université PARIS-SACLAY

CentraleSupélec

UNIVERSITÉ PARIS SUD
Comprendre le monde, construire l'avenir

NNT : 2017SACLE362

Fast and Accurate 3D X-ray Image Reconstruction for Non Destructive Testing Industrial Applications

Thèse de doctorat de l'Université Paris-Saclay
préparée au Laboratoire des Signaux et Systèmes (L2S)

École doctorale No. 580 : Sciences et technologie de l'information et de la communication
Spécialité de doctorat : traitement du signal et des images

Thèse présentée et soutenue à Gif-sur-Yvette, le 01/12/2017, par

Li Wang

Composition du Jury :

Bruno Sixou, MCF (HDR), INSA Lyon Laboratoire CREATIS	Rapporteur
Jan Sijbers, Professeur, Université Antwerp Department Physique	Rapporteur
Udo von Toussaint, DR Institut de Max-Planck	Examinateur
Thomas Rodet, Professeur, ENS Cachan Laboratoire SATIE	Président
Ali Mohammad-Djafari, DR, CNRS Laboratoire des Signaux et Systèmes (L2S)	Directeur de thèse
Nicolas Gac, MCF, Université Paris-Sud Laboratoire des Signaux et Systèmes (L2S)	Co-Encadrant de thèse

Thèse de doctorat

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France

Acknowledgement

How time flies ! It has been more than six years since I came to France ! After one year of college, two years of master and three years of Ph.D., the little ”Orsay Ville” has become my second hometown.

The three years of Ph.D. finished in the blink of an eye. Still, I’ve learned a lot from my supervisors and my colleagues and have met a lot of friends in the warm family of ”Laboratoire des Signaux et Système”. I would like to express my deepest gratitude and appreciation to all the members of the big L2S family and all my friends.

First and foremost I would like to express my deep gratitude to my advisors Prof. Ali MOHAMMAD-DJAFARI and Dr. Nicolas GAC. It has been an honor to be one of their Ph.D. students. I want to firstly thank Prof. DJAFARI for his continuous support and patient guidance, for leading me to the charm and profound Bayesian methods. I want to gratefully thank Dr. GAC, for his kindness to always try his best to help me solve the implementation problems and opens the door to the ”GPU world” for me.

I would also like to thank the members of the Groupe Problème Inverse (GPIers). I am grateful to Dr. Aurelia Fraysse and Prof. Charles Soussen for their suggestions and help for my presentation of my Ph.D. defense. I would like to thank all the GPIers : Dr. François Orieux, Dr. Mathieu Kowalski, Dr. Mircea Dumitru, Camille Chapdelaine, Amine Hadjyoucef, Alina Meresescu, Dr. Fangchen Feng, Guillaume Revillon, Thi Thanh Nguyen, Maxime Martelli and the ancient GPIers who have already left but will be members of the GPI family forever : Youssouf Sameut, Dr. Thomas Boulay, Dr. Long Chen, Dr. Yuling Zheng and Dr. Leila Gharshalli, Dr. Caifang Cai, for their friendship and help, and a lot of interesting and helpful discussions during these years.

I would like to express my deep gratitude for the reviewers of my thesis, Prof. Jan Sijbers and Dr. Bruno Sixou and the examiners of my Ph.D. defense, Prof. Thomas Todet and Prof. Udo von Toussaint. It was a great honor to have them as jury members and I am deeply grateful for their suggestions for my work and the time they invested for reading my manuscript and for coming to my defense.

My sincere thanks also goes to all my friends and colleagues from CentraleSupélec who are very kind with me and take cares of me all the time : Maryvonne Giron, Frédéric Desprez, Céline Labrude, Laurence Antunes, Delphine Maucherat, Dr. Patrice Brault, Dr. Giuseppe Valenzise, Prof. Dominique Lesselier, Prof. Gilles Duc, Prof. Michel Kieffer, Dr. Zicheng Liu, Yanqiao Hou, Chen Kang, Dr. Wenjie Li, Shanshan Wang, Jian Song, Xiaojun Xi, Dr. Chi Jin, Weichao Liang, Zhenyu Liao, Peipei Ran, Paulo Prezotti, Sara Berri, Dr. Zheng Chen, Dr. Chao He...

I also deeply thank all my dearest Chinese friends, who have been together with me for already ten years : Hui Hua, Dr. Jie Wei, Mengsu Guo, Yao Liu, Kai Wan, Dr. Yunsong Wang,

Siqi Wang, Dr. Xin Wang, Nan Guan, Dr. Xiang Liu. No matter where we are in the world I will always remember them forever and always.

I want to thank all those I know and all those who know me. . .

But the most important, I devote this thesis to my parents, Feng Wang and Xiuying Li, for their endless support and love ! At last, this thesis should be devoted to my husband, Dr. Chao Zhang for always being by my side whenever I'm happy or sad.

Great thanks to the China Scholarship Council (CSC), this work would not have been possible without their support.

Table des matières

Table des figures

Liste des tableaux

Table of abbreviations

ADMM	Alternating Direction Method of Multipliers
ART	Algebraic Reconstruction Technique
BP	Back-projection
CG	Conjugate Gradient
CT	Computed Tomography
DT-CWT	Dual-Tree Complex Wavelet Transformation
DWT	Discrete Wavelet Transformation
EM	Expectation-Maximization
FBP	Filtered Back-projection
FT	Fourier Transformation
GCV	Generalized Cross Validation
GD	Gradient Descent
GG	Generalized Gaussian
GPU	Graphics Processing Unit
HHBM	Hierarchical Haar based Bayesian Method
HT	Haar transformation
ISNR	Improvement of Signal to Noise Ratio
JMAP	Joint Maximum A Posterior
KL	Kullback–Leibler divergence
LASSO	Least Absolute Shrinkage and Selection Operator
LS	Least Square
ML	Maximum Likelihood
NDT	Non Destructive Testing
OS-EM	Ordered Subset Expectation Maximization
PET-CT	Positron Emission Tomography–Computed Tomography
pdf	probability density function
PM	Posterior Mean
PSNR	Peak Signal to Noise Ratio
QR	Quadratic Regularization
RMSE	Relative Mean Square Error
ROCC	Reconstruction of Object by Considering Contours
RT	Radon Transform
SART	Simultaneous Algebraic Reconstruction Technique
SEM	Stochastic Expectation Maximization
SIRT	Simultaneous Iterative Reconstruction Technique
SNR	Signal to Noise Ratio
SSIM	Structural SIMilarity
St-g	Generalized Student-t distribution
St-t	Student-t distribution
TV	Total Variation
VBA	Variational Bayesian Approach
VS	Variable Splitting

Notations :

