elles nécessitent une amélioration de leurs performances en terme de temps de calcul et d'espace mémoire requis pour pouvoir être plus couramment utilisées.

Ce projet a pour but d'optimiser les performances globales des techniques itératives de reconstruction en 3D en utilisant une représentation alternative de l'objet ainsi que d'étudier l'impacte d'une telle représentation sur la qualité des images reconstruites. L'utilisation de la représentation de l'objet en coordonnées cylindriques a un impact significatif sur les performances globales des techniques itératives car elle permet une réduction considérable de la mémoire utilisée mais ce modèle est encore très lent, son effet sur la qualité de l'image méconnu et ses possibilités d'amélioration très peu étudiées. ce projet se concentre sur les améliorations possibles avec cette représentation cylindrique permettant l'augmentation de la vitesse de calcul en utilisant la forme particulière de la matrice de projection en 3D.

Ce projet consiste dans un premier temps à développer une nouvelle technique de parallélisation efficace de la méthode qui utilise la structure bloc-circulante de la matrice de projection, mais aussi à améliorer la vitesse de convergence grâce à la création de préconditionneurs utilisant la propriété diagonale-dominante de la matrice de projection bloc-diagonalisée dans le domaine de Fourier.

La deuxième partie de ce projet consiste en une étude poussée de la qualité des images résultant de l'utilisation de cette représentation cylindrique de l'objet, une application d'un modèle polychromatique afin de réduire les artéfacts métalliques et la mise en oeuvre de la méthode pour la reconstruction de données réelles.

ABSTRACT

Computed tomography usually uses analytical techniques for image reconstruction. But Iterative, statistical reconstruction methods can be good alternative to this conventional analytical techniques. The main problem with these methods is that they require an improvement in their performances in terms of amount of computation and memory footprint for wider adoption of the approach.

This project aims to optimize the overall performance of iterative reconstruction techniques in 3D using an alternative representation of the object and to study the impact of such representation on the quality of the reconstructed images. The use of object representation in cylindrical coordinates has a strong impact on the overall performance of iterative techniques because it allows a considerable reduction of the memory footprint but this model is still very slow, its effect on the quality of the unknown image and opportunities for improvement aren't really studied. This project focuses on possible improvements to increase the speed of calculation using the special form of the projection matrix in 3D.

First, this project is to develop new techniques for efficient parallelization of the method that uses block-circulating structure of the projection matrix, but also improve the speed of convergence through the creation of preconditioners using the property diagonal-dominant property of the block projection-diagonalized projection matrix in the Fourier domain. The second part of this project consists of a detailed study of the image quality resulting from the use of this cylindrical representation of the object, an application of a polychromatic model to reduce metal artifacts and the implementation of the method for reconstruction of real data.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vi
TABLE DES MATIÈRES	vii
LISTE DES FIGURES	xi
LISTE DES SIGLES ET ABRÉVIATIONS	xiii
CHAPITRE 1 INTRODUCTION	1
1.1 Reconstruction de l'image en tomographie	1
1.2 Objectifs de recherche	2
1.3 Plan du mémoire	2
CHAPITRE 2 REVUE DE LITTÉRATURE	4
2.1 Modèle de formation des données	4
2.1.1 Acquisition des données	4
2.1.2 Modélisation déterministe de la projection	5
2.1.3 Discrétisation	6
2.1.4 Calcul de la matrice de projection	7
2.2 Description des méthodes de reconstruction analytiques	9
2.3 Description des algorithmes de reconstruction algébriques	9
2.3.1 Modèle des données	10
2.3.1.1 Modélisation monochromatique	10
2.3.1.2 Modélisation polychromatique	10
2.3.2 Modélisation du bruit	12
2.3.3 Estimateur	13
2.3.3.1 Maximum de vraisemblance	13
2.3.3.2 Maximum <i>a posteriori</i>	14
2.3.3.3 Fonction de pénalisation	14

2.3.4	Algorithme d'optimisation	16
2.3.4.1	Généralités	16
2.3.4.2	Méthodes d'optimisation de la plus forte pente, de Newton et de quasi-Newton	17
2.3.4.3	Préconditionnement	19
2.4	Description de l'algorithme de reconstruction algébrique en coordonnées cylin- driques	20
2.4.1	Représentation de l'objet	21
2.4.2	Propriétés de la matrice normale du système linéaire	23
2.4.2.1	Propriétés mathématiques	23
2.4.2.2	Propriétés et simplification des opérations de projection et rétroprojection	24
2.4.2.3	Présentation des types de représentation de matrices creuses	25
2.4.3	Préconditionnement	27
2.4.3.1	Principe du préconditionnement	27
2.4.3.2	Structure de la matrice normale	27
2.4.3.3	Préconditionnement dans le domaine spatial	28
2.4.3.3.a	Préconditionneur diagonal	28
2.4.3.3.b	Préconditionneur de Cholesky incomplet	28
2.4.3.4	Préconditionnement dans le domaine de Fourier	29
2.4.3.4.a	Principe	29
2.4.3.4.b	Préconditionneur diagonal	30
2.4.3.4.c	Préconditionneur de Cholesky incomplet	30
2.4.3.5	Adaptation à la représentation cylindrique	31
2.4.4	Avantages et inconvénients de la représentation cylindrique	31
2.5	Mesure de la qualité d'image	31
2.5.1	Rappel des enjeux	31
2.5.2	Méthodes d'évaluation théorique de qualité d'image	32
2.5.3	Mesure du contraste	32
2.5.4	Mesure de la résolution	33
2.6	Objectifs détaillés	33
CHAPITRE 3 MÉTHODOLOGIE		35
3.1	Estimation	35
3.2	Parallélisation	36
3.2.1	Possibilités de parallélisation	36

3.2.1.1	Parallélisations possibles avec représentation de l'objet en coordonnées cartésiennes	37
3.2.1.2	Parallélisations possibles avec représentation de l'objet en coordonnées cylindriques	38
3.2.2	Mise en oeuvre de la parallélisation	39
3.3	Preconditionnement	41
3.3.1	Structure de l'estimateur en 3D	41
3.3.2	Préconditionneur diagonal dans le domaine de Fourier	45
3.3.3	Préconditionneur de Cholesky incomplet dans le domaine de Fourier	46
3.4	Mesure de qualité d'image	47
3.4.1	Rappel des enjeux	47
3.4.2	Calcul de la RMSE	49
3.4.3	Calcul de la MTF	49
3.4.4	Méthode d'amélioration de la qualité de l'image : modélisation polychromatique	50
CHAPITRE 4	RÉSULTATS	52
4.1	Modèle	52
4.1.1	Modélisation de la source de rayons X	52
4.1.2	Modélisation des rayons et de la géométrie du scanner	53
4.1.3	Modélisation de l'objet	54
4.2	Génération des données simulées	54
4.2.1	Création du fantôme et erreur de représentation	54
4.2.2	Génération du sinogramme	56
4.3	Estimation	57
4.4	Implémentation	58
4.5	Taille de la matrice de projection	58
4.6	Parallélisation	59
4.6.1	Étude du nombre de coeurs et du schéma de stockage	60
4.6.2	Étude de la taille de l'image et de la nature de la multiplication	62
4.6.3	Conclusion sur la parallélisation	65
4.7	Préconditionnement	67
4.7.1	Mesure du temps par itération	68
4.7.2	Mesure de la distance à la solution exacte	69
4.7.3	Conclusions sur le préconditionnement	71
4.8	Qualité d'image	72

4.8.1	Mesure de la RMSE avec un fantôme réaliste	72
4.8.2	Mesure de la qualité d'image avec un fantôme spécifique	74
4.8.2.1	Mesure de la MTF	74
4.8.2.2	Mesure de la RMSE locale	79
4.8.2.3	Conclusion sur la mesure de qualité d'image	80
4.8.3	Amélioration de la qualité des images avec l'ART polaire : réduction des artefacts métalliques	81
CHAPITRE 5 CONCLUSION		84
5.1	Synthèse des travaux	84
5.2	Limitations de la solution proposée	86
5.3	Améliorations futures	87
RÉFÉRENCES		89

LISTE DES FIGURES

Figure 2.1	Acquisition de données en tomographie ©Goldman (2007)	4
Figure 2.2	(a) méthode <i>pixel-driven</i> . (b) méthode <i>ray-driven</i> . (c) méthode <i>distance-driven</i> ©De Man et Basu (2004)	7
Figure 2.3	représentation polaire en 2D ©Jian <i>et al.</i> (2007)	22
Figure 2.4	représentation polaire alternative en 2D ©Rodriguez-Alvarez <i>et al.</i> (2011)	22
Figure 2.5	représentation cylindrique (à gauche) et sphérique (à droite) ©Thibaudeau <i>et al.</i> (2013)	23
Figure 3.1	Répartition du temps de calcul moyen par itération de l'estimation d'un objet en coordonnées cylindrique	37
Figure 3.2	Matrice normale du problème d'optimisation bloc diagonalisée pour un problème en 3D de très petite taille	45
Figure 4.1	Fantôme XCAT utilisé pour le préconditionnement et la parallélisation	55
Figure 4.2	Fantôme XCAT utilisé pour la mesure de la MTF	57
Figure 4.3	Modèle polaire	61
Figure 4.4	Modèle cartésien	61
Figure 4.5	avec 290 projections	63
Figure 4.6	avec 146 projections	63
Figure 4.7	avec 580 projections	64
Figure 4.8	avec 290 projections	64
Figure 4.9	avec 146 projections	64
Figure 4.10	temps de calcul par itération, SNR = 25 dB	68
Figure 4.11	temps de calcul par itération, SNR = 35 dB	68
Figure 4.12	Distance normalisée à la solution exacte, SNR = 25 dB	70
Figure 4.13	Distance normalisée à la solution exacte, SNR = 35 dB	70
Figure 4.14	RMSE, SNR = 25 dB	73
Figure 4.15	RMSE, SNR = 35 dB	73
Figure 4.16	SNR=25dB	75
Figure 4.17	SNR=35dB	75
Figure 4.18	sans bruit	75
Figure 4.19	SNR=25dB	75
Figure 4.20	SNR=35dB	75
Figure 4.21	sans bruit	75
Figure 4.22	$\lambda = 0.1$	78

Figure 4.23	$\lambda = 0.01$	78
Figure 4.24	$\lambda = 0.001$	78
Figure 4.25	$\lambda = 0.1$	78
Figure 4.26	$\lambda = 0.01$	78
Figure 4.27	$\lambda = 0.001$	78
Figure 4.28	RMSE, SNR = 25 dB	79
Figure 4.29	RMSE, SNR = 35 dB	79
Figure 4.30	rétroprojection filtrée	82
Figure 4.31	modèle monochromatique	82
Figure 4.32	modèle polychromatique	82

LISTE DES SIGLES ET ABRÉVIATIONS

BC	Blocs-circulant
CT	Computed tomography
2D	deux dimensions
3D	trois dimensions
FBP	Filtered Back-Projection (rétroprojection filtrée)
ART	Algebraic Reconstruction Technique (technique algébrique de reconstruction)
FFT	Fast Fourier Transform (transformée de Fourier rapide)
ML	Maximum likelihood (maximum de vraisemblance)
MAP	Maximum a posteriori
EM	algorithme d'Espérance Maximisation
L-BFGS-B	Limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm Bounded (Broyden–Fletcher–Goldfarb–Shanno algorithme Borné à mémoire Limitée)
SNR	Signal Noise Ratio (rapport signal à bruit)
CNR	Contrast Noise Ratio (rapport contraste à bruit)
RMSE	Root Mean Square Error (racine de l'erreur quadratique moyenne)
MTF	Modulation Transfer Function (fonction de modulation de transfert)
PSF	Point Spread Function (fonction d'étalement de point)
LSF	Line Spread Function (fonction d'étalement de ligne)
GPU	Graphics Processing Unit (processeur graphique)
keV	kilo-electron-volt
TS	Triplet Storage (stockage en triplets)
CRS	Compressed Row Storage (stockage par lignes compressées)
ICRS	Incremental Compressed Row Storage (stockage incrémental par lignes compressées)
BICRS	Bi-directional Incremental Compressed Row Storage (stockage incrémental bi-directionnel par lignes compressées)
HTS	Hilbert-curve Triplet Storage (stockage en triplet suivant la courbe d'Hilbert)
PD	Préconditionneur diagonal
IC	préconditionneur de Cholesky Incomplet

CHAPITRE 1

INTRODUCTION

L'imagerie médicale regroupe un ensemble de techniques non-invasives permettant de visualiser l'intérieur du corps humain. Elle est utilisée comme outil diagnostique en médecine ou dans la recherche biomédicale afin d'améliorer la connaissance du corps humain. Parmi les techniques d'imagerie, la tomодensitométrie ou tomographie calculée par ordinateur (abrégée CT-scan avec CT : *computed tomography*), ou communément appelée tomographie, consiste à mesurer l'absorption des rayons X par les tissus de l'objet à imager en faisant varier l'angle de projection des rayons. Les images 2D ou 3D de l'objet sont ensuite reconstruites grâce à un traitement informatique.

Cette technique d'imagerie est très couramment utilisée et ses applications sont nombreuses. Elle sert notamment à détecter et suivre l'évolution des tumeurs et hémorragies dans la boîte crânienne, à imager les artères coronaires et à diagnostiquer les maladies abdominales et le cancer. Le problème est que les rayons X sont cancérigènes, il est alors préférable de diminuer au maximum la dose de rayons X utilisée à chaque examen ce qui va diminuer la qualité des mesures en augmentant le bruit. Une application plus récente de cette technique est l'imagerie intra-opératoire lors de chirurgies cardiaques, vasculaires et de la moelle épinière. Pour cette dernière application, la qualité d'image et le temps de reconstruction sont deux critères très importants et améliorables.

Ces diverses applications de la tomодensitométrie nécessitent donc d'avoir des méthodes de reconstruction rapides d'image de bonne qualité et efficaces dans la réduction du bruit.

1.1 Reconstruction de l'image en tomographie

Pour obtenir des images à partir des mesures d'atténuations de rayons X, il est nécessaire d'effectuer un traitement informatique. Pour cela, il existe deux grands types de méthodes : les méthodes analytiques, abrégées FBP pour *Filtered Back-Projection*, et les méthodes algébriques, abrégées ART pour *Algebraic Reconstruction Technique*.

Les scanners à rayons X actuellement exploités pour des applications cliniques utilisent les méthodes analytiques ou des méthodes semi-itératives car elles sont rapides du fait de la simplicité du modèle utilisé et de la méthode de calcul nécessaire à la reconstruction. Le prin-

principal problème de ces méthodes est que le modèle qu'elles utilisent ne prend pas en compte le bruit. Cette approximation du modèle va avoir un impact important sur la qualité des images reconstruites à partir de mesures très bruitées provenant de protocole expérimental basse-dose.

Dans le cas de données bruitées, les méthodes algébriques itératives sont désormais reconnues pour avoir des résultats de reconstruction de meilleure qualité que les méthodes analytiques. Ces méthodes permettent la prise en compte d'informations à priori sur l'image dans leur modèle telles le spectre de la source de rayons X utilisée et lors de la reconstruction telles le type de bruit ou de structure attendu. Ces informations permettent d'améliorer considérablement la qualité des images reconstruites. En revanche la taille importante des objets et données à manipuler pour une reconstruction en 3D rend le temps de calcul très long et l'espace mémoire requis très important. Ces propriétés des méthodes algébriques rend irréalizable leur utilisation dans le domaine clinique.

1.2 Objectifs de recherche

L'objectif général de ce travail a été de développer une méthode algébrique itérative de reconstruction efficace permettant d'améliorer la vitesse d'exécution et l'espace mémoire requis tout en conservant la qualité d'image supérieure des méthodes algébriques.

Pour cela, on a choisi de développer une représentation alternative de l'objet à imager en coordonnées cylindriques afin de diminuer la taille des structures à manipuler lors de la reconstruction avec une méthode algébrique. De plus, on a cherché à augmenter la vitesse de la reconstruction itérative en diminuant le temps de calcul de chaque itération et le nombre d'itérations nécessaire.

1.3 Plan du mémoire

Ce mémoire est présenté de la manière suivante :

Dans le chapitre 2, on présente les différents modèles de formation de données, on décrit plus en profondeur les deux grands types de méthodes de reconstruction FBP et ART. On présente plus particulièrement l'ART en coordonnées cylindriques qui nous semble être la solution la plus adaptée pour répondre à nos objectifs. Enfin, on présente les techniques de mesure qui permettront de tester l'impact du nouvel algorithme sur la qualité de l'image.

Dans le chapitre 3, on décrit les méthodes développées pour atteindre notre objectif et plus

particulièrement, la méthode d'optimisation du calcul de chaque itération mise en oeuvre, la méthode de diminution du nombre d'itérations utiles et les méthodes pour calculer et améliorer la qualité de l'image.

Dans le chapitre 4, on décrit d'abord les paramètres (modèle utilisé, géométrie et fantôme utilisés pour la génération des données, paramètres d'estimation) et le protocole expérimental utilisés pour la validation des méthodes développées. Les résultats de ces expériences sont ensuite présentés et interprétés.

Les résultats du chapitre précédent sont résumés et discutés dans le chapitre 5 et sont suivis d'une conclusion sur le travail effectué et une esquisse des travaux restants à faire.

CHAPITRE 2

REVUE DE LITTÉRATURE

2.1 Modèle de formation des données

2.1.1 Acquisition des données

La tomographie à rayons X est une technique d'imagerie consistant à reconstruire le volume d'un objet à partir d'une série de mesures faites de l'extérieur de cet objet en utilisant les propriétés de pénétration de la matière des rayons X. Le montage d'acquisition classique comprend une source de rayons X et un panneau de barrettes de détecteurs entre lesquels est placé l'objet à imager. La source et le panneau sont placés de part et d'autre de l'objet et se déplacent conjointement dans le même sens le long d'une trajectoire circulaire, émettant des rayons X et effectuant des mesures à incréments angulaires définis. (Fig. 2.1). On se restreint au cadre de la tomographie axiale, on considère que l'ensemble source détecteur ne se déplace pas le long de l'axe de rotation.

Les détecteurs mesurent le nombre de photons reçus pour chaque angle de projection de

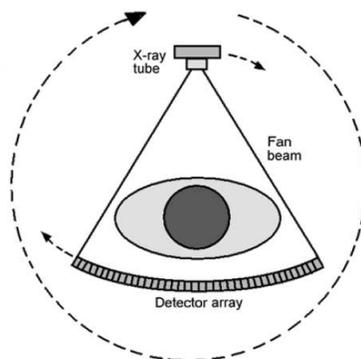


Figure 2.1 Acquisition de données en tomographie ©Goldman (2007)

rayons X, qui, soustrait au nombre de photons émis par la source, donne une mesure de l'atténuation des rayons X par les matériaux traversés. Cette atténuation des rayons X par un matériau est définie par le coefficient d'atténuation de ce matériau et dépend de l'énergie des rayons. Les mesures sont appelées mesures de projection et forment ce qui est appelé le sinogramme. Ces données sont traitées mathématiquement pour reconstituer des images ou coupes bidimensionnelles de l'objet qui représentent le coefficient d'atténuation de chaque

point de l'espace suivant un plan donné. Ce plan est choisi généralement perpendiculaire ou parallèle à l'axe de rotation.

2.1.2 Modélisation déterministe de la projection

Le traitement mathématique des données recueillies utilise une simplification du phénomène physique responsable de l'atténuation des rayons afin de modéliser ce que représente une mesure de projection.

Les mesures de projections d'un objet μ , de coordonnées en trois dimensions (x_1, x_2, x_3) , sont modélisées par un ensemble d'intégrales de lignes d'angles différents. Chaque intégrale de ligne représente la propagation et l'atténuation d'un rayon de la source de rayons X à un détecteur à travers l'objet pour un angle de projection donné et est la somme des coefficients d'atténuations des matériaux traversés en ligne droite par les rayons. Les coefficients d'atténuation des matériaux dépendent de l'énergie des rayons qui les traversent.

Les sources de rayons X des scanners CT sont polychromatiques donc émettent des photons dans un spectre continu d'énergies (ϵ) entre 5 et 150 keV. On considère i une paire angle de projection-détecteur. L'intensité photonique I_i mesurée par ce détecteur à l'angle de projection associé suit la relation suivante (Elbakri et Fessler (2002)) :

$$I_i = \int I_0(\epsilon) \exp\left(-\int_{L_i} \mu(x_1, x_2, x_3, \epsilon) dl\right) d\epsilon \quad (2.1)$$

où I_i est l'intensité photonique mesurée, I_0 est l'intensité photonique émise, μ le coefficient d'atténuation linéaire de l'objet, L_i les trajectoires des rayons atteignant le détecteur à travers l'objet pour un angle de projection donné (couple angle de projection-détecteur i) et ϵ l'énergie des rayons X. On émet l'hypothèse que le coefficient est mesuré en cm^{-1} , on a donc $\mu \geq 0$.

On définit la projection comme la fonction $\tilde{y} : \mathbb{N} \rightarrow \mathbb{R}$ telle que $\tilde{y} = I_i / \int_{\epsilon} I_0(\epsilon) d\epsilon$, on a alors :

$$\tilde{y} = \int b(\epsilon) \cdot \exp\left(-\int_{L_i} \mu(x_1, x_2, x_3, \epsilon) dl\right) d\epsilon \quad (2.2)$$

avec $b(\epsilon)$ le spectre d'émission de la source de rayons X.

lorsqu'on observe le spectre de la source de rayons X utilisée par les scanners clinique, on s'aperçoit que l'intensité des rayons X est beaucoup plus élevée pour une énergie de 70 keV que pour le reste du spectre. Il est donc possible de simplifier les calculs en négligeant la polychromaticité de la source et en utilisant un modèle monochromatique. le coefficient d'atténuation ne dépend plus de l'énergie. Dans ce cadre la loi de Beer-Lambert peut être

appliquée et l'équation 2.1 peut être simplifiée :

$$I_i = I_0 \exp \left(- \int_{L_i} \mu(x_1, x_2, x_3) dl \right) \quad (2.3)$$

En définissant la projection monochromatique comme la fonction $y : \mathbb{N} \rightarrow \mathbb{R}$ telle que $y(i) = -\log(I_i/I_0)$, on obtient :

$$y(i) = \int_{L_i} \mu(x_1, x_2, x_3) dl \quad (2.4)$$

2.1.3 Discrétisation

En réalité, afin de pouvoir effectuer les calculs de reconstruction et représenter le résultat numériquement, l'image est discrétisée en un nombre fini de voxels. Ceci implique une discrétisation de l'objet $\boldsymbol{\mu}(x_1, x_2, x_3)$ qu'on nommera $\boldsymbol{\mu}$.

L'intégrale le long de la droite L_i passant entre la source positionnée selon l'angle et le détecteur de la paire angle de projection-détecteur i peut être approximée par la somme des contributions p_{ij} des voxels j de $\boldsymbol{\mu}(x_1, x_2, x_3)$ à l'atténuation mesurée I_i . Il existe plusieurs méthodes de calcul des p_{ij} qui seront détaillées dans la section suivante. On regroupe les coefficients p_{ij} de cette somme dans un vecteur ligne \mathbf{p}_i de la taille de l'objet discrétisé. Cette approximation permet de représenter $\int_{L_i} \mu(x_1, x_2, x_3, \epsilon) dl$ de la forme suivante : $\mathbf{p}_i \boldsymbol{\mu}(\epsilon)$.

Les vecteurs \mathbf{p}_i , associés respectivement à chacune des paires angle de projection-détecteur, sont rassemblés en tant que lignes d'une matrice \mathbf{P} appelée matrice de projection, ce qui permet de représenter les intensités photoniques mesurées par tous les détecteurs pour tous les angles de projections de la manière suivante :

$$\mathbf{I}(\epsilon) = \int I_0(\epsilon) \exp(-\mathbf{P}\boldsymbol{\mu}(\epsilon)) d\epsilon \quad (2.5)$$

où \mathbf{P} est l'opérateur de projection discret composé de l'ensemble des \mathbf{p}_i .

Si on reprend la fonction de projection avec l'opérateur discret dans le cadre monochromatique, on obtient :

$$\mathbf{y} = \mathbf{P}\boldsymbol{\mu} \quad (2.6)$$

où \mathbf{y} représente le sinogramme

Et dans le cadre polychromatique, on obtient :

$$\tilde{\mathbf{y}} = \int b(\epsilon) \cdot \exp(-\mathbf{P}\boldsymbol{\mu}(\epsilon)) d\epsilon \quad (2.7)$$

2.1.4 Calcul de la matrice de projection

L'opérateur de projection qui représente la contribution de chaque pixel à l'atténuation mesurée par chaque paire angle de projection-détecteur va conditionner le système à résoudre. Il est donc nécessaire de choisir la méthode de calcul la plus adaptée à notre problème, c'est à dire ayant le meilleur compromis entre la facilité de manipulation de la décomposition et le réalisme du modèle.

Afin de faciliter le calcul de l'opérateur de projection, on attribue un coefficient égal à zéro à tous les pixels qui ne sont pas traversés par un des rayons considérés.

Il existe différentes méthodes de calcul de l'opérateur de projection. Parmi ces méthodes les plus courantes sont les méthodes dites *pixel-driven*, *ray-driven* et *distance-driven* (Fig. 2.2).

La méthode de calcul de l'opérateur de projection *pixel-driven* consiste à faire passer une

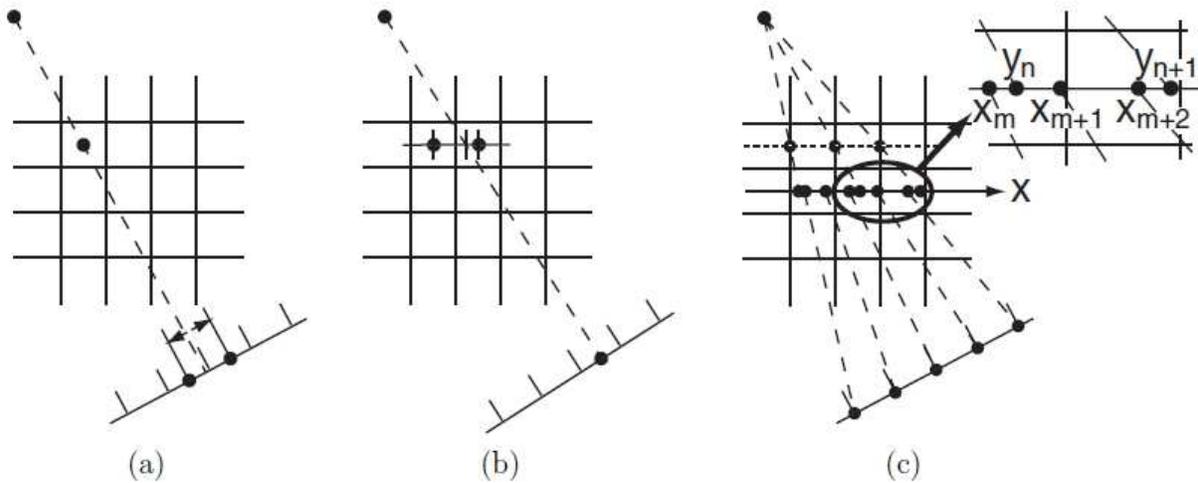


Figure 2.2 (a) méthode *pixel-driven*. (b) méthode *ray-driven*. (c) méthode *distance-driven*
©De Man et Basu (2004)

ligne de la source à rayons X vers le détecteur de telle sorte qu'elle passe par le centre du pixel étudié. La valeur du pixel sera alors calculée par interpolation linéaire et application de facteurs de pondération relatifs à l'angle de projection de la ligne et au rapport entre la taille des pixels et des détecteurs. L'inconvénient de cette méthode est que les détecteurs doivent être régulièrement espacés. Cette méthode provoque l'apparition d'artéfacts de hautes fréquences. Ces artéfacts peuvent être prévenus en utilisant une pondération plus sophistiquée qui va augmenter la complexité de calcul de cette méthode résultant à un temps de calcul rallongé (De Man et Basu (2004)).

La méthode dite *ray-driven* consiste à tracer une ligne reliant la source au centre du

détecteur concerné. L'opérateur est ensuite défini en calculant la contribution de chaque voxel à l'atténuation de chaque rayon. Cette méthode est la plus facile à utiliser car elle comporte le moins de contraintes, mais introduit des artéfacts lors de la rétroprojection et nécessite un accès à la mémoire non séquentiel qui augmente le temps de calcul. Ces artéfacts peuvent être grandement atténués en utilisant la technique de rayons-multiples. L'augmentation du temps de calcul peut être contournée en pré-calculant et stockant l'ensemble de la matrice de projection, mais cette solution est inadaptée pour des jeux de données de grande taille.

La troisième méthode *distance-driven* est définie dans le cadre 2D par De Man et Basu (2004). Elle consiste à représenter sur un même axe les détecteurs et les pixels pour ensuite effectuer un ensemble d'opérations sur toutes les frontières. la boucle principale de calcul est sur la ligne commune d'interceptions plutôt que sur les pixels ou les détecteurs, ce qui rend les calculs rapides. De plus, De Man et Basu (2002) montrent aussi que cette méthode concilie les avantages des deux méthodes précédentes : un accès séquentiel à la mémoire et une bonne vitesse de calcul. En outre, elle ne considère plus les détecteurs comme ponctuels, ce qui permet une meilleure représentation du phénomène de projection. Cette méthode est malheureusement plus complexe que les deux précédentes et nécessite une grille de pixels orthonormée afin de pouvoir établir cet axe commun aux détecteurs et aux pixels.

En pratique, plus d'un rayon atteint chaque détecteur car ceux ci ne sont pas ponctuels. Et l'intégrale ligne utilisée dans les formules 2.1 et 2.2 est une approximation qui peut faire apparaître des artéfacts sur les images reconstruites. Afin de surmonter ce problème, on utilise la technique dite de rayons-multiples qui considère que U rayons atteignent chaque détecteur lors du calcul des coefficients de la matrice de projection. on utilisera alors la moyenne sur les U rayons pour le calcul de chaque coefficient de la matrice de projection. Le nombre de rayons U sera choisi en fonction du type de scanner utilisé pour l'acquisition des données.

Le principal problème qui se pose lors de cette étape est le temps de calcul de cette matrice. Ce problème peut être contourné en pré-calculant la matrice de projection. L'autre problème qui se pose alors est la taille de cette matrice qui contient un nombre de valeurs égal à la taille de l'objet à reconstruire multipliée par le nombre total de paires détecteur-angle de projection. Ces valeurs sont, pour la majorité, des zéros. Il est donc nécessaire de trouver un moyen de résoudre ce problème en trouvant un format de stockage adapté et, si possible, des méthodes pour réduire la taille de cette matrice.

Une fois la projection modélisée, pour retrouver l'image à partir des mesures de projection obtenues, il est nécessaire d'utiliser une méthode d'inversion afin de retrouver l'image à partir

des projections. Pour ce faire, il existe deux grands types de méthodes : les méthodes de reconstruction analytiques et les méthodes de reconstruction algébriques.