$b(x, y)$	Back-projection result
$\mathcal{B}$	Back-projection operator
$\boldsymbol{D}$	Linear transformation operator
	Chap4 : Multilevel Haar transformation operator
$\boldsymbol{d}_1$	Convolution kernel for 1D derivation
$\boldsymbol{d}_2$	Convolution kernel for 2D derivation
$\boldsymbol{D}_1 \in \mathbb{R}^{M\times M}$	Convolution operator for 1D derivation
$\boldsymbol{D}_2 \in \mathbb{R}^{N\times N}$	Convolution operator for 2D derivation
$f(x, y)$	attenuation coefficient in a 2D image
$f(x, y, z)$	attenuation coefficient in a 3D object
$\widehat{f}$	Estimation of object
$\boldsymbol{f} \in \mathbb{R}^{N\times 1}$	The original object
f_j	the j-th voxel in object $\boldsymbol{f}$
$\widehat{\boldsymbol{f}}$	Estimation of ojbect
$\ddot{\boldsymbol{f}} \in \mathbb{R}^{N\times 1}$	Gradient of $\boldsymbol{f}$
$\boldsymbol{f}_{-r}$	Vector $\boldsymbol{f}$ except the element f_r
$F(u, v)$	2D Fourier transformation of $f(x, y)$
$\mathcal{F}_{1D}$	1D Fourier Transformation
$\mathcal{F}_{2D}$	2D Fourier Transformation
$\mathcal{F}_{1D}^{-1}$	1D inverse Fourier Transformation
$\mathcal{F}_{2D}^{-1}$	2D inverse Fourier Transformation
$g(r, \phi)$	the observed projection data
$\boldsymbol{g} \in \mathbb{R}^{M\times 1}$	The projection sinogram
g_i	The data observed by the i-th detector
$\boldsymbol{g}_0 \in \mathbb{R}^{M\times 1}$	Projection data without noise
$\ddot{\boldsymbol{g}} \in \mathbb{R}^{M\times 1}$	Gradient of projection $\boldsymbol{g}$
$G(\rho, \phi)$	1D Fourier Transformation of $g(r, \phi)$
$\mathcal{G}(\cdot)$	Gamma distribution
$g_0(n)$, $g_1(n)$	Low-pass/high-pass filter pair for lower filter bank of DT-CWT
$\boldsymbol{H} \in \mathbb{R}^{M\times N}$	projection operator
H_{ij}	The length of the i-th projection line in pixel j
$h_0(n)$, $h_1(n)$	Low-pass/high-pass filter pair for upper filter bank of DT-CWT
I_0	incident beam intensity
I	attenuated radiation intensity
$\boldsymbol{I}_n \in \mathbb{R}^{n\times n}$	Identity matrix
$J(\cdot)$	Optimization criterion
$J(\boldsymbol{f})$	Optimization criterion on $\boldsymbol{f}$
$J(\cdot)$	Optimization criterion
$\nabla J(\cdot)$	Derivative of criterion $J(\cdot)$
l	trajectory of a radiation
	Chap 4 : Level of Haar transformation
$\mathcal{L}(\cdot)$	Laplacian distribution
$\boldsymbol{m} \in \mathbb{R}^{N\times 1}$	Mean of distribution of Markovian model for $\boldsymbol{f}$
$\boldsymbol{N}_r(\cdot)$	Set of the neighbour voxels
$\mathcal{N}(\cdot)$	Normal or Gaussian distribution
$p(\cdot)$	Probability distribution function
$q(\cdot)$	Probability distribution function
$Q(\boldsymbol{g}, \boldsymbol{f})$	Discrepancy of observed data
$\boldsymbol{q} \in \mathbb{R}^{N\times 1}$	Normalized gradient of object $\ddot{\boldsymbol{f}}$
r	distance of l from the origin
$\boldsymbol{r}$	Current voxel in the Markovian model
$\boldsymbol{r}'$	Neighbour voxels in teh Markovian model
$r \in [1, l+1]$	Rank of coefficent in l-level Haar transformation
$R(\boldsymbol{f})$	Regularization term on $\boldsymbol{f}$
$\mathcal{R}$	Radon transform
$\mathcal{R}^{\#}$	Dual Radon transform
$\mathcal{S}t(\cdot)$	Student-t distribution
$\mathcal{S}t_g(\cdot)$	Generalized Student-t distribution
$\boldsymbol{v}_\epsilon \in \mathbb{R}^{M\times 1}$	Variance of the additive noise $\boldsymbol{\epsilon}$
$\boldsymbol{V}_\epsilon \in \mathbb{R}^{M\times M}$	Covariance matrix of the additive noise $\boldsymbol{\epsilon}$
$\boldsymbol{v}_f \in \mathbb{R}^{N\times 1}$	Variance of $\boldsymbol{f}$
$\boldsymbol{V}_f \in \mathbb{R}^{N\times N}$	Covariance matrix of $\boldsymbol{f}$
$\boldsymbol{v}_z \in \mathbb{R}^{N\times 1}$	Variance of $\boldsymbol{z}$
$\boldsymbol{V}_z \in \mathbb{R}^{N\times N}$	Covariance matrix of $\boldsymbol{z}$
$\boldsymbol{v}_\xi \in \mathbb{R}^{N\times 1}$	Variance of $\boldsymbol{\xi}$
$\boldsymbol{V}_\xi \in \mathbb{R}^{N\times N}$	Covariance matrix of $\boldsymbol{\xi}$
$\boldsymbol{v}_\rho \in \mathbb{R}^{M\times 1}$	Variance of $\boldsymbol{\rho}$
$\boldsymbol{V}_\rho \in \mathbb{R}^{N\times N}$	Covariance matrix of $\boldsymbol{\rho}$
$\boldsymbol{v}_\tau \in \mathbb{R}^{M\times 1}$	Variance of $\boldsymbol{\tau}$
$\boldsymbol{V}_\tau \in \mathbb{R}^{N\times N}$	Covariance matrix of $\boldsymbol{\tau}$
$\boldsymbol{v}_c \in \mathbb{R}^{M\times 1}$	Variance of $\ddot{\boldsymbol{f}}$
$\boldsymbol{V}_c \in \mathbb{R}^{N\times N}$	Covariance matrix of $\ddot{\boldsymbol{f}}$
$\boldsymbol{z} \in \mathbb{R}^{M\times 1}$	Hidden variable
(α, β)	Pair of parameters in $\mathcal{S}t_g$
$(\alpha_{\epsilon_0}, \beta_{\epsilon_0})$	Pair of parameters for $\mathcal{S}t_g$ distribution of $\boldsymbol{v}_\epsilon$
$(\alpha_{\xi_0}, \beta_{\xi_0})$	Pair of parameters for $\mathcal{S}t_g$ distribution of $\boldsymbol{v}_\xi$
$(\alpha_{z_0}, \beta_{z_0})$	Pair of parameters for $\mathcal{S}t_g$ distribution of $\boldsymbol{v}_z$
$(\alpha_{\rho 0}, \beta_{\rho 0})$	Pair of parameters for $\mathcal{S}t_g$ distribution of $\boldsymbol{v}_\rho$
$(\alpha_{\tau_0}, \beta_{\tau_0})$	Pair of parameters for $\mathcal{S}t_g$ distribution of $\boldsymbol{v}_\tau$
$(\alpha_{c_0}, \beta_{c_0})$	Pair of parameters for $\mathcal{S}t_g$ distribution of $\boldsymbol{v}_c$
λ	Regularization parameter
$\boldsymbol{\epsilon} \in \mathbb{R}^{M\times 1}$	Additive noise of projection system
$\boldsymbol{\xi} \in \mathbb{R}^{N\times 1}$	Uncertainties of transformation
$\boldsymbol{\rho} \in \mathbb{R}^{M\times 1}$	Error of projection
ϵ_i	The noise of the i-th detection
ϕ	the angle the normal vector to l makes with the x axis
ν	Degrees of freedom in $\mathcal{S}t$
$\boldsymbol{\theta}$	Parameters in the Bayesian model
$\boldsymbol{\kappa}$	Hyper-parameters in the Bayesian model
$\gamma_\cdot, \gamma_f, \gamma_z$	Gradient descent step length
$\Gamma(\cdot)$	Gamma function
$\boldsymbol{\eta}$	Unit vector of r
$\boldsymbol{\eta}_1$	Projection of $\boldsymbol{\eta}$ onto axis x
$\boldsymbol{\eta}_2$	Projection of $\boldsymbol{\eta}$ onto axis y
$\boldsymbol{\tau} \in \mathbb{R}^{M\times 1}$	Noise of projection of the gradient $\ddot{\boldsymbol{f}}$

1 Résumé en français

1.1 Motivation et contexte

Cette thèse est une contribution au problème de reconstruction de tomographie en 3D. Le rayonnement X après sa découverte par Wilhelm Rontgen a été rapidement utilisé dans le domaine de la tomographie. Au début, on utilise uniquement les images radio permettant de voir la projection de l'objet traversé par les rayons X sur la plaque radio. Mais la projection des rayons X dans une seule direction perdait des informations, notamment l'information de profondeur.En 1917, Radon propose un algorithme afin de reconstruire les images, coupes 2D du patient mais qui ne pouvait etre appliqué en pratique avant les années 1970 et l'arrivée des scanners CT (Computed Tomography). Cette technique d'acquisition avec les projections appliquées dans différentes directions est développée avec de nombreuses publications sur les méthodes analytiques et itératives. Dans ces techniques, les données mesurées à différents angles sont utilisées pour reconstuire le volume 3D. À la fin des années 1970, les méthodes de reconstruction à faisceau en éventail (fan beam) ont été développées. La technique de tomographie en rayons X a d'abord été utilisée dans les applications médicales rapidement après son invention puis a été utilisée dans des nombreux autres domaines : la science marine, les géosciences, les applications industrielles, etc.

La base de la technique CT de rayons X est basée sur l'atténuation des rayons X en passant par un objet. Une vue de dessus d'un système de projection de rayons X est illustrée dans la figure 2.1. Différents matériaux avec une densité différente dans l'objet ont un taux d'atténuation différent pour le rayon X. Cette atténuation est définie par la loi de Beer-Lambert :

$$I = I_0 e^{-\int f(l)\,\mathrm{d}l}, \tag{1.1}$$

où l est le rayon X passant par un objet, I_0 est l'intensité du faisceau incident, $f(l)$ est le coefficient d'atténuation linéaire local sur le rayonnement l et I est l'intensité de rayonnement atténuée.

L'un des défis de la CT est de diminuer le nombre de projections. L'absence de données complètes est justifiée par des raisons différentes selon les applications. Le rayon X a été utilisé pour la première fois dans les applications médicales depuis son introduction dans les années 1970. Il a été utilisé dans le diagnostic de la tête, des poumons, de l'angiographie pulmonaire, du cardiaque, de l'abdomen et du pelvis, des extrémités, etc. En utilisant la tomographie en rayon X, le diagnostic des organes peut être effectué avec une haute résolution afin de distinguer plus de détails. Cependant, cette technique présente de nombreux inconvénients. Le plus important est le risque pour les patients. Le rayon X utilisé peut endommager les cellules du corps, y compris les molécules d'ADN, qui peuvent conduire à cancer, ou à un autre effet secondaire indésirable. Par conséquent, les efforts de recherche récents se concentrent sur les techniques de reconstruction CT avec une dose plus faible de rayonnement, également appelée ALARA (As Low As Reasonably Achievable) et moins de temps d'exposition pour chaque projection. Dans les applications industrielles, le CT à rayons X est souvent utilisé dans le Contrôle Non Destructif (CND). Il est utilisé pour produire des représentations 3D de composants à l'externe et à l'interne, pour la détection des défauts, l'analyse des pannes, etc. Dans les applications industrielles, la dose de rayonnement n'est pas nécessairement faible. Le défi principal de l'application industrielle est la détection dynamique. Dans certains cas, l'objet examiné n'est pas statique, par exemple l'objet est imagé avec un liquide qui s'écoule. Dans certains autres examens industriels, les objets sont placés sur un convoyeur. Dans ces cas, la projection devrait se faire en peu de temps, donc le nombre de projections doit être réduit. Par conséquent, les reconstructions devraient être réalisées avec moins de projections.

Dans l'imagerie médicale et le CND industriel, un autre problème auquel nous pouvons faire face est l'angle limité d'acquisition. Dans les cas général, le scanner CT teste un volume en obtenant des rayonnements en projection pour des angles uniformément répartis entre $[0°, 180°]$. Mais dans certaines applications, la couverture angulaire totale n'a pas pu être obtenue. Un exemple dans la demande médicale est la tomographie dentaire où toutes les projections ne sont pas faites en raison d'encombrement du tomographe. Dans les applications industrielles, ce problème apparaît également lorsque le composant testé a une grande taille et est supérieure à la portée du scanner, on obtient alors un angle de projections limité. Dans l'angle angulaire limité, les projections sont réparties entre $[0°, \theta]$, où $\theta < 180°$.

Dans les travaux précédents, un grand nombre d'articles sont publiés sur la reconstruction de CT de rayons X, en s'appuyant sur différentes méthodes (analytique, algébrique, régularisation, Bayesian, etc.) pour différents types de systèmes de projections (faisceau parallèle, faisceau en éventail, faisceau conique, etc.), compte tenu de différents modèles de bruits (bruit de Poisson, bruit gaussien) et en utilisant différentes méthodes d'estimation (Maximum A Posterior, Posterior Mean, etc). À partir de ces travaux, nous pouvons voir que les méthodes sont plus robustes au caractère incomplet des données lorsque des informations complémentaires sont prises en compte.

Dans la tomodensitométrie 3D, en raison de la grande taille des données, les coûts de calcul sont un frein aux développements de nouvelles méthodes. Au début de la tomographie en rayon X, les méthodes statistiques n'étaient généralement pas utilisées en raison du calcul itératif trop coûteux. Avec le développement du matériel informatique et l'utilisation des aspects de l'accélération du calcul, les méthodes itératives deviennent abordables. Les processeurs many cores de type GPU (Graphic Processing Unit) qui permettent une parallélisation massive des calculs sont à présent largement utilisés dans la reconstruction d'objets 3D.

Un autre défi dans le rayon X est constitué par les artefacts de reconstruction. Il existe de

nombreux types d'artefacts causés par des raisons différentes : les artefacts métalliques, les artefacts en anneaux, les artefacts dû au mouvement, etc. Le bruit d'acquisition est également un défi communément rencontré dans les problèmes de reconstruction de tomographie. Les méthodes itératives sont adaptées pour éliminer ces différents types d'artefacts. Au cours de ma thèse, je ne considérerai que le bruit lié à l'acquisition.

Malgré le fait que les méthodes itératives obtiennent des résultats plus précis et peuvent supprimer les artefacts dans des conditions mal posées avec un ensemble de données insuffisantes, limitées ou tronquées, il y a un autre problème crucial à considérer : la décision d'initialisation des hyper-paramètres. Il existe de nombreux chercheurs qui travaillent sur ce problème. Avec ces techniques, on peut obtenir une valeur optimale pour les paramètres qui conduisent à des bons résultats de reconstruction, mais les coûts de calcul de ces méthodes sont coûteux. Dans la plupart des cas, les personnes choisissent la valeur des paramètres de manière manuelle ou empirique lors de simulations préalables. Dans cette thèse, la méthode bayésienne est utilisée et les paramètres sont estimés simultanément pendant la reconstruction.

L'inférence bayésienne est une façon couramment utilisée de prendre en compte l'information préalable et de la modélisée par une distribution a priori choisie afin de complexifier le monis possible les calculs. Dans les méthodes bayésiennes, le modèle markovien est fréquemment utilisé pour définir la variable continue par morceaux. La valeur de chaque voxel n'est pas indépendante des autres, mais dépendantes des voxels voisins.

Objet examiné dans la tomographie à rayons X étant composé de plusieurs matériaux homogènes, la propriété continue par morceaux de l'objet est l'une des informations les plus fréquemment considérées dans la reconstruction tomographie de rayons X. Par ailleurs, la parcimonie d'une variable cachée qui est liée à la variable continue par morceaux est également fréquemment considérée. Cette variable cachée qui a une propriété de parcimonie peut être, par exemple, le gradient du volume, le coefficient de transformation d'ondelettes du volume ou un coefficient de la transformée du volume à l'aide d'un dictionnaire. Le mot ≪parcimonie≫ implique que le vecteur ne contient que quelques éléments non nuls. De nombreuses transformations parcimonieuses sont réversibles. On peut reconstruire l'objet avec les coefficients parcimonieux qui contiennent beaucoup moins de valeurs non nulles que l'objet lui-même. Le problème inverse mal posé lié à la reconstruction de l'objet revient alors à l'estimation des coefficients parcimonieux.

1.2 Contributions

Un aperçu de la contribution de cette thèse figure à l'annexe B.

Nous proposons dans cette thèse de doctorat principalement deux méthodes dans le contexte bayésien. Ces deux méthodes ont un modèle de système hiérarchique, une variable cachée étant considérée.

En reconstruction tomographique à rayons X, l'objet est typiquement continu par morceaux et, dans certains cas, constant par morceaux. Ainsi, deux aspects du choix de la variable cachée sont considérés dans notre travail : la transformation des ondelettes qui est utilisée dans la méthode HHBM au chapitre 4, et l'operateur Laplacien qui est utilisé dans la méthode ROCC au chapitre 5. Pour ces deux variables cachées, une distribution parcimonieuse leur est associée. Une distribution Student-t généralisée est choisie en raison de sa propriété à queue lourde,

modélisant ainsi la propriété de parcimonie, et elle s'exprime comme la loi marginale d'une distribution Gamma Inverse. La structure marginale de deux distributions conjuguées facilite l'estimation des variables dans les méthodes bayésiennes.

Parmi les transformations d'ondelettes, nous choisissons d'utiliser la transformation de Haar discrète. Pour un objet continu par morceaux ou constante par morceaux, la transformation Haar donne un coefficient faible pour la variable. Etant une transformée en ondelettes basique, la transformation de Haar n'est pas cher en termes de calcul et peut être accélérée efficacement à l'aide de processeurs GPU. Le Laplacian de l'objet est considéré comme une variable cachée dans une autre méthode proposée, présentée au chapitre 5. Un modèle markovien non homogène est utilisé pour définir la variable.

Dans cette thèse, le mot "**variables**" représente les inconnues qui doivent être estimées, par exemple les valeurs des pixels ou des voxels d'objet. Le mot "**variables cachées**" représente les variables intermédiaires qui sont souvent nécessaires pour modéliser l'objet comme les contours, les étiquettes ou les classifications des régions ou les coefficients de transformation dans les modèles hiérarchiques. Le mot "**paramètres**" représente les paramètres qui apparaissent dans le modèle a priori des variables telles que la variance du bruit, et ils sont également estimés pendant les itérations. Le mot "**hyper paramètres**" représente les paramètres des distributions a priori des paramètres, et ils doivent être fixés pour l'initialisation. Dans cette thèse, l'initialisation des hyper paramètres est discutée. Au lieu de tester des différents hyper-paramètres et de choisir le meilleur, nous discutons de la façon de fixer les hyper paramètres directement en condiserant les modèles a priori et les informations préalables sur les variables.

1.3 Organisation de thèse

Au chapitre 2, nous présentons le modèle continu et discret du système CT, et le modèle direct de l'opération de projection. Les méthodes conventionnelles, y compris les méthodes analytiques, les méthodes algébriques, les méthodes de régularisation, les méthodes statistiques et l'inférence bayésienne sont présentées. Les avantages et les défauts de chaque type de méthode sont discutés. Ensuite, nous montrons quelques résultats de simulation pour la reconstruction CT en utilisant deux des méthodes les plus représentatives : la technique de rétro-projection filtrée (FBP) et la méthode de Variation Totale (TV). Les résultats de la simulation en utilisant ces deux méthodes avec différents paramètres de projection seront donnés dans ce chapitre. Et je vais montrer l'influence du paramètre dans la méthode de regularisation et démontrer la difficulté de choisir cette valeur.

Dans le chapitre 3, nous présentons le contexte de l'inférence bayésienne. Les propriétés intrinsèques de l'objet (informations a priori) dans les applications industrielles sont discutées. Dans ce chapitre, la propriété de la distribution Student-t généralisée est présentée et l'avantage de l'utiliser est discuté.

Dans la deuxième partie de cette thèse, nous présentons les principales contributions de mon travail.

Au chapitre 4, une méthode HHBM est présentée. Dans cette méthode, tout d'abord, un modèle de synthèse est considéré avec un modèle de bruit non stationnaire. Dans ce modèle, seulement les coefficients de transformation de Haar sont estimés et l'objet est obtenu par un traitement a posteriori à partir des coefficients estimés. Ensuite, dans un modèle plus général,

l'objet et les coefficients de la transformation de Haar sont simultanément estimés avec un modèle de bruit non stationnaire. Comme le volume se compose de plusieurs blocs homogènes, le coefficient de transformation est faible. Nous avons proposé d'utiliser la distribution Student-t généralisée pour modéliser la répartition préalable du coefficient. Une fois que nous obtenons la distribution a posteriori, un estimateur du Maximum A Posteriori (JMAP) est utilisé pour estimer les variables inconnues. Certaines extensions de cette méthode sont également discutées. Une technique de scission des variables est appliquée sur cette méthode et il est démontré qu'en l'utilisant, on peut considérer le bruit comme le mélange d'un bruit gaussien et d'un bruit parcimonieux. Grâce à cela, les résultats de la reconstruction dépassent la méthode originale. Une autre extension consiste à comparer la transformation de Haar avec la transformation d'ondelettes complexes à double arbre (Dual Tree Complexe Wavelet Transformation DT-CWT). En comparant avec la transformation de Haar, la transformation DT-CWT prend en compte plus de directions d'ondelettes et est plus robuste lorsque les coefficients sont contaminés. Lorsqu'ils sont utilisés dans la méthode bayésienne proposée, le DT-CWT offre une meilleure qualité de reconstruction que la transformation de Haar.

Au chapitre 5, une méthode ROCC est proposée. Dans cette méthode, on considère les relations entre le Laplacien de projection et le Laplacien de l'objet. En utilisant le Laplacien de l'objet comme une variable cachée, nous bénéficions de sa propriété de parcimonie pour préserver les contours de l'objet. Dans cette méthode, les voxels sont modélisés par un modèle markovien non homogène. Dans ce modèle markovien non homogène, chaque voxel est modélisé par une distribution qui dépend des valeurs laplaciennes et des valeurs des voxels voisins. Lorsque son voxel voisin a une valeur Laplacienne plus grande, la valeur voxel voisine aura moins d'influence sur le modèle du voxel actuel.