2.2 Description des méthodes de reconstruction analytiques

Les méthodes de reconstruction actuelles les plus courantes sont adaptées d'une représentation déterministe et continue du problème. Les projections sont assimilées à des intégrales et peuvent être représentées par la transformée de Radon dans le cadre de rayons parallèles. Cette transformée de Radon est ensuite inversée pour effectuer le calcul de la rétroprojection. Il est ensuite nécessaire de discrétiser le problème pour pouvoir effectuer les calculs numériquement et l'adapter aux rayons en éventail. La méthode analytique la plus fréquemment utilisée est la rétroprojection filtrée ou FBP (*filtered backprojection*). Cette méthode est décrite par Kak et Slaney (1987) et ne sera pas détaillée dans ce mémoire. Elle utilise une modélisation monochromatique de la source et consiste à reconstruire l'image en calculant la rétroprojection du sinogramme filtré par un filtre rampe. Elle peut être décrite avec la formule suivante :

$$\hat{\boldsymbol{\mu}} = \mathbf{P}^T \mathbf{F}_{FBP} \mathbf{y} \quad (2.8)$$

où $\hat{\boldsymbol{\mu}}$ est l'objet reconstruit, \mathbf{P}^T la transposée de la matrice de projection et \mathbf{F}_{FBP} le filtre rampe.

Les avantages principaux de cette méthode sont sa rapidité et le faible coût en mémoire. En effet la transposée de la matrice de projection peut être calculée à la volée car elle n'est utilisée qu'une fois lors de la reconstruction ce qui permet de minimiser au maximum l'espace mémoire requis pour l'opérateur de rétroprojection.

Les inconvénients de cette méthode viennent de la modélisation des données et sont :

- l'utilisation d'un modèle déterministe qui ne modélise pas le bruit, cette approche ne considère pas la nature quantique de l'atténuation des rayons X ou le bruit introduit par les appareils de mesure, ce qui la rend très sensible au niveau de bruit des mesures ;
- l'approximation monochromatique de la source est inexacte. En effet, en réalité la source de rayons X est polychromatique et l'approximation monochromatique provoque notamment l'apparition d'artéfacts métalliques (Hamelin (2009)).

2.3 Description des algorithmes de reconstruction algébriques

Ces algorithmes consistent à résoudre, à l'aide de méthodes numériques d'estimation itératives, un système basé sur 2.7 ou 2.6, si on utilise l'hypothèse de monochromaticité, et un modèle stochastique des données.

2.3.1 Modèle des données

2.3.1.1 Modélisation monochromatique

Le modèle le plus simple utilisé par les algorithmes de reconstruction algébrique est le modèle monochromatique. En discrétisant la loi de Beer-Lambert (2.3), le modèle s'écrit :

$$\mathbf{y} = \mathbf{P}\boldsymbol{\mu} \quad (2.9)$$

Cette modélisation de la source a comme avantage de diminuer la complexité de calcul de la reconstruction et donc de l'accélérer. L'inconvénient de ce modèle est qu'il ne prend pas en compte la dépendance des coefficients d'atténuations au niveau d'énergie, ce qui provoque la création d'artéfacts notamment lors de la reconstruction de matériaux de forts coefficients d'atténuation tels les métaux.

Le modèle monochromatique est le plus utilisé car il permet les reconstructions les plus rapides et nécessite le moins d'espace mémoire.

2.3.1.2 Modélisation polychromatique

Il est possible de ne pas utiliser l'hypothèse de monochromaticité des photons X captés afin de garder un modèle plus réaliste qui permet entre autre une meilleure réduction des artéfacts métalliques. On peut utiliser le modèle de projection 2.7 qui tient compte du spectre d'émission de la source et de la variation des coefficients d'atténuation selon l'énergie des photons qui traversent l'objet. Un tel modèle de projection est non linéaire et se prête mal à l'inversion analytique. Il convient donc mieux aux approches statistiques de reconstruction.

Ce type de modèle est principalement utilisé pour la réduction d'artéfacts métalliques. Une des application de ce modèle est l'imagerie par tomographie à rayons X des vaisseaux périphériques traités par angioplastie avec implantation d'un tuteur endovasculaire métallique ou stent. Cette application nécessite la reconstruction d'images de haute résolution. La présence d'un stent métallique provoque l'apparition d'artéfacts métalliques qui nuisent à la précision de la mesure dans les images reconstruites, la réduction de ces artéfacts devient alors nécessaire. Dans ce cadre, l'utilisation d'un modèle polychromatique peut s'avérer très utile pour remplacer les prétraitements du sinogramme utilisés habituellement par les scanner cliniques qui restent peu efficaces pour réduire les artéfacts métalliques.

Pour construire un tel modèle, il faut repartir de l'équation 2.7. Dans ce modèle on considère que la valeur des coefficients d'atténuation varie selon l'énergie des photons X. Pour des raisons évidentes d'implémentation du modèle, il est obligatoire d'adopter un modèle discret du spectre représenté par un vecteur $\mathbf{b} = [b_1, b_2, \dots, b_K]^t \in \mathbb{R}^K$. Ce qui donne l'équation

2.7 discrétisée :

$$\tilde{y} = \sum_{k=0}^K b_k \cdot \exp(-\mathbf{P}\boldsymbol{\mu}_k) \quad (2.10)$$

L'objet est alors représenté par K images correspondantes aux K niveaux d'énergie échantillonnés. Il est donc nécessaire de trouver une simplification pour avoir une seule image. Pour cela la décomposition de Alvarez-Macovski est utilisée. Celle-ci permet de décrire la variation de la valeur du coefficient d'atténuation en fonction du niveau d'énergie suivant deux effets : l'effet photoélectrique et l'effet Compton. La décomposition s'écrit :

$$\boldsymbol{\mu}_k = \Phi_k \cdot \phi(\mu_0) + \Theta_k \cdot \theta(\mu_0) \quad (2.11)$$

où μ_0 l'objet à l'énergie de référence E_0 (e.g., $E_0 = 70keV$).

La contribution de l'effet photoélectrique est composée de :

1. Φ_k : constante photoélectrique à l'énergie k , calculée avec la formule suivante : $\Phi_k = \frac{1/E_k}{1/E_0^3}$ avec E_0 énergie de référence,
2. ϕ : coefficient photoélectrique.

Et la contribution de l'effet Compton est composée de :

1. Θ_k : constante de Compton à l'énergie k , calculée avec la formule suivante : $\Theta_k = \frac{f_{KN}(E_k)}{f_{KN}(E_0)}$ et f_{KN} : fonction de Klein-Nishina,
2. θ : coefficient de Compton.

Le problème se situe maintenant dans l'évaluation des coefficients ϕ et θ . Pour cela, il existe deux méthodes.

La première méthode, développée par De Man *et al.* (2001), utilise des tables de mesures de $\boldsymbol{\mu}_k$, ϕ et θ en fonction de l'énergie des photons X, pour établir de manière empirique les fonctions non linéaires $\phi = f(\mu_{70keV})$ et $\theta = f(\mu_{70keV})$.

La deuxième méthode, développée par Menvielle (2004), consiste à définir : $\theta = \alpha \cdot \phi + \beta$ et calculer α et β en utilisant les données de De Man *et al.* (2001).

Ces méthodes permettent d'estimer la variation de l'effet photoélectrique et de Compton en fonction de la valeur du coefficient à l'énergie de référence.

Le modèle polychromatique a comme avantage de proposer une représentation plus réaliste de la projection qui entraîne des images reconstruites de meilleure qualité (Hamelin *et al.* (2008)), mais aux dépens de la simplicité et du temps de calcul.

2.3.2 Modélisation du bruit

En réalité, la nature quantique de l'atténuation du faisceau de rayons X, les limitations du matériel de détection et l'inadéquation du modèle de projection à la nature du rayonnement sont des sources d'incertitudes qui altèrent les mesures d'atténuations en y ajoutant du bruit. Le rapport signal à bruit du sinogramme obtenu avec un scanner clinique actuel est d'environ 30 dB mais celui-ci varie généralement entre 20 et 40 dB suivant la dose de rayons X utilisée, l'objet à imager et la qualité des détecteurs du scanner.

On pose l'hypothèse que l'incertitude sur \tilde{y} est principalement due aux effets quantiques liés à l'atténuation des rayons X dans l'objet (Guan et Gordon (1996)). Les mesures d'atténuations sont alors adéquatement modélisées par une loi de Poisson multivariée (Macovski (1983)) de moyenne décrite par $\tilde{y} = \exp(-\mathbf{y}) = \exp(-\mathbf{P}\boldsymbol{\mu})$ pour le monochromatique et 2.10 pour le polychromatique. Le lecteur peut se rapporter à la section 1.1.4 de Hamelin (2009) pour une explication détaillée de la modélisation.

Pour la suite de ce chapitre, afin de faciliter la compréhension, on ne parlera que du modèle monochromatique.

On peut simplifier le problème en posant l'hypothèse que de grands nombre de photons ont été captés lors des mesures. Il est alors possible d'approximer la loi de Poisson par une loi gaussienne multivariée (Bouman et Sauer (1993)), ce qui nous permet de poser :

$$\tilde{y} = \exp(-\mathbf{P}\boldsymbol{\mu}) + \tilde{b} \quad (2.12)$$

où $\tilde{b} \hookrightarrow \mathbf{N}(0, \text{diag}[\exp(-\mathbf{P}\boldsymbol{\mu})])$. On développe alors :

$$\log \tilde{y} = \exp(-\mathbf{P}\boldsymbol{\mu}) + \log(1 + [\exp(-\mathbf{P}\boldsymbol{\mu})]^{-1}\tilde{b}) \quad (2.13)$$

Pourvu que \tilde{b} soit petit par rapport à $\exp(-\mathbf{P}\boldsymbol{\mu})$, le dernier terme est adéquatement approché par un développement de Taylor du premier ordre, d'où :

$$\mathbf{y} = -\log(\tilde{y}) = \mathbf{P}\boldsymbol{\mu} + \mathbf{b} \quad (2.14)$$

avec $\mathbf{b} \hookrightarrow \mathbf{N}(0, \Sigma_b)$.

On peut modéliser ce bruit de plusieurs manière différentes :

- soit en utilisant une variance uniforme. On peut représenter la variance du bruit \mathbf{b} de l'équation 2.14 par $\Sigma_b = \sigma^2 \mathbf{I}$ où σ^2 est la variance du bruit,
- soit en utilisant une variance dépendant du nombre de photons pour tenir compte de la distribution poissonnienne du bruit. Pour cela, on utilise l'approximation développée

par Sauer et Bouman (1993)

$$\Sigma_b = \text{diag}\{\exp(-y_1), \exp(-y_2), \dots, \exp(-y_{N_Y})\} \quad (2.15)$$

La deuxième source de bruit la plus importante est le bruit électronique introduit par les limitations du matériel de détection dans le sinogramme, ce bruit est modélisé par un bruit blanc gaussien qui peut être négligé pour des comptes de photons suffisamment importants Ma *et al.* (2012).

2.3.3 Estimateur

Dans le cadre d'un modèle stochastique, la reconstruction correspond à l'estimation du paramètre $\boldsymbol{\mu}$ du modèle probabiliste des projections présenté dans la section précédente. Il est donc nécessaire d'établir des estimateurs cohérents avec le modèle statistique. En tomographie, les deux principaux estimateurs utilisés sont : l'estimation par maximum de vraisemblance (ML) et l'estimation *a posteriori* (MAP) car ce sont les estimateurs ponctuels les plus simples. La différence entre les deux est que l'estimateur MAP prend en compte les informations connues de l'objet contrairement à l'estimateur MV, l'avantage que présente cet ajout est détaillé dans 2.3.3.2.

Lors de l'estimation, on admet une contrainte de non-négativité, c'est à dire que l'on suppose qu'aucun matériau traversé durant les projection n'amplifie l'intensité des rayons X ou ne produit de photon.

2.3.3.1 Maximum de vraisemblance

Le maximum de vraisemblance (ML) réduit la reconstruction à la maximisation de la vraisemblance de la loi de Poisson dont la densité est :

$$f(\tilde{y}; \boldsymbol{\mu}) = \frac{\exp[-p(\boldsymbol{\mu})] p(\boldsymbol{\mu})^{\tilde{y}}}{\tilde{y}!} \quad (2.16)$$

où $p(\boldsymbol{\mu}) = \exp(-\mathbf{P}\boldsymbol{\mu})$. Ce qui revient à résoudre le problème d'optimisation non linéaire avec contrainte de non-négativité (Hamelin (2009)) :

$$\hat{\boldsymbol{\mu}} = \underset{\boldsymbol{\mu} \geq 0}{\operatorname{argmin}} \sum_{i=0}^N [p_i(\boldsymbol{\mu}) - \tilde{y}_i \log p_i(\boldsymbol{\mu})] \quad (2.17)$$

où N le nombre de paires détecteur-angle de projection.

Si on reprend l'approximation gaussienne de loi de Poisson de la section précédente, la reconstruction par maximum de vraisemblance revient alors à maximiser la log-vraisemblance

gaussienne, ce qui revient à résoudre le problème des moindres carrés suivant (Sauer et Bouman (1993)) :

$$\hat{\boldsymbol{\mu}} = \underset{\boldsymbol{\mu} \geq 0}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{y} - \mathbf{P}\boldsymbol{\mu}\|_{\boldsymbol{\Sigma}_b}^2 \quad (2.18)$$

2.3.3.2 Maximum *a posteriori*

L'estimateur du maximum *a posteriori* est une approche bayésienne d'estimation qui maximise la log-vraisemblance de l'objet conditionnellement aux données. Ce type d'approche a comme avantage de produire des algorithmes de reconstruction plus robustes que ceux obtenus par estimation de la ML dans le cas où les données sont peu informatives car l'apport d'une information *a priori* peut pallier le manque relatif d'information apporté par les mesures.

La log-vraisemblance à maximiser provient de la densité conditionnelle $f(\boldsymbol{\mu} | \mathbf{y})$ proportionnelle au produit de la densité conditionnelle des données $f(\mathbf{y} | \boldsymbol{\mu})$ et de la densité *a priori* $f(\boldsymbol{\mu})$ de l'objet. Cette densité *a priori* décrit certaines propriétés supposées de l'objet. On pose l'hypothèse que cette densité suit une loi exponentielle de la forme :

$$f(\boldsymbol{\mu}) = K \exp \left(-\lambda \sum_{i=1}^J \nu_i \psi(\delta_i^t \boldsymbol{\mu}) \right) \quad (2.19)$$

où $\psi(u)$ est une fonction de pénalisation, δ_i détermine une propriété de l'image à pénaliser, $\lambda \nu_i$ pondère la pénalisation de cette propriété et K normalise la densité. D'après Idier (2001), l'estimateur MAP s'écrit :

$$\hat{\boldsymbol{\mu}} = \underset{\boldsymbol{\mu} \geq 0}{\operatorname{argmin}} \Phi(\boldsymbol{\mu}) + \lambda R(\boldsymbol{\mu}) \quad (2.20)$$

où $\Phi(\boldsymbol{\mu})$ est la log-vraisemblance poissonienne 2.17 ou gaussienne 2.18 et λ est un scalaire contrôlant l'équilibre entre la fidélité de la solution aux données mesurées et l'information *a priori*. λ est appelé paramètre de régularisation. $R(\boldsymbol{\mu})$ correspond à la fonction de régularisation (voir 2.3.3.3) et s'écrit :

$$R(\boldsymbol{\mu}) = \sum_{i=1}^J \nu_i \psi(\delta_i^t \boldsymbol{\mu}) \quad (2.21)$$

2.3.3.3 Fonction de pénalisation

La fonction de pénalisation sert à encourager les comportements désirables de l'image reconstruite. Ces propriétés sont déterminées par les hypothèses faites à partir des connaissances *a priori* de l'objet.

En reconstruction tomographique, la plupart des fonction de régularisation peuvent être

représentées sous la forme suivante (Fessler (2000)) :

$$R(\boldsymbol{\mu}) = \frac{1}{2} \sum_{i=1}^J \psi_i([\mathbf{D}\boldsymbol{\mu}]_i) \quad (2.22)$$

où \mathbf{D} est la matrice de pénalisation de taille $J \times N$ avec N le nombre d'éléments de $\boldsymbol{\mu}$ et $J \in \mathbb{N}$ dépend du type de pénalisation choisi. Cette forme est une version simplifiée de 2.21. La matrice \mathbf{D} est définie en fonction ce que l'on veut pénaliser. Par exemple, si on veut pénaliser la valeur des pixels directement on choisit $\mathbf{D} = \mathbf{I}$.

On verra dans la section suivante que les techniques de minimisation classiques convergent vers une solution globale si et seulement si la fonction à minimiser est dérivable et convexe, ce qui veut dire qu'on choisit la fonction de régularisation convexe et dérivable afin de faciliter la convergence.

En tomographie à rayons X, on suppose que l'image est constituée de zones homogènes, c'est à dire des structures dont les coefficients d'atténuation intrinsèques varient peu ou prou. Ces zones sont séparées par des frontières nettes ou des espaces vides.

On cherche d'abord à pénaliser les variations entre pixels voisins. Si on ne considère que les différences premières entre pixels voisins horizontalement et verticalement, on définira \mathbf{D} de manière à avoir $\forall i \in \mathbb{N}, [\mathbf{D}\boldsymbol{\mu}]_i = \sum_k (\boldsymbol{\mu}_i - \boldsymbol{\mu}_k)$ avec k indice des pixels voisins de $\boldsymbol{\mu}_i$.

Une des fonctions de pénalisation la plus simple et la plus utilisée en tomographie est la pénalisation quadratique convexe : $\psi(x) = x^2/2$ appelée L2. l'avantage de cette fonction est que dans le contexte du modèle gaussien 2.18, l'estimateur devient un problème de moindres carrés linéaires de la forme :

$$\operatorname{argmin}_{\boldsymbol{\mu} \geq 0} \frac{1}{2} \|\mathbf{y} - \mathbf{P}\boldsymbol{\mu}\|^2 + \frac{\lambda}{2} \boldsymbol{\mu}^T \mathbf{D}^T \mathbf{D} \boldsymbol{\mu} \quad (2.23)$$

Il existe des méthodes numériques rapides pour résoudre ce problème 2.23 (Nocedal et Wright (1999)). Si on néglige la contrainte de non-négativité, l'estimateur résout le système linéaire suivant :

$$(\mathbf{P}^T \mathbf{P} + \lambda \mathbf{D}^T \mathbf{D}) \hat{\boldsymbol{\mu}} = \mathbf{P}^T \mathbf{y} \quad (2.24)$$

La solution exacte de ce système devient calculable même si elle reste très coûteuse en temps de calcul et espace mémoire.

L'inconvénient majeur de cette pénalité L2 est qu'elle pénalise fortement les grandes variations de valeurs entre pixels voisins, ce qui va conduire à un lissage des limites des différentes zones homogènes.

Ces limites sont mieux reconstruites avec l'utilisation d'une fonction de pénalité convexe

asymptotiquement linéaire type L2L1 (Charbonnier *et al.* (1997)). Un exemple de fonction L2L1 est une fonction qui correspond à la branche supérieure d'une hyperbole verticale $\psi(x) = \sqrt{x^2 + \delta^2} - \delta$. δ contrôle le point de transition entre le comportement quadratique proche de l'origine et le comportement linéaire de la fonction. Ce type de fonction va pénaliser quadratiquement les petites oscillations et linéairement les grosses différences d'atténuations. Cette méthode permet d'obtenir une solution ayant un meilleur équilibre entre correction du bruit et perte de résolution que L2 mais au prix d'une augmentation de la complexité de calcul. Il n'existe alors plus de solution explicite ce qui va rendre impossible le calcul de la solution exacte.

2.3.4 Algorithme d'optimisation

Maintenant qu'on a l'estimateur ou fonction de coût, il est nécessaire d'utiliser des algorithmes d'optimisation pour trouver son minimum. Lors du choix de cet algorithme, il est nécessaire de tenir compte de nos objectifs que sont la réduction du temps de calcul et de l'espace mémoire à utiliser. On ne s'intéresse dans cette section qu'aux algorithmes d'optimisation à directions de descente car ce sont, de loin, les plus utilisés en tomographie. On ne décrira pas l'algorithme espérance-maximisation (EM) (Dempster *et al.* (1977)) car il est reconnu pour avoir un faible taux de convergence. De plus on ne présentera que les algorithmes travaillant sur l'ensemble de l'image plutôt que sur ceux qui considèrent les pixels séparément tels ICD (Sauer et Bouman (1993)) et OS-EM (Hudson et Larkin (1994)) qui ont des problèmes de convergence globale.

2.3.4.1 Généralités

On cherche à résoudre : $\underset{x}{\operatorname{argmin}} f(x)$ avec $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et $f \in C^1$.

Les algorithmes à directions de descente génèrent une suite d'itérées d'équation suivante :

$$x_{k+1} = x_k + \alpha_k \mathbf{d}_k \quad (2.25)$$

Le déplacement est calculé en deux étapes :

- identification d'une direction de descente \mathbf{d}_k ,
- recherche linéaire d'une longueur de pas α_k dans cette direction.

Une direction de descente \mathbf{d}_k est une direction selon laquelle un déplacement au voisinage immédiat du point courant résulte en une diminution de la valeur de la fonction objectif. \mathbf{d}_k est une direction de descente pour f en x_k si et seulement si $\nabla f(x_k)^T \mathbf{d}_k < 0$. Sauf si celle-ci est nulle, ce qui indique un point stationnaire, la direction opposée au gradient est celle qui résulte en la plus forte descente.

L'étape de recherche linéaire consiste à trouver les suites de couples (α_k, \mathbf{d}_k) qui permettent d'atteindre le minimum de f . Pour cela, il est possible d'imposer une décroissance minimum de f en utilisant une recherche linéaire qui satisfasse la condition d'Armijo :

$$f(x_k + \alpha_k \mathbf{d}_k) \leq f(x_k) + \gamma_{Armijo} \alpha_k \nabla f(x_k)^T \mathbf{d}_k \quad (2.26)$$

où $\gamma_{Armijo} \in]0, 1[$.

Un algorithme à direction de descente utilisant la recherche linéaire d'Armijo converge vers un minimum global ssi $f \in \mathbb{C}^2$ et f est convexe.

Cette étape peut revenir à obtenir une solution α_k au sous problème de minimisation à une variable en effectuant une recherche linéaire de α_k le long de \mathbf{d}_k lorsque le calcul de $f(x_k + \alpha_k \mathbf{d}_k)$ est possible :

$$\alpha_k = \underset{\alpha}{\operatorname{argmin}} f(x_k + \alpha \mathbf{d}_k) \quad (2.27)$$

avec $\alpha > 0$.

2.3.4.2 Méthodes d'optimisation de la plus forte pente, de Newton et de quasi-Newton

Ces méthodes calculent une direction de descente \mathbf{d}_k en résolvant un système linéaire de la forme :

$$\mathbf{M}_k \mathbf{d}_k = -\nabla f(x_k) \quad (2.28)$$

avec \mathbf{M}_k définie positive.

La direction calculée correspond à une transformation linéaire de l'opposé du gradient :

$$\mathbf{d}_k = -\mathbf{M}_k^{-1} \nabla f(x_k) \quad (2.29)$$

Il existe trois principales méthodes d'optimisation par descente de gradient : algorithmes d'optimisation de la plus forte pente, de Newton et de quasi-Newton.

Le premier et le plus simple est la méthode de la plus forte descente. Elle utilise $\mathbf{M}_k = \mathbf{Id}$ dans 2.29. La descente est calculée avec la formule suivante :

$$\mathbf{d}_k = -\nabla f(x_k) \quad (2.30)$$

cette méthode est utilisée par les algorithmes type SPS (Erdoğan et Fessler (1999)). Elle a comme avantage d'avoir une complexité de calcul très faible mais a comme inconvénient une vitesse de convergence linéaire.

La deuxième méthode est la méthode dite de Newton. Cette méthode utilise le développement en série de Taylor du deuxième ordre pour approximer la fonction autour du point courant. Ce qui donne l'équation suivante :

$$\nabla^2 f(x_k) \mathbf{d}_k = -\nabla f(x_k) \quad (2.31)$$

Cette méthode permet d'avoir une vitesse de convergence quadratique mais elle demande de calculer et le gradient et l'inverse du hessien de f à chaque itération. Il est nécessaire d'avoir un hessien inversible (i.e. défini positif vu qu'il s'agit d'un minimum) pour pouvoir appliquer cette méthode afin de pouvoir aisément déterminer \mathbf{d}_k . Le calcul de \mathbf{d}_k à chaque itération demande de nombreux et longs calculs ce qui augmente la complexité de calcul de la méthode et la rend inutilisable en pratique.

La troisième méthode dite méthode quasi-Newton, calcule la direction de descente de la manière suivante :

$$\mathbf{d}_k = -\mathbf{B}_k^{-1} \nabla f(x_k) \quad (2.32)$$

où \mathbf{B}_k est une approximation définie positive inversible du hessien. Cette approximation et son inverse sont généralement mises à jour à chaque itération. \mathbf{B}_k est calculée de manière à satisfaire l'équation sécante ou condition de quasi-Newton :

$$\mathbf{B}_k(x_k - x_{k-1}) = \nabla f(x_k) - \nabla f(x_{k-1}) \quad (2.33)$$

Avec cette approximation du hessien, on n'a pas l'assurance que \mathbf{d}_k est une direction de descente par recherche linéaire d'Armijo. Pour surmonter ce problème on remplace la condition d'Armijo par la condition forte de Wolfe :

$$\begin{aligned} f(x_k + \alpha_k \mathbf{d}_k) &\leq f(x_k) + \gamma_{Armijo} \mathbf{B}_k \nabla f(x_k)^T \mathbf{d}_k \\ |f(x_k + \alpha_k \mathbf{d}_k)^T \mathbf{d}_k| &\leq \gamma_{Wolfe} |\nabla f(x_k)^T \mathbf{d}_k| \end{aligned} \quad (2.34)$$

où $\gamma_{Wolfe} \in]\gamma_{Armijo}, 1[$.

La méthode de quasi-Newton propose un bon compromis entre les méthodes de plus forte pente et de Newton, c'est à dire au prix d'une complexité de calcul moyenne : on ne calcule pas l'inverse du hessien mais on effectue des calculs matriciels pour trouver une approximation aisément inversible du hessien, elle permet d'avoir une vitesse de convergence super-linéaire. Cette méthode est notamment utilisée par les algorithmes BFGS et L-BFGS (Nocedal et Wright (1999)).

Hamelin *et al.* (2011) montre que l'algorithme L-BFGS-B est une des meilleures méthodes d'optimisation dans le cadre d'une reconstruction de tomographie à rayons X en terme de vitesse de calcul et de qualité de l'image reconstruite.

2.3.4.3 Préconditionnement

En reprenant la fonction objectif à minimiser 2.23, on peut la réécrire :

$$f(\boldsymbol{\mu}) = \frac{1}{2}\boldsymbol{\mu}^T(\mathbf{P}^T\mathbf{P} + \lambda\mathbf{D}^T\mathbf{D})\boldsymbol{\mu} - \mathbf{y}^T\mathbf{P}\boldsymbol{\mu} \quad (2.35)$$

On peut simplifier cette équation en :

$$f(\boldsymbol{\mu}) = \frac{1}{2}\boldsymbol{\mu}^T\mathbf{M}\boldsymbol{\mu} - \mathbf{v}\boldsymbol{\mu} \quad (2.36)$$

avec $\mathbf{M} = \mathbf{P}^T\mathbf{P} + \lambda\mathbf{D}^T\mathbf{D}$ et $\mathbf{v} = \mathbf{y}^T\mathbf{P}$.

En reprenant l'équation à résoudre 2.24, on observe qu'elle peut être réécrite sous la forme :

$$\mathbf{M}\boldsymbol{\mu} = \mathbf{v} \quad (2.37)$$

réduisant le problème à un système linéaire à n équations avec \mathbf{M} la matrice normale.

D'après Nocedal et Wright (1999), la vitesse de convergence de la minimisation de f avec \mathbf{M} matrice définie positive par un algorithme de direction de descente dépend de la distribution des valeurs propres de \mathbf{M} . En effet les valeurs propres de \mathbf{M} conditionnent \mathbf{M} et donc f , moins les valeurs propres seront étalées meilleure sera la direction de descente calculée. Si le lecteur désire la démonstration complète de cette affirmation, il pourra se reporter au paragraphe 3.3 de Nocedal et Wright (1999).

Le nombre de condition d'une matrice est une bonne représentation de la distribution des valeurs propres d'une matrice et permet d'en évaluer le conditionnement. Pour une matrice \mathbf{M} non-singulière il est défini tel que (Chen (2005)) :

$$\kappa(\mathbf{M}) = \|\mathbf{M}\|\|\mathbf{M}^{-1}\| \quad (2.38)$$

Où $\|\cdot\|$ est la norme de la matrice.

Pour une matrice symétrique, non-singulière cette relation est réduite à :

$$\kappa(\mathbf{M}) = \frac{|\lambda|_{max}}{|\lambda|_{min}}$$

où $|\lambda|_{max}$ et $|\lambda|_{min}$ correspondent respectivement à la plus grande et la plus petite valeur

propre de la matrice du système. $\kappa(\mathbf{M}) \geq 1$ et idéalement $\kappa(\mathbf{M}) = 1$. Plus le nombre de condition est proche de 1 et moins le spectre des valeurs propres est étendu.

En modifiant ou préconditionnant le système 2.37, on peut rendre cette distribution de valeurs propres plus favorable et améliorer significativement la vitesse de convergence des méthodes d'optimisation à direction de descente. Le préconditionnement a pour but de réduire l'étalement des valeurs propres de la matrice du système et par conséquent le nombre de conditions. Cette opération consiste en un changement de la variable $\boldsymbol{\mu}$ dans la fonction f de 2.36 en $\mathbf{C}\boldsymbol{\gamma}$ avec \mathbf{C} une matrice inversible appelée le préconditionneur. Cette opération en reprenant 2.36 et en effectuant le changement de variable $\boldsymbol{\mu} = \mathbf{C}\boldsymbol{\gamma}$, f devient :

$$f(\mathbf{C}\boldsymbol{\gamma}) = \frac{1}{2}\boldsymbol{\gamma}^t\mathbf{C}^t\mathbf{M}\mathbf{C}\boldsymbol{\gamma} - b\mathbf{C}\boldsymbol{\gamma} \quad (2.39)$$

L'estimateur résout alors le système suivant :

$$\mathbf{C}^t(\mathbf{P}^t\mathbf{P} + \lambda\mathbf{D}^t\mathbf{D})\mathbf{C}\hat{\boldsymbol{\gamma}} = \mathbf{C}^t\mathbf{P}^t\mathbf{y} \quad (2.40)$$

Le but des préconditionneurs est de réduire l'étalement des valeurs propres de $\mathbf{C}^t(\mathbf{P}^t\mathbf{P} + \lambda\mathbf{D}^t\mathbf{D})\mathbf{C}$ ce qui équivaut à se rapprocher le plus possible de l'inverse du hessien de f , idéalement : $\mathbf{C}^t\mathbf{C} = (\mathbf{P}^t\mathbf{P} + \lambda\mathbf{D}^t\mathbf{D})^{-1}$.

Il existe un grand nombre de préconditionneurs différents dont certains seront exposés dans 2.4.3.

2.4 Description de l'algorithme de reconstruction algébrique en coordonnées cylindriques

À chaque itération k , on effectue le calcul de $\nabla f(x_k) = \mathbf{P}^t(\mathbf{y} - \mathbf{P}x_k) + \lambda\mathbf{D}^t\mathbf{D}x_k$ qui nécessite une projection et une rétroprojection pour la recherche de direction de descente et on calcule la valeur de $f(x_k) = \frac{1}{2} \|\mathbf{y} - \mathbf{P}x_k\|^2 + \frac{\lambda}{2}x_k^t\mathbf{D}^t\mathbf{D}x_k$ qui nécessite une projection pour la recherche linéaire du pas. On remarque qu'on effectue la même projection pour le calcul de $\nabla f(x_k)$ et $f(x_k)$. À chaque itération, on calcule donc une projection et une rétroprojection.