Les résultats de la reconstruction, correspondant à un fantôme Shepp Logan simulé en 2D ou 3D et un objet Head en 3D, sont présentés au chapitre 4 et 5. Nous avons comparé les résultats obtenues avec nos méthodes et les méthodes de l'état de l'art (FBP, QR et TV). Différents paramètres pour les simulations sont également pris en compte. Le jeu de données avec un nombre insuffisant de projections, avec un angle de projection limité, ou avec un SNR élevé sur l'ensemble de données, sont testés.

Les conclusions et certaines perspectives sont données dans le chapitre 6. Dans des travaux à venir, tout d'abord, d'autres estimateurs peuvent être utilisés, par exemple l'estimateur de la Moyenne a Posteriori (PM). Les estimateurs PM sont généralement réalisés par les approches bayésiennes variationnelles (VBA), avec les détails donnés dans l'annexe A. En utilisant la méthode VBA, les coûts de calcul sont très coûteux pour les simulations de tomographie en 3D, mais notre groupe a récemment trouvé un moyen de simplifier le calcul. Cet algorithme pourrait être mis en œuvre dans nos travaux à venir. L'analyse de l'influence des hyper-paramètres et des explications de la robustesse des hyper-paramètres dans la méthode HHBM sera poursuivie. D'autres distributions de mélange de variance gaussienne seront également utilisées et comparées avec la distribution de $\mathcal{S}t_g$. Pour la méthode ROCC, seul le cas de projection parallèle est discuté. Nous chercherons les solutions pour étendre cette méthode à d'autres types de projections.

1.4 La méthode HHBM

Dans cette méthode, le modèle direct du système est considéré comme suivant :

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}, \tag{1.2}$$

où $\boldsymbol{g} \in \mathbb{R}^{M\times 1}$ représente la projection estimé, $\boldsymbol{f} \in \mathbb{R}^{N\times 1}$ représente l'objet, $\boldsymbol{H} \in \mathbb{R}^{M\times N}$ représente l'operateur linear de projection, et $\boldsymbol{\epsilon} \in \mathbb{R}^{M\times 1}$ représente le bruit additive. Dans notre travail, on considère un bruit Gaussian dans cette modèle, donc on a :

$$p(\boldsymbol{\epsilon}|\boldsymbol{v}_\epsilon) = \mathcal{N}(\boldsymbol{\epsilon}|0, \boldsymbol{V}_\epsilon), \text{ where } \boldsymbol{V}_\epsilon = \text{diag}\left[\boldsymbol{v}_\epsilon\right], \tag{1.3}$$

où $\boldsymbol{v}_\epsilon \in \mathbb{R}^{M\times 1}$ représente la variance de bruit, qui est considéré comme un inconnues dans cette méthode semi-supervisé. Considerant la propriété de variance de bruit, il est modélisé par la loi Gamma Inverse :

$$p(\boldsymbol{v}_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) = \mathcal{IG}(\boldsymbol{v}_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}). \tag{1.4}$$

Dans cette méthode, une autre relation linear est consideré :

$$\boldsymbol{f} = \boldsymbol{D}\boldsymbol{z} + \boldsymbol{\xi}, \tag{1.5}$$

où $\boldsymbol{z} \in \mathbb{R}^{N\times 1}$ représente le coefficient parcimonieux de $\boldsymbol{f}$ via la transformation représenté par l'opérateur $\boldsymbol{D} \in \mathbb{R}^{N\times N}$. $\boldsymbol{\xi} \in \mathbb{R}^{N\times 1}$ représente les incertitude de cette transformation.

En choisissant le modèle précédent pour $\boldsymbol{\xi}$, l'objectif est que $\boldsymbol{f}$ et $\boldsymbol{D}\boldsymbol{z}$ sont le plus approximatifs possible. Par conséquent, nous définissons $\boldsymbol{\xi}$ comme un bruit parcimonieux, avec la plupart des valeurs proches de zéro. La distribution Student-t généralisée est utilisée pour modéliser $\boldsymbol{\xi}$. La distribution de $\mathcal{S}t_g$ pour $\boldsymbol{\xi}$ est définie par le modèle hiérarchique :

$$p(\boldsymbol{\xi}|0, \boldsymbol{v}_\xi) = \mathcal{N}(\boldsymbol{\xi}|0, \boldsymbol{V}_\xi), \text{ where } \boldsymbol{V}_\xi = \text{diag}\left[\boldsymbol{v}_\xi\right], \tag{1.6}$$

$$p(\boldsymbol{v}_\xi|\alpha_{\xi_0}, \beta_{\xi_0}) = \mathcal{IG}(\boldsymbol{v}_\xi|\alpha_{\xi_0}, \beta_{\xi_0}). \tag{1.7}$$

À partir de la définition de la loi $\mathcal{S}t_g$:

$$\mathcal{S}t_g(x|\alpha, \beta) = \int \mathcal{N}(x|v)\mathcal{IG}(v|\alpha, \beta)\, \mathrm{d}v, \tag{1.8}$$

on peut trouver que tous les deux bruit $\boldsymbol{\epsilon}$ et $\boldsymbol{\xi}$ sont modélisé par la loi $\mathcal{S}t_g$. Un autre variable modélisé par la loi $\mathcal{S}t_g$ est la variable cachée $\boldsymbol{z}$:

$$p(\boldsymbol{z}|\boldsymbol{v}_z) = \mathcal{N}(\boldsymbol{z}|0, \boldsymbol{V}_z), \text{ where } \boldsymbol{V}_z = \text{diag}\left[\boldsymbol{v}_z\right], \tag{1.9}$$

$$p(\boldsymbol{v}_z|\alpha_{z_0}, \beta_{z_0}) = \mathcal{IG}(\boldsymbol{v}_z|\alpha_{z_0}, \beta_{z_0}). \tag{1.10}$$

Cette loi est utilisé considerant que la variable $\boldsymbol{z}$ est parcimonieuse.

Avec tous les loi définies, la distribution a postériori peut être obtenu par le théorème de Bayes :

$$p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_\epsilon, \boldsymbol{v}_\xi, \boldsymbol{v}_z|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{v}_\epsilon)p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{v}_\xi)p(\boldsymbol{z}|\boldsymbol{v}_z)p(\boldsymbol{v}_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0})p(\boldsymbol{v}_\xi|\alpha_{\xi_0}, \beta_{\xi_0})p(\boldsymbol{v}_z|\alpha_{z_0}, \beta_{z_0}). \tag{1.11}$$

Puis, l'estimation de Maximum a postériori joindre (JMAP) est utilisé pour estimer tous les inconnus.

1.5 La méthod ROCC

Dans cette méthode, les Laplacian de projection $\ddot{\boldsymbol{g}}$ et Laplacian d'object $\ddot{\boldsymbol{f}}$ sont consideré, avec les définitions :

$$\ddot{\boldsymbol{f}} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} * \boldsymbol{f} = \boldsymbol{d}_2 * \boldsymbol{f}, \tag{1.12}$$

$$\ddot{\boldsymbol{g}} = [-1\ 2\ -1] * \boldsymbol{g} = \boldsymbol{d}_1 * \boldsymbol{g}, \tag{1.13}$$

où $\boldsymbol{d}_1$ et $\boldsymbol{d}_2$ sont des noyau de convolution, et les convolutions avec eux peuvent représenté par multiplication avec matrices correspondants de la même taille de l'objet, qui sont présenté comme suivant :

$$\ddot{\boldsymbol{g}} = \boldsymbol{D}_1 \boldsymbol{g}, \tag{1.14}$$

$$\ddot{\boldsymbol{f}} = \boldsymbol{D}_2 \boldsymbol{f}. \tag{1.15}$$

Dans cette method, on considère les relations suivant :

$$\boldsymbol{f} \rightarrow \mathrm{RT} \rightarrow \boldsymbol{g}, \tag{1.16}$$

$$\ddot{\boldsymbol{f}} \rightarrow \mathrm{RT} \rightarrow \ddot{\boldsymbol{g}}. \tag{1.17}$$

En considerant les bruit du système, on a :

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}, \tag{1.18}$$

$$\ddot{\boldsymbol{g}} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\tau}. \tag{1.19}$$

En supposant que tous les deux sont bruit Gaussian, on a la loi de vraisemblance suivant :

$$p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{v}_\epsilon) = \mathcal{N}(\boldsymbol{g}|\boldsymbol{H}\boldsymbol{f}, \boldsymbol{V}_\epsilon),\ \boldsymbol{V}_\epsilon = \mathrm{diag}\left[\boldsymbol{v}_\epsilon\right], \tag{1.20}$$

$$p(\ddot{\boldsymbol{g}}|\ddot{\boldsymbol{f}}, \boldsymbol{v}_\tau) = \mathcal{N}(\ddot{\boldsymbol{g}}|\boldsymbol{H}\ddot{\boldsymbol{f}}, \boldsymbol{V}_\tau),\ \boldsymbol{V}_\tau = \mathrm{diag}\left[\boldsymbol{v}_\tau\right], \tag{1.21}$$

où les variances des bruit sont considéré comme inconnus et il sont modélisés par la loi de Gamma Inverse :

$$p(\boldsymbol{v}_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) = \mathcal{IG}(\boldsymbol{v}_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}), \tag{1.22}$$

$$p(\boldsymbol{v}_\tau|\alpha_{\tau_0}, \beta_{\tau_0}) = \mathcal{IG}(\boldsymbol{v}_\tau|\alpha_{\tau_0}, \beta_{\tau_0}). \tag{1.23}$$

Dans cette méthode, le parcimonie de la variable cachée $\ddot{\boldsymbol{f}}$ est considerée, et il est modélisé par la loi Student-t généralisé via le structure de Gaussian-Gamma Inverse :

$$p(\ddot{\boldsymbol{f}}|\boldsymbol{v}_c) = \mathcal{N}(\ddot{\boldsymbol{f}}|0, \boldsymbol{V}_c),\ \boldsymbol{V}_c = \mathrm{diag}\left[\boldsymbol{v}_c\right], \tag{1.24}$$

$$p(\boldsymbol{v}_c|\alpha_{c_0}, \beta_{c_0}) = \mathcal{IG}(\boldsymbol{v}_c|\alpha_{c_0}, \beta_{c_0}). \tag{1.25}$$

Dans cette modèle, la variable $\boldsymbol{f}$ est modélisé par une modèle Markovian, qui depend de la valeur de $\ddot{\boldsymbol{f}}$ estimés. Dans cette modèle, la loi de distribution pour chaque voxel f_r est :

$$p(f_r|\boldsymbol{f}_r, \boldsymbol{q}, v_{f_r}) = \mathcal{N}(f_r|m_r, v_{f_r}) \propto v_{f_r}^{-1} \exp\left\{-\frac{1}{2} v_{f_r}^{-1} \left(f_r - \frac{\sum_{r' \in \boldsymbol{N}_{r(r)}} (1 - q_{r'})\, f_{r'}}{\sum_{r' \in \boldsymbol{N}_{r(r)}} (1 - q_{r'})}\right)^2\right\}, \tag{1.26}$$

où $\boldsymbol{f}_{-r}$ représente tous les voxels sauf f_r, v_{f_r} est la variance de voxel r, et :

$$m_r = \frac{\sum_{r' \in \boldsymbol{N}_{r(r)}} (1 - q_{r'}) f_{r'}}{\sum_{r' \in \boldsymbol{N}_{r(r)}} (1 - q_{r'})}. \tag{1.27}$$

La variance $\boldsymbol{v}_f$ est considérée comme inconue et être modélisé par la loi de Gamma Inverse :

$$p(\boldsymbol{v}_f | \alpha_{f_0}, \beta_{f_0}) = \mathcal{IG}(\boldsymbol{v}_f | \alpha_{f_0}, \beta_{f_0}). \tag{1.28}$$

Dans ce méthode, nous présentons une méthode bayésienne qui utilise un modèle markovien pour la variable continue par morceaux. De la propriété de la transformée de Radon, le Laplacien de l'objet $\ddot{\boldsymbol{f}}$ est lié au Laplacien de la projection sinogram $\ddot{\boldsymbol{g}}$ par un modèle direct. En utilisant simultanément les deux fonctions directes $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$ et $\ddot{\boldsymbol{g}} = \boldsymbol{H}\ddot{\boldsymbol{f}} + \boldsymbol{\tau}$, ces deux variables sont estimées. Pendant ce temps, la valeur localisée Laplacienne est utilisée comme coefficient de poids dans le modèle antérieur de $\boldsymbol{f}$ dans chaque itération. Enfin, toutes les variables et les paramètres sont estimés simultanément via la distribution postérieure en utilisant l'algorithme d'estimation JMAP. L'initialisation des hyper paramètres est proposée, et leur relative insensibilité à la reconstruction est prouvée par les simulations. À partir des simulations, nous concluons que la méthode ROCC proposée s'améliore pour préserver les contours en comparaison avec les méthodes conventionnelles de QR et de TV. Lorsque l'ensemble de données est biaisé avec un bruit de SNR=$40dB$, l'erreur carrée moyenne relative de l'objet reconstruit est plus petite que les méthodes QR et TV, et l'avantage est plus évident lorsqu'il y a moins de projections. La valeur RMSE de la méthode ROCC en utilisant 36 projections est au même niveau que la méthode TV en utilisant 90 projections. Cette méthode est adaptée à la reconstruction CT de projection de faisceau parallèle, où la transformée de radon est utilisée. Le travail futur se concentre sur l'extension du faisceau parallèle aux projections du faisceau conique.

1.6 Conclusions et persipectives

Dans cette thèse, nous avons proposé principalement deux méthodes bayésiennes pour la reconstruction 3D CT. Ces méthodes sont basées sur le contexte du Contrôle Non Destructif (CND) dans les applications industrielles. La grande taille des données dans les applications 3D et le coût de calcul sont pris en compte dans les algorithmes.

Dans le chapitre 4, nous avons proposé une méthode bayésienne basée sur le modèle direct du système $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$. Dans cette méthode, la nouveauté est qu'on considère une structure hiérarchique dans le modèle du système, avec $\boldsymbol{f} = \boldsymbol{D}\boldsymbol{z} + \boldsymbol{\xi}$. Avec le niveau supplémentaire du système, la variable $\boldsymbol{f}$ est représentée par une variable cachée $\boldsymbol{z}$. Pour le choix du modèle précédent, la propriété considérée passe de la modélisation de la variable continue par morceaux à la modélisation de la variable cachée qui est parcimonieux. Une distribution Student-t généralisée ($\mathcal{S}t_g$) est présentée dans ce chapitre. Grâce au fait que la distribution de $\mathcal{S}t_g$ est marginale d'une distribution Gaussian-Gamma Inverse, les algorithmes d'estimation sont faciles à appliquer.

Dans les algorithmes itératifs, y compris l'inférence bayésienne, l'initialisation des paramètres est cruciale pour les résultats finaux de la reconstruction. Dans la section 4.4.2 et la section 5.6.3, nous avons présenté la stratégie d'initialisation des hyper-paramètres pour l'algorithme d'optimisation dans l'inférence bayésienne avec le modèle proposé. Cette stratégie

d'initialisation est démontrée expérimentalement. Dans l'analyse des hyper paramètres, nous avons observé une influence des hyper-paramètres relativement faible sur le comportement de l'algorithme itératif correspondant. L'intérêt de cette dépendance faible relative est qu'il offre un moyen pratique d'initialisation qui n'est généralement pas trivial.

Dans la section 4.5, nous avons présenté une autre transformation orthonormée, la transformation des ondes ondulées à double arbre (DT-CWT). Cette transformation est comparée à la transformation de Haar discrète. Dans la comparaison, nous constatons que la transformation DT-CWT ne provoquera pas les artefacts du bloc dans les figures reconstruites et elle est plus robuste à la contamination des coefficients. Lors de l'utilisation dans la méthode bayésienne proposée, le DT-CWT a une meilleure performance de reconstruction que HT. L'inconvénient est que le calcul du DT-CWT est relativement plus coûteux que l'utilisation de la transformation Haar.

Dans la section 4.8, nous avons proposé d'utiliser un autre modèle avancé dans cette approche bayésienne. Dans ce modèle, le bruit additif est séparé en deux parties, l'un est modélisé par une distribution Gaussien et un autre modélisé par une distribution $\mathcal{S}t_g$. Le modèle direct considéré pour le système de projection est

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\rho} + \boldsymbol{\epsilon}. \tag{1.29}$$

Dans ce système direct, on considère un modèle de bruit plus compliqué. Les résultats de la simulation montrent que ce modèle, en tenant compte d'un modèle de bruit plus complexe, obtient de meilleurs résultats de reconstruction.

Au chapitre 5, une autre stratégie pour la reconstruction d'objets CT à rayons X est présentée. Dans cette méthode, on considère la relation entre l'objet $\boldsymbol{f}$, la projection de l'objet $\boldsymbol{g}$, le Laplacien de l'objet $\ddot{\boldsymbol{f}}$ et le Laplacien de la projection sinogram $\ddot{\boldsymbol{g}}$. Selon le système de transformation du radon, les relations du système de projection directe que nous prenons en compte sont :

$$\ddot{\boldsymbol{g}} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}, \tag{1.30}$$

$$\ddot{\boldsymbol{g}} = \boldsymbol{H}\ddot{\boldsymbol{f}} + \boldsymbol{\tau}, \tag{1.31}$$

avec quelques relations entre $\boldsymbol{g}$ et $\ddot{\boldsymbol{g}}$ et entre $\boldsymbol{f}$ et $\ddot{\boldsymbol{f}}$.

En considérant ces relations, le Laplacien de l'objet $\ddot{\boldsymbol{f}}$ est considéré comme une variable cachée dans le modèle du système. Un modèle Markovien non homogène est utilisé pour l'objet continu par morceaux $\boldsymbol{f}$. Dans ce modèle Markovien, la variable cachée $\ddot{\boldsymbol{f}}$ est utilisée comme paramètre. En utilisant ce modèle markovien en considérant la valeur de $\ddot{\boldsymbol{f}}$, les contours et les zones homogènes sont modélisés avec le même modèle avec des paramètres différents, et donc la continuité des zones homogènes et la discontinuité des contours sont à la fois considérées. Cette stratégie préserve les bords lors de la reconstruction de l'objet.

Pour l'avenir, plusieurs aspects sont pris en considération. Tout d'abord, le temps de calcul devrait prendre en compte. Le processeur GPU n'est utilisé que pour l'opérateur de projection et l'opérateur de rétroprojecteur dans cette thèse. Afin d'accélérer le calcul, plus de calculs dans le programme peuvent être mis en œuvre en utilisant un processeur GPU. Par exemple, la transformation DT-CWT présentée au chapitre 4 présente de très bonnes propriétés en comparaison avec la transformation de Haar, mais il a un calcul plus compliqué. Une fois que nous implémentons la transformation en utilisant le processeur GPU, l'utilisation du DT-CWT ou des nombreuses autres transformations sera moins coûteuse.

Dans cette thèse, nous avons utilisé l'estimateur Maximum A Posterior (MAP) pour l'estimation lorsque la distribution postériori est obtenue. Il existe d'autres algorithmes d'estimation, et certains donnent plus d'informations pour les variables. Par exemple, l'estimation de la moyenne postériori peut être utilisée pour estimer les variables. En utilisant la méthode Variation Bayesian Approach (VBA), la distribution a postériori est abordée par une autre distribution séparable :

$$p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}) = q(\boldsymbol{f}, \boldsymbol{\theta}) = q_1(\boldsymbol{f})q_2(\boldsymbol{\theta}), \tag{1.32}$$

où $\boldsymbol{\theta}$ représente toutes les variables cachées et les paramètres du modèle. La moyenne et la variance de la répartition approximative sont estimées.