Les algorithmes de reconstruction algébriques itératifs effectuent plusieurs itérations lors de l'utilisation d'un algorithme d'optimisation de descente du gradient avec une estimation MV ou MAP. Le stockage en mémoire de la matrice de projection et sa transposée si-possible peut limiter le temps de calcul.

On a mentionné dans la partie 2.1.4 qu'un des problèmes de la matrice de projection en tomographie à rayon X est sa taille importante. Afin de surmonter ce problème on peut utiliser un modèle de représentation de l'objet non pas en coordonnées cartésiennes classiques

mais en coordonnées polaires en 2D et cylindriques ou sphériques en 3D avec une approche basée sur la rotation de l'objet (Zeng et Gullberg (1992); Zeng *et al.* (1994)) et non plus sur la rotation de la source conjointement aux détecteurs.

On considère que la source est fixe et que l'objet tourne sur lui même pour chaque angle de projection. Si la représentation de l'objet est invariante par rotation, alors, le projecteur est composé d'une matrice de projection pour un seul angle de projection et d'opérateurs de rotation de l'objet. L'objet est ensuite échantillonné sur une grille polaire suivant les coordonnées radiale ρ et tangentielle θ . Si on choisit un nombre d'échantillons tangentiels (θ) proportionnel au nombre de projection, la représentation polaire est invariante par rotation et permet d'assimiler les opérateurs de rotation de l'objet à de simples changements d'indices en jouant sur les symétries naturelles de la représentation. Ceci va conduire à une diminution significative de l'opérateur de projection constitué d'une matrice de projection pour un angle projection et d'opérations de changements d'indices.

2.4.1 Représentation de l'objet

La grille représentant l'objet est en coordonnées polaire en deux dimensions. Le problème avec ce type de représentation est le choix de la forme des voxels. En effet, contrairement à la représentation cartésienne les voxels ne sont plus cubiques et sont de tailles différentes, il est donc nécessaire de faire des compromis quand à leur forme et leur taille afin de satisfaire au mieux les objectifs recherchés que sont la diminution de la taille de la matrice de projection à stocker et la limitation de la complexité de calcul. il existe plusieurs types de représentation polaire ou cylindrique dont la taille et la forme des pixels ou voxels sont différentes. On présente d'abord les représentations en 2D qui ont été développées puis les représentations en 3D

En deux dimensions, la représentation la plus simple est celle décrite par Jian *et al.* (2007) et présentée par la Fig 2.3. les pixels sont de taille identiques sur un même cercle mais leur taille augmente avec leur distance au centre de l'image et le nombre de pixels tangentiels est un multiple du nombre de projections. L'avantage de cette représentation est que les frontières des pixels sont très simples à calculer. L'inconvénient de cette représentation est l'inhomogénéité des pixels qui va entraîner des difficultés de conditionnement du problème et de régularisation.

Rodriguez-Alvarez *et al.* (2011) a présenté deux autres représentations polaires qui ont pour but de garder des pixels d'aires équivalentes afin de limiter les difficultés de régularisation et de conditionnement. La première (Fig 2.4-A) considère des pixels circulaires et la deuxième (Fig 2.4-B) des pixels en trapèze. Ces deux représentations sont moins utiles car elles compliquent le calcul du projecteur et augmentent sa taille. De plus la validité physique

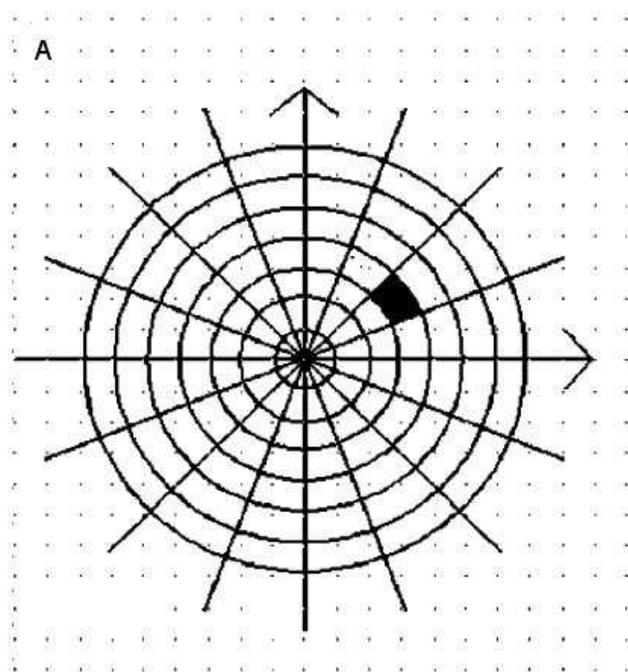


Figure 2.3 représentation polaire en 2D ©Jian *et al.* (2007)

des pixels circulaires est discutable car ils ne recouvrent pas la totalité de l'objet.

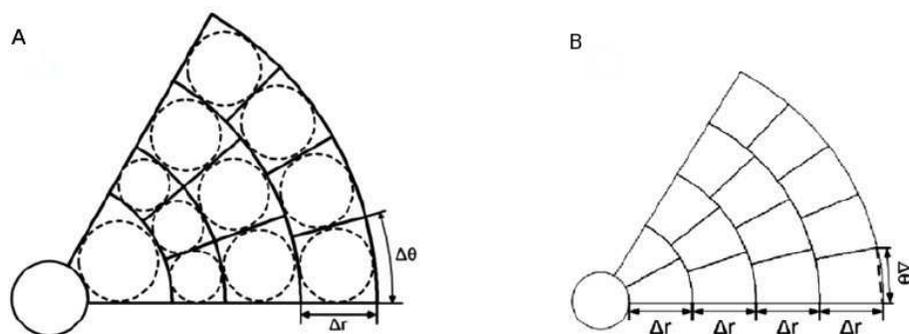


Figure 2.4 représentation polaire alternative en 2D ©Rodriguez-Alvarez *et al.* (2011)

Thibaudeau *et al.* (2013) a développé deux représentations polaires en trois dimensions, la première consiste en une représentation cylindrique de l'objet et la deuxième en une représentation sphérique de l'objet (Fig 2.5). Ces représentations conservent l'avantage de la représentation polaire 2D de Jian *et al.* (2007), c'est à dire qu'elles ne nécessitent le stockage de la projection que pour un seul angle de projection. L'avantage de la représentation cylindrique par rapport à la représentation sphérique est qu'elle correspond mieux aux objets à imager (corps humain) et permet une modélisation du décalage de la source perpendiculai-

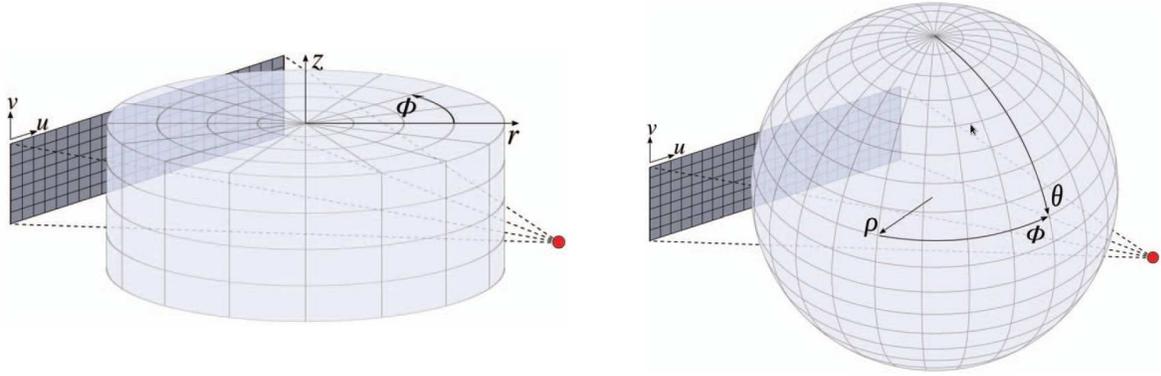


Figure 2.5 représentation cylindrique (à gauche) et sphérique (à droite) ©Thibaudeau *et al.* (2013)

rement à l'axe source-détecteur beaucoup plus aisée.

2.4.2 Propriétés de la matrice normale du système linéaire

La représentation de l'objet en coordonnées polaires introduit un certain nombre de propriétés que ce soit au niveau de la matrice de projection ou de la multiplication matrice de projection-vecteur (projection) ou bien encore la multiplication transposée de la matrice de projection-vecteur (rétroprojection).

2.4.2.1 Propriétés mathématiques

L'objet est représenté suivant les coordonnées (ρ, θ, z) suivant la représentation cylindrique de la Fig.2.5 avec z l'axe de rotation de l'objet.

Soit N_p le nombre de projections ou angles de projection. On reprend la représentation de Jian *et al.* (2007) étendue en trois dimensions avec la représentation cylindrique de Thibaudeau *et al.* (2013). On choisit dans cette représentation un nombre de voxels tangentiels égal au nombre de projection N_p et on admet que l'incrément de l'angle de projection est régulière. L'objet est donc discrétisé en N_p secteurs circulaires égaux.

Soit \mathbf{y}_n les atténuations mesurées pour l'angle de projection θ_n avec $0 \leq n \leq N_p - 1$ et \mathbf{P}_0 la matrice de projection pour l'angle de référence θ_0 . Le projecteur pour l'angle de projection θ_n est obtenu en faisant tourner l'objet d'un angle $-\theta_n$ puis en le multipliant par \mathbf{P}_0 pour obtenir \mathbf{y}_n . La rotation de l'objet avec cette représentation cylindrique peut être effectuée par une manipulation d'index et ne nécessite pas de calcul complexe ou coûteux en temps. on appelle \mathbf{S}_μ^{-1} , la rotation de l'objet d'un incrément d'angle. \mathbf{S}_μ^{-1} est en fait un opérateur de

décalage circulaire de K échantillons avec K le nombre de voxels tangentiels multiplié par le nombre de voxels dans la direction z .

La matrice de projection complète \mathbf{P} peut être écrite avec la relation suivante (Goussard *et al.* (2013)) :

$$\mathbf{P} = [\mathbf{P}_0^T \mid (\mathbf{P}_0 \mathbf{S}_\mu^{-1})^T \mid \dots \mid (\mathbf{P}_0 \mathbf{S}_\mu^{-N_p+1})^T]^T \quad (2.41)$$

On peut représenter \mathbf{P}_0 de la manière suivante :

$$\mathbf{P}_0 = [\mathbf{P}_{0,0} \mid \mathbf{P}_{0,1} \mid \dots \mid \mathbf{P}_{0,N_p-1}] \quad (2.42)$$

Où chaque $\mathbf{P}_{0,n}$; $0 \leq n \leq N_p - 1$ est un bloc (M, K) avec M le nombre total de détecteurs c'est à dire le nombre de détecteur par barrette multiplié par le nombre de barrettes et K le nombre de voxels tangentiels multiplié par le nombre de voxels dans la direction z .

Si on reprend 2.41 et qu'on y intègre 2.42, on obtient la relation suivante :

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{0,0} & \mathbf{P}_{0,1} & \dots & \mathbf{P}_{0,N_p-1} \\ \mathbf{P}_{0,N_p-1} & \mathbf{P}_{0,0} & \dots & \mathbf{P}_{0,N_p-2} \\ \vdots & & \ddots & \\ \mathbf{P}_{0,1} & \mathbf{P}_{0,2} & \dots & \mathbf{P}_{0,0} \end{bmatrix} = \begin{bmatrix} \mathbf{P}_0 \\ \mathbf{P}_0 \mathbf{S}_\mu^{-1} \\ \vdots \\ \mathbf{P}_0 \mathbf{S}_\mu^{-N_p+1} \end{bmatrix} \quad (2.43)$$

La matrice de projection est donc bloc-circulante et à un taux de remplissage très faible, en effet pour chaque ligne de \mathbf{P} , c'est à dire chaque couple angle de projection-détecteur, le nombre de voxels contribuant à l'atténuation mesuré est très faible. Par exemple pour une matrice de projection 2D avec 672 détecteurs, 1160 projections et 256 pixels tangentiels c'est à dire la géométrie standard 2D d'un scanner Siemens SOMATOM, son taux de remplissage est d'environ $3 * 10^{-3}$. D'après l'étude faite par Golkar (2013), $\mathbf{P}^T \mathbf{P}$ est de plus bloc-diagonale dominante.

2.4.2.2 Propriétés et simplification des opérations de projection et rétroprojection

Afin de diminuer le temps de calcul des opérations de projection et rétroprojection, on effectue une parallélisation de ces opérations. Pour effectuer une parallélisation efficace, il est nécessaire d'avoir des opérations indépendantes les unes des autres.

L'opération de projection consiste en une simple multiplication de la matrice de projection \mathbf{P} avec un vecteur objet. En regardant la décomposition 2.43 de \mathbf{P} , on peut voir que les projections (y_n) aux différents angles de projection peuvent être calculées de manière indé-

pendante. En effet on peut appliquer, de manière indépendante l'opérateur de rotation (\mathbf{S}_μ^{-n}), donc la manipulation d'index du vecteur à multiplier, et la multiplication avec l'opérateur de projection partiel, ou ligne de blocs de \mathbf{P} , correspondant. Ce calcul peut donc être parallélisé facilement.

La structure orthogonale de la matrice de rotation \mathbf{S}_μ , i.e. : $\mathbf{S}_\mu^T = \mathbf{S}_\mu^{-1}$ peut être utilisée pour exprimer la rétroprojection de \mathbf{y} en $\boldsymbol{\mu}^{(b)}$ de la manière suivante (Goussard *et al.* (2013)) :

$$\boldsymbol{\mu}^{(b)} = \sum_{n=0}^{N-1} \mathbf{S}_\mu^n \mathbf{P}_0^T \mathbf{y}_n \quad (2.44)$$

Cette formule est facile à implémenter mais elle ne permet aucune parallélisation. Pour surmonter cette difficulté, on utilise la structure bloc-circulante de la matrice \mathbf{P} en la partitionnant en colonnes de blocs de la manière suivante :

$$\mathbf{P} = [\mathbf{Q}_0 \mid \mathbf{S}_y^{-1} \mathbf{Q}_0 \mid \dots \mid \mathbf{S}_y^{-N+1} \mathbf{Q}_0] \quad (2.45)$$

où,

$$\mathbf{Q}_0 = [\mathbf{P}_{0,0}^T \mid \mathbf{P}_{0,N-1}^T \mid \mathbf{P}_{0,N-2}^T \mid \dots \mid \mathbf{P}_{0,1}^T]^T \quad (2.46)$$

et \mathbf{S}_y l'opérateur de rotation inverse qui est en fait un opérateur de décalage circulaire de M échantillons. 2.46 peut être écrite directement à partir de la structure détaillée de l'opérateur de projection 2.43. \mathbf{P}^T s'écrit :

$$\mathbf{P}^T = \begin{bmatrix} \mathbf{Q}_0^T \\ (\mathbf{S}_y^{-1} \mathbf{Q}_0)^T \\ \vdots \\ (\mathbf{S}_y^{-N+1} \mathbf{Q}_0)^T \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_0^T \\ \mathbf{Q}_0^T \mathbf{S}_y \\ \vdots \\ \mathbf{Q}_0^T \mathbf{S}_y^{N-1} \end{bmatrix} \quad (2.47)$$

On observe dans 2.47 que \mathbf{P}^T a la même structure que \mathbf{P} dans 2.41 et donc la rétroprojection peut être aussi aisément parallélisée aux prix du stockage de \mathbf{Q}_0 en plus de \mathbf{P}_0 .

2.4.2.3 Présentation des types de représentation de matrices creuses

D'après la section 2.4.2.1, l'équation 2.41 montre que l'ensemble de la matrice de projection peut être retrouvé à partir de \mathbf{P}_0 . De plus cette matrice est creuse et a un taux de remplissage proche de $3 * 10^{-3}$. On ne stocke donc que la matrice \mathbf{P}_0 sous forme de matrice creuse pour gagner de l'espace.

Il existe quatre principaux types de représentation de matrice creuse qui permettent une compression plus ou moins importante et dont les méthodes de multiplication matrice creuse-

vecteur sont différentes :

- triplet storage (TS) : la matrice creuse est stockée sous forme de trois vecteurs : le premier vecteur contient les indices de colonne, le deuxième les indices de ligne et le dernier les valeurs des coefficients non nuls de la matrice (Gendron (2008)) ;
- compressed row storage (CRS) : la matrice creuse est représentée par trois vecteurs. Le premier contient les coefficients non nuls de la matrice obtenus en balayant la matrice ligne par ligne. Pour chaque composante de ce vecteur, on stocke dans un deuxième vecteur l'indice de colonne du coefficient. Le dernier vecteur est de la taille du nombre de lignes plus un. Chacune des composantes de ce vecteur indique la position, dans le premier vecteur, du coefficient débutant la ligne concernée. La dernière composante permet de détecter la fin du vecteur des coefficients et est définie comme le nombre de coefficients non nuls plus un (Haase *et al.* (2007), Gendron (2008)) ;
- Hilbert-curve ordered triplet storage (HTS) : la matrice est stockée avec le même type de vecteurs que pour le TS mais est ordonnée de manière spécifique, fractale suivant la courbe d'Hilbert (Haase *et al.* (2007)) ;
- incremental compressed row storage (ICRS) : la matrice creuse est représentée par au plus trois vecteurs. Les valeurs sont classées par ligne puis par colonne ensuite les trois vecteurs sont construits comme suit (Koster (2002)) :
 - le premier vecteur stocke les valeurs des coefficients non nuls ;
 - le deuxième vecteur, nommé vecteur d'indices de colonne incrémental ΔJ , est construit avec la formule suivante :

$$\Delta J_k = J_k - J_{k-1} + \begin{cases} 0 & \text{si } I_k = I_{k-1} \\ n & \text{sinon} \end{cases}, \text{ avec } J_k \text{ indice de la colonne du coefficient non nul } k \text{ et } I_k \text{ son indice de ligne ;}$$
 - le troisième vecteur contient le nombre de ligne qu'il faut sauter pour passer d'une ligne contenant des valeurs non nulles à la suivante ne contenant pas que des zéros, sa première valeur est l'indice de la première ligne ne contenant des coefficients non nuls. Si toutes les lignes contiennent des valeurs des valeurs non nulles le vecteur sera de la taille du nombre de lignes et ne contiendra que des uns, il n'est pas utile donc il ne sera pas stocké.
- bi-directional incremental compressed row storage (BICRS) : Ce système de stockage est le même que le précédent à la différence que les deux derniers vecteurs peuvent contenir des valeurs négatives et les valeurs non nulles n'ont pas à être classées au début de l'opération de stockage (Yzelman et Bisseling (2011)).

On stocke donc la matrice de projection avec un de ces formats, quand à l'objet reconstruit discrétisé, il est stocké sous forme d'une matrice pleine dont chaque élément représente un

voxel de l'objet.

2.4.3 Préconditionnement

Golkar (2013) a montré que la représentation en coordonnées polaires va détériorer considérablement le conditionnement du système linéaire à résoudre et par conséquent va diminuer la vitesse de convergence des méthodes d'optimisation. Il est donc primordial de préconditionner le système linéaire afin de faciliter la convergence vers une solution satisfaisante.

2.4.3.1 Principe du préconditionnement

Le principe du préconditionnement pour un algorithme en coordonnées polaire est exposé dans 2.3.4.3. L'aspect deux dimensions du préconditionnement avec la représentation polaire est étudié en détails dans Golkar (2013). D'après les travaux de Golkar (2013), la modélisation du bruit avec le modèle gaussien donne une meilleure approximation du hessien de f que la modélisation poissonnienne du bruit et simplifie les calculs.

On rappelle que le préconditionnement consiste à trouver une matrice \mathbf{C} qui diminuera le nombre de conditions du système 2.40. Pour cela, il s'agit d'approximer l'inverse de la matrice normale du système selon $\mathbf{C}^T \mathbf{C} \approx \mathbf{M}^{-1}$ afin de trouver le meilleur compromis vitesse de convergence, simplicité de calcul et donc temps de calcul par itération. En tomographie on sait que la matrice de projection est de grande taille donc la matrice normale est aussi de grande taille ce qui amènera à des préconditionneurs de grande taille. Afin de limiter au maximum l'espace mémoire utilisé et de rendre possible la reconstruction, il est nécessaire de considérer en plus de la vitesse de convergence et de calcul, l'espace mémoire utilisé par le préconditionneur dans le choix de celui-ci.

2.4.3.2 Structure de la matrice normale

Dans le cadre d'une approximation gaussienne de la distribution poissonnienne des photons, avec une estimation MAP utilisant une fonction de régularisation L2 (2.23), la matrice normale du système \mathbf{M} est composée de deux sous matrices :

- la matrice des différences de la fonction de régularisation \mathbf{D} , celle-ci est diagonale dominante et bloc-circulante (BC), sa transposée est aussi BC donc $\mathbf{D}^t \mathbf{D}$ est BC. On renverra le lecteur vers l'annexe C de Golkar (2013) pour la démonstration.
- la matrice de projection \mathbf{P} qui est aussi BC tout comme sa transposée, la multiplication des deux est aussi BC (Golkar (2013)).

La matrice normale est donc à caractère BC. De plus cette matrice est diagonale dominante, définie positive et hermitienne (Golkar (2013)). Cette propriété n'est vraie que pour un

compte de photons suffisamment grand pour permettre la modélisation du bruit des données par un bruit blanc gaussien de variance $\Sigma = \mathbb{I}$.

2.4.3.3 Préconditionnement dans le domaine spatial

Le préconditionnement est exposé dans 2.3.4.3. L'utilisation d'une représentation polaire ne change en rien le principe : il s'agit de trouver le préconditionneur \mathbf{C} tel que $\mathbf{C}^t\mathbf{C}$ soit le plus proche possible de l'inverse de la matrice normale \mathbf{M}^{-1} .

2.4.3.3.a Préconditionneur diagonal

On a vu dans 2.4.3.2 que la matrice normale est diagonale dominante. Le choix de préconditionneur le plus simple est donc de faire une approximation diagonale de \mathbf{M} . Le préconditionneur \mathbf{C} aura la forme :

$$\mathbf{C} = (\text{diag}(\mathbf{P}^T\mathbf{P}) + \lambda\text{diag}(\mathbf{D}^T\mathbf{D}))^{-\frac{1}{2}} \quad (2.48)$$

Ce préconditionneur a comme avantage d'être très simple à calculer, nécessite peu d'espace mémoire pour être stocké et sa multiplication avec un vecteur ou une matrice est très rapide. Par contre, il approxime pauvrement la matrice normale et donc augmentera vraisemblablement peu la vitesse de convergence.

2.4.3.3.b Préconditionneur de Cholesky incomplet

La matrice normale \mathbf{M} est creuse, définie-positive et hermitienne. On peut donc utiliser la décomposition de Cholesky incomplète. Cette méthode consiste à faire la décomposition de Cholesky de la matrice creuse $\hat{\mathbf{M}}$ suivante : $\hat{\mathbf{M}} = \mathbf{L}\mathbf{L}^T = \mathbf{R}^T\mathbf{R}$ avec \mathbf{L} matrice triangulaire inférieure et \mathbf{R} matrice triangulaire supérieure. Afin de simplifier les opération et de diminuer le temps de calcul il est nécessaire de faire une approximation de \mathbf{M} .

La méthode utilisée par Golkar (2013) est d'approximer \mathbf{M} par une matrice creuse blocs diagonale dont on ne conserve que les diagonales de chaque bloc ainsi que les valeurs supérieures à une fraction spécifique de la valeur maximum de la diagonale pour chaque bloc. On calcul ensuite la décomposition de la matrice normale approximée $\hat{\mathbf{M}} = \mathbf{L}\mathbf{L}^T$. Le préconditionneur est ensuite défini : $\mathbf{C} = \mathbf{L}^{-1}$.

L'avantage de cette méthode est qu'il y a moins d'approximations faites pour le calcul du préconditionneur par rapport au préconditionneur diagonal simple. Le taux de remplissage de l'approximation de \mathbf{M} peut être contrôlé par la valeur de la fraction de la valeur maximum retenue pour la conservation des valeurs qui ne sont pas sur la diagonale. La diminution du

nombre d'itération peut être supérieure. Les inconvénients de ce préconditionneur sont une approximation pauvre de \mathbf{M}^{-1} , un espace de stockage du préconditionneur plus important et un temps de multiplication, et donc de calcul, plus long.

2.4.3.4 Préconditionnement dans le domaine de Fourier

2.4.3.4.a Principe

On a montré que la matrice normale est bloc-circulante dans 2.4.3.2. Chen (2005) décrit une méthode qui permet de transformer une matrice bloc-circulante en une matrice bloc-diagonale en passant dans le domaine de Fourier. Soit la matrice normale \mathbf{M} ou Hessien de notre système, blocs-circulante, on peut donc écrire :

$$\mathbf{M} = \mathbf{F}_n^{-1} \mathbf{M}_{bd} \mathbf{F}_n \quad (2.49)$$

avec \mathbf{M}_{bd} définie positive, bloc-diagonale et \mathbf{F}_n la matrice de la transformée de Fourier discrète par bloc. Une difficulté est que chaque bloc diagonal de \mathbf{M}_{bd} correspond à une fréquence, il n'est donc pas possible d'utiliser l'algorithme de FFT classique pour calculer \mathbf{M}_{bd} .

Le préconditionnement dans le domaine de Fourier revient à chercher une approximation de \mathbf{M}_{bd} aisément inversible de manière à pouvoir construire le préconditionneur \mathbf{C} tel que :

$$\mathbf{C}^T \mathbf{C} = \mathbf{F}_n^{-1} \text{approx}\{\mathbf{M}_{bd}^{-1}\} \mathbf{F}_n \quad (2.50)$$

On suppose que $\mathbf{C} \in \mathbb{R}^N$, cette hypothèse a été vérifiée empiriquement. On a $\mathbf{F}_n^{-1} = \frac{1}{N} \mathbf{F}_n^*$ avec N le nombre d'éléments de \mathbf{M} et $\mathbf{Q}^T \mathbf{Q} = \text{approx}\{\mathbf{M}_{bd}^{-1}\}$. On peut écrire :

$$\begin{aligned} \mathbf{C}^* \mathbf{C} &= \mathbf{C}^T \mathbf{C} \\ &= \frac{1}{N} \mathbf{F}_n^* \text{approx}\{\mathbf{M}_{bd}^{-1}\} \mathbf{F}_n \\ &= \frac{1}{N} \mathbf{F}_n^* \mathbf{Q}^T \mathbf{Q} \mathbf{F}_n \\ &= \frac{1}{N} \mathbf{F}_n^* \mathbf{Q}^T N \mathbf{F}_n^{-*} \mathbf{F}_n^{-1} \mathbf{Q} \mathbf{F}_n \\ \mathbf{C}^* \mathbf{C} &= (\mathbf{F}_n^{-1} \mathbf{Q} \mathbf{F}_n)^* (\mathbf{F}_n^{-1} \mathbf{Q} \mathbf{F}_n) \end{aligned}$$

d'où :

$$\mathbf{C} = \mathbf{F}_n^{-1} \mathbf{Q} \mathbf{F}_n \quad (2.51)$$

avec \mathbf{Q} telle que $\mathbf{Q}^T \mathbf{Q} = \text{approx}\{\mathbf{M}_{bd}^{-1}\}$.

2.4.3.4.b Préconditionneur diagonal

L'approximation la plus simple de \mathbf{M}_{bd} est de ne garder que ses éléments diagonaux. Cela permet d'avoir une matrice aisément inversible. Le preconditionneur aura la forme :

$$\mathbf{C} = \mathbf{F}_n^{-1} \text{diag}\{\mathbf{M}_{bd}\}^{-\frac{1}{2}} \mathbf{F}_n \quad (2.52)$$

avec $\text{diag}\{\mathbf{M}_{bd}\}^{-\frac{1}{2}}$ la racine carrée de l'inverse de la matrice $\text{diag}\{\mathbf{M}_{bd}\}$. L'avantage de ce preconditionneur est qu'il nécessite aussi peu d'espace de stockage que le preconditionneur diagonal dans le domaine spatial mais qu'il y a moins d'approximations faites, au lieu d'approximer l'ensemble d'une matrice bloc-circulante, on approxime une matrice bloc-diagonale dont les blocs sont à diagonale dominante (cette propriété sera vérifiée dans 3.3). Il devrait donc avoir de meilleures performances. L'inconvénient c'est qu'il nécessite un passage dans le domaine de Fourier pour les calculs qui alourdit de manière significative la complexité et donc le temps de calcul par itération.

2.4.3.4.c Préconditionneur de Cholesky incomplet

On a \mathbf{M}_{bd} définie positive, blocs-diagonale, dont les blocs diagonaux ne sont pas creux mais ont un taux de remplissage inférieur à 0.5 pour la géométrie 2D avec 672 détecteurs, 1160 projections et 256 pixels tangentiels. Il est donc nécessaire de faire une première approximation des blocs diagonaux et les rendre creux pour pouvoir faire une décomposition de Cholesky incomplète de chaque bloc sur le même modèle que dans le domaine spatial. C'est à dire ne conserver que les diagonales de chaque bloc ainsi que les valeurs supérieures à une fraction spécifique de la valeur maximum de la diagonale pour chaque bloc. On effectue ensuite la décomposition de Cholesky incomplète sur la matrice $M_{bd} \approx L_{bd} L_{bd}^T$ ce qui amène à $M_{bd}^{-1} \approx L_{bd}^{-T} L_{bd}^{-1}$

Le preconditionneur aura donc la forme :

$$\mathbf{C} = \mathbf{F}_n^{-1} \mathbf{L}_{bd}^{-1} \mathbf{F}_n \quad (2.53)$$

Avec \mathbf{L}_{bd} matrice triangulaire inférieure résultant de la décomposition de Cholesky incomplète de \mathbf{M}_{bd} .

L'avantage de ce preconditionneur est qu'il nécessite un nombre limité et contrôlable d'approximations via le nombre de valeurs gardées pour l'approximation de M_{bd} . Ceci va permettre d'augmenter la vitesse de convergence par rapport au preconditionneur diagonal dans le domaine de Fourier. L'inconvénient est qu'il prend plus d'espace mémoire et nécessite plus de temps de calcul pour l'obtenir et l'utiliser.

2.4.3.5 Adaptation à la représentation cylindrique

Tous les préconditionneurs présentés ont été testés pour des problèmes de tomographie polaire en deux dimensions (Golkar (2013)) mais n'ont pas encore été testés en trois dimensions. Il est donc nécessaire d'adapter ces préconditionneurs à la 3D.

2.4.4 Avantages et inconvénients de la représentation cylindrique

Les avantages de la représentation polaire ou cylindrique de l'objet sont :

- l'espace mémoire requis pour le stockage de la matrice de projection est environ $N_p/4$ fois plus faible que celui requis pour les méthodes algébriques utilisant une représentation cartésienne de l'objet. Ce facteur peut aller jusqu'à N_p , lorsque la méthode algébrique n'est pas optimisée en utilisation de l'espace mémoire ;
- la rotation de la source ou de l'objet demande des opérations très simples et peu coûteuses en temps de calcul ;
- la projection et la rétroprojection sont facilement parallélisable.