Dans la méthode VBA, les distributions approches sont calculées en minimisant la divergence de Kullback Leibler. Au cours du calcul de la méthode VBA, un calcul des éléments diagonaux de $\boldsymbol{H}^T\boldsymbol{H}$ de la grande taille est nécessaire, montré à l'Annexe A. Certains résultats de simulation de la méthode VBA sont présentés dans [WMDGD16] pour la reconstruction du fantôme 2D Shepp Logan en considérant un bruit stationnaire. En raison de la taille énorme de la matrice $\boldsymbol{H}$ dans le problème 3D, il est très coûteux de calculer directement par projection et rétroprojection pour obtenir chaque élément diagonal. Maintenant, notre groupe est en train de étudier ce sujet et rechercher des algorithmes qui calculent approximativement les éléments diagonaux en utilisant les propriétés de projection et de rétroprojection.

Dans cette thèse, la distribution $\mathcal{S}t_g$ est utilisée pour définir les variables parcimonieuses. Comme nous l'avons mentionné, la distribution de $\mathcal{S}t_g$ a une structure à queue lourde. Dans le même temps, il peut avoir une variance très faible grâce au fait qu'il existe deux paramètres pour contrôler la forme de cette distribution. Une autre propriété très utile dans le contexte bayésien est que, elle peut être exprimée comme un marginal de la distribution de Gaussian-Gamma Inverse bivariée. Comme la distribution normale et la distribution Gamma Inversée sont conjuguées, l'estimation des variables dans les méthodes bayésiennes est plus simple.

Dans le cadre de travaux futurs, au lieu d'utiliser les distributions du mélange Gaussian-Gamma Inverse, on cherche d'autres distributions similaires, par exemple les distributions du mélange de variance normale. On va comparer les différentes distributions de mélange et comparer leurs propriétés, concernant la propriété de la distribution marginale et l'influence des hyper paramètres des distributions marginales.

L'un des aspects les plus importants de notre travail futur est d'étudier l'influence de l'initialisation des hyper-paramètres. Dans cette thèse, nous avons observé une influence par semaine relative des hyper paramètres du modèle bayésien selon les modèles antérieurs. Dans le futur travail, nous étudierons la raison de cette faible influence sur les hyper paramètres de ce modèle et voir si la même propriété apparaît lorsqu'on utilise les autres modèles de mélange de variance normale.

Dans un futur travail à long terme, nous chercherons de nouveaux modèles a priori pour la reconstruction de CDT. Nous nous concentrerons sur les cas mal posé où le nombre de projections est extrêmement limité. Pour cela, nous devrions optimiser le modèle a priori et éviter les artefacts apparaissant dans les résultats de la reconstruction. Nous allons également étudier la relation entre les méthodes de régularisation et les méthodes bayésiennes. Il existe des nombreuses méthodes de régularisation qui fonctionnent très bien pour le problème de reconstruction mal posé si un paramètre de régularisation approprié est choisi. Nous pouvons étudier, du point de vue bayésien, le principe de la régularisation. Dans le même temps, nous pouvons bénéficier des avantages de la méthode bayésienne selon laquelle les paramètres sont estimés

pendant les itérations et les hyper paramètres sont relativement insensibles aux résultats de la reconstruction.

2 Introduction

2.1 Motivation and Context

This thesis is a contribution to the 3D X-ray Computed Tomography (CT) reconstruction problem. After the discovery of the X radiation by Wilhelm Röntgen, it was quickly used in the domain of tomography. At first, people uses the attenuation projection images to see the object passed by the X radiation. In 1917, the proposition of the Radon transformation [Dea07] solves the problem of the reconstruction. However, X-ray projection in one direction would lose information, for example the depth information. Then the technique of acquisition with the projections applied in different directions is developed in the 1970s, with many publications on both analytic and iterative methods [AGU76, OMA+76]. In these techniques, the data measured at different angles are used to reconstruct the 3D volume. At the end of 1970s, the fan-beam reconstruction methods were developped. The X-ray CT technique was firstly used in the medical applications soon after its invention [GBDBM74, LWGR74, PA74]. Then it began to be used in many other domains, for example the marine science [BLO95], the geosciences [WV+87], the industrial applications [HME+81], etc. With the development of the reconstruction techniques with full projection data, the interest changed to the study of **incomplete data problems**.

The foundation of the X-ray CT technique is based on the attenuation of the X rays when passing through an object. A top view of a X-ray projection system is shown in Figure 2.1. Different materials with different density in the object have different attenuation rate for the radiation. This attenuation is defined by the Beer's law :

$$I = I_0 e^{-\int f(l)\,\mathrm{d}l}, \tag{2.1}$$

where l is the monochromatic X-ray passing an object, I_0 is the incident beam intensity, $f(l)$ is the local linear attenuation coefficient along the radiation l and I is the attenuated radiation intensity.

One of the challenges of the X-ray CT is to decrease the number of projections. X-ray CT

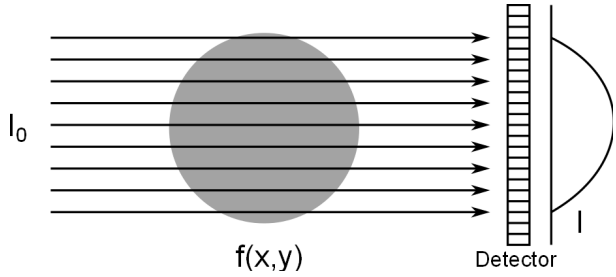


FIGURE 2.1 – Top view of the X-ray CT projections system.

was firstly used in medical applications since its introduction in the 1970s. It was used in the diagnosis of head, lungs, pulmonary angiogram, cardiac, abdominal and pelvic, extremities, etc [Arr99]. By using the X-ray CT, the diagnosis of the organs can be done with high resolution, in order to distinguish more details. However, this technique has many drawbacks. The most important one is the risks for the patients. The radiation used in CT scans can damage body cells, including DNA molecules, which can lead to cancer, or some other adverse side effect. Therefore recent research efforts focus on techniques for the CT reconstruction with lower dose of radiation, also known as ALARA(As Low As Reasonably Achievable), and less expose time for each projection. In the industrial applications, the X-ray CT is often used in the Non Destructive Testing (NDT) [HFU08]. It is used to produce 3D representations of components both externally and internally, for the flaw detection, failure analysis, etc. In the industrial applications, the reduction of the radiation dose is not a crucial problem. The main challenge of the industrial application is in the dynamic detection. In some cases the object under examination is not static, for example the object with flowing liquid. In some other industrial examinations, the objects are placed on a conveyor. In these cases, the projection should be done in a short time, hence the number of projections need to be reduced. Hence the reconstructions are expected to be done with **less projections**.

In both medical imaging and industrial NDT, another problem that we may face to is the limited angle of projections. In typical works, the CT scanner tests a volume by obtaining projecting radiations for angles evenly distributed between $[0°, 180°]$, with full angle coverage for the volume. But in some applications, the full angle coverage could not be obtained. An example in the medical application is the dentistry CT. The projection is sometimes not done for all the angles because of the special structure of the mouth. In industrial applications, this problem normally appears when the component under testing has a large size. When the size of the component is larger than the scope of the scanner, one obtains a limited angle of projections. In the **limited angle** CT, the projections are distributed between $[0°, \theta]$, where $\theta < 180°$.

In the previous works, a great number of articles are published on the X-ray CT reconstruction, basing on different methods (analytical [FDK84], algebraic [GBH70, AK84, BYL+00], regularization [CMRS00, GO09, SP08], Bayesian [GCSR14, BMD08, MD96, KVS+06], etc), for different types of projections systems (parallel-beam [CLMT02], fan-beam [AZ05], cone-beam [FDK84], etc), considering different noise models (Poisson noise [SB93, FDK84], Gaussian noise [RLR+12]), and by using different estimation methods (Maximum A Posterior [QL00], Posterior Mean [Eri69], etc). From these works we can see that the methods are more robust in the ill-posed cases when supplementary prior information are considered.

In 3D X-ray CT, because of the large data size, the computational costs is crucial. At the beginning of the development of X-ray CT, the statistical methods were not typically used because of the very costly iterative computation. With the development of the computational hard-

ware and the utilization of the computation acceleration aspects, the iterative methods become feasible in the X-ray CT reconstruction. The GPU processor is now commonly used in the reconstruction of 3D CT objects.

Another challenge in the X-ray CT is the artifacts. There are many types of artifacts caused by different reasons. For example the streak artifacts, ring artifact, motion artifact, etc. The noise of acquisition is also a commonly faced challenge in the X-ray CT reconstruction problems. The iterative methods are adapted to removing these different types of artifacts.

Despite the fact that the iterative methods get more precise results and can reduce the artifacts in ill-posed conditions with insufficient or limited or truncated dataset, there is another crucial problem to consider : the decision for the **initialization of hyper-parameters**. There are many researchers working on this problem [FSH15, MDA15, SC91, ZLQ97, SBS98]. With these techniques, one can obtain an optimal value for the parameters which lead to good reconstruction results, but the computational costs of these methods are expensive. In most cases, people choose the value of the parameters manually, or empirically with some beforehand simulations. In this thesis, the Bayesian inference is considered and the parameters and unknown variables are estimated simultaneously during the reconstruction.

Bayesian inference is a commonly used way to take into consideration the prior information. In the Bayesian context the prior information is modeled by a prior distribution. The convenience by using the Bayesian inference is that the prior properties could be taken into consideration with a variaty of choice of prior distributions, and among the feasible prior distributions we could choose those who simplifies the computation. In Bayesian inference, the **Markovian model** is frequently used to define the **piecewise continuous** variable. The value of each voxel is not independent from the others, but depending on the neighbor voxels.