Les inconvénients de cette représentation sont :

- les voxels d'une grille sphérique ou cylindrique ne sont pas tous de la même taille, et c'est l'inconvénient principal. Il est donc nécessaire d'étudier l'impact de cette inhomogénéité de la représentation sur la qualité de l'image reconstruite et sur la vitesse de convergence des algorithmes d'optimisation utilisés ;
- le stockage de la matrice de projection pour une seule projection nécessite l'application d'un opérateur de rotation pour reconstruire la matrice de projection complète, ce qui augmente la quantité d'opération et donc le temps de calcul ;
- cette approche nécessite une interpolation de l'image reconstruite en coordonnées polaires pour pouvoir l'afficher et l'utiliser en coordonnées cartésiennes. Cette opération diminue la qualité de l'image et augmente très légèrement le temps de reconstruction.

2.5 Mesure de la qualité d'image

2.5.1 Rappel des enjeux

Afin de comparer l'efficacité de deux algorithmes, il est primordial de définir des critères de qualité de l'image. En effet, le temps de convergence seul ne suffit pas car l'algorithme de reconstruction peut converger vers une image de plus ou moins bonne qualité selon le modèle utilisé, l'estimateur, la fonction de coût et les variables y étant liées. Donc, si on veut comparer les performances de reconstruction avec les représentations polaires et cartésiennes, il est nécessaire d'avoir des critères de qualité d'image afin de définir le même objectif à

atteindre pour chaque algorithme et pouvoir comparer leur rapidité et l'espace mémoire utilisé pour cela. La qualité d'une image reconstruite en tomographie peut être évaluée selon deux critères : sa résolution et son contraste. On décrira les différentes méthodes existantes pour évaluer ces critères.

2.5.2 Méthodes d'évaluation théorique de qualité d'image

Avec l'utilisation des ART avec des estimateurs comme le maximum de vraisemblance, il est possible de prédire de manière théorique certains paramètres de la reconstruction comme la variance (Fessler (1996)) ou la résolution (Shi et Fessler (2009)) des images reconstruites. L'inconvénient de ces méthodes d'évaluation théoriques de la qualité de l'image est qu'elles ne prennent en compte que le modèle théorique utilisé pour la reconstruction et par conséquent, les phénomènes physiques non considérés par le modèle ne sont pas pris en compte. Ceci peut provoquer des erreurs entre la qualité prédite et la qualité réelle des images. L'avantage de ces méthodes est qu'elles nécessitent un temps d'exécution bien plus faible que les méthodes empiriques. Pour ce projet, on privilégiera la faible erreur des mesures à la vitesse de calcul élevée de l'évaluation théorique de la qualité d'image, c'est pourquoi on ne détaillera que les méthodes empiriques.

2.5.3 Mesure du contraste

Le contraste est la propriété intrinsèque de l'image qui quantifie la différence de coefficients d'atténuation observable. Il est le critère de qualité d'image le plus utilisé dans la littérature pour des raisons de simplicité de calcul et de mise en oeuvre. L'approche la plus simple de mesure de contraste consiste à utiliser un fantôme contenant des zones homogènes voisines de coefficients d'atténuation plus ou moins différents et ainsi établir la différence minimale de coefficients permettant l'observation de deux zones distinctes (Hamelin *et al.* (2011)).

La mesure du contraste peut aussi être associée à la mesure du bruit résiduel des images reconstruites. Le calcul de ce critère est alors réduit à la mesure du rapport signal à bruit (SNR) (Bergner *et al.* (2012)). Cette mesure peut être perfectionnée en prenant le rapport contraste à bruit (CNR) (Wang *et al.* (2013)). Une autre approche peut être de prendre l'erreur quadratique moyenne (RMSE) d'une zone homogène reconstruite (Champley et Martz (2013)) par rapport au fantôme initial utilisé pour l'acquisition des données. On utilise une zone homogène afin de pouvoir négliger l'impact de tout autre dégradation de l'objet que le bruit sur la valeur de la RMSE.

Cette dernière méthode semble la plus utilisée actuellement en tomographie à rayons X car elle est plus simple d'utilisation que la méthode utilisant un fantôme contenant des

zones homogènes de coefficients d'atténuation différents et donne une idée mieux définie de la différence de coefficients d'atténuation discernable que le SNR ou CNR.

2.5.4 Mesure de la résolution

La résolution d'une image peut être définie de manière aussi simple que le contraste en prenant un fantôme adapté, c'est à dire comportant des petites zones ayant un coefficient d'atténuation très différent de l'arrière plan. Ces zones sont de plus en plus rapprochées. On établit alors la résolution comme la plus petite distance entre deux petites zones de forte atténuation discernables. Cette méthode nécessite l'établissement d'un critère définissant si deux zones sont discernables et un fantôme avec des zones de forte atténuation assez proches pour établir ce seuil, ce qui n'est pas toujours aisé lors de mesures de haute résolution (Staude et Goebbels (2011)).

Pour palier ces inconvénients, il existe une autre approche avec la mesure de la fonction de transfert de modulation (MTF) (Rossman (1969)) qui est calculée par la transformée de Fourier de la fonction d'étalement de point (PSF). Cette méthode nécessite un fantôme avec des impulsions de forte atténuation par rapport à l'arrière plan pour permettre la mesure de la PSF. La MTF est actuellement la méthode la plus utilisée pour le calcul de la résolution en tomographie (Ziegler *et al.* (2006)) et tomographie polaire (Thibaudeau *et al.* (2013), Soriano *et al.* (2012)).

2.6 Objectifs détaillés

Par rapport aux algorithmes de reconstruction analytiques, les algorithmes de reconstruction algébrique classiques permettent une meilleure modélisation du bruit. Ceci peut amener à une meilleure reconstruction de données bruitées. Les inconvénients de ce type d'algorithmes par rapport aux méthodes analytiques sont une vitesse de reconstruction beaucoup plus lente et un espace mémoire requis beaucoup plus important. En effet les algorithmes algébriques sont des méthodes itératives nécessitant un grand nombre d'opérations coûteuses et un espace mémoire très important.

L'algorithme algébrique avec représentation polaire ou cylindrique permet de surmonter le problème de l'espace mémoire requis pour une reconstruction mais garde comme inconvénients : un long temps de calcul pour une reconstruction et une représentation avec des voxels de tailles différentes dont on ne connaît pas précisément l'impact sur la qualité de l'image reconstruite.

Ce projet a comme objectifs :

1. de diminuer le temps de calcul nécessaire à une reconstruction avec l'algorithme algébrique utilisant une représentation cylindrique par rapport aux ART standard ;
2. d'évaluer la qualité des images reconstruites toujours par rapport à la qualité d'image obtenue avec un ART standard et si-possible de l'améliorer ;
3. de limiter l'espace mémoire nécessaire à la réalisation des deux premiers objectifs ;
4. de garder un algorithme le plus transportable possible, c'est à dire un algorithme adaptable au plus grand nombre de scanners, géométries et systèmes.

Afin de diminuer le temps de calcul, on a utilisé la parallélisation et le préconditionnement. Il est nécessaire de tenir compte de la taille du préconditionneur lors de l'étude du meilleur compromis diminution du temps de calcul-taille du préconditionneur afin de garder un espace mémoire requis pour la reconstruction limité. Afin d'évaluer la qualité d'image des reconstruction faites avec les méthodes développées, on a étudié la résolution et le contraste des images reconstruites.

Lors des divers développement on a essayé de garder l'algorithme le plus transportable possible en réduisant au maximum le nombre de conditions à l'application de notre méthode, ceci afin de permettre l'adaptation au modèle polaire de techniques de reconstruction particulières développées en coordonnées cartésiennes telles qu'une modélisation polychromatique de la source pour réduire les artéfacts métalliques.

CHAPITRE 3

MÉTHODOLOGIE

3.1 Estimation

On utilise pour ce projet le modèle monochromatique suivant dérivé de 2.14 :

$$\mathbf{y} = \mathbf{P}\boldsymbol{\mu} + \mathbf{b} \quad (3.1)$$

avec $\mathbf{b} \hookrightarrow \mathbf{N}(0, \Sigma_b)$. On considère le nombre de photons captés assez grand pour pouvoir utiliser la modélisation du bruit par un bruit blanc gaussien, d'où $\Sigma_b = \mathbf{I}$.

L'estimation de l'image est faite avec un estimateur du maximum *a posteriori* (MAP) afin de pouvoir intégrer les informations *a priori* qu'on a de l'objet et qui sont résumés dans la section 2.3.3.3. L'*a priori* est essentiel pour obtenir une reconstruction de bonne qualité. Afin de prendre en compte la présence de grandes zones homogènes dans l'objet, la fonction de pénalisation utilise une pénalisation quadratique L2 sur les différences premières des plus proches voisins. Ce type de fonction de pénalisation couplé au choix d'une variance uniforme du bruit a comme avantage de rendre la matrice normale bloc-circulante, ce qui permet le calcul du hessien et facilite le préconditionnement. On a vu dans 2.3.3.3 que l'estimateur devient le problème des moindres carrés linéarisés 2.23.

Le choix d'une fonction de pénalisation L2 et d'une variance uniforme du bruit n'est pas le choix le plus classique. Généralement, on utilise une pénalisation L2L1 avec une variance non uniforme du bruit. En revanche, notre choix permet de garder un modèle simple afin d'avoir une bonne base de travail et un maximum d'outils pour l'étudier. Il sera par la suite nécessaire d'étendre l'étude aux modèles couramment utilisés.

Pour l'optimisation, on choisit l'algorithme d'optimisation L-BFGS (Nocedal et Wright (1999)) utilisant la méthode quasi-Newton qui a comme avantage d'être très rapide pour résoudre ce type de problème sans calculer directement l'inverse du hessien qui est impossible avec nos moyens pour l'ART cartésien. La méthode de quasi-Newton est décrite dans 2.3.4. On rappelle qu'un algorithme de direction de descente nécessite, à chaque itération, le calcul de la valeur de la fonction de coût et de son gradient, ce qui requiert une projection et une rétroprojection.

Le critère d'arrêt de l'algorithme choisi doit être indépendant du type de représentation de l'objet afin de pouvoir comparer les représentations cartésienne et polaire. On s'est aperçu expérimentalement que la norme L_∞ du gradient de la fonction de coût de l'estimateur MAP

décrit précédemment est de même ordre de grandeur pour un objet représenté en coordonnées cartésiennes ou cylindriques. La recherche de minimum avec BFGS nécessite de faire tendre la norme du gradient de la fonction de coût vers 0. On a donc défini une valeur minimum de la norme L_∞ du gradient comme critère d'arrêt et, en cas de convergence trop lente ou inexistante de l'algorithme, un nombre maximum d'itérations.

3.2 Parallélisation

Pour diminuer le temps de calcul d'une reconstruction avec l'algorithme algébrique utilisant la représentation cylindrique, on a commencé par utiliser la parallélisation, une technique qui a fait ses preuves dans de nombreux domaines.

Avec les ordinateurs d'aujourd'hui possédant plusieurs coeurs, la parallélisation est une manière simple et efficace d'accélérer la vitesse de calcul. Encore faut-il avoir des calculs parallélisables c'est à dire indépendants les uns des autres et que ceux-ci nécessitent un temps de calcul total important.

3.2.1 Possibilités de parallélisation

On cherche à diminuer le temps d'une reconstruction avec l'algorithme algébrique itératif utilisant la représentation de l'objet en coordonnées cylindriques et à comparer ce temps de calcul à celui obtenu avec une représentation de l'objet en coordonnées cartésiennes. Il est donc nécessaire de regarder les possibilités de parallélisation des deux approches. Dans les deux cas, on s'intéressera uniquement à la parallélisation des opérations nécessitant le plus long temps de calcul et dont le gain potentiel en temps sera le plus grand. D'après la Fig. 3.1 résultant d'expériences préliminaires, on a déterminé que pour l'algorithme utilisant la représentation cylindrique de l'objet, lors d'une itération, les deux opérations nécessitant le plus long temps de calcul sont la projection et la rétroprojection. L'utilisation de la représentation en coordonnées cartésiennes de l'objet donne le même résultat.

Ces deux opérations sont des produits matriciels, opération qui est aisément parallélisable. Le problème, dans le contexte étudié, vient du fait que les matrices utilisées sont de trop grande taille pour être stockées entièrement mais contiennent beaucoup de zéros. Donc pour un gain d'espace mémoire, les matrices sont sauvegardées sous forme de matrice creuse dont les différents types de stockage ont été décrits dans 2.4.2.3. L'inconvénient de cette forme de stockage est qu'il rend le produit matriciel beaucoup plus difficilement parallélisable, il n'existe pas de bibliothèque parallélisée performante pour ce type d'opération. Il est donc nécessaire de trouver d'autres moyens simples de parallélisation adaptés à la forme de représentation de l'objet.

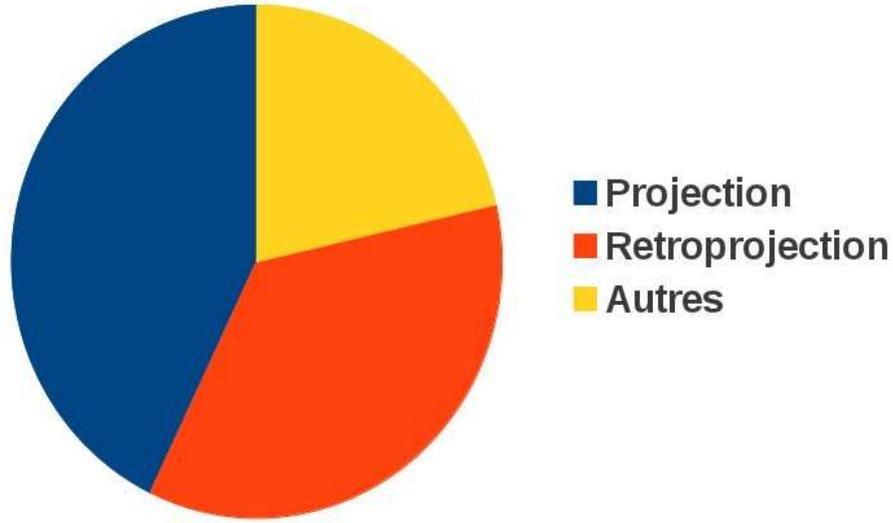


Figure 3.1 Répartition du temps de calcul moyen par itération de l'estimation d'un objet en coordonnées cylindrique

3.2.1.1 Parallélisations possibles avec représentation de l'objet en coordonnées cartésiennes

Avec la représentation cartésienne, on utilise une adaptation trois dimensions de la méthode utilisant les symétries exposées en deux dimensions par Gendron (2008). Gendron (2008) montre qu'avec un nombre de projection N multiple de 4, on peut calculer l'opérateur de projection partiel $\mathbf{P}_{\pi/2}$ pour les projections d'angle 0 à $\pi/2$ puis retrouver l'opérateur de projection \mathbf{P} par simple rotation d'angle $-\pi/2$ de l'objet. On a donc l'opérateur de projection suivant :

$$\mathbf{P} = \left[\mathbf{P}_{\pi/2}^T \mid (\mathbf{P}_{\pi/2} \mathbf{S}_{\pi/2}^{-1})^T \mid \dots \mid (\mathbf{P}_{\pi/2} \mathbf{S}_{\pi/2}^{-3})^T \right]^T \quad (3.2)$$

Avec $\mathbf{S}_{\pi/2}$ l'opérateur de rotation d'angle $\pi/2$. Cette méthode permet de séparer la projection en quatre opérations indépendantes et parallélisables de la forme :

$$\mathbf{y}_{k\pi/2} = \mathbf{P}_{\pi/2} \mathbf{S}_{\pi/2}^{-k} \boldsymbol{\mu} \quad (3.3)$$

où $k \in [0, 3]$ et $\mathbf{y}_{k\pi/2}$ le sinogramme pour les paires angles de projection-détecteurs avec un angle de projection $\in [k\pi/2, (k+1)\pi/2[$.

La rétroprojection est parallélisable de la même manière en considérant l'équation de la rétroprojection :

$$\boldsymbol{\mu} = \mathbf{P}^T \mathbf{y} = (\mathbf{y}^T \mathbf{P})^T \quad (3.4)$$

En réutilisant 3.2 et la propriété $S_{\pi/2}^T = S_{\pi/2}^{-1}$, on peut décomposer l'opération 3.4 en quatre opérations indépendantes et parallélisables :

$$\mu = \left(\sum_{k=0}^3 \mathbf{y}_{k\pi/2}^T \mathbf{P}_{\pi/2} \mathbf{S}_{\pi/2}^{-k} \right)^T = \sum_{k=0}^3 \mathbf{S}_{\pi/2}^k (\mathbf{y}_{k\pi/2}^T \mathbf{P}_{\pi/2})^T \quad (3.5)$$

En utilisant les symétries, on peut paralléliser la projection et la rétroprojection en 4 opérations totalement indépendantes.

L'inconvénient de cette méthode, concernant la rétroprojection, est qu'il est obligatoire d'effectuer une synchronisation des opérations dans un accumulateur. Cette opération risque de diminuer les performances de la parallélisation.

3.2.1.2 Parallélisations possibles avec représentation de l'objet en coordonnées cylindriques

Avec la représentation cylindrique, on utilise la méthode développée par Goussard *et al.* (2013) et exposée dans 2.4.2.1. Cette méthode nécessite d'utiliser la représentation cylindrique de Thibaudeau *et al.* (2013) et d'avoir un nombre de voxels tangentiels (voxels dans la direction θ) de la représentation cylindrique égal au nombre de projections. Cette méthode utilise la formule 2.41 de l'opérateur de projection qui permet de diviser l'opération de projection en N_p opérations indépendantes et parallélisables, N_p étant le nombre d'angles de projection :

$$\mathbf{y}_i = \mathbf{P}_0 \mathbf{S}_\mu^{-i} \boldsymbol{\mu} \quad (3.6)$$

où $i \in [0, N_p - 1]$ est l'indice de l'angle de projection, \mathbf{y}_i représente le sinogramme pour l'angle de projection d'indice i et \mathbf{S}_μ^{-1} est l'opérateur de rotation de l'objet d'un incrément d'angle $2\pi/N_p$.

Avec la représentation polaire, contrairement à la représentation cartésienne, comme \mathbf{P}_0 est de très petite taille, il est possible de stocker explicitement \mathbf{P} et \mathbf{P}^T ou plutôt \mathbf{P}_0 ou \mathbf{P}_0^T et \mathbf{Q}_0 ou \mathbf{Q}_0^T . Ceci permet à la rétroprojection d'être parallélisable de la même manière que la projection. En utilisant la formule 2.44 exposée dans 2.4.2.2, on peut décomposer la rétroprojection en N_p opérations indépendantes et parallélisables :

$$\boldsymbol{\mu}_i = (\mathbf{S}_y^{-i} \mathbf{Q}_0)^T \mathbf{y} \quad (3.7)$$

avec $i \in [0, N_p - 1]$, l'indice de l'angle de projection, $\boldsymbol{\mu}_i$, la portion i de l'objet polaire et \mathbf{S}_y^{-1} , l'opérateur de rotation d'un incrément d'angle $2\pi/N_p$.

On doit effectuer une multiplication matrice-vecteur en parallèle. En passant par la forme transposée des deux expressions, on peut transformer 3.6 en :

$$\mathbf{y}_i^T = (\mathbf{S}_\mu^{-i} \boldsymbol{\mu})^T \mathbf{P}_0^T \quad (3.8)$$

ainsi que 3.7 en :

$$\boldsymbol{\mu}_i^T = (\mathbf{S}_y^{-i} \mathbf{y})^T \mathbf{Q}_0 \quad (3.9)$$

Ceci permet de faire une multiplication vecteur-matrice plutôt que matrice-vecteur, ce qui peut avoir un impact sur la vitesse de calcul de l'opération.

Une parallélisation est donc possible au niveau de la projection et de la rétroprojection. Contrairement à la rétroprojection avec la représentation cartésienne, la parallélisation avec la représentation cylindrique ne va pas nécessiter de synchronisation dans un accumulateur et devrait donc être beaucoup plus rapide.

3.2.2 Mise en oeuvre de la parallélisation

Une fois les possibilités de parallélisation cernées, s'est posée la question de l'implémentation de cette parallélisation. Le 4^e objectif du projet (voir 2.6) étant d'optimiser un ART pour la tomographie à rayons X au sens large et non pas adapté à un seul type de scanner, de géométrie ou de système, on choisit une méthode de parallélisation simple et souple (donc sans utilisation de GPU).

On a implémenté les opérations de projection et rétroprojection en C++ selon les formules mathématiques détaillées dans 3.2.1.1 pour la représentation cartésienne et dans 3.2.1.2 pour la représentation cylindrique. Par rapport aux autres langages de programmation, le C++ bénéficie d'une meilleure gestion de l'allocation et de la gestion de la mémoire via les pointeurs et les déclarations de types de variables. Ceci permet d'optimiser l'espace mémoire utilisé lors de la reconstruction. De plus, il existe pour ce langage la librairie de parallélisation simple et libre d'accès, OpenMP, adaptable à un grand nombre de systèmes.

Pour la parallélisation de la représentation cartésienne, on a utilisé les formules 3.3 pour la projection et 3.5.

La projection est faite en trois étapes parallélisées :

1. rotation de l'objet : $\mathbf{S}_{\pi/2}^{-k} \boldsymbol{\mu}$, pour cette étape il est nécessaire de dupliquer et stocker quatre fois l'objet afin de pouvoir effectuer les rotations simultanément ;
2. multiplication par la matrice de projection partielle $\mathbf{P}_{\pi/2}$;
3. assemblage de la solution dans un vecteur \mathbf{y} .

La rétroprojection est faite en trois étapes aussi, seule la dernière étape n'est pas parallélisée :

1. multiplication du sinogramme partiel avec la matrice de projection partielle : $\mathbf{y}_{k\pi/2}^T \mathbf{P}_{\pi/2}$ avec $k \in [0, 3]$, pour cette étape et la suivante, il est nécessaire de stocker les résultats de ces multiplications dans quatre vecteurs séparés de la taille de $\boldsymbol{\mu}$ qu'on appellera μ_i ;
2. rotation de chaque μ_k selon $\mathbf{S}_{\pi/2}^k \mu_k$;
3. synchronisation et accumulation des μ_i dans le vecteur $\boldsymbol{\mu}$.

La parallélisation de la représentation cartésienne nécessite donc l'augmentation de l'espace mémoire utilisé par quatre fois la taille de l'objet $\boldsymbol{\mu}$.

Pour la parallélisation de la représentation cylindrique, on a utilisé les formules 3.6 pour la projection et 3.7 ou leurs transposées.

La projection est faite en trois étapes parallélisées :

1. rotation de l'objet : $\mathbf{S}_{\mu}^{-i} \boldsymbol{\mu}$, avec \mathbf{S}_{μ}^{-i} un décalage circulaire de K échantillons. Afin de faciliter ce décalage circulaire, on stocke deux fois l'objet dans le même vecteur $\boldsymbol{\mu}_2 = [\boldsymbol{\mu}^T | \boldsymbol{\mu}^T]^T$. Le décalage circulaire devient alors juste un décalage de K des indices de $\boldsymbol{\mu}$ dans $\boldsymbol{\mu}_2$ pour obtenir le résultat de la rotation de $\boldsymbol{\mu}$ ce qui permet de ne nécessiter aucune copie supplémentaire. Cette technique a l'avantage de ne nécessiter qu'une seule copie de $\boldsymbol{\mu}$ pour l'ensemble des décalages d'indices ;
2. multiplication par la matrice de projection partielle \mathbf{P}_0 ;
3. assemblage de la solution dans un vecteur \mathbf{y} .

La rétroprojection est faite de la même manière :

1. rotation du sinogramme : $\mathbf{S}_y^i \mathbf{y}$, avec \mathbf{S}_y^i un décalage circulaire de M échantillons. Afin de faciliter ce décalage circulaire, on stocke deux fois l'objet dans le même vecteur $\mathbf{y}_2 = [\mathbf{y}^T | \mathbf{y}^T]^T$. Le décalage circulaire devient alors juste un décalage de M des indices de \mathbf{y} dans \mathbf{y}_2 pour obtenir le résultat de la rotation de \mathbf{y} ;
2. multiplication par la matrice de rétroprojection partielle \mathbf{Q}_0^T ;
3. assemblage de la solution dans un vecteur $\boldsymbol{\mu}$.

La parallélisation de la représentation cylindrique nécessite donc l'augmentation de l'espace mémoire utilisé par deux fois la taille maximale entre l'objet et le sinogramme.

La parallélisation avec la représentation cartésienne et cylindrique est effectuée par une implémentation en C++ des opérations de projection et de rétroprojection avec utilisation de la librairie OpenMP. Mais elle a comme inconvénient d'augmenter l'espace mémoire requis de quatre fois la taille de l'objet avec la représentation cartésienne et de deux fois la taille maximale entre l'objet et le sinogramme avec la représentation cylindrique.

3.3 Preconditionnement

La deuxième méthode développée pour diminuer le temps de calcul d'une reconstruction avec l'algorithme algébrique utilisant la représentation cylindrique est le preconditionnement. Le preconditionnement permet de réduire le nombre d'itérations mais augmente la durée de chacune d'elles ainsi que l'espace mémoire nécessaire.

Pour une représentation cylindrique de l'objet avec les conditions expérimentales qu'on détaille dans le chapitre suivant, il s'agit donc de trouver le preconditionneur proposant le meilleur compromis entre la diminution du temps de calcul global et l'augmentation de l'espace mémoire requis. Il n'existe pas de preconditionneur simple et efficace pour une représentation cartésienne, on a donc utilisé les résultats obtenus en deux dimensions et en coordonnées polaires par Golkar (2013). Ceux-ci montrent que le passage dans le domaine de Fourier permet d'avoir des preconditionneurs ayant une vitesse de convergence plus rapide en nombre d'itérations mais aussi un temps total de calcul plus court pour atteindre une qualité d'image voulue.

On ne s'est donc intéressé qu'aux preconditionneurs développés par Golkar (2013) en 2D dans le domaine de Fourier : le preconditionneur diagonal et le preconditionneur de Cholesky incomplet. Le passage à la 3D amène des changements au niveau de la structure et de la taille de l'estimateur qu'il a fallu prendre en compte dans l'adaptation de la construction et du calcul des preconditionneurs en 3D

On présentera d'abord la structure de l'estimateur en 3D et les changements qu'on y a apportés afin de rendre possible le preconditionnement dans le domaine de Fourier. On décrira ensuite les preconditionneurs diagonal et de Cholesky incomplet adaptés à la 3D.

3.3.1 Structure de l'estimateur en 3D

On cherche à approximer l'inverse du Hessian de l'estimateur $\mathbf{H} = \mathbf{P}^T\mathbf{P} + \lambda\mathbf{D}^T\mathbf{D}$ bloc-diagonalisé dans le domaine de Fourier avec \mathbf{D} la matrice de pénalisation des différences premières des plus proches voisins. Afin de permettre une bloc diagonalisation d'une matrice dans le domaine de Fourier, elle doit être BC. On rappelle la structure 2D de la matrice de projection, d'après l'équation 2.43, on a :

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{0,0} & \mathbf{P}_{0,1} & \dots & \mathbf{P}_{0,N-1} \\ \mathbf{P}_{0,N-1} & \mathbf{P}_{0,0} & \dots & \mathbf{P}_{0,N-2} \\ \vdots & & \ddots & \\ \mathbf{P}_{0,1} & \mathbf{P}_{0,2} & \dots & \mathbf{P}_{0,0} \end{bmatrix} = \begin{bmatrix} \mathbf{P}_0 \\ \mathbf{P}_0\mathbf{S}_\mu^{-1} \\ \vdots \\ \mathbf{P}_0\mathbf{S}_\mu^{-N+1} \end{bmatrix}$$

où chaque $\mathbf{P}_{0,n}$; $0 \leq n \leq N - 1$ est un bloc (M, K) avec M le nombre total de détecteurs et K le nombre de voxels tangentiels (voxels dans la direction θ).

Telle que décrite ci-dessus, \mathbf{P} est BC. Traditionnellement, on stocke chaque tranche de l'objet par un bloc consécutif en mémoire. Pour conserver l'invariance par rotation et donc la propriété BC de la matrice, on stocke consécutivement en mémoire l'objet complet suivant une direction θ pour chaque angle de projection, c'est à dire qu'on représente l'objet (N_r, N_s, N_t) avec N_r , nombre de voxels radiaux, N_s , nombre de voxels axiaux et N_t le nombre de voxels tangentiels plutôt que (N_r, N_t, N_s) qui est l'extension logique de la représentation polaire en 3D. Cette représentation de l'objet permet de garder la même forme de \mathbf{P} que dans l'équation 2.43 mais avec des $\mathbf{P}_{0,n}$; $0 \leq n \leq N_t - 1$ blocs (M, K) avec M le nombre de détecteurs multiplié par le nombre de barrettes et $K = N_r \times N_s$. Cette réorganisation de l'estimateur permet de conserver la propriété BC de \mathbf{P} .

Avec ce changement de représentation de l'objet, on a $\mathbf{D} = [\alpha_r \mathbf{D}_r | \alpha_t \mathbf{D}_t | \alpha_z \mathbf{D}_z]$ avec \mathbf{D}_r la matrice de pénalisation des différences premières des plus proches voisins dans la direction ρ , \mathbf{D}_t la matrice de pénalisation dans la direction θ , \mathbf{D}_z la matrice de pénalisation dans la direction z et $\alpha_r, \alpha_t, \alpha_z$, les pondérations des différences dans chaque direction, on choisit $\alpha_r = \alpha_z = \mathbf{I}$ car les centres des pixels comparés sont toujours à la même distance et α_t matrice diagonale de pondération en fonction de la distance entre les centres des pixels comparés. Golkar (2013) a montré que cette pondération améliore le conditionnement de la matrice en 2D, on émet l'hypothèse que le même effet est retrouvé en 3D. Les sous matrices $\mathbf{D}_r, \mathbf{D}_t$ et \mathbf{D}_z ont les formes suivantes :

Matrice des différences premières dans la direction radiale :

$$\mathbf{D}_r = \left[\begin{array}{ccc} \underbrace{\begin{bmatrix} 1 & -1 & & \\ & 1 & -1 & \\ & & \ddots & \\ & & & \dots & 1 & -1 \\ & & & & & 0 \end{bmatrix}}_{K \times K} & & 0 \\ & \ddots & \\ & & \underbrace{\begin{bmatrix} 1 & -1 & & \dots \\ & 1 & -1 & \\ & & \ddots & \\ & & & \dots & 1 & -1 \\ & & & & & 0 \end{bmatrix}}_{K \times K} \end{array} \right]_{KN_p \times KN_p}$$

Matrice des différences premières dans la direction tangentielle :

$$\mathbf{D}_t = \left[\begin{array}{ccc} \underbrace{\begin{bmatrix} 1 & & \\ & 1 & \\ & & \ddots \\ & & & 1 \end{bmatrix}}_{K \times K} & \underbrace{\begin{bmatrix} -1 & & \\ & -1 & \\ & & \ddots \\ & & & -1 \end{bmatrix}}_{K \times K} & 0 \\ & \ddots & \\ & & \underbrace{\begin{bmatrix} 1 & & \\ & 1 & \\ & & \ddots \\ & & & 1 \end{bmatrix}}_{K \times K} \end{array} \right]_{KN_p \times KN_p}$$

On a $\mathbf{D}_r, \mathbf{D}_t$ et \mathbf{D}_z BC.

D'après les expressions de $\mathbf{D}_r, \mathbf{D}_t$ et \mathbf{D}_z , on peut en déduire que $\mathbf{D}^T \mathbf{D} = \alpha_r \mathbf{D}_r^T \mathbf{D}_r + \mathbf{D}_t^T \alpha_t^T \alpha_t \mathbf{D}_t + \alpha_z \mathbf{D}_z^T \mathbf{D}_z$ est BC. On rappelle que $\mathbf{P}^T \mathbf{P}$ est aussi BC. Ceci implique que le Hessien de l'estimateur est aussi BC, ce qui est la propriété fondamentale pour sa bloc-diagonalisation et son préconditionnement dans le domaine de Fourier.

Lors du passage à la 3D, d'autres difficultés apparaissent : l'augmentation de la taille de l'estimateur va provoquer l'augmentation de la taille des préconditionneurs. Ceci va rendre impossible, avec l'espace mémoire à notre disposition, le calcul de la transformée de Fourier de l'ensemble de la matrice normale ou le calcul de l'inverse de la matrice normale bloc-diagonalisée. Il est donc primordial de développer des méthodes pour surmonter ces problèmes et de modifier le compromis vitesse de l'estimation/taille du préconditionneur à faire lors du choix du préconditionneur en 3D.

3.3.2 Préconditionneur diagonal dans le domaine de Fourier

La figure Fig. 3.2 présente le hessien de la matrice normale bloc diagonalisé dans le domaine de Fourier pour un problème 3D de très petite taille avec une représentation cylindrique. On peut penser que chaque bloc de la diagonale du hessien dans le domaine de Fourier, est à diagonale dominante.

L'approximation la plus facile et rapide est la réduction du hessien à sa seule diagonale.

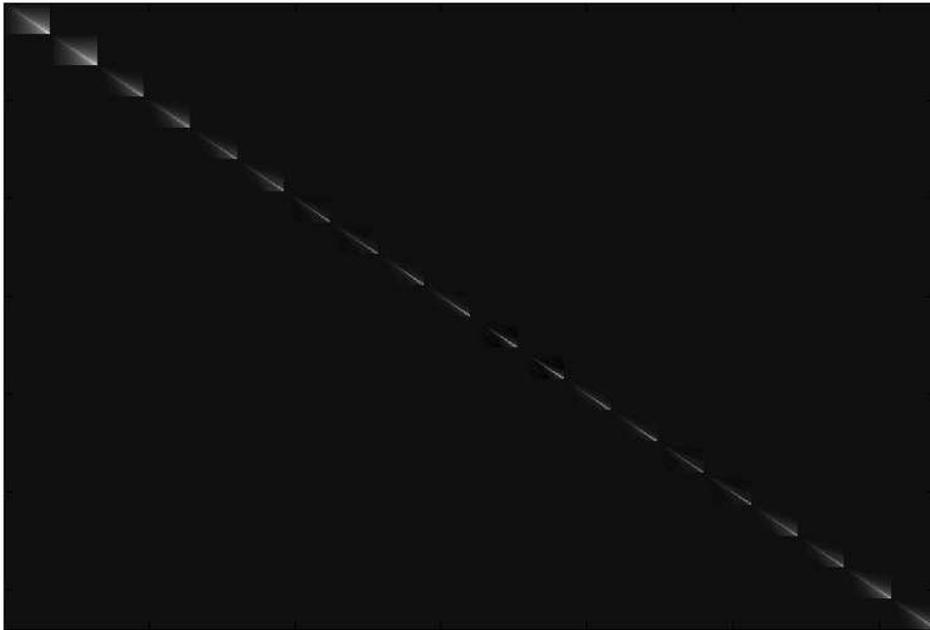


Figure 3.2 Matrice normale du problème d'optimisation bloc diagonalisée pour un problème en 3D de très petite taille

On emploie le même type d'algorithme qu'en 2D dont l'équation est 2.52. On cherche à avoir $\mathbf{C} = \mathbf{F}_n^{-1} \text{diag}\{\mathbf{M}_{bd}\}^{-\frac{1}{2}} \mathbf{F}_n$ avec $\text{diag}\{\mathbf{M}_{bd}\}^{-\frac{1}{2}}$ la racine carrée de l'inverse de la matrice $\text{diag}\{\mathbf{M}_{bd}\}$. On peut noter que l'inverse de $\text{diag}\{\mathbf{M}_{bd}\}$ est simplement l'inverse de chacun de ses éléments diagonaux.

Pour le passage dans le domaine de Fourier, on ne peut pas simplement utiliser l'algorithme de FFT car chaque bloc du hessien bloc-diagonalisé correspond à une fréquence. Avec la taille importante des objets à manipuler en 3D, les calculs de toutes les fréquences et de l'inverse du hessien sont impossibles avec l'espace mémoire à notre disposition.

Pour surmonter les problèmes de saturation de la mémoire lors du calcul de la transformée de Fourier et de l'inverse de la matrice normale, on effectue le calcul du préconditionneur en deux étapes :

1. calcul de la transformée de Fourier des éléments diagonaux du hessien uniquement (le calcul de la transformée de Fourier de la totalité des éléments étant trop long et nécessitant trop d'espace mémoire)
2. inversion de l'approximation diagonale du hessien dans le domaine fréquentiel.

Ce préconditionneur a de bonnes performances en 2D du fait du faible nombre d'approximations, on espère un comportement similaire en 3D.

3.3.3 Préconditionneur de Cholesky incomplet dans le domaine de Fourier

L'objectif de ce préconditionneur est de réduire le nombre d'approximations à faire par rapport au préconditionneur diagonal de la section précédente, ce qui revient à garder une plus grande partie du hessien bloc-diagonalisée pour construire le préconditionneur. Ceci a pour but de mieux approcher l'inverse du hessien ce qui va amener à une convergence nécessitant moins d'itérations. A contrario, chaque itération va nécessiter un temps de calcul plus long.

Le préconditionneur est développé en utilisant une méthode similaire à celle présentée en 2.4.3.4.c et décrite par Golkar (2013), mais se différenciant sur deux points : une application à la 3D et une approche différente de l'approximation du hessien.

Afin de surmonter les problèmes de saturation de mémoire exposés dans la section précédente, le calcul du préconditionneur est fait en cinq étapes :

1. Calcul des N_p blocs dans le domaine de Fourier.

Pour cette première étape, pour effectuer une transformation de Fourier de l'ensemble des éléments de la matrice normale, il est impossible d'utiliser l'algorithme de FFT, on compense en utilisant une transformée de Fourier discrète particulière mise en oeuvre en

C. on sépare ces N_p blocs en sous-ensembles et on effectue une transformée de Fourier de chacun de ces sous-ensembles pris séparément afin de pouvoir construire le hessien bloc diagonalisé au complet avec l'espace mémoire dont on dispose.

2. Pour chacun des N_p blocs du Hessien bloc diagonalisé séparément : recherche des éléments dont la valeur absolue est supérieure à une fraction de la valeur absolue de la valeur maximale du bloc.

Il faut se rappeler que ces blocs ne sont pas creux et de grande taille. Il est donc nécessaire d'approximer ces blocs par des blocs creux dont on contrôle la taille via le taux de remplissage afin de réduire l'espace mémoire requis pour le stockage du préconditionneur 3D.

3. Calcul du taux de remplissage total, s'il est en dessous du taux voulu, on diminue la fraction utilisée et on revient à l'étape 2.
4. Calcul pour chaque bloc diagonal de sa décomposition de Cholesky incomplète $\mathbf{B}_n \approx \mathbf{R}_n^T \mathbf{R}_n$ avec \mathbf{R}_n la matrice triangulaire supérieure issue de la décomposition de Cholesky incomplète du bloc \mathbf{B}_n . L'ensemble des \mathbf{R}_n est regroupé dans la matrice bloc-diagonale \mathbf{R} , on obtient $\mathbf{R}^T \mathbf{R} \approx \mathbf{M}_{bd}$.
5. On définit ensuite le préconditionneur $\mathbf{C} = \mathbf{F}_n^{-1} \mathbf{R}^{-T} \mathbf{F}_n$ sans calculer explicitement \mathbf{R}^{-T} ou \mathbf{R}^{-1} car ce calcul n'est pas toujours faisable. Lors de la multiplication avec le préconditionneur, on résout alors un système triangulaire creux, ce qui est plus rapide et moins coûteux que le calcul et le stockage de la matrice pleine \mathbf{R}^{-1} .

Ce préconditionneur a de meilleures performances en 2D que le préconditionneur diagonal du fait du moins grand nombre d'approximations, on espère un comportement similaire en 3D.

3.4 Mesure de qualité d'image

3.4.1 Rappel des enjeux

On utilise une nouvelle représentation, des préconditionneurs et la parallélisation pour augmenter les performances d'un algorithme algébrique au niveau temps de calcul et espace mémoire utilisé. Afin de valider les méthodes développées pour diminuer le temps de calcul et l'espace mémoire requis, il est nécessaire de trouver une méthode de comparaison de deux algorithmes de reconstruction itératifs utilisant des représentations de l'objet différentes.

Une difficulté se pose alors : le choix du critère d'arrêt des deux algorithmes. En effet si on veut comparer le temps de reconstruction, il est nécessaire de définir le même objectif à atteindre pour les deux algorithmes. L'objectif doit donc être indépendant de la méthode.

La méthode la plus simple est de définir un temps de calcul soit en nombre d'itérations, soit en volume de calcul, puis de comparer la qualité des reconstructions correspondantes. Le problème de cette méthode est de savoir comment comparer ces reconstructions. Une autre méthode, plus usuelle, est de définir une qualité de reconstruction à atteindre et de mesurer le temps pris par chaque algorithme pour l'atteindre. Ceci présuppose d'être capable d'établir un lien entre le critère d'arrêt de chaque algorithmes et la qualité d'image. On retrouve dans ces deux méthodes le même problème de comparaison de qualité d'image de deux reconstructions d'algorithmes avec des représentations différentes.

La comparaison de deux reconstructions d'images se fait avec la mesure de la qualité de celles-ci.

Le problème du choix de la méthode de mesure de qualité d'image se pose alors. La première difficulté est la différence de représentation des images. Pour surmonter cette difficulté, on ne mesure la qualité d'image que sur les images en coordonnées cartésiennes car ce sont les seules images de scanner à rayons X qu'on utilise réellement en radiologie. Les images en coordonnées polaires sont donc traduites en coordonnées cartésienne par interpolation linéaire.

Les objets imagés en tomographie à rayons X sont composés de grandes zones homogènes séparées par des frontières nettes, une reconstruction de bonne qualité correspond à une image ayant un bruit très faible, permettant de distinguer des faibles différences de coefficients d'atténuation et d'observer la séparation nette entre deux structures très proches.

Le contraste est la propriété intrinsèque de l'image qui quantifie la différence de coefficients d'atténuation observable. La résolution est la distance minimale observable entre deux points.

On peut alors définir la qualité d'une image par son niveau de bruit, son contraste et son intensité. Le niveau de bruit et le contraste d'une reconstruction peuvent être évalués grâce à la RMSE et la résolution peut être mesurée à l'aide de la fonction de transfert de modulation (MTF).

Les deux sections suivantes exposent les méthodes de mesures pour les deux critères retenus : la RMSE et la MTF. Ces mesures s'effectuent sur les images interpolées en cartésien pour la méthode en coordonnées cylindrique car ce sont les seules images utilisées en pratique.

Le modèle en coordonnées cylindrique a une grille non-homogène avec des plus petits voxels au centre de l'image qu'en périphérie. Cette représentation peut avoir un impact significatif sur la qualité locale de l'image. Il est donc nécessaire d'établir une méthode de mesure de la qualité d'image locale. Pour cette étude, on présuppose qu'il est possible d'obtenir des reconstructions de qualité globale équivalentes et que les méthodes de mesure de qualité d'image sont applicable sur une petite portion de l'image. On choisit alors comme critère

d'arrêter un critère de qualité d'image globale de la reconstruction afin de pouvoir comparer l'évolution des valeurs de mesures de qualité d'image locale.

3.4.2 Calcul de la RMSE

La première mesure de qualité d'image est la racine de l'erreur quadratique moyenne (RMSE) de l'image par rapport au fantôme utilisé pour obtenir le sinogramme.

Le calcul de la RMSE sur la globalité de l'image est très simple. Il se fait en utilisant la formule ci-dessous :

$$RMSE = \sqrt{\frac{1}{N_{tot}} \sum_{i=1}^{N_{tot}} (\tilde{\mu}_i - \mu_i)^2} \quad (3.10)$$

où N_{tot} le nombre total de voxels de l'image, $\tilde{\mu}$ l'image reconstruite et μ le fantôme initial.

Le calcul local de la RMSE se fait en utilisant la même formule 3.10 avec des petites zones homogènes prédéfinies de l'image globale disposées à des distances différentes de son centre. On utilise pour cela un fantôme constitué d'une seule zone homogène afin de restreindre la RMSE à la mesure du niveau de bruit.

Cette technique a les inconvénients suivants :

1. elle n'est applicable que si on connaît l'objet à reconstruire, ce qui est une limitation pratique importante ;
2. la RMSE ne rend compte que très partiellement de la qualité de l'image à reconstruire.

3.4.3 Calcul de la MTF

Lors de la reconstruction d'un objet par tomographie, les discontinuités sont lissées. Pour mesurer de manière simple la résolution, on étudie donc la réponse impulsionnelle représentée par la PSF que l'on représente généralement dans le domaine fréquentiel par la MTF qui est en fait la transformée de Fourier de la PSF. Meilleure est la résolution, plus la PSF est proche d'un dirac et donc plus la MTF est plate.

La MTF est calculée localement sur des petites parties prédéfinies de l'image. Ces parties sont centrées sur des lésions circulaires quasi-ponctuelles de très forte atténuation au milieu d'une zone de faible atténuation. Elles sont situées à des distances différentes du centre de l'image. On a effectué le calcul de la MTF de la manière suivante :

1. estimation MAP de la fonction d'étalement de point (PSF) ou réponse impulsionnelle 2D avec le modèle direct suivant :

$$\tilde{\mu}_p = (\mu_p * PSF) + b \quad (3.11)$$

où μ_p est la partie étudiée du fantôme initial, $\tilde{\mu}_p$ la partie étudiée de l'image reconstruite et b bruit blanc gaussien ;

2. calcul de la MTF à partir de la PSF estimée avec la formule suivante :

$$MTF = \frac{1}{2} |FFT(LSF_x) + FFT(LSF_y)| \quad (3.12)$$

où LSF_x est la fonction d'étalement de ligne suivant l'axe des abscisses calculée suivant la formule $LSF_x(j) = \sum_i PSF(j, i)$ et LSF_y , la fonction d'étalement de ligne suivant l'axe des ordonnées calculée $LSF_y(j) = \sum_i PSF(i, j)$. La fonction d'étalement de ligne correspond à la projection de la réponse impulsionnelle sur l'axe x ou y .

3.4.4 Méthode d'amélioration de la qualité de l'image : modélisation polychromatique

On a vu dans le chapitre 2 que la source de rayons X est naturellement polychromatique et que l'approximation monochromatique entraîne des artéfacts en cas de présence de matériaux de coefficients d'atténuation élevés tels des métaux dans l'objet à imager. L'objectif de l'utilisation d'un modèle polychromatique est de diminuer les artéfacts présents lors de la reconstruction d'objets métalliques avec des données réelles afin d'améliorer la qualité des images reconstruites.

Le passage au polychromatique augmente la complexité du calcul mais diminue le nombre d'approximations faites donc va augmenter le temps d'estimation pour un gain éventuel en qualité d'image reconstruite et en réduction d'artéfacts métalliques.

Les objectifs de cette partie sont de montrer :

1. qu'il est possible d'adapter, sans difficulté, à une représentation cylindrique, les méthodes polychromatiques développées antérieurement pour une représentation cartésienne ;
2. que le comportement des méthodes de reconstruction est analogue à celui du modèle monochromatique ;
3. que l'amélioration de la qualité d'image par rapport à l'utilisation d'un modèle monochromatique reste significative.

On adapte le modèle polychromatique décrit dans 2.3.1.2 à une représentation cylindrique de l'objet. On fait l'estimation de μ_0 comme l'a faite Hamelin *et al.* (2008) dans le cadre d'une représentation cartésienne de l'objet, avec le maximum *a posteriori* (MAP) d'un problème

des moindres carrés régularisé qui s'écrit tel que :

$$\hat{\boldsymbol{\mu}} = \underset{\mu_0 \geq 0}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{y} + \ln \mathbf{p}(\mu_0, b)\|^2 + \lambda R(\mu_0) \quad (3.13)$$

où $\mathbf{p}(\mu_0, b) = \sum_{k=0}^K b_k \cdot \exp(-\mathbf{P}\mu_k)$ et $\mu_k = \Phi_k \cdot \phi(\mu_0) + \Theta_k \cdot \theta(\mu_0)$.

L'adaptation de ces formule à la représentation cylindrique nécessite de prendre la matrice de projection \mathbf{P} calculée pour l'objet en coordonnées cylindriques μ_0 suivant la méthode présentée précédemment. Par rapport au monochromatique, l'utilisation du polychromatique nécessite :

1. une fonction de coût spécifique à la représentation cylindrique,
2. une modélisation et un échantillonnage de la source de rayons X pour :
 - calculer les b_k qui permettent de quantifier la contribution de chaque énergie à la mesure d'atténuation du détecteur,
 - calculer les constantes photoélectrique Φ_k et de Compton Θ_k pour chaque énergie échantillonnée,
3. la définition des fonctions non linéaires : $\phi = f(\mu_0)$ et $\theta = f(\mu_0)$.

On calcule μ_0 à l'énergie de référence $E_0 = 70keV$. Les fonctions $\phi = f(\mu_0)$ et $\theta = f(\mu_0)$ sont modélisées par, respectivement, une hyperbole verticale et une hyperbole horizontale paramétrées pour passer par l'origine et correspondre au mieux aux tables de mesures de De Man *et al.* (2001).

La principale difficulté de l'adaptation à la représentation cylindrique de la méthode utilisant une représentation cartésienne est la taille inhomogène des voxels. Cette propriété de la représentation cylindrique détériore le conditionnement du problème et nécessite l'adaptation de la fonction de pénalisation en conséquence.

Pour ce faire, on choisit la même fonction de pénalisation que le modèle en coordonnées cartésiennes, c'est à dire une fonction de pénalisation L2L1 sur les différences premières avec les plus proches voisins. Afin de considérer les différences de taille de voxels, on utilise une pondération similaire à celle utilisée avec la fonction de régularisation L2 du modèle monochromatique en coordonnées cylindriques.

Afin de contourner les problèmes de convergence dûs à la détérioration du conditionnement, on choisit comme image de départ pour l'algorithme de reconstruction le résultat de la FBP. On s'est aperçu empiriquement que cette technique permettait de palier les problèmes de convergence. Il sera nécessaire par la suite de développer des préconditionneurs adaptés au modèle polychromatique afin de proposer une solution rigoureuse à ce problème de conditionnement et permettre de diminuer le temps de reconstruction d'images de bonne qualité.

CHAPITRE 4

RÉSULTATS

Le but de ce chapitre est de tester si les méthodes développées dans le chapitre précédent permettent de remplir les objectifs initiaux. Ces objectifs sont :

1. de diminuer le temps de calcul nécessaire à une reconstruction avec l'algorithme algébrique utilisant une représentation cylindrique par rapport aux ART standard ;
2. d'évaluer la qualité des images reconstruites toujours par rapport à la qualité d'image obtenue avec un ART standard et si-possible de l'améliorer ;
3. de limiter l'espace mémoire nécessaire à la réalisation des deux premiers objectifs ;
4. de garder un algorithme le plus transportable possible, c'est à dire un algorithme adaptable au plus grand nombre de scanners, géométries et systèmes.

Afin d'effectuer les expériences de validation de nos objectifs, il est nécessaire de définir un cadre de travail, c'est à dire un modèle de formation des données et un protocole d'acquisition des données qui servent aux reconstructions.

L'objectif général de ce projet est de développer une méthode de reconstruction algébrique compatible avec les scanner cliniques, ce qui implique qu'on définisse un cadre expérimental le plus proche possible des conditions réelles d'application de l'algorithme. On utilise essentiellement des données simulées lors de la réalisation de nos expériences. En plus d'une référence à l'objet réel, ceci a pour avantage de permettre d'avoir, très rapidement et à moindres frais, des sinogrammes de fantômes correspondant à chaque protocole expérimental défini.

4.1 Modèle

Cette partie a pour but de décrire les différents modèles utilisés pour ce projet.

4.1.1 Modélisation de la source de rayons X

la source est modélisée de deux façons suivant le but recherché :

Pour tester la diminution du temps de calcul et pour comparer la qualité d'image suivant la représentation de l'objet utilisée, on réduit à une seule énergie le spectre d'émission afin de simplifier le modèle et d'augmenter la simplicité et la vitesse du calcul de la reconstruc-

tion, la source est alors dite monochromatique. C'est le modèle majoritairement utilisé en tomographie.

Une alternative utilisée seulement pour des cas particuliers est la modélisation polychromatique de la source. Afin d'améliorer la qualité des images reconstruites avec le modèle utilisant une représentation cylindrique de l'objet, on effectue un échantillonnage du spectre d'émission des rayons X afin d'améliorer le réalisme du modèle et donc de mieux prendre en compte les matériaux de forte atténuation ce qui permet de diminuer les artéfacts lors de la reconstruction.

Pour résumer :

- source monochromatique : la source est réduite à un seul niveau d'énergie d'intensité égale à l'ensemble des intensités pour tous les niveaux d'énergie du spectre d'émission. C'est le modèle couramment utilisé, il permet l'accélération du temps de calcul ;
- source polychromatique : la modélisation de la source se fait par un échantillonnage du spectre d'émission tous les KeV entre 20 et 120 KeV. Ce modèle permet la réduction des artéfacts métalliques.

4.1.2 Modélisation des rayons et de la géométrie du scanner

Les rayons sont modélisés avec la méthode *ray-driven* par des segments infiniment fins allant de la source de rayons X au panneau de détecteurs afin de discrétiser directement l'intégrale ligne d'une projection. Des méthodes présentées dans 2.1.4, ce modèle de rayon a comme avantage d'être le plus simple à calculer avec la représentation cylindrique de l'objet. Afin de réduire l'impact sur la qualité de l'image reconstruite de l'approximation des détecteurs ponctuels faite avec cette méthode, on utilise la technique de rayons multiples. On choisit de modéliser dix rayons fins par détecteurs espacés régulièrement sur la totalité de la surface du détecteur.

On choisit comme géométrie du scanner, la géométrie d'un scanner clinique couramment utilisé et pour lequel nous avons des données réelles à étudier, le scanner SOMATOM de Siemens dont les caractéristiques sont détaillées ci dessous :

- panneau de détecteurs en arc de cercle de 16 barettes,
- 672 détecteurs de 1 mm de largeur et 0.75 mm de hauteur par barette,
- distance source de rayons X-centre de l'image : 570 mm,
- distance centre de l'image-panneau de détecteurs : 470 mm.

On choisit deux protocoles d'acquisition différents :

- un protocole basse-dose, c'est à dire dont la dose de rayons X à chaque projection est faible générant un sinogramme bruité de SNR 25dB ;
- un protocole haute-dose, c'est à dire dont la dose de rayons X à chaque projection

est élevée afin de permettre l'obtention d'images de haute résolution générant un sinogramme faiblement bruité de SNR 35dB.

On choisit de ne pas fixer le nombre de projection pour chaque protocole afin de garder un certain contrôle sur la taille totale du sinogramme.

4.1.3 Modélisation de l'objet

On choisit de représenter l'objet par une grille cylindrique à l'image du travail de Thiabaudeau *et al.* (2013) plutôt qu'une représentation alternative telle celles développées par Rodriguez-Alvarez *et al.* (2011). Afin de pouvoir appliquer les méthodes développées, on définit un nombre de voxels circulaires identique au nombre de mesures de projection effectuées.

Pour ce qui est des résultats de l'approche cartésienne, on utilise la représentation en coordonnées cartésiennes de l'objet la plus courante. L'objet est modélisé par une grille homogène de voxels cubiques. L'avantage de cette représentation est la simplicité même de celle-ci facilitant le calcul de la matrice de projection.

4.2 Génération des données simulées

Le crime inverse se produit lorsque les données pour un problème inverse sont générées avec le même (ou presque le même) modèle que celui utilisé lors du processus d'inversion.

Afin d'éviter le crime inverse, on utilise pour la projection un simulateur de projections indépendant de notre modèle : CT Project (Segars *et al.* (2008)). Ce simulateur nécessite la création d'un fantôme via le logiciel XCAT pour pouvoir produire le sinogramme réaliste correspondant.

4.2.1 Création du fantôme et erreur de représentation

Des expériences préliminaires ont montré que la nature de l'objet a une influence significative sur la vitesse de convergence. En particulier, la présence d'inhomogénéités et leur amplitude a un impact conséquent sur la vitesse de convergence. Pour faire nos expériences, on a pris un cadre générique qui représente la majorité des applications

La génération des fantômes initiaux est donc faite avec le programme XCAT. Afin de travailler sur la reconstruction d'image correspondant à l'utilisation clinique de nos méthodes, on choisit un fantôme réaliste d'une portion du tronc d'un individu mâle sans les bras (Fig. 4.1) pour toutes les expériences sauf pour la mesure de la qualité d'image par rapport au centre. Le fantôme généré par XCAT est continu et est découpé en 4 tranches d'épaisseur 1.5 mm représentant des images de $384mm \times 384mm$

Pour les expériences de parallélisation et de préconditionnement, chaque tranche est discrétisée en une image de 256×256 pixels de taille 1.5 mm afin d'imager le tronc de l'individu dans toute sa largeur tout en gardant un objet à reconstruire et une matrice de projection de taille suffisamment faible pour être utilisés avec nos moyens et un ART en coordonnées cartésienne simple.

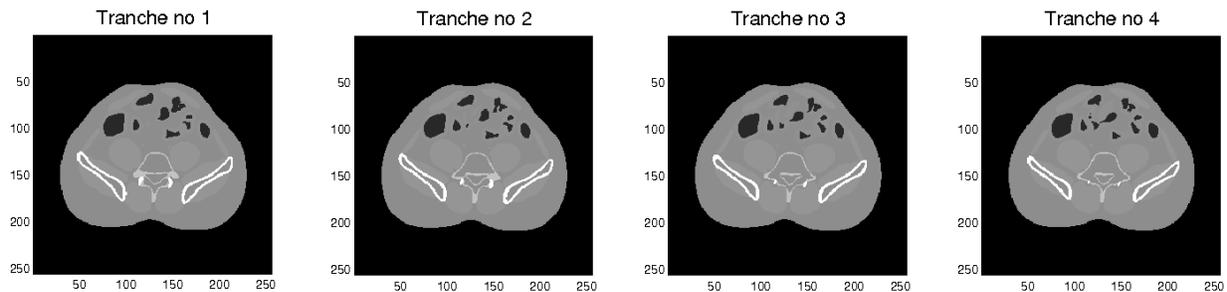


Figure 4.1 Fantôme XCAT utilisé pour le préconditionnement et la parallélisation

Pour la mesure de la qualité de l'image locale, on utilise un fantôme continu d'une tranche de 1.5 mm d'épaisseur représentant une image de $384mm \times 384mm$. Le fantôme est composé d'une région circulaire de coefficient d'atténuation égal aux muscles du corps humain contenant de multiples billes circulaires d'aluminium de 1.5 mm de diamètre pour la mesure de MTF afin de représenter des impulsions (Fig. 4.2) ou sans ces billes pour la mesure de la RMSE pour ne pas biaiser cette mesure avec des phénomènes d'étalement. Les billes d'aluminium sont placées à 5 distances différentes au centre de l'image, espacées d'au moins 37.5 mm dans le but de faciliter le calcul de la PSF et de pouvoir comparer la MTF en fonction de la distance au centre pour plusieurs mesures.

Le fantôme est ensuite discrétisé en une image de 512×512 pixels de $0.75 \text{ mm} \times 0.75 \text{ mm}$ en coordonnées cartésiennes pour pouvoir mesurer la qualité de l'image sur une image de résolution numérique quatre fois plus grande que les images utilisées pour la parallélisation et le préconditionnement afin d'améliorer la précision des mesures de qualité d'image. Lors de l'utilisation de l'algorithme en coordonnées cylindrique, on utilise une image de 256×1160 pixels allant de 0.0015 mm^2 au centre à 0.78 mm^2 en périphérie. On a donc des pixels polaires 375 fois plus petits que les pixels cartésiens au centre de l'image et 1.4 fois plus grands en périphérie.

Cette différence significative de taille des pixels avec la représentation cylindrique est à l'origine de trois difficultés : une erreur de représentation variable, un choix à faire au niveau de la régularisation et une détérioration du conditionnement du problème qui va entraîner des problèmes de convergence.

Un bon préconditionnement peut permettre de surmonter le problème de détérioration du conditionnement du problème et l'influence de la régularisation sera étudié avec les mesures de qualité d'image.

L'aire couverte par les pixels polaires correspond à 80 % de l'aire couverte par les pixels cartésiens, les pixels polaires sont donc en moyenne plus petits. L'aire restante se situe dans les coins de l'image et l'information qu'elle contient n'est généralement pas nécessaire car les zones d'intérêt se situent normalement au centre de l'image. Maintenant concernant la partie de l'image couverte par les deux représentations, on observe une meilleure représentation donc une discrétisation avec des pixels plus petits sur 73 % de l'aire concernée avec la représentation cylindrique. La proportion de l'image qui est moins bien représentée se situe en périphérie de l'image, zone qui contient normalement de grandes zones homogènes et où il est moins important d'avoir une bonne représentation. Contrairement à ce qu'on aurait pu penser, la représentation en coordonnées cylindrique est donc meilleure que la représentation en coordonnées cartésiennes en terme d'erreur de représentation.

4.2.2 Génération du sinogramme

Le sinogramme est généré avec le programme CT Project à partir du fantôme créé avec XCAT afin d'éviter le crime inverse et de générer des données réaliste tenant compte du bruit de Poisson introduit par la nature quantique des photons.

On choisit de prendre une géométrie réaliste, c'est à dire la même taille de détecteur et de barrette qu'un scanner Siemens Somatom qui a des détecteurs de taille 1.5 mm \times 1.0 mm. On ne garde que 4 barrettes et 580 projections afin de limiter le temps de calcul et l'espace mémoire requis pour de pouvoir faire des reconstructions avec le modèle cartésien avec les moyens à notre disposition.

On effectue la mesure de la qualité de l'image dans un cadre 2D dans un premier temps pour simplifier les mesures. Dans la troisième dimension, la taille des voxels ne changeant pas, on peut supposer que la qualité de l'image a le même comportement avec la représentation cylindrique et cartésienne. La réduction au cadre 2D permet aussi de diminuer le temps de calcul et l'espace mémoire requis, ce qui facilite les expériences. On prend 1160 projections et une seule barrette afin de garder un contexte correspondant au scan 2D d'un fantôme par un scanner Siemens Somatom toujours dans l'esprit de se rapprocher le plus possible des conditions d'utilisation réelles.

Le bruit poissonien est introduit via CT Project, on choisit un SNR de 25 dB correspondant à un protocole de reconstruction basse dose de rayons X et un SNR de 35 dB correspondant à un protocole avec une dose classique de rayons X.