Thanks to the fact that the object under examination in X-ray CT is composed of several homogeneous materials, the piecewise continuous property of the object is one of the most commonly considered prior information in the X-ray CT reconstruction. Besides, the sparsity of a **hidden variable** which is related with the piecewise continuous variable is also considered frequently. This hidden variable which has a sparsity property can be, for example, the gradient of the volume, the wavelet transformation coefficient of the volume or a dictionary basis coefficient of the volume. The word ”sparsity” implies that the vector contains only few non zero entries. Many sparse transformations are reversible. One can reconstruct the object with the sparse coefficient which contains much less non zero values than the object itself. The ill-posed inverse problem of reconstructing the object can then be transferred to the reconstruction of the sparse coefficients.

2.2 Contributions

An outline of the contribution of this thesis is given in Appendix B.

We propose in this PhD thesis mainly two methods in the Bayesian context. Both of these two methods have a hierarchical system model, with a sparse hidden variable being considered. A sparsity enforcing prior distribution is used to model the hidden variable.

In the X-ray CT reconstruction, the object is typically piecewise continuous, and in some cases piecewise constant. So, two aspects for choosing the sparse hidden variable are considered in our work : the wavelet transformation which is used in the HHBM method, Chapter 5,

and the Laplacian which is used in the ROCC method, Chapter 6. Both of them are sparse, and considered as a hidden variable, associated with a prior distribution which enforces the sparse structure of the hidden variable. A generalized Student-t distribution [Dum16, DLGMD17] is chosen because of its heavy-tailed property, modeling the sparsity property, and it is expressed as the marginal of a Normal-Inverse Gamma distribution. The marginal structure of two conjugate distributions facilitate the estimation of the variables in the Bayesian methods.

Among the wavelet transformations, we choose to use the discrete Haar transformation. For a piecewise constant or piecewise continuous object, the Haar transformation gives a sparse coefficient for the variable. As one of the basic wavelet transformation, the Haar transformation is computationally cheap, and can be efficiently accelerated using the GPU processor. The gradient of the object is considered as a hidden variable in another proposed method, presented in Chapter 6, and a non-homogeneous Markovian model is used to define the variable.

In this thesis, the word "**variables**" represents the unknowns which need to be estimated, for example the object's pixels or voxels values. The word "**hidden variables**" represents the intermediate variables which are often needed to model the object such as contours, region or classification labels or the transformation coefficients in the hierarchical models. The word "**parameters**" represents the parameters which appear in the prior model of the variables such as the variance of the noise, and they are also estimated during the iterations. The word "**hyper-parameters**" are the parameters of the prior distributions of the parameters, and they need to be fixed for the initialization. In this thesis, the initialization of the hyper-parameters is discussed. Instead of testing for different hyper-parameters and choosing the best one, we discuss the way to fix the hyper-parameters directly by taking into consideration the prior models and the prior information on the variables.

2.3 Organization of thesis

In Chapter 3, we present the continuous and discrete model of the CT system, and the direct model of the projection operation. The conventional methods, including the analytical methods, the algebraic methods, the regularization methods, the statistical methods, and the Bayesian inference are presented. The advantages and shortcomings for each different type of methods are discussed. Then we show some simulation results for the CT reconstruction by using two of the most representative methods : the Filtered Back Projection (FBP) technique and the Total Variation (TV) method. The simulation results by using these two methods under different settings of the projection system will be given in this chapter. And I am going to show the influence of the parameter in TV and demonstrate the difficulty to choose this parameter value.

In Chapter 4 we present the context of the Bayesian inference. The prior properties of the object in the industrial applications are discussed. In this chapter, the property of the generalized Student-t distribution is presented and the advantage of using it is discussed.

In the second part of this thesis, we present our main contributions.

In Chapter 5, a **HHBM method** is presented. In this method, first of all a synthesis model is considered with a non-stationary noise model. In this model, only the **Haar transformation** coefficients are estimated, and the object is obtained by a post-processing from the estimated coefficients. Then, in a more general model, the object and the Haar transformation coefficients

are simultaneously estimated with a non-stationary noise model. As the volume consists of several homogeneous blocks, the transform coefficient is sparse. We proposed to use the **generalized Student-t distribution** to model the prior distribution for the coefficients. Once we get the posterior distribution, a **Joint Maximum A Posterior estimator (JMAP)** is used to estimate the unknown variables. Some extensions of this method are also discussed. A **variable splitting** technique is applied on this method and it is shown that by using it, we can consider the noise as the mixture of a Gaussian noise and a sparse noise. Thanks to that, reconstruction results surpasses the original method. Another extension is to compare the Haar transformation with the **Dual-Tree Complex Wavelet Transform (DT-CWT)**. Comparing with the Haar transformation, the DT-CWT transformation takes into account more wavelet directions, and is more robust when the coefficient is contaminated. When they are used in the proposed Bayesian method, the DT-CWT offers a better quality of reconstruction than the Haar transformation.

In Chapter 6, a **ROCC method** is proposed. In this method, the relations between the Laplacian of projection sinogram and the Laplacian of the object is considered. By using the **Laplacian** of the object as a hidden variable, we benefit from its sparse property to preserve the contours of the object. In this method the voxels are modeled by a **non-homogeneous Markovian model**. In this non-homogeneous Markovian model, each voxel is modeled by a distribution which depends on the laplacian values and the voxel values of its neighbor voxels. When its neighbor voxel has a bigger laplacian value, the neighbor voxel value will have less influence on the model of the current voxel.

The reconstruction results, corresponding to a simulated 2D or 3D Shepp Logan phantom, and a 3D Head object, are shown in Chapter 5 and Chapter 6. We compared the proposed methods with the FBP, QR and TV methods. Different settings for the simulations are also considered. Observations with insufficient number of projections, or limited angle of projections, with both high and low SNR, are tested.

The conclusions and some perspectives are given in Chapter 7. In the future work, firstly, other estimators can be use, for example the Posterior Mean (PM) estimator. PM are usually realized by the Variational Bayesian Approach (VBA) method, with details given in Appendix A. By using the VBA method, the computational costs is very expensive for the 3D CT simulations, but our group have recently find a way to simplify the computation. This algorithm could be implemented in our future work. The analysis of the influence of hyper-parameters and the explanations of the robustness of the hyper-parameters in the HHBM method will be continued. Some other Gaussian variance mixture distributions [DGWMD17] will also be used and compare with the $\mathcal{S}t_g$ distribution. For the ROCC method, only the parallel projection case is discussed. We will search for the solutions to extend this method to other projection geometries.

3

X-ray CT and the state-of-the-art methods

3.1 Introduction

The 3D CT reconstruction is treated as a typical inverse problem but with a huge data size. With the increase of the computing speed of computer hardware, the research interest changed from the analytical methods to the iterative methods. The iterative methods are now more and more frequently used in real applications.

This chapter is devoted to the presentation of the reconstruction methods for X-ray CT reconstruction. The Filtered Back-Projection (FBP) based on the Radon transform is still the most commonly used reconstruction algorithm in real applications. Research has been focused on the reconstruction with fewer projections, which makes it an ill-posed inverse problem. In this case, the iterative methods are used to optimize the reconstruction results. Algebraic Reconstruction Technique (ART) [BKK12, BKK12, CM15, GBH70], Simultaneous algebraic reconstruction technique (SART) [AK84, RSD09] and Simultaneous Iterative Reconstruction Technique (SIRT) [BYL+00, WLL14] are a few of those firstly proposed iterative methods. Generally we get an under-determined system of equations for the reconstruction of X-ray CT. In this context, the regularization methods are used such that complementary prior information is combined in the criterion. The statistical method are frequently used as the statistical models can be used to define the unknowns and be estimated by using the statistical estimator. The Maximum Likelihood (ML) methods are one type of the statistical methods. Together with their different algorithms such as the Expectation Maximization (EM) algorithm [Moo96], the stochastic EM (SEM) [TET+04] and the Ordered subsets-EM (OS-EM) [HL94], the ML methods are commonly used in PET-CT reconstruction problems. Another widely used type of statistical methods for PET or X-ray CT reconstruction are the Bayesian methods [AMD10, BMD08, FR12, KVS+06, ZHL+16], and they are attracting more and more attentions in the field of tomography as the computation time is significant reduced by using the GPU processor.

In this chapter, we are going to present the basic analytical methods, the regularization

methods and the Bayesian methods for the X-ray CT reconstruction.

3.2 The X-ray Computed Tomography

The technique of the X-ray CT has been mainly used for medical imaging and for Non Destructive Testing (NDT). Figure 3.1 and Figure 3.2 illustrate some examples of the X-ray CT reconstruction images in industrial and medical applications.

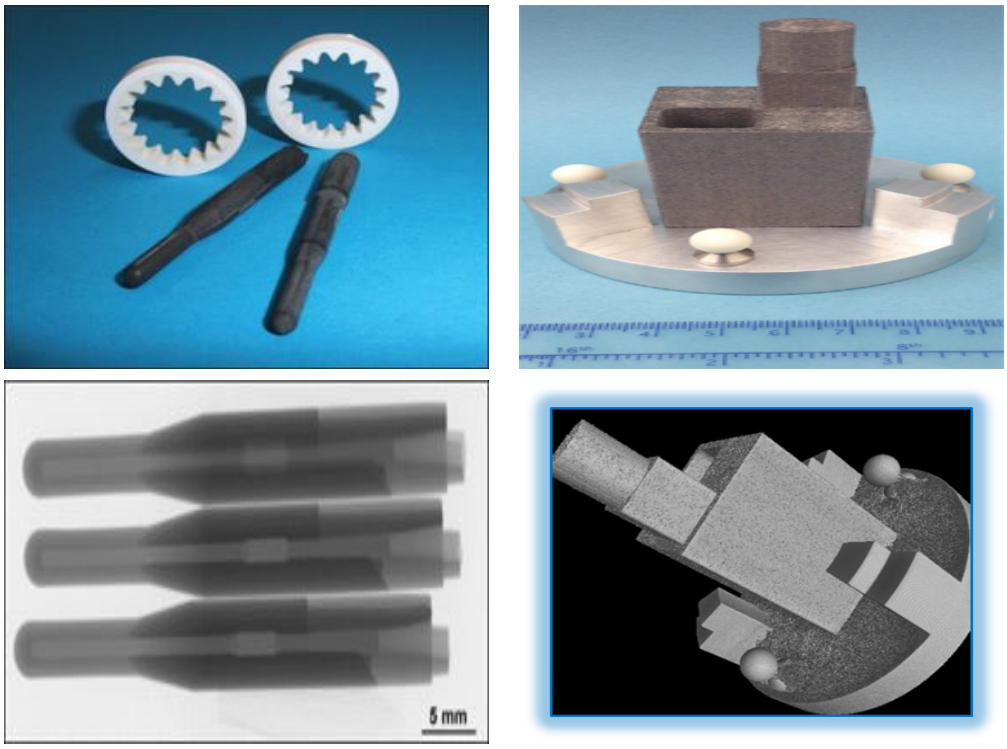


FIGURE 3.1 – The industrial components tested by X-ray CT reconstruction. Top : the outward appearance of the components, bottom : the corresponding reconstructed objects. (Image on the left : http ://www.materialstoday.com/hardmetals-and-ceramics/features/2c-ceramics-moves-into-the-industrial-reality-zone/ ; image on the right : http ://www.npl.co.uk/science-technology/dimensional/x-ray-computed-tomography.)