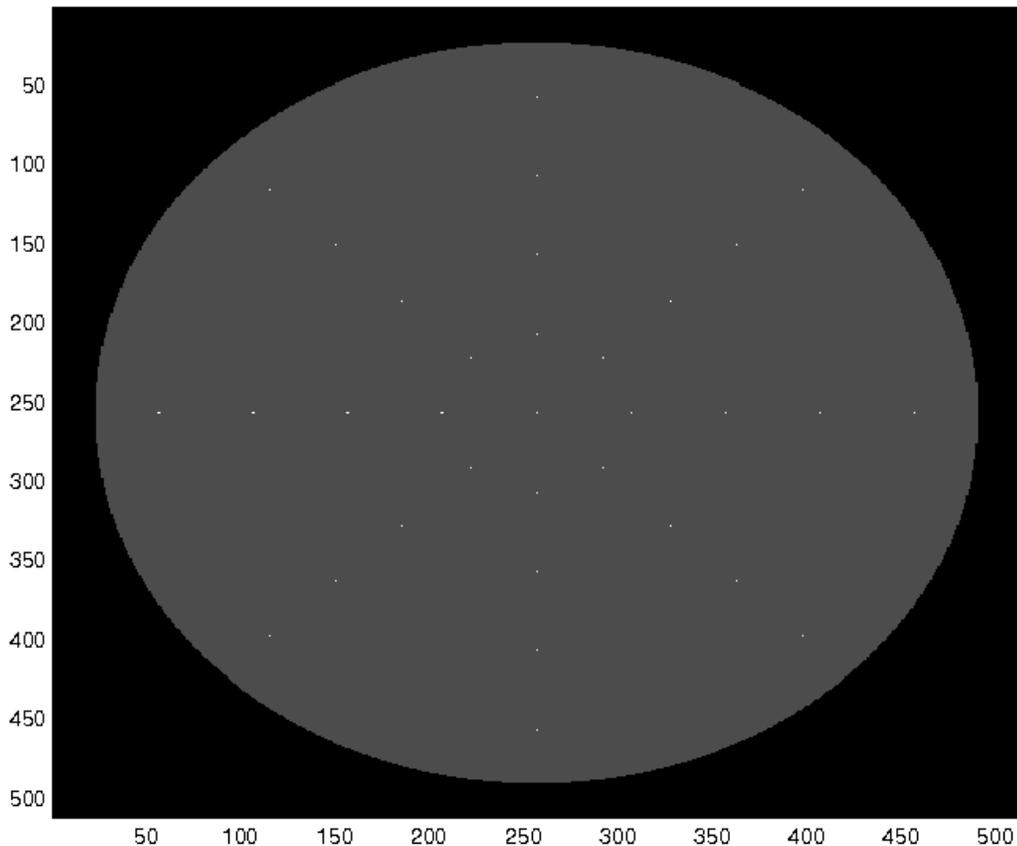


Figure 4.2 Fantôme XCAT utilisé pour la mesure de la MTF

4.3 Estimation

L'estimation est faite avec le logiciel MATLAB, en utilisant l'algorithme d'optimisation L-BFGS-B tel que décrit dans la section 3.1. On utilise un estimateur MAP et une fonction de régularisation L2 sur les différences premières afin de réduire l'impact du bruit sur la qualité des images reconstruites et rendre applicables les préconditionneurs développés dans le chapitre précédent.

Le paramètre λ de la fonction de régularisation a été fixé empiriquement de manière à avoir la meilleure qualité d'image au bout de 100 itérations sur un jeu de données indépendant de ceux utilisés pour les comparaisons, car empiriquement on a observé que les changements effectués sur l'estimée passées 100 itérations sont indiscernables lors de l'observation des images reconstruites. Une méthode de comparaison des performances des deux algorithmes

de reconstruction plus rigoureuse a été développée par Hamelin *et al.* (2011) mais cette validation complète nécessite l'utilisation de techniques de type Monte-Carlo et dépasse le cadre de ce travail.

L'initialisation de l'algorithme d'optimisation se fait avec une image nulle afin de traiter de la convergence dans un cadre général.

Dans le cadre de l'évaluation de la parallélisation et du préconditionnement, l'estimation se fait avec un maximum de 100 itérations pour la raison citée dans le paragraphe précédent, ou moins si l'algorithme a convergé, ce qui voudrait dire que le minimum de notre problème de minimisation est atteint. Ceci a pour but de permettre de voir la réduction maximale du critère pour chaque méthode.

Pour les mesures de qualité d'image, on utilise la solution exacte dans le cas de la représentation cylindrique de l'objet et la valeur de l'estimée après 100 itération avec la représentation cartésienne car on veut comparer les qualités maximales atteignables avec les deux types de représentations pour le même algorithme et seule la représentation cylindrique permet le calcul de la solution exacte avec nos moyens.

4.4 Implémentation

Le script de l'estimation est codé en MATLAB ce qui permet de profiter d'outils simples d'utilisation de calcul algébrique.

Afin d'optimiser les fonctions de projection et de rétroprojection, ces deux opération sont écrites en C++ pour les raisons citées dans 3.2.2 et introduites dans le code MATLAB via l'interface Mex. L'implémentation en C++ permet, en outre, d'utiliser la librairie d'accélération de calcul avec des matrices creuses SparseLib (<http://people.cs.kuleuven.be/~albert-jan.yzelman/software.php#SL>). Afin d'être aisément manipulable en C++, les matrices de projection et de rétroprojection sont stockées dynamiquement après leur construction avec MATLAB.

Toutes les estimations sont faites sur un ordinateur Intel avec 12 coeurs d'1,5 GHz, 24 Go de RAM, ayant un système d'exploitation Linux.

4.5 Taille de la matrice de projection

L'objectif de cette section est de vérifier que la représentation polaire nous permet bien de diminuer l'espace mémoire requis pour le stockage de la matrice de projection.

Dans cette partie, on compare la taille des matrices de projections pour les géométries définie dans 4.2.

Dans le cadre 2D, on a un objet de taille 512×512 , avec $N_p = 1160$ projections et un scanner d'une barette de 672 détecteurs. Dans le cadre 3D, on a un objet de taille $256 \times 256 \times 4$, avec $N_p = 580$ projections et un scanner de 4 barettes de 336 détecteurs. Pour faciliter nos expériences, on a choisi un faible nombre de tranches et de barettes de détecteur pour tester la 3D afin de diminuer la taille de chaque bloc de la matrice de projection sans changer la structure de cette matrice, ce qui permet de garder la même structure algébrique du problème et donc observer les mêmes améliorations tout en diminuant l'espace mémoire requis pour le stockage de la matrice de projection et le temps de calcul de chacune des sous opérations indépendantes lors de la projection et de la rétroprojection sans en diminuer le nombre qui dépend du nombre de projections choisi, en outre, on utilise les symétries exposées par Gendron (2008) afin de diviser par quatre la taille des matrices de projection en coordonnées cartésiennes. Les tailles des matrices de projection sont présentées dans le tableau suivant :

représentation	cadre 2D	cadre 3D
cylindrique	0.011 Go	0.013 Go
cartésienne	3.497 Go	2.117 Go

On observe que la représentation cylindrique permet de diminuer la taille de la matrice de projection par $N_p/4$. Ceci est conforme à la théorie car la représentation cylindrique permet de ne stocker la matrice de projection que pour un angle de projection en utilisant l'invariance par rotation, alors que l'utilisation de la représentation cartésienne nécessite le stockage du quart des angles de projection si on utilise l'invariance par rotation d'un angle $\pi/2$ pour un nombre de projections multiple de quatre.

On peut en conclure que la représentation cylindrique est très efficace pour diminuer l'espace mémoire requis pour le stockage de la matrice de projection et donc pour la reconstruction.

4.6 Parallélisation

L'objectif de cette section est de montrer l'impact de la parallélisation sur les représentations cartésienne et cylindrique en terme de vitesse de calcul d'une estimation. D'après la figure 3.1 et des expériences préliminaires, on sait que les opérations de projection et de rétroprojection sont à l'origine de plus de 75% du temps de calcul de chaque itération de l'estimation que ce soit en coordonnées cartésiennes ou polaires. On ne s'est donc intéressé qu'à ces deux opérations pour l'étude de l'impact de la parallélisation.

La méthode de calcul de la multiplication matrice creuse-vecteur dépend du schéma de stockage de la matrice creuse de projection et de rétroprojection utilisé. Les différentes méthodes mises en oeuvre ayant des complexités de calcul différentes, il est nécessaire de prendre

en compte le schéma de stockage des matrices creuses utilisées lors de l'évaluation du temps de calcul.

La taille de l'objet peut avoir un impact sur le temps nécessaire à la répartition des données sur les différents coeurs et sur le temps de calcul de la multiplication. Cette propriété doit donc être étudiée.

Cette étude se compose donc de deux expériences qui permettent de tester deux facteurs qui peuvent influencer les performances de la parallélisation autres que le nombre de coeurs utilisés : le schéma de stockage de la matrice de projection et la taille de l'objet à utiliser.

Avec la première expérience, on étudie l'impact du schéma de stockage de la matrice de projection et du nombre de coeurs utilisés pour la projection et la rétro-projection. La seconde expérience consiste à étudier l'impact de la taille de l'objet sur la parallélisation en fonction du nombre de coeurs pour la projection et la rétro-projection.

Pour réaliser les deux expériences, on utilise le fantôme et la géométrie décrits dans 4.2. Le temps de calcul de chaque opération est mesuré à l'aide du logiciel MATLAB et de ses fonctions tic-toc. La taille de l'objet à reconstruire est définie comme le nombre d'éléments totaux de cet objet afin de s'affranchir du type de représentation.

4.6.1 Étude du nombre de coeurs et du schéma de stockage

Dans un premier temps, on étudie l'impact du nombre de coeurs utilisés ainsi que du type de stockage de la matrice de projection ou de rétro-projection sur les opérations de projection et de rétro-projection. Pour chaque opération, 30 répétitions sont faites et seul le temps de calcul moyen est gardé. Les données utilisées sont celles présentées dans 4.2. La projection est en fait une multiplication matrice-vecteur quand à la rétro-projection, c'est une multiplication vecteur-matrice. Les résultats sont présentés sous forme de deux graphiques 4.3 et 4.4.

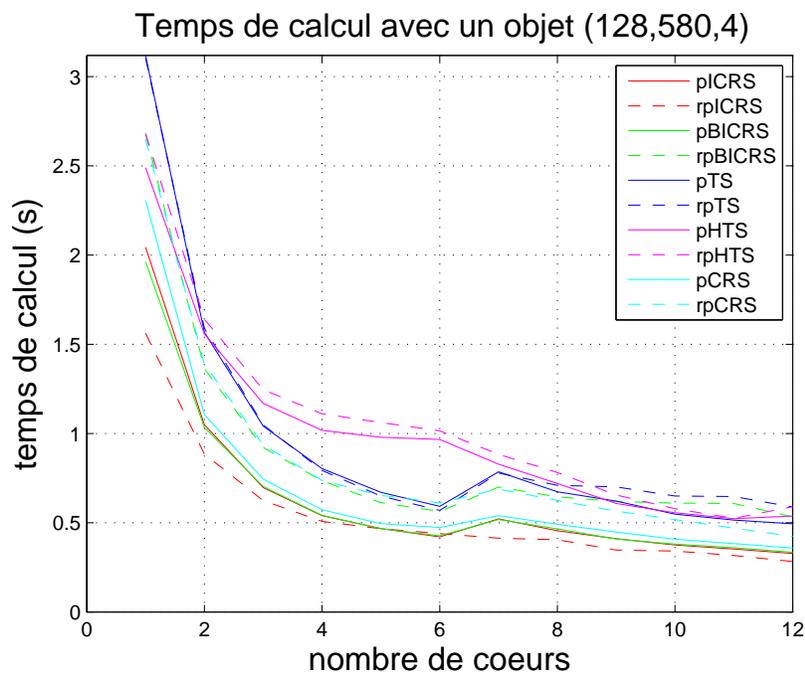


Figure 4.3 Modèle polaire

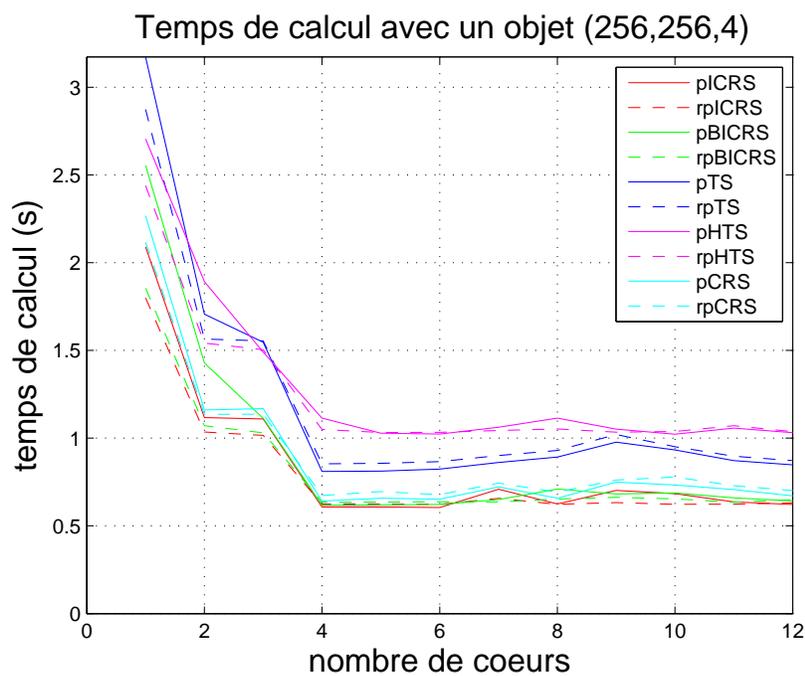


Figure 4.4 Modèle cartésien

Les deux figures ci-dessus représentent les moyennes du temps de calcul d'une projection (p-) et d'une rétroprojection (rp-) en fonction du type de stockage de la matrice (TS, CRS, HTS, ICRS, BICRS) et du nombre de coeurs utilisés

On observe avec la représentation cartésienne une diminution d'un facteur 2 à 3,5 du temps de calcul pour un nombre de coeurs utilisés supérieur à 3. De plus, on peut voir que le schéma de stockage le plus rapide pour la multiplication est le schéma ICRS et le plus lent est le HTS, et ce, quel que soit le nombre de coeurs utilisés.

Pour la représentation cylindrique, on observe une diminution du temps de calcul par un facteur de 2.5 à 4 suivant le schéma de stockage pour un nombre de coeurs utilisés supérieur à 5. Comme pour le cartésien, le schéma ICRS est le plus rapide et le schéma HTS le plus lent.

La représentation en coordonnées cylindrique permet un meilleur gain avec la parallélisation que la représentation cartésienne, ce qui amène à une meilleure vitesse de calcul des deux opérations étudiées avec la représentation cylindrique. Il est nécessaire de prendre en compte dans l'interprétation de ces résultats le fait qu'un travail d'optimisation plus important a été réalisé sur les opérations utilisant la représentation cylindrique. Il est sans doute possible d'améliorer la vitesse de calcul des opérations parallélisées utilisant la représentation cartésienne. La comparaison entre les deux représentations ne peut donc être prise en compte que dans le cadre de l'implémentation des deux opérations spécifique à ce projet. On peut tout de même conclure que la parallélisation permet une augmentation significative de la reconstruction avec le modèle en coordonnées cylindriques.

4.6.2 Étude de la taille de l'image et de la nature de la multiplication

Avec l'expérience précédente, on a vu qu'il était possible de diminuer le temps de calcul avec la représentation cylindrique en utilisant la parallélisation. On a vu que cette diminution du temps de calcul dépend du type de stockage de la matrice de projection et de rétroprojection.

Cette deuxième expérience a pour but d'étudier l'impact de la taille de l'objet à reconstruire ainsi que de la nature de la multiplication que constituent la projection et la rétroprojection sur la diminution du temps de calcul avec l'utilisation de la parallélisation.

On utilise les données obtenues dans 4.2, redimensionnées pour diviser la taille de l'objet à reconstruire par 4 et par 8. On doit aussi diviser le nombre de projections par 2 et par 4 pour avoir une bonne équivalence entre les deux représentations. On ne garde que le schéma le plus rapide ICRS, le plus lent HTS et le plus simple TS. pour l'objet en coordonnées cylindriques, on calcule la projection et la rétroprojection en utilisant une multiplication matrice-vecteur ou vecteur-matrice. La projection et la rétroprojection en coordonnées cartésiennes sont respectivement une multiplication matrice-vecteur et vecteur-matrice. Comme pour l'expérience précédente, on effectue une moyenne sur 30 itérations de chaque opération que l'on présente sous forme des graphiques : Fig. 4.5 et Fig. 4.6 pour le modèle cartésien et Fig. 4.7, Fig. 4.8

et Fig. 4.9 pour le modèle polaire.

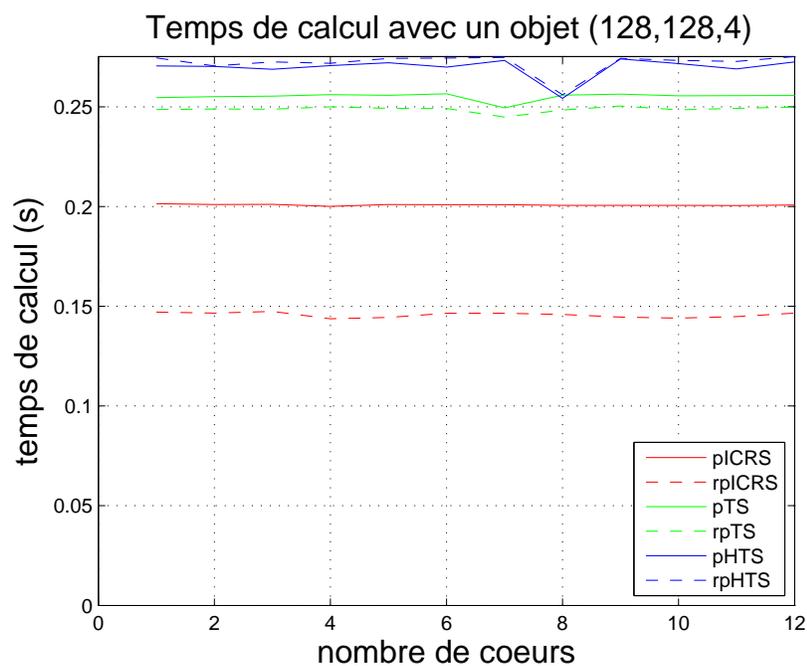


Figure 4.5 avec 290 projections

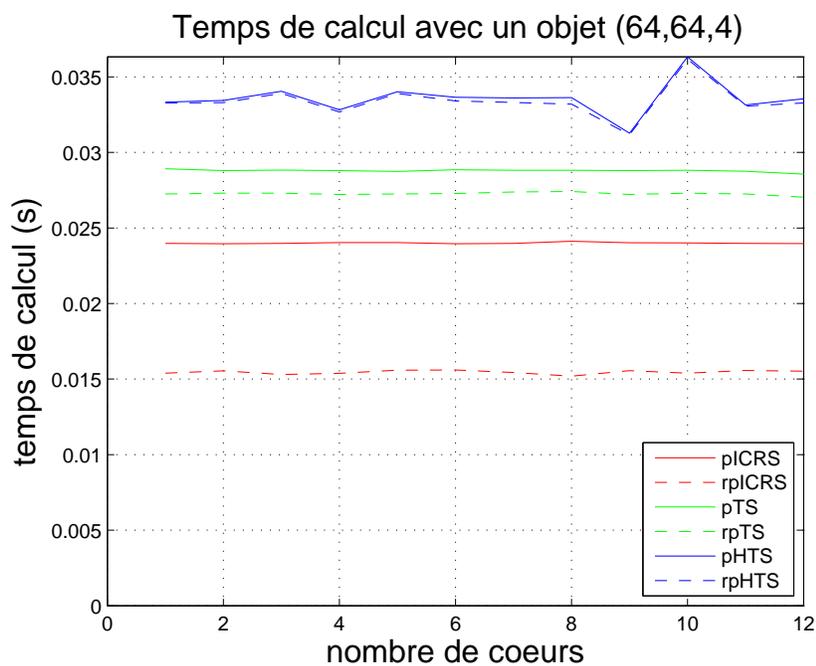


Figure 4.6 avec 146 projections

Les deux figures ci-dessus représentent les moyennes du temps de calcul d'une projection (p) et d'une rétroprojection (rp) en fonction du type de stockage de la matrice et du nombre de coeurs utilisés avec le modèle cartésien

Les trois figures ci-contre présentent les moyennes du temps de calcul d'une projection (p-) et d'une rétroprojection (rp-) en fonction du type de stockage de la matrice, du nombre de coeurs utilisés et du type de multiplication matrice-vecteur ou vecteur-matrice (-t) avec le modèle polaire

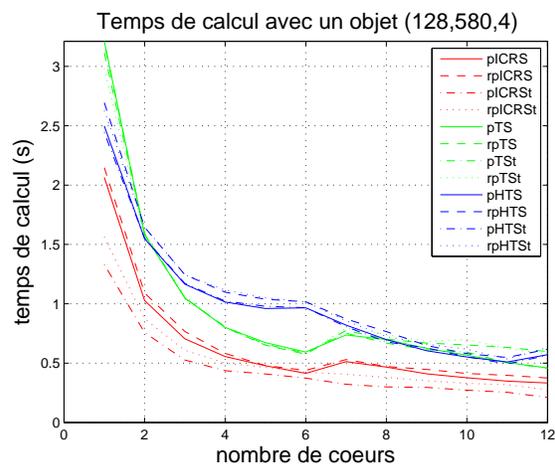


Figure 4.7 avec 580 projections

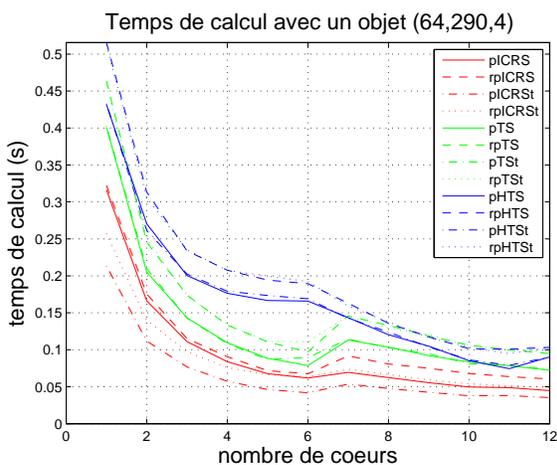


Figure 4.8 avec 290 projections

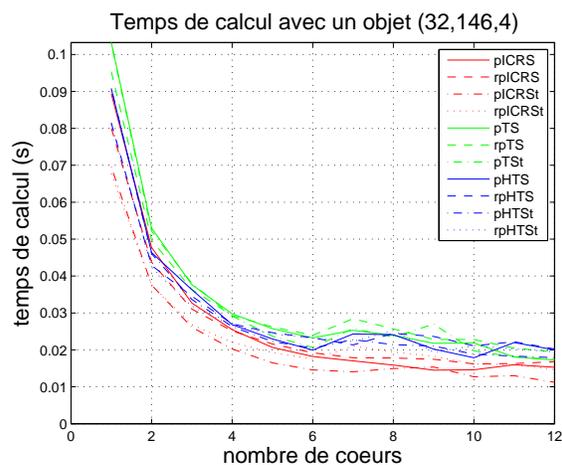


Figure 4.9 avec 146 projections

Tout d'abord concernant la réduction de la taille de l'objet, on observe que le schéma ICRS est toujours le plus rapide et globalement HTS reste le plus lent sauf pour un nombre de coeurs faible où TS peut être plus lent. Avec la représentation cylindrique, quelle que soit la taille de l'objet étudiée, les gains maximaux de vitesse de calcul restent d'un facteur 2,5 à 4. Contrairement à la représentation cylindrique, avec la représentation cartésienne et des objets de moins grosse taille, il n'y a plus d'augmentation de vitesse de calcul avec la parallélisation. À première vue, on peut penser que la distribution des données entre les coeurs compense le temps gagné par la parallélisation ce qui empêche tout gain de vitesse de calcul, mais il faut se rappeler que les opérations usant la représentation cartésienne ont été moins bien optimisées. Ces résultats montrent que l'implémentation de l'ART avec une

représentation cartésienne peut être améliorée.

Dans le cadre de ce projet, la représentation cylindrique parallélisée représente un net avantage par rapport à la représentation cartésienne parallélisée car elle permet de diminuer le temps de calcul de la projection et de la rétroprojection même pour de petits objets.

En comparant, au niveau de la représentation cylindrique, la vitesse de calcul de la multiplication matrice-vecteur et vecteur-matrice, on s'aperçoit que la multiplication vecteur-matrice est plus rapide que l'inverse autant au niveau de la projection que de la rétroprojection. Mais la différence entre les deux types de multiplication reste très faible par rapport au temps de calcul global de chacune des opérations.

4.6.3 Conclusion sur la parallélisation

Au vu de ces premiers résultats, pour la suite des expériences, on utilise un nombre de coeurs supérieur à 5 ainsi que les multiplications vecteur-matrice pour la projection et la rétroprojection en coordonnées cylindriques.

On a vu avec ces deux expériences que la parallélisation de l'algorithme utilisant la représentation cylindrique permet de diminuer considérablement le temps de calcul par itération jusqu'à atteindre une vitesse de projection et de rétroprojection supérieure à celle obtenue avec le même algorithme mais en utilisant la représentation cartésienne de l'objet. De plus, cette vitesse de calcul est affectée par le nombre de coeurs, la taille de l'objet et le type de stockage des matrices de projection et rétroprojection.

Lorsque l'on compare nos résultats en 3D et les résultats obtenus précédemment en 2D par Goussard *et al.* (2013), on observe que les matrices de projection et de rétroprojection ont la même propriété bloc-circulante avec le même nombre de bloc pour un même nombre de projection N_p . En effet, les matrices de projection 2D et 3D sont bloc-circulantes et chaque ligne de blocs est composée de $N_p = N_t$ blocs où N_t est le nombre de pixels ou voxels tangentiels. Ces blocs sont de taille (M, K) . Dans le cas de la matrice 2D, M est le nombre de détecteurs utilisés et $K = N_r$ avec N_r le nombre de pixels radiaux. Pour la matrice 3D, M est le nombre de détecteur multiplié par le nombre de barrettes utilisées et $K = N_r * N_s$ avec N_r le nombre de voxels radiaux et N_s le nombre de voxels axiaux, on a donc des blocs plus grands. Or on a vu dans 2.4.2.2 que la parallélisation est faite en fonction du nombre de blocs, ce qui explique que l'efficacité de cette parallélisation en 3D est la même qu'en 2D. Il en va de même avec l'utilisation de la matrice de rétroprojection. Pour un même scanner, le passage du 2D au 3D ne changera *a priori* pas l'efficacité de la parallélisation mais augmentera l'espace mémoire utilisé et le temps de calcul de par l'augmentation de la taille de chaque bloc des matrices de projection et de rétroprojection.

Ces premières expériences permettent de remplir partiellement le premier objectif énoncé

au début du chapitre en diminuant le temps de chaque itération. Pour remplir totalement l'objectif, il faut maintenant comparer le nombre d'itérations que nécessitent deux algorithmes utilisant chacun un des deux types de représentations exposés pour reconstruire une image de même qualité et si possible le diminuer. Ce point est étudié avec le préconditionnement.

4.7 Préconditionnement

L'objectif qu'on cherche à atteindre avec le preconditionnement est le même qu'avec la parallélisation, c'est à dire la diminution du temps de calcul de l'algorithme de reconstruction, mais avec un approche différente. Au lieu de diminuer le temps de calcul de chaque itération, la méthode consiste à diminuer le nombre d'itérations au prix d'une légère augmentation de la complexité et du temps de calcul de chaque itération (cf 2.4.3). Il est donc nécessaire d'étudier les compromis nombre d'itérations-temps de calcul par itération pour chaque type de preconditionneur afin de pouvoir déterminer lequel permet la plus grande diminution du temps de calcul. Ce n'est pas le seul objectif, on sait que, dans le cas de la représentation cylindrique, le problème est détérioré par les tailles variables des voxels. Ceci provoque des problèmes de convergence qui rendent le preconditionnement primordial.

Lors de cette étude il est aussi nécessaire de garder à l'esprit l'objectif qui consiste à limiter l'espace mémoire utilisé durant la reconstruction lors de l'étude des preconditionneurs.

Dans cette section, on étudie les capacités de diminution du temps de calcul des preconditionneurs diagonal (PD) et de Cholesky incomplet (IC) dans le domaine de Fourier en trois dimensions. On ne prend en compte que ces deux types de preconditionneurs car ce sont ceux qui ont donnés les meilleurs résultats en 2D.

Cette étude consiste donc d'une part à évaluer l'augmentation du temps de calcul par itération et de l'espace mémoire utilisé pour chaque preconditionneur et d'autre part à étudier l'augmentation de la vitesse de convergence (la diminution du nombre d'itérations) que peut apporter chaque preconditionneur. Il n'existe pas à ce jour de preconditionneur simple et efficace pour la représentation cartésienne, on effectue alors à titre indicatif une comparaison des résultats avec la méthode cartésienne non-preconditionnée afin de voir si, avec le preconditionnement, on peut égaler la vitesse de convergence et la vitesse de calcul de la méthode cartésienne.

On sait que le niveau de bruit du sinogramme ou SNR affecte la vitesse de convergence des ART, on peut donc supposer que l'efficacité des preconditionneurs en dépend aussi. On fait donc varier le SNR des données lors de nos expériences.

Afin de limiter la taille mémoire des preconditionneurs IC, on a utilisé plusieurs taux de remplissage (1/5 et 1/100). Pour garder l'approximation creuse de la matrice normale bloc-diagonalisée définie positive afin de pouvoir appliquer la décomposition de Cholesky incomplète, on a du pondérer sa diagonale par les valeurs α (2 et 10). On a testé ces deux paramètres avec deux valeurs afin d'être en mesure d'étudier leur impact sur les performance du preconditionnement. Les critères de comparaison mesurés sont le temps de calcul par itération, la distance normalisée par rapport à la solution exacte et la RMSE globale normalisée.

Les résultats sont représentés sous forme de graphiques.

Pour les expériences sur le préconditionnement, on a utilisé la géométrie détaillés dans 4.2 et on a gardé les paramètres permettant la meilleure diminution du temps de calcul de la projection et rétroprojection parallélisées, c'est à dire un nombre de coeurs égal à 12 ainsi que les multiplications vecteur-matrice pour la projection et la rétroprojection en coordonnées cylindriques.

4.7.1 Mesure du temps par itération

Le temps par itération (Fig. 4.10 et Fig. 4.11) est mesuré via les fonctions tic et toc de MATLAB en secondes pour chaque itération et chaque préconditionneur. On réalise cette expérience avec les préconditionneurs décrits dans la section précédente et un SNR de 25 dB et de 35 dB.

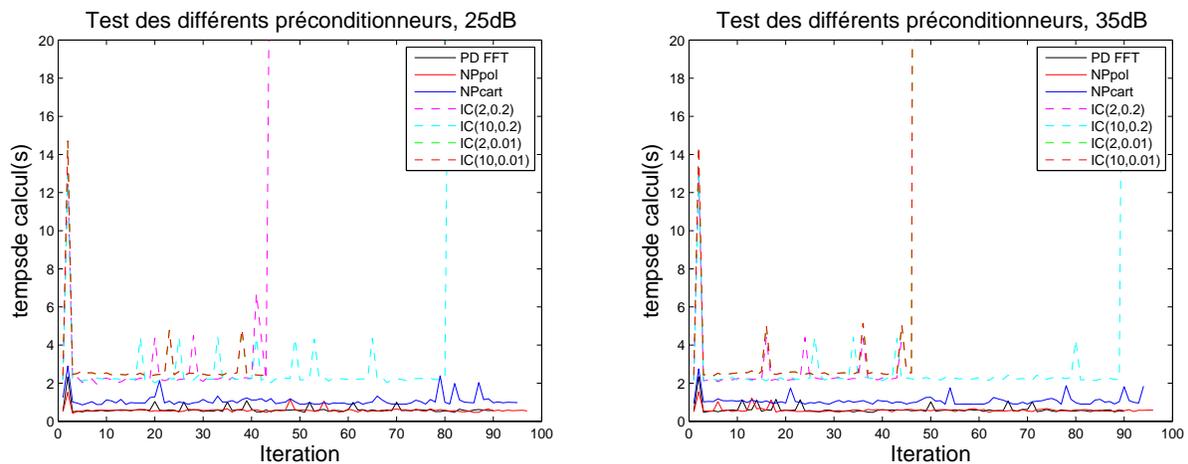


Figure 4.10 temps de calcul par itération, SNR = 25 dB Figure 4.11 temps de calcul par itération, SNR = 35 dB

PD FFT : préconditionneur diagonal dans le domaine de Fourier ; NPpol : sans préconditionneur, modèle polaire ; NPcart : sans préconditionneur, modèle cartésien ; IC : préconditionneur de Cholesky incomplet dans le domaine de Fourier suivi de la valeur de α et du taux de remplissage

On observe que, quel que soit le rapport signal à bruit (SNR), le temps par itération ne change pas ce qui est cohérent car le SNR n'agit que sur la difficulté de converger et non pas la complexité de calcul par itération. Au contraire, on observe que le temps par itération varie beaucoup suivant le préconditionneur. Le préconditionneur IC augmente le temps de calcul par itération de manière significative. On remarque que quels que soient les paramètres

de ce préconditionneur (choix de α et taux de remplissage) le temps par itération varie très peu. Au contraire, le préconditionneur PD augmente très peu voir pas du tout le temps de calcul.

Cette augmentation du temps de calcul est due à la différence de structure entre les préconditionneurs. Le préconditionneur PD, moins complexe, est représenté sous forme d'un seul vecteur alors que le préconditionneur de Cholesky incomplet est une matrice creuse bloc-diagonale. Lors d'une itération on effectue une multiplication préconditionneur-objet estimé et une multiplication transposée du préconditionneur-gradient du critère à minimiser. Avec le préconditionneur PD ces deux opérations se résument à la multiplication terme à terme de deux vecteurs qui est une opération très rapide, alors qu'avec le préconditionneur IC, ces deux opérations représentent deux systèmes triangulaires à résoudre à chaque itération. Ces deux opérations sont bien plus complexes et ont un temps de calcul beaucoup plus long, ce qui naturellement va augmenter le temps de calcul d'une itération.

Enfin, sans préconditionnement, une itération avec le modèle polaire est plus rapide qu'avec le modèle cartésien. Cela confirme les résultats de temps de calcul de la projection et la rétroprojection parallélisées et la meilleure optimisation de l'ART polaire développé par rapport au cartésien étudié.

4.7.2 Mesure de la distance à la solution exacte

Pour cette deuxième expérience, on utilise les mêmes conditions expérimentales que l'expérience précédente. On définit la vitesse de convergence comme le nombre d'itérations nécessaires à la réduction de la distance à la solution exacte du problème à optimiser, autrement dit, moins un préconditionneur nécessite d'itérations pour se rapprocher de la solution exacte, plus sa vitesse de convergence est rapide. Cette mesure de convergence a comme avantage d'être très simple, si on arrive à calculer la solution exacte, et explicite. En effet, elle permet de savoir si notre algorithme a convergé et si non, s'il est proche de la solution exacte de l'estimation.

L'inconvénient majeur de ce critère est que la solution exacte ne peut être calculée qu'en coordonnées cylindriques en passant dans le domaine de Fourier. Ce dernier aspect permet de calculer l'inverse de N_p sous-matrices (les blocs de la diagonale) plutôt que l'inverse d'une seule grande matrice qui est impossible à faire avec les moyens à notre disposition vu la taille du Hessien de notre estimateur. La solution exacte est ensuite interpolée en coordonnées cartésiennes pour être utilisée avec le modèle cartésien. Ce critère n'est pas optimum pour comparer le polaire et le cartésien par contre il est très efficace pour comparer l'efficacité des différents préconditionneurs du modèle polaire.

La mesure de la distance normalisée à la solution exacte à chaque itération (Fig. 4.12 et

Fig. 4.13) met en évidence plusieurs propriétés intéressantes des préconditionneurs.

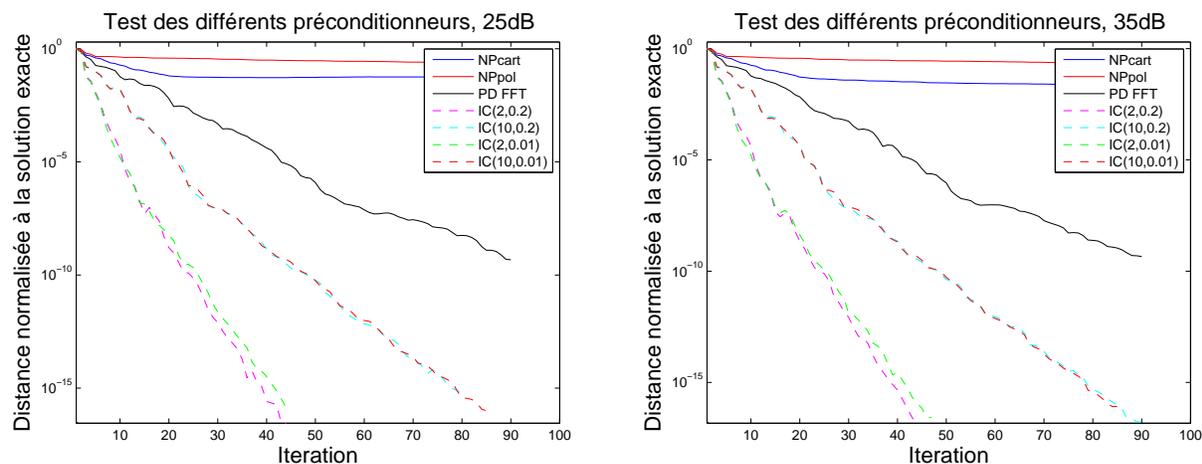


Figure 4.12 Distance normalisée à la solution exacte, SNR = 25 dB

Figure 4.13 Distance normalisée à la solution exacte, SNR = 35 dB

PD FFT : préconditionneur diagonal dans le domaine de Fourier; NPpol : sans préconditionneur, modèle polaire; NPcart : sans préconditionneur, modèle cartésien; IC : préconditionneur de Cholesky incomplet dans le domaine de Fourier suivi de la valeur de α et du taux de remplissage

Tout d'abord on observe que sans préconditionneur la convergence du modèle polaire est beaucoup plus lente que celle du modèle cartésien et reste après 100 itérations à une distance plus grande à la solution exacte que le modèle cartésien, ceci montre encore que le préconditionnement du modèle polaire est obligatoire pour obtenir au moins les mêmes résultats que le modèle cartésien.

Ensuite on observe que le préconditionnement permet effectivement d'améliorer la vitesse de convergence. Le préconditionneur PD diminue moins cette vitesse de convergence que le préconditionneur IC.

Concernant le préconditionneur IC, on observe une augmentation de la vitesse de convergence avec l'augmentation du taux de remplissage pour une pondération α petite de la diagonale de l'approximation creuse de la matrice de projection. De plus, cette vitesse de convergence augmente avec le taux de remplissage, même si l'augmentation est faible. Ceci montre que lorsqu'on augmente le taux de remplissage, avec le modèle et la géométrie utilisés pour cette étude, on multiplie par 20 la taille du préconditionneur en rajoutant des valeurs qui n'auront que peu d'impact par rapport à celles déjà présentes. Il n'est donc pas nécessaire d'avoir un taux de remplissage élevé. Dans les conditions expérimentales, 1/100 suffit à avoir de bonnes performances du préconditionneur. Pour un α grand, c'est à dire un poids important de la diagonale par rapport au reste des valeurs, le préconditionneur IC tend à se

comporter comme le préconditionneur PD et la réduction de la distance à la solution exacte est moins grande. En outre, dans cette configuration la différence de taux de remplissage aura peu d'impact c'est à dire que la diagonale aura un poids très important par rapport au reste des valeurs.

Comme prédit dans la section 3.3 : moins le préconditionneur nécessite d'approximations, plus l'augmentation de la vitesse de convergence est importante.

Enfin, le SNR, pour les valeurs étudiées, n'a pas d'influence significative sur les performances des préconditionneurs à s'approcher de la solution exacte.

4.7.3 Conclusions sur le préconditionnement

On peut conclure de ces expériences que quel que soit le SNR, le préconditionnement augmente la vitesse de convergence et permet de diminuer le temps de calcul d'une estimation par une distance voulue à la solution exacte de l'estimée suffisamment grande. Dans les conditions expérimentales utilisées, si la distance à la solution exacte voulue dépasse 10^{-1} , le préconditionnement est efficace.

Quand on compare le préconditionneur PD avec le préconditionneur IC, on observe qu'une itération est approximativement quatre fois plus rapide, que l'espace mémoire nécessaire au stockage de la matrice de préconditionnement est beaucoup plus faible (dans le cas présent entre 6 et 120 fois plus faible), et la vitesse de convergence en nombre d'itérations est seulement 1.5 ou deux fois plus élevée. Le préconditionneur PD nécessite donc un temps de calcul moins long pour atteindre la même qualité d'image en utilisant moins de mémoire. On peut dire que le préconditionneur PD est plus efficace et remplit mieux les objectifs de diminution du temps de calcul et d'espace mémoire que le préconditionneur IC.

4.8 Qualité d'image

Avec l'utilisation de la représentation cylindrique on a pu diminuer la taille de la matrice de projection, avec le préconditionnement et la parallélisation, on a pu remplir les objectifs qu'étaient la diminution du temps de calcul avec l'algorithme utilisant cette représentation.

L'objectif de cette partie est d'évaluer l'impact de la représentation cylindrique sur la qualité de l'image globale et locale des reconstructions. Pour cela, on utilise comme référence la qualité d'image obtenue avec l'algorithme de reconstruction cartésien.

Dans la partie 3.4, on définit la qualité d'image par sa résolution et son contraste. De plus, on voit que la résolution peut être mesurée à l'aide de la MTF et que le contraste est lié au niveau de bruit de l'image qui peut être mesuré par la RMSE sur un fantôme homogène. En outre, on ne mesure la qualité d'image que sur les images en coordonnées cartésiennes car ce sont les seules images de scanner à rayons X qu'on utilise réellement en radiologie.

On sait que la qualité d'image dépend du niveau de bruit des données, c'est pourquoi on étudie la qualité d'image pour deux niveaux de SNR du sinogramme, 25 dB et 35 dB, qui correspondent à une valeur basse et une valeur standard du SNR des données obtenues avec les scanners actuels.

On a d'abord étudié la qualité globale de l'image à travers l'étude de la RMSE sur un fantôme réaliste.

4.8.1 Mesure de la RMSE avec un fantôme réaliste

Pour cette partie, on utilise le même fantôme et la même géométrie que ceux utilisés pour l'étude des préconditionneurs.

La mesure de la RMSE à chaque itération (Fig. 4.14 et Fig. 4.15) montre que sans préconditionneur l'estimée du modèle polaire est de moins bonne qualité que celle du modèle cartésien mais qu'avec préconditionneur, on obtient une meilleure RMSE globale de l'image qu'avec la représentation cartésienne au bout de 100 itérations. Ceci est dû à la différence de vitesse de convergence beaucoup plus lente du modèle cartésien.

On observe que les préconditionneurs atteignent un seuil. Ce seuil correspond à la distance entre le fantôme et la solution exacte du modèle et ne peut être dépassé sans changer les paramètres de l'estimateur. Le préconditionneur IC l'atteint plus vite que le préconditionneur PD entre 10 et 20 itérations au lieu de 30 du fait de sa vitesse de convergence plus élevée, ce qui confirme les résultats de l'étude de vitesse de convergence.

Le SNR est corrélé avec la qualité de l'image globale atteinte, en effet plus le SNR est important meilleure sera la qualité de l'image maximale atteinte, ce qui est conforme à la théorie et à nos attentes. De plus le SNR n'a pas d'impact sur les performances des deux

préconditionneurs à atteindre la qualité d'image maximale.

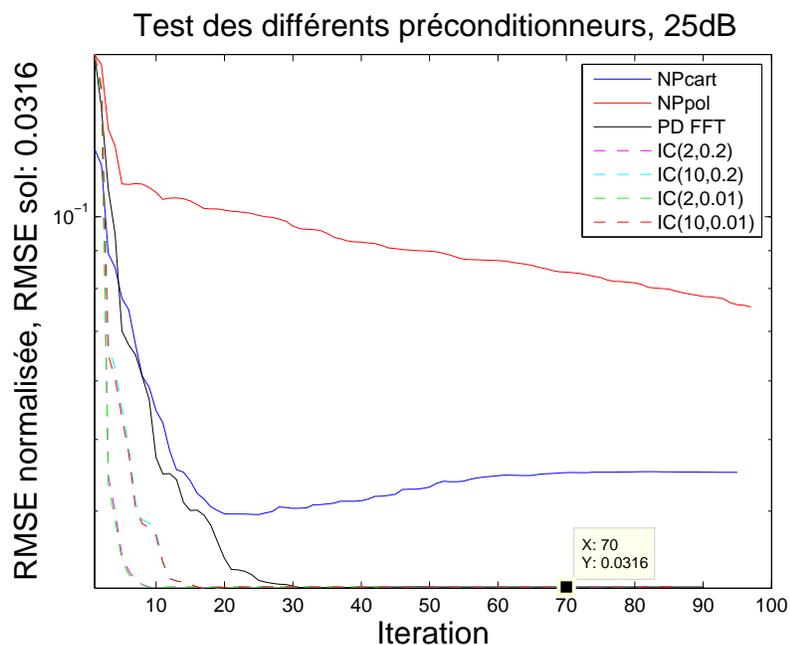


Figure 4.14 RMSE, SNR = 25 dB

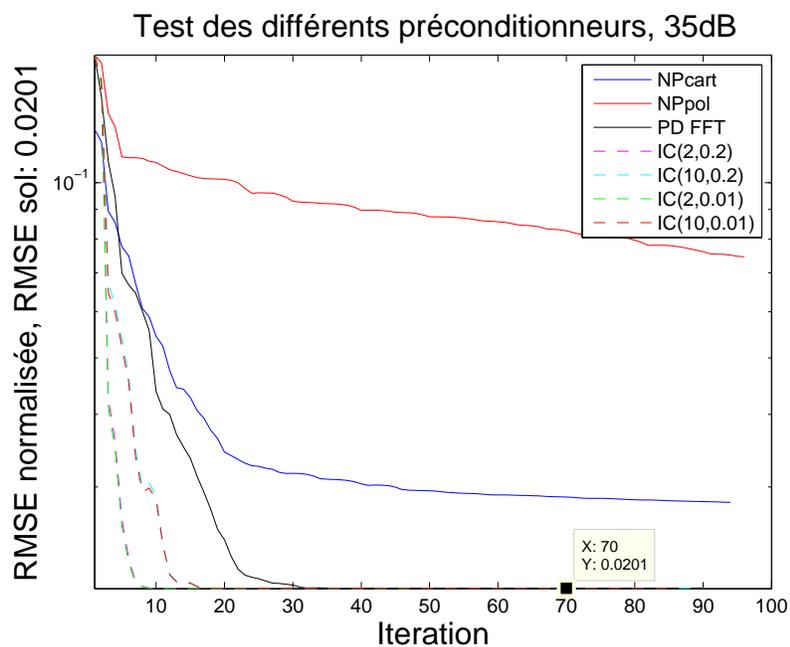


Figure 4.15 RMSE, SNR = 35 dB

PD FFT : preconditionneur diagonal dans le domaine de Fourier ; NPpol : sans preconditionneur, modèle polaire ; NPcart : sans preconditionneur, modèle cartésien ; IC : preconditionneur de Cholesky incomplet dans le domaine de Fourier suivi de la valeur de α et du taux de remplissage

Cette expérience montre que l'utilisation de préconditionneur pour le modèle polaire permet d'atteindre une RMSE globale de même ordre de grandeur qu'avec le modèle cartésien.

4.8.2 Mesure de la qualité d'image avec un fantôme spécifique

Nos premières observations d'images reconstruites avec la représentation cylindrique nous ont fait penser que la qualité de l'image locale peut être affectée par la distance au centre de l'image. De plus, la taille des voxels augmente avec cette distance au centre, ce qui détériore le conditionnement de l'image, il est possible que cette propriété rende la qualité de l'image inhomogène. C'est pourquoi la MTF et la RMSE sont mesurées localement en prenant en compte la distance au centre de l'image.

4.8.2.1 Mesure de la MTF

La MTF a été calculée en utilisant le fantôme présenté dans 4.2. Pour chaque impulsion représentée par une bille métallique très petite, la PSF a été estimée localement par MAP dans une fenêtre de l'image reconstruite centrée sur l'impulsion étudiée. La MTF est ensuite calculée suivant la formule présentée dans 3.4.

L'estimée utilisée est, pour le modèle cartésien, le résultat de l'estimation au bout de 100 itérations, seuil au delà duquel les modifications apportées à l'estimée sont indiscernables, et, pour le modèle polaire, la solution exacte calculée avec l'inverse de la matrice normale du système.

La taille de la fenêtre a été établie empiriquement à 50×50 pixels afin de recouvrir totalement l'étalement de l'impulsion ainsi qu'une partie de l'arrière-plan pour faciliter l'estimation de la PSF.

Afin de tester l'impact du bruit sur la qualité des images reconstruites, on a généré des sinogrammes ayant un SNR de 25 et 35 dB qui sont une valeur basse et une valeur moyenne du SNR des données d'un scanner clinique actuel. Le paramètre de régularisation est défini empiriquement afin d'avoir la meilleure qualité d'image par rapport au niveau de bruit, c'est à dire 10 avec 35 dB et 100 avec 25 dB. Afin d'éviter la singularité de la matrice normale du système et faciliter le calcul de son inverse, on a dû fixer un $\lambda = 1$ lors de l'estimation sans bruit.

Les résultats sont présentés sous forme de graphiques représentant la moyenne des MTF calculées pour les huit impulsions à la même distance au centre pour les cinq distances choisies : pour le modèle polaire : 4.16, 4.17 et 4.18 et pour le modèle cartésien : 4.19, 4.20 et 4.21.

Mesure de la MTF avec le modèle polaire

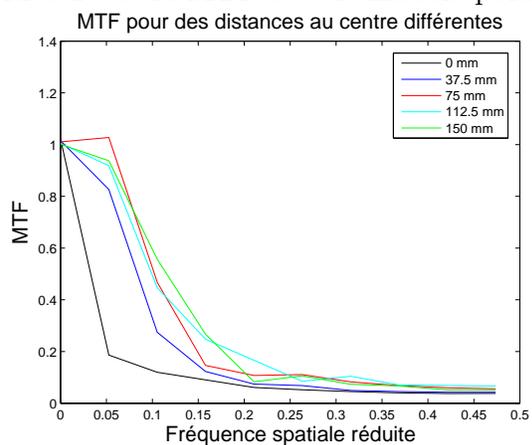


Figure 4.16 SNR=25dB

Mesure de la MTF avec le modèle cartésien

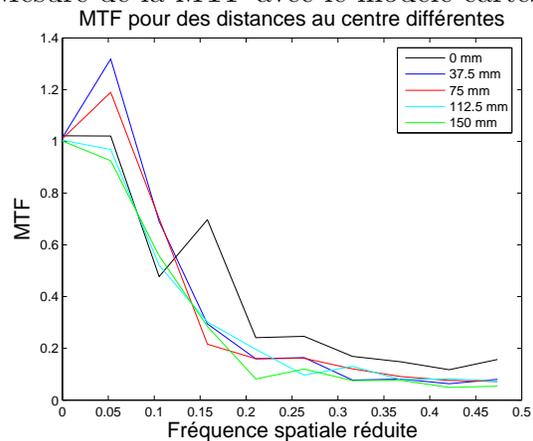


Figure 4.19 SNR=25dB

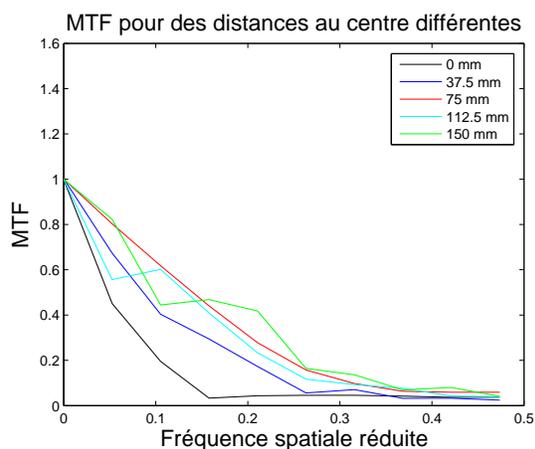


Figure 4.17 SNR=35dB

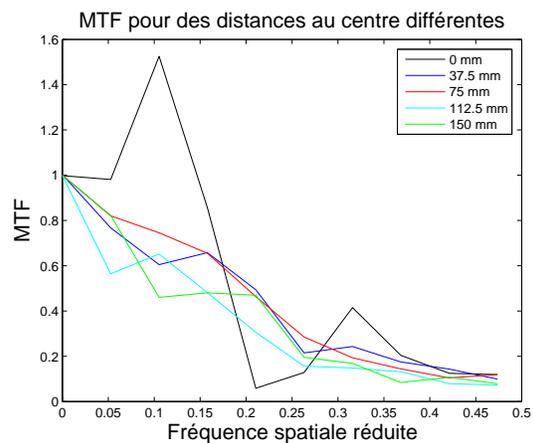


Figure 4.20 SNR=35dB

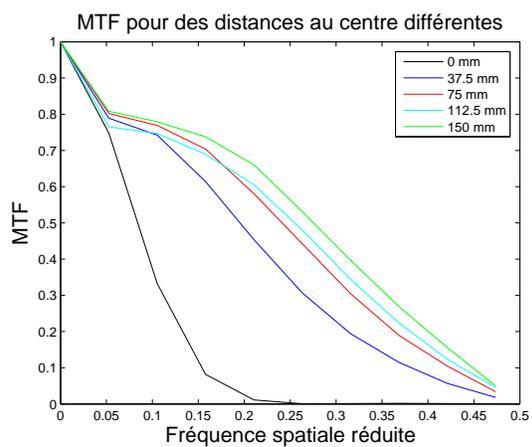


Figure 4.18 sans bruit

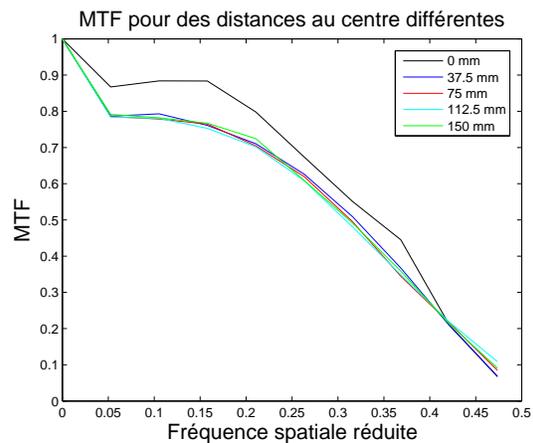


Figure 4.21 sans bruit

On observe que dans le cas du modèle polaire, la MTF est la moins bonne au centre de l'image (c'est à dire qu'elle est la moins proche de l'horizontale d'équation $y = 1$ représentant la transformée de Fourier d'un dirac, la PSF idéale) et s'améliore plus on s'écarte du centre quel que soit le niveau de bruit. À l'inverse, avec le modèle cartésien la MTF est la meilleure au centre de l'image et varie très peu avec la distance au centre. La résolution de l'image est différente selon le modèle, elle est plutôt homogène dans le cas du modèle cartésien alors que dans le cas du modèle polaire la résolution est meilleure en périphérie qu'au centre de l'image. Avec l'estimateur utilisé, la taille variable des pixels de la représentation polaire rend la résolution de l'image inhomogène.

On a donc fait une deuxième expérience en prenant les mêmes paramètres que précédemment avec un sinogramme non bruité mais en diminuant λ à 0.1, 0.01 et 0.001 afin de diminuer l'effet de la régularisation de l'estimateur utilisé donc le poids donné à l'information connue *a priori*.

Les résultats sont présentés sous forme de graphiques représentant la moyenne des MTF calculées pour les huit lésions à la même distance au centre pour les cinq distances choisies : 4.22, 4.23 et 4.24 pour le modèle polaire et 4.25, 4.26 et 4.27 pour le modèle cartésien.

On observe qu'avec un λ très faible ce qui équivaut à une absence de régularisation, la MTF pour le modèle polaire varie beaucoup avec la distance au centre contrairement au cartésien où la MTF est la même quelle que soit cette distance. Ce premier résultat met en évidence l'inhomogénéité de la qualité de l'image avec l'introduction de la représentation cylindrique du modèle.

La deuxième observation est que, avec la représentation cylindrique, plus on augmente le poids de la régularisation λ , plus la MTF est mauvaise et plus la diminution de la qualité de la MTF avec la distance au centre est marquée contrairement au modèle cartésien où ça n'a aucune influence. On en déduit que l'inhomogénéité de la résolution de l'image dépend fortement de la régularisation. En effet avec la fonction de régularisation utilisée, plus on s'éloigne du centre, moins l'image est lissée et donc meilleure est la résolution. La fonction de régularisation utilisée est donc mieux adaptée pour les grands pixels périphériques que pour les petits pixels centraux en terme de résolution de l'image reconstruite. Il est donc nécessaire de faire dépendre la régularisation de la distance au centre et donc de la taille des voxels.

Lors du choix de la fonction de régularisation pour le modèle polaire, on a utilisé une formulation purement discrète du passage du cartésien au cylindrique en négligeant le jacobien du passage des coordonnées cartésiennes aux coordonnées polaires. Les travaux complémentaires de Morzadec (2014) montrent que l'inclusion du jacobien et donc d'une pondération radiale dans la fonction de régularisation permet de rendre la résolution des images recons-

truites avec la représentation cylindrique uniforme.

Mesure de la MTF pour la représentation
cylindrique sans bruit

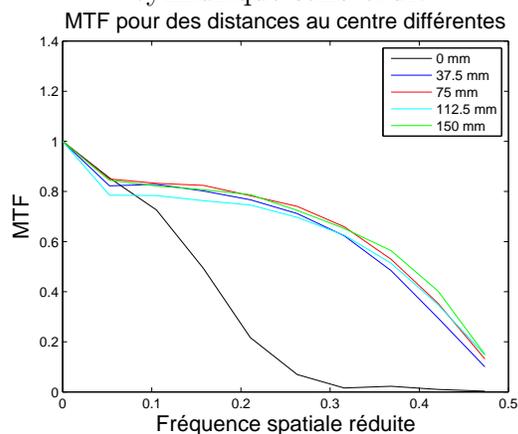


Figure 4.22 $\lambda = 0.1$

Mesure de la MTF pour la représentation
cartésienne sans bruit

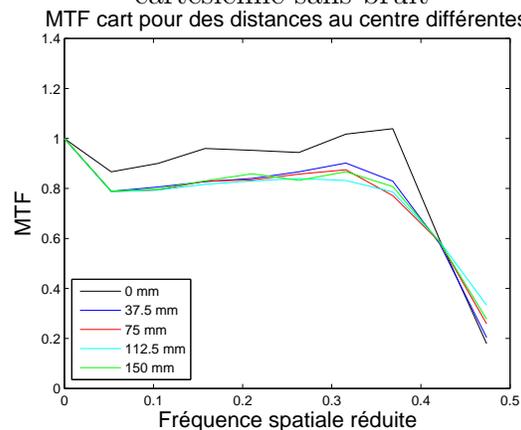


Figure 4.25 $\lambda = 0.1$

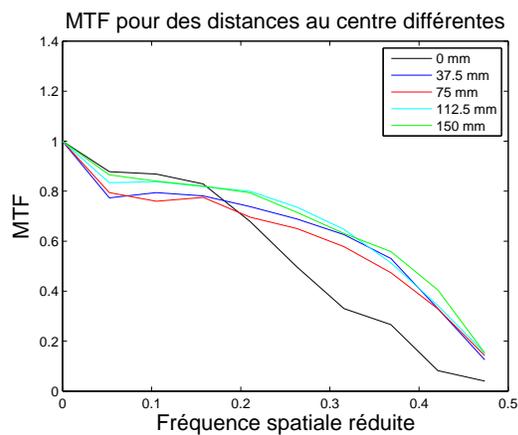


Figure 4.23 $\lambda = 0.01$

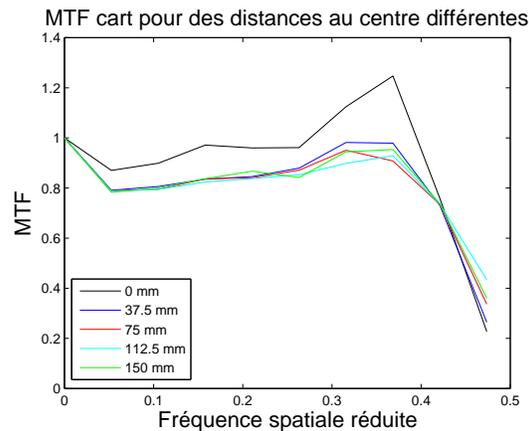


Figure 4.26 $\lambda = 0.01$

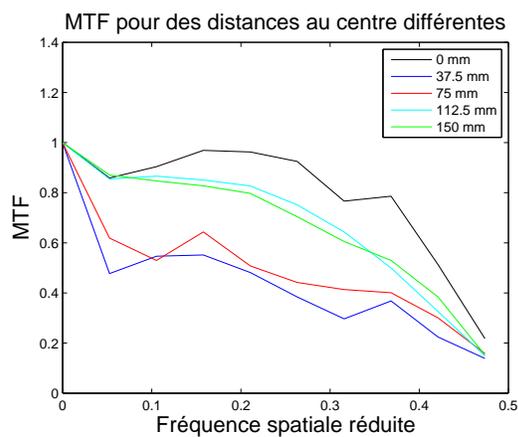


Figure 4.24 $\lambda = 0.001$

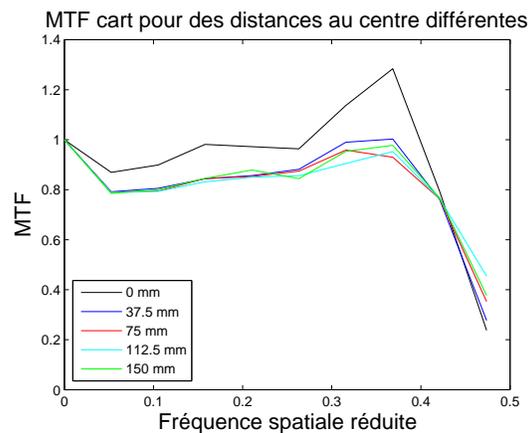


Figure 4.27 $\lambda = 0.001$

4.8.2.2 Mesure de la RMSE locale

La RMSE a été mesurée en utilisant les mêmes paramètres expérimentaux, les mêmes fenêtres de 50×50 pixels et le même fantôme mais sans impulsion que pour la MTF.

On rappelle qu'empiriquement, on a observé que les changements effectués sur l'estimée après 100 itérations sont indiscernables lors de l'observation des images reconstruites que ce soit avec le modèle polaire ou cartésien, c'est pourquoi, l'estimée utilisée est pour le modèle cartésien le résultat de l'estimation au bout de 100 itérations. Pour le modèle polaire, on utilise soit la solution exacte calculée avec l'inverse de la matrice normale du système, soit, le résultat de l'estimation au bout de 100 itérations sans préconditionneur afin d'observer l'impact des problèmes de convergence sur le niveau de bruit de l'image reconstruite.

Les résultats sont présentés sous forme de graphiques représentant la moyenne des RMSE calculées pour les huit fenêtres à la même distance au centre pour les cinq distances choisies pour un SNR de 25 dB (4.28) et pour un SNR de 35 dB (4.29).

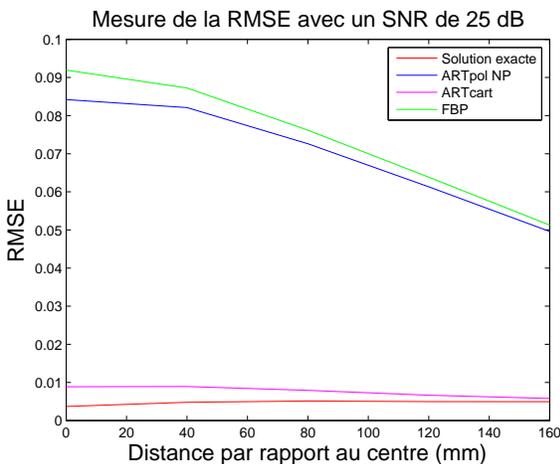


Figure 4.28 RMSE, SNR = 25 dB

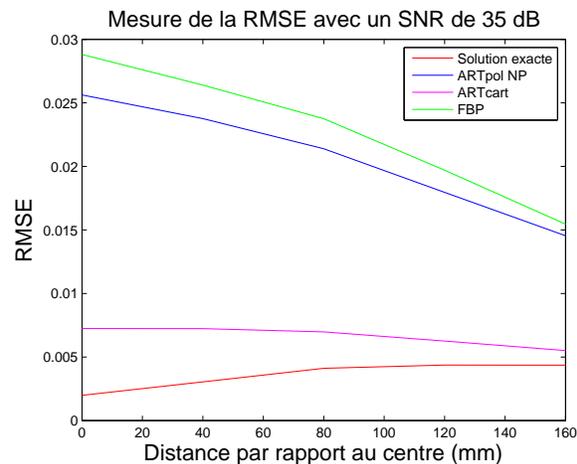


Figure 4.29 RMSE, SNR = 35 dB

Solution Exacte : mesure avec la solution exacte du système avec le modèle polaire ; ARTpol NP : mesure avec l'estimée à la 100e itération, modèle polaire sans préconditionnement ; ARTcart : mesure avec l'estimée à la 100e itération, modèle cartésien ; FBP : mesure avec le résultat de la rétroprojection filtrée

On observe dans le cas du modèle polaire sans préconditionnement, une RMSE presque aussi élevée que celle de la reconstruction par FBP et qui diminue avec la distance au centre. On peut dans un premier temps penser à un choix de coefficient de régularisation inadapté qui abouti a un mauvais compromis résolution/bruit résiduel. Dans le cas de la solution exacte avec la représentation cylindrique, la RMSE est beaucoup plus faible que dans le cas précédent, le choix du coefficient de régularisation n'est donc pas la cause de cette RMSE

élevée sans préconditionnement. Cette observation illustre les problèmes de convergence du modèle au centre de l'image dûs à la taille variable des pixels qui altère le conditionnement du système. Sans préconditionnement, la méthode algébrique avec le modèle polaire ne permet pas d'avoir une réduction significative du bruit par rapport à la méthode analytique.

Dans le cadre de la solution exacte, la RMSE augmente avec la distance au centre, c'est à dire, la capacité à distinguer de faibles contrastes diminue avec la distance au centre.

Concernant le modèle cartésien, la RMSE a tendance à diminuer avec la distance au centre, le contraste est un peu moins élevé au centre qu'en périphérie de l'image. Ces tendances ne sont pas modifiées avec la modification du SNR.

Avec l'estimateur utilisé, la taille variable des pixels de la représentation cylindrique rend la réduction du bruit non-uniforme. En effet, contrairement à la résolution, on a une meilleure réduction du niveau de bruit au centre de l'image avec de petits pixels qu'en périphérie avec de grands pixels, alors qu'avec le modèle cartésien la différence est beaucoup moins marquée. De la même manière que pour la résolution, il est ici aussi important de faire varier la pénalisation en fonction de la taille des pixels, les travaux de Morzadec (2014) montrent que l'ajout du jacobien du passage des coordonnées cartésiennes aux coordonnées polaires dans la fonction de régularisation modifie ce comportement. Avec cette pondération radiale, on a une meilleure réduction de bruit globale et la réduction de bruit sera plus importante en périphérie qu'au centre.

4.8.2.3 Conclusion sur la mesure de qualité d'image

Lorsque l'on s'intéresse à la qualité d'image globale, on observe qu'avec la représentation cylindrique et l'utilisation d'un préconditionneur, on obtient une meilleure RMSE globale qu'avec la représentation cartésienne. On peut alors supposer que la qualité de la reconstruction est meilleure avec le modèle polaire préconditionné. Cette hypothèse est à vérifier avec une étude précise des paramètres d'estimation et de représentation afin d'avoir la meilleure qualité d'image dans les deux cas et d'éviter toute erreur de représentation.

L'étude de la qualité locale des reconstructions avec les deux types de représentation permet de mettre en évidence une différence très importante entre le modèle cartésien et polaire : la qualité de l'image avec le modèle cartésien est homogène contrairement au modèle polaire où la qualité d'image varie beaucoup suivant la distance au centre.

Concernant la représentation cylindrique, l'utilisation d'une grille inhomogène de pixels nécessite un préconditionnement pour obtenir une qualité d'image fiable mais inhomogène. En effet, la mesure de la MTF permet d'observer une meilleure résolution en périphérie qu'au centre de l'image. Les contours des formes imagées sont inégalement lissés. Au contraire, la mesure de la RMSE nous montre que la réduction du bruit est plus importante au centre

qu'en périphérie. On a utilisé pour cette étude une fonction de régularisation L2 négligeant le jacobien de la transformation cartésien polaire. L'inclusion de ce jacobien comme l'a fait Morzadec (2014) permet d'obtenir une qualité d'image beaucoup plus homogène.

4.8.3 Amélioration de la qualité des images avec l'ART polaire : réduction des artéfacts métalliques

Ces expériences ont deux objectifs :

- vérifier que le comportement qualitatif de l'approche de réduction d'artéfacts métalliques avec la représentation cartésienne est bien conservé avec l'approche avec la représentation cylindrique lors de la reconstruction de données réelles avec l'utilisation du modèle polychromatique ;
- tester qualitativement la possibilité de reconstruction d'une image en 3D à partir de données réelles.

Pour cela on utilise des données réelles pré-existantes résultant du scan d'un fantôme contenant des billes de fer par un scanner SIEMENS Somatom utilisant le protocole expérimental InnerEar. Ce dernier est un protocole haute résolution de 1160 projections aboutissant à des données de SNR d'environ 35 dB . Afin de diminuer le volume des données, on ne considère que 2 des 8 barrettes de 672 détecteurs pour effectuer la reconstruction.

On choisi des paramètres réaliste de l'objet, c'est à dire deux tranches 512×512 pixels de 0.5 mm soit avec la représentation cylindrique un nombre de pixels radiaux (dans la direction ρ) égal à 226 afin de couvrir l'ensemble du fantôme et garder une espace mémoire requis assez bas pour que l'estimation soit calculable avec nos moyens.

L'estimation est faite avec l'algorithme L-BFGS-B en utilisant la représentation cylindrique et la méthode décrite dans 3.4.4 sans préconditionnement ou le modèle monochromatique.

On modélise la source en utilisant le spectre d'émission donné par Siemens pour ce type de scanner. L'intensité de la source de rayon X y est échantillonnée entre 20 et 120 keV tout les keV. Les fonctions $\phi = f(\mu_0)$ et $\theta = f(\mu_0)$ sont définies respectivement par une hyperbole verticale et une hyperbole horizontale dont les paramètres sont calculés pour modéliser le mieux possible les mesures des tables de De Man *et al.* (2001). L'énergie de référence est définie à 70 keV qui est la valeur la plus communément utilisée pour les reconstructions monochromatiques et correspond à un pic de haute énergie.

On utilise une fonction de régularisation L2L1 plus complexe. Ceci permet de tester la possibilité d'utiliser un autre type de fonction de régularisation que L2 avec le modèle cylindrique. Les paramètres sont déterminés empiriquement afin d'avoir la meilleure réduction d'artéfacts possible sur l'image reconstruite.

L'absence de préconditionneur développé pour le modèle polychromatique nous empêche de résoudre le problème de convergence de l'algorithme. Afin de contourner ce problème, on prend comme image de départ pour l'estimation le résultat de la FBP. On s'est aperçu empiriquement qu'avec cette image de départ, les problèmes de convergence n'apparaissent pas sur l'image reconstruite.

Les résultats des reconstructions sont représentés sous forme d'image avec la FBP (Fig. 4.30), le modèle polaire monochromatique (Fig. 4.31) et le modèle polaire polychromatique (Fig. 4.32).

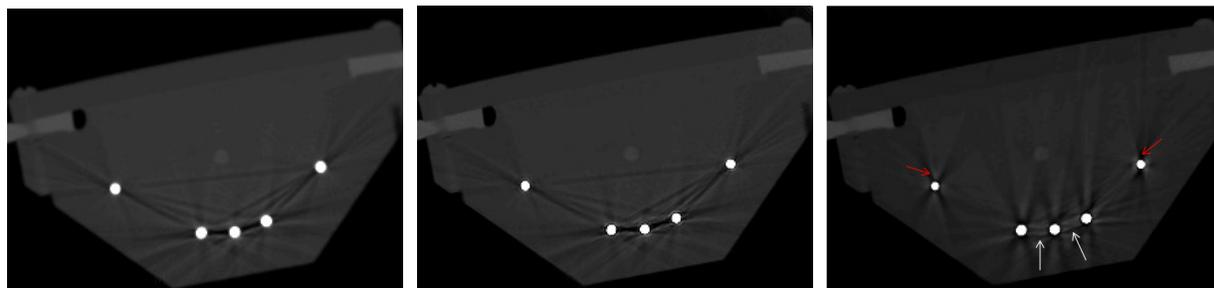


Figure 4.30 rétroprojection filtrée Figure 4.31 modèle monochromatique Figure 4.32 modèle polychromatique

Si l'on compare la reconstruction avec le modèle monochromatique ou la FBP et la reconstruction avec le modèle polychromatique représentées avec la même échelle de gris, on observe une réduction des artéfacts métalliques entre les billes de fer indiqués par les flèches blanches et les contours des billes semblent plus nets, cette deuxième hypothèse nécessite la représentation du profil 1D des différentes images reconstruites afin d'être validée mais ces graphiques n'ont pas pu être construits faute de temps. En revanche, il y a une accentuation d'autres artéfacts partant des billes et indiqués par les flèches rouges qui peut entraîner une détérioration de certaines parties de l'image. Ces résultats sont cohérents avec ceux obtenus par Hamelin (2009) avec la méthode utilisant la représentation cartésienne qu'on a adaptée à notre modèle polaire. Le modèle polychromatique permet donc une diminution globale des artéfacts métalliques et par conséquent, une augmentation de la qualité globale de la reconstruction. L'inconvénient de ce modèle est sa complexité qui va augmenter le temps de calcul et l'espace mémoire supplémentaire requis, c'est pourquoi on limite son utilisation à la reconstruction d'images d'objets contenant des matériaux de coefficient d'absorption forts, comme les métaux, provoquant des artéfacts de reconstruction.

Cette expérience a permis de montrer que le modèle en coordonnées cylindrique permet de faire des reconstructions d'images à partir de données cliniques réelles de petite taille et

est compatible avec des méthodes d'augmentation de qualité d'image développées pour la représentation cartésienne de l'objet telle la modélisation polychromatique de la source.

CHAPITRE 5

CONCLUSION

5.1 Synthèse des travaux

L'objectif de ce projet était d'optimiser les performances, espace mémoire utilisé et vitesse de calcul d'une technique de reconstruction algébrique itérative en tomographie à rayons X afin de pouvoir reconstruire efficacement des images à partir de données cliniques.

Afin de diminuer l'espace mémoire requis, on a utilisé une représentation de l'objet en coordonnées cylindriques, ce qui introduit une grande redondance des coefficients de la matrice de projection. En jouant sur les symétries qui en résultent, en stockant la matrice sous forme de matrice creuse et en usant de l'approche par rotation qui consiste à considérer la rotation de l'objet et non de la source, on a pu simplifier la matrice de projection à une projection et donc réduire drastiquement la taille de celle-ci en mémoire.

Cette représentation particulière a certaines propriétés intéressantes et d'autres gênantes. En effet, outre la réduction importante de l'espace mémoire requis, la représentation cylindrique de l'objet a comme avantage de rendre la matrice de projection bloc-circulante. Cette structure remarquable de la matrice de projection amène des possibilités de parallélisation simple des opérations de projection et rétroprojection et des possibilités de calcul du hessien sous certaines hypothèses, ce qui permet le calcul de la solution exacte. A contrario, l'inconvénient de cette représentation est la forme inhabituelle et la taille variable des voxels. Ces propriétés provoquent des problèmes de pondération de la régularisation et de convergence de l'algorithme d'optimisation. Afin de surmonter cette dernière difficulté, il est primordial de développer des préconditionneurs efficaces pour la résolution de ce système.

Pour étudier tous ces éléments, on s'est placé dans un cadre très simple. On a choisi un estimateur MAP et une fonction de régularisation L2 qui permet d'obtenir un hessien constant de la matrice normale du système. On a modélisé le bruit des données par un bruit blanc gaussien, ce qui a permis de ne pas appliquer de pondération sur le terme des moindres carrés du problème à résoudre. Ce cadre restrictif a comme avantage de rendre la matrice normale du système et son hessien bloc-circulants et de permettre le calcul de la solution exacte.

La parallélisation est un moyen simple et efficace d'amélioration de la vitesse de calcul de chaque itération. En utilisant la propriété bloc-circulante et la petite taille de la matrice

de projection avec la représentation cylindrique, on a pu diviser les deux opérations les plus coûteuses de chaque itération : la projection et la rétroprojection, en N opérations indépendantes, où N est le nombre total de projections. Avec la mise en oeuvre développée et les moyens à notre disposition, la parallélisation de ces opérations a permis de réduire le temps de calcul d'un facteur 4 et d'avoir une vitesse de calcul plus rapide que celle obtenue avec un ART cartésien parallélisé en utilisant les symétries. Ce résultat montre qu'il est possible d'augmenter significativement la vitesse de calcul de la méthode développée de manière très simple.

L'inconvénient majeur de cette représentation de l'objet est la forme inhabituelle et la taille variable des voxels.

Ces propriétés des voxels de la représentation cylindrique détériorent le conditionnement du problème, ce qui entraîne des problèmes de convergence de l'algorithme d'estimation. Afin de rendre possible l'utilisation de notre modèle, on a donc dû améliorer le conditionnement du problème en développant des préconditionneurs. Pour cela, on a utilisé les propriétés du hessien de la matrice normale du système qui est constant et bloc circulant, afin d'établir des préconditionneurs simples adaptés au problème. On a bloc-diagonalisé le hessien dans le domaine de Fourier. En utilisant la propriété diagonale dominante de chaque bloc de la diagonale de cette matrice, on a ensuite construit deux préconditionneurs dans le domaine de Fourier : un préconditionneur diagonal et un préconditionneur de Cholesky incomplet. On a pu ainsi améliorer le conditionnement du problème et la vitesse de convergence de l'estimation. Malgré une vitesse de convergence en nombre d'itérations plus rapide avec le préconditionneur de Cholesky incomplet, le préconditionneur diagonal a le meilleur compromis entre l'augmentation du volume de calculs, de la vitesse de convergence et de l'espace mémoire utilisé. On a donc développé avec le préconditionneur diagonal, un moyen rapide, efficace et n'augmentant que très peu l'espace mémoire requis, pour diminuer le temps de calcul de l'estimation et surmonter les problèmes de convergence.

Outre la détérioration du conditionnement du problème, la taille variable des voxels de la représentation cylindrique entraîne aussi des difficultés de régularisation notamment dans la pondération des différences entre pixels. Ceci rend la qualité de l'image reconstruite inhomogène. On a donc étudié la qualité des images obtenues afin de connaître l'impact réel de la représentation cylindrique sur la qualité de la reconstruction. La distance au centre étant proportionnelle à la taille des voxels, l'évaluation de la qualité des images reconstruites avec notre approche a été faite suivant deux critères : le bruit résiduel et la résolution, en faisant varier cette distance. On a observé que l'utilisation de notre modèle entraîne des variations de la qualité de l'image en fonction de la distance au centre. En effet, à la différence de la représentation cartésienne, avec la représentation cylindrique, le bruit résiduel et la résolu-

tion augmentent avec la distance au centre. Cette augmentation est liée à une pondération inadaptée de la pénalisation des différences entre voxels voisins qu'on a choisi homogène dans les directions θ et z et dépendante de la distance au centre dans la direction ρ . Avec la représentation cylindrique, On a donc une image de qualité inhomogène contrairement à un ART classique avec une représentation cartésienne de l'objet. Ce problème peut être surmonté en incluant le jacobien de la transformation coordonnées cartésiennes polaires dans la fonction de régularisation.

On a terminé notre étude par une expérience visant à tester la possibilité de réduire les artéfacts métalliques lors de la reconstruction d'images à partir de données réelles en utilisant la représentation cylindrique. On a donc adapté la modélisation polychromatique de la source aux coordonnées cylindrique. On a pu observer que la reconstruction avec notre modèle avait le même comportement concernant la réduction des artéfacts que le modèle cartésien classique.

On a montré que le modèle en coordonnées cylindrique permet de faire des reconstructions d'images à partir de données cliniques réelles et est compatible avec des méthodes d'augmentation de qualité d'image développées pour la représentation cartésienne de l'objet comme le modèle polychromatique de la source.

En conclusion, le modèle développé a permis de remplir les objectifs initiaux qui étaient de diminuer l'espace mémoire, d'augmenter la vitesse de calcul requis à l'utilisation d'un ART et permettre la reconstruction d'images à partir de données cliniques mais aux dépends de l'homogénéité et du conditionnement du problème.

5.2 Limitations de la solution proposée

Dans ce projet, on a vu qu'il était possible de reconstruire des images à partir de données réelles en un temps acceptable mais l'algorithme reste plus lent et demande plus d'espace mémoire que les algorithmes analytiques.

Le cadre utilisé est très restrictif. La représentation cylindrique est pensée pour une reconstruction tomographique axiale et n'est pas adaptée à une reconstruction hélicoïdale. De plus, on a modélisé le bruit avec une variance uniforme et on a utilisé une fonction de régularisation L2 afin de faciliter le préconditionnement. Ceci a comme inconvénient de ne pas prendre en compte la dépendance de la variance du bruit à la valeur de l'atténuation lors de l'assimilation de la distribution poissonnienne du bruit à une distribution gaussienne.

Concernant la qualité des images reconstruites, on a une sur-régularisation au centre et

une sous-régularisation en périphérie qui nuit à la qualité globale de l'image et dépendent du poids du terme de régularisation, dépendance qui n'a pas été étudiée. En outre, l'image reconstruite en coordonnées cylindriques doit être interpolée en coordonnées cartésienne, cette opération peut détériorer la qualité de l'image finale.

Finalement, l'impact de la modélisation polychromatique de la source reste à être testée, la qualité des images en résultant n'étant pas étudiée et les préconditionneurs n'étant pas bien adaptés à ce modèle, l'application de méthodes d'augmentation de qualité d'image développées pour la représentation cartésienne est possible mais nécessite des adaptations qui peuvent être contraignantes.

5.3 Améliorations futures

Les résultats encourageants avec notre modèle poussent à persévérer dans cette direction et à trouver des solutions aux problèmes de rapidité d'exécution et de qualité d'image afin de rendre utilisable la méthode développée pour la reconstruction d'image en tomographie provenant de données bruitées ou sujettes à la création d'artéfacts où les avantages d'un algorithme itératif rapide est un avantage.

En gardant le cadre restrictif qu'on a établi, il reste des améliorations à faire concernant la parallélisation et la qualité de l'image reconstruite.

Il serait possible d'avoir une meilleure parallélisation en utilisant le *Cloud computing* ou la parallélisation sur GPU, d'augmenter la vitesse de calcul en implémentant l'algorithme entier en langage compilé tel le C++. De plus, une étude précise de l'impact du nombre de blocs et de la taille de chaque bloc sur la diminution de la vitesse de calcul avec la parallélisation serait utile.

Concernant la qualité inégale des images, l'objectif serait de développer une fonction de régularisation permettant d'homogénéiser la qualité des reconstructions. De plus, une étude rigoureuse de la qualité des reconstructions et du préconditionnement avec le modèle polychromatique pourrait être envisagée afin de connaître précisément les performances de ce type de modélisation.

Il faudrait enfin étudier plus précisément l'effet du relâchement des hypothèses utilisées dans notre projet, en particulier l'utilisation de fonctions différentes de pénalisation et la pondération du terme de moindres carrés.

En effet, si on utilise la fonction de pénalisation L2L1 qui est la fonction la plus couramment utilisée au lieu de L2, le hessien n'est alors plus constant. Il devient alors nécessaire de

l'approximer par une matrice constante bloc-circulante afin de rendre le préconditionnement possible. L'essai d'un préconditionneur dérivé du hessien avec une solution uniforme nous a permis de surmonter les problèmes de convergence, même si la vitesse de convergence reste plus faible qu'avec L2. Il reste à faire une étude rigoureuse de cette approche.

De même, si on pondère le terme de moindres carrés en utilisant une modélisation poissonienne du bruit avec une variance Σ_b non-uniforme dépendant de l'atténuation, le hessien n'est alors plus bloc-circulant. Ici aussi, il faut définir une approximation bloc-circulante du hessien pour le calcul de préconditionneurs. Ceci peut être fait avec une approximation circulante du sinogramme afin d'uniformiser la pondération du terme des moindres carrés du problème.

RÉFÉRENCES

- BERGNER, F., BRENDEL, B., NOEL, P. B., DOBRITZ, M. et KOEHLER, T. (2012). Robust automated regularization factor selection for statistical reconstructions. *2nd International Conference on Image Formation in X-Ray Computed Tomography*. Salt Lake City, USA.
- BOUMAN, C. A. et SAUER, K. D. (1993). A generalized Gaussian image model for edge-preserving MAP estimation. *IEEE Trans. Image Process.*, IP-2, 296–310.
- CHAMPLEY, K. M. et MARTZ, H. E. (2013). Statistical-analytic regularized reconstruction for X-ray CT. *Proc. Int. Meeting on Fully 3D Image Reconstr. in Rad. and Nucl. Med.* Lake Tahoe, CA, 173–176.
- CHARBONNIER, P., BLANC-FÉRAUD, L., AUBERT, G. et BARLAUD, M. (1997). Deterministic edge-preserving regularization in computed imaging. *IEEE Trans. Image Process.*, IP-6, 298–311.
- CHEN, K. (2005). *Matrix preconditioning techniques and applications*, vol. 19 de *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press.
- DE MAN, B. et BASU, S. (2002). Distance-driven projection and backprojection. *IEEE Nucl. Sci. Symp. Conf. Record.* IEEE, vol. 3, 1477–1480.
- DE MAN, B. et BASU, S. (2004). Distance-driven projection and backprojection in three dimension. *Phys. Med. Biol.*, 49, 2463–2475.
- DE MAN, B., NUYTS, J., DUPONT, P. et SUETENS, P. (2001). An iterative maximum-likelihood polychromatic algorithm for CT. *IEEE Trans. Medical Imaging*, 20, 999–1008.
- DEMPSTER, A. P., LAIRD, N. M. et RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. B*, 39, 1–38.
- ELBAKRI, I. A. et FESSLER, J. A. (2002). Statistical image reconstruction for polyenergetic X-ray computed tomography. *IEEE Trans. Medical Imaging*, 21, 89–99.
- ERDOĞAN, H. et FESSLER, J. A. (1999). Ordered subsets algorithms for transmission tomography. *Phys. Med. Biol.*, 44, 2835–2851.
- FESSLER, J. A. (1996). Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood) : Applications to tomography. *IEEE Trans. Image Process.*, 5, 493–506.
- FESSLER, J. A. (2000). Statistical image reconstruction methods for transmission tomography. J. M. Fritzpatrick et M. Sonka, éditeurs, *Handbook of Medical Imaging*, SPIE Press, Bellingham, WA, vol. 2, chapitre 1. 1–70.

- GENDRON, D. (2008). *Méthodes Algébriques pour la Reconstruction en Tomographie à Rayons X*. Mémoire de maîtrise, École Polytechnique de Montréal.
- GOLDMAN, L. W. (2007). Principles of CT and CT technology. *J Nucl. Med. Technol.*, 35, 115–128.
- GOLKAR, M. (2013). *Fast iterative reconstruction in X-ray tomography using polar coordinates*. Thèse de doctorat, École Polytechnique de Montréal.
- GOUSSARD, Y., GOLKAR, M., WAGNER, A. et VOORONS, M. (2013). Cylindrical coordinate representation for statistical 3D CT reconstruction. *Proc. Int. Meeting on Fully 3D Image Reconstr. in Rad. and Nucl. Med.* Lake Tahoe, CA, 138–141.
- GUAN, H. et GORDON, R. (1996). Computed tomography using algebraic reconstruction techniques (ARTs) with different projection access schemes : a comparison study under practical situations. *Phys. Med. Biol.*, 41, 1727–1743.
- HAASE, G., LIEBMANN, M. et PLANK, G. (2007). A Hilbert-order multiplication scheme for unstructured sparse matrices. *Int. J. Parallel Emerg. Distrib. Syst.*, 22, 213–220.
- HAMELIN, B. (2009). *Accélération d'une Approche Régularisée de Reconstruction en Tomographie à Rayons X avec Réduction des Artéfacts Métalliques*. Thèse de doctorat, École Polytechnique de Montréal.
- HAMELIN, B., GOUSSARD, Y., DUSSAULT, J.-P., SOULEZ, G., BEAUDOIN, G. et CLOUTIER, G. (2011). Performance comparison on nonlinear solvers for statistical reconstruction in transmission tomography. Submitted to *IEEE Trans. Medical Imaging*, 1st rev.
- HAMELIN, B., GOUSSARD, Y., GENDRON, D., DUSSAULT, J.-P., CLOUTIER, G., BEAUDOIN, G. et SOULEZ, G. (2008). Iterative CT reconstruction of real data with metal artifact reduction. *Proc. IEEE ISBI*. 1453–1456.
- HUDSON, H. M. et LARKIN, R. S. (1994). Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans. Medical Imaging*, 13, 601–609.
- IDIER, J., éditeur (2001). *Approche Bayésienne pour les Problèmes Inverses*. Traitement du Signal et de l'Image. Hermes Science – Lavoisier, Paris, France.
- JIAN, L., LITAO, L., PENG, C., QI, S. et ZHIFANG, W. (2007). Rotating polar-coordinate ART applied in industrial CT image reconstruction. *NDT & E International*, 40, 333–336.
- KAK, A. C. et SLANEY, M. (1987). *Principles of Computerized Tomographic Imaging*. IEEE Press, New York, NY.
- KOSTER, J. (2002). *Parallel Templates for Numerical Linear Algebra, a High-Performance Computation Library*. Mémoire de maîtrise, Utrecht University, The Netherlands.

- MA, J., LIANG, Z., FAN, Y., LIU, Y., HUANG, J., CHEN, W. et LU, H. (2012). Variance analysis of X-ray CT sinograms in the presence of electronic noise background. *Med. Phys.*, 39, 4051–4065.
- MACOVSKI, A. (1983). *Medical Imaging Systems*. Prentice-Hall information and system sciences series. Prentice-Hall, Englewood Cliffs, NJ.
- MENVIELLE, N. (2004). *Réduction des Artéfacts Métalliques en Tomographie à Rayons X*. Mémoire de maîtrise, École Polytechnique, Montréal, Canada.
- MORZADEC, A. (2014). Uniformisation de la qualité des images de reconstruction tomographique en utilisant les coordonnées cylindriques. *Rapport de stage, Polytechnique Montréal*.
- NOCEDAL, J. et WRIGHT, S. J. (1999). *Numerical Optimization*. Operations Research. Springer Verlag, New York, NY.
- RODRIGUEZ-ALVAREZ, M.-J., SANCHEZ, F., SORIANO, A., IBORRA, A. et MORA, C. (2011). Exploiting symmetries for weight matrix design in CT imaging. *Math. Comput. Model.*, 54, 1655–1664.
- ROSSMAN, K. (1969). Point spread-function, line spread-function, and modulation transfer function. *Radiology*, 93, 257–272.
- SAUER, K. D. et BOUMAN, C. A. (1993). A local update strategy for iterative reconstruction from projections. *IEEE Trans. Signal Process.*, SP-41, 534–548.
- SEGARS, W. P., MAHESH, M., BECK, T. J., FREY, E. C. et TSUI, B. M. W. (2008). Realistic CT simulation using the 4D XCAT phantom. *Medical Physics*, 35, 3800–8.
- SHI, H. et FESSLER, J. A. (2009). Quadratic regularization design for 2d ct. *IEEE Trans. Medical Imaging*, 28, 645–656.
- SORIANO, A., RODRIGUEZ-ALVAREZ, M., IBORRA, A., SANCHEZ, F., CARLES, M., CONDE, P., GONZALEZ, A., HERNANDEZ, L., MOLINER, L., ORERO, A., VIDALA, L. et BENLLOCHA, J. (2012). EM tomographic image reconstruction using polar voxels. *Int. Worksh. Rad. Imag. Detectors*. Portugal.
- STAUDE, A. et GOEBBELS, J. (2011). Determining the spatial resolution in computed tomography - Comparison of MTF and line-pair structures. *International Symposium on Digital Industrial Radiology and Computed Tomography*.
- THIBAudeau, C., LEROUX, J.-D., FONTAINE, R. et LECOMTE, R. (2013). Fully 3D iterative CT reconstruction using polar coordinates. *Med. Phys.*, 40, 111904.
- WANG, A. S., WEBSTER STAYMAN, J., OTAKE, Y., KLEINSZIG, G., VOGT, S., KHANNA, A. J., GOKASLAN, Z. L. et SIEWERDSEN, J. H. (2013). Statistical recons-

truction for soft tissue imaging with low dose C-arm cone beam CT. *Proc. Int. Meeting on Fully 3D Image Reconstr. in Rad. and Nucl. Med.* Lake Tahoe, CA, 201–204.

YZELMAN, A. N. et BISSELING, R. H. (2011). Two-dimensional cache-oblivious sparse matrix-vector multiplication. *Parallel Computing*, 37, 806–819.

ZENG, G. L. et GULLBERG, G. T. (1992). Frequency domain implementation of the three-dimensional geometric point response correction in SPECT imaging. *IEEE Trans. Nucl. Sci.*, 39, 1444–1453.

ZENG, G. L., HSIEH, Y.-L. et GULLBERG, G. T. (1994). A rotating and warping projector/backprojector for fan-beam and cone-beam iterative algorithm. *IEEE Trans. Nucl. Sci.*, 41, 2807–2811.

ZIEGLER, A., KÖHLER, T. et PROKSA, R. (2006). Noise and resolution in images reconstructed with FBP and OSC algorithms for CT. *Med. Phys.*, 34, 585–598.