3.3 The Radon Transform and Analytical reconstruction techniques

The Radon transform is based on the work of the mathematician Johann Radon in 1917 [Rad17]. The problem of reconstructing a function $f(x, y)$ (in 2D case) or $f(x, y, z)$ (in 3D case) from the integrals along different directions is becoming widely used in many domains, for example the seismology, the medical imaging, the astronomy, the industrial Non Destructive testing, etc.

The Radon transformation is expressed by operator $\mathcal{R}$, and g is the corresponding function

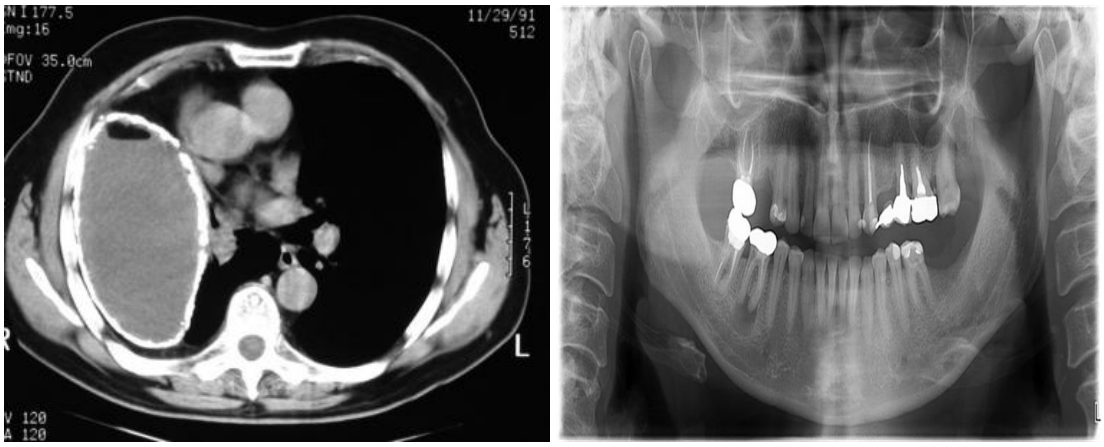


FIGURE 3.2 – The medical images detected by X-ray CT reconstruction. Brain, bone, teeth, etc. (Image on the top : http ://www.southlakeregional.org/Default.aspx ?cid=784 ; image on the bottom left : http ://emedicine.medscape.com/article/355892-overview ; image on the bottom right : http ://texas-dental-implants.com/wp-content/uploads/2015/02/PAN-X-rays-Small.jpeg.)

in the transformation domain :

$$g = \mathcal{R}f. \tag{3.1}$$

In tomography, the function $f(x, y, z)$ represents an unknown density, and the integrals of the function is called the projection, or sinogram.

From the Beer's law, Eq.(2.1), by applying the logarithm in both size, we obtain :

$$\log I = \log I_0 - \int f(l)\, \mathrm{d}l, \tag{3.2}$$

and it leads to the Radon transform :

$$g(r, \phi) = -\log \frac{I}{I_0} = \int_{-\infty}^{+\infty} f(x, y)\, \mathrm{d}x\, \mathrm{d}y. \tag{3.3}$$

Figure 3.3 illustrate the Radon transform of a continuous function $f(x, y)$. For any point (x, y) on the integral line we have :

$$r = x \cos\phi + y \sin\phi. \tag{3.4}$$

And we obtain :

$$g(r, \phi) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\delta\left(r - x\cos\phi - y\sin\phi\right)\, \mathrm{d}x\, \mathrm{d}y. \tag{3.5}$$

Mathematically, the adjoint Radon transform, also known as the dual Radon transform, is expressed as :

$$\widehat{f}(x, y) = \mathcal{R}^{\sharp}\left[g(r, \phi)\right] = \frac{1}{2\pi} \int_0^{\pi} \int_{-\infty}^{+\infty} g(r, \phi)\delta(x\cos\phi + y\sin\phi - r)\, \mathrm{d}\phi\, \mathrm{d}r. \tag{3.6}$$

Figure 3.4 illustrate the Radon transform and its dual transform pair. The dual Radon transform is also called the backprojection (BP) in CT, which is considered as the adjoint operator of the projection. The relation of the backprojection and the real object is :

$$b(x, y) = f(x, y) * *h(x, y), \tag{3.7}$$

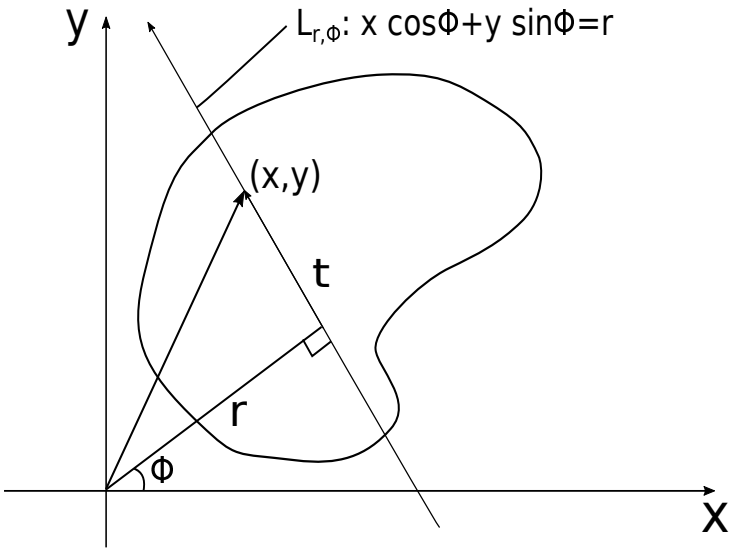


FIGURE 3.3 – The Radon transform of continuous function.

where $**$ is the 2D convolution and

$$h(x,y) = \frac{1}{\sqrt{x^2+y^2}}. \tag{3.8}$$

Geometrically, it is simply a propagation of the measured sinogram back into the image space along the projection paths.

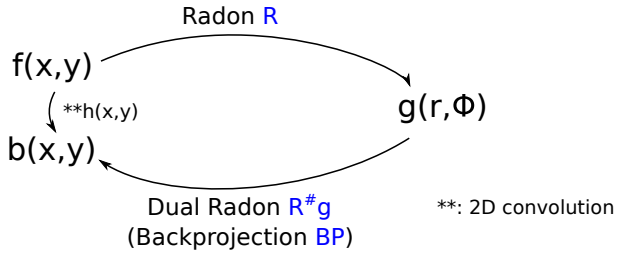


FIGURE 3.4 – The Radon transform and Backprojection operator.

In the following sections, I'm going to describe in a short and synthetic way the analytical relations between $f(x,y)$ and $g(r,\phi)$ which are the basis of many analytical reconstruction methods.

3.3.1 The central slice theorem

The central slice theorem indicates that : the 1D Fourier Transformation (FT) of a projection line $G(\rho,\phi) = \mathcal{F}_{1D(r)}\{g(r,\phi)\}$ is equal to the 2D FT of $F(u,v) = \mathcal{F}_{2D}\{f(x,y)\}$ evaluated at angle ϕ :

$$G(\rho,\phi) = F(u,v)|_{u=\rho\cos\phi, v=\rho\sin\phi} = F(\rho,\phi). \tag{3.9}$$

The relations between the figure 2D $f(x,y)$, the projection sinogram $g(r,\phi)$ and their Fourier Transformations $F(u,v)$, $G(\rho,\phi)$ are illustrated in Figure 3.5.

This property leads to a direct reconstruction strategy : the direct Fourier interpolation method. By using this method the object is reconstructed by the following steps :

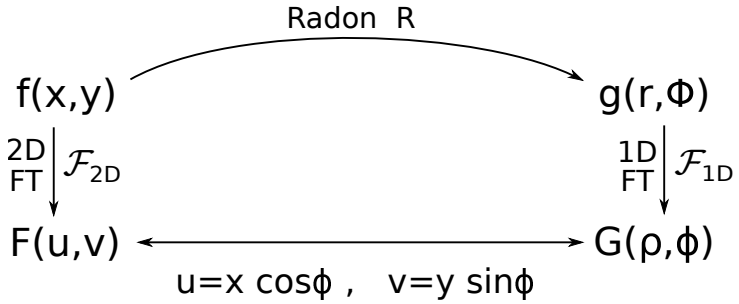


FIGURE 3.5 – The central slice theorem.

1. Apply 1D FT for each projection.
2. Convert from polar coordinates $F(\rho, \phi)$ to Cartesian coordinates $\widehat{F}(u, v)$ using $u = \rho \cos \phi$ and $v = \rho \sin \phi$.
3. Apply inverse 2D FT to $\widehat{F}(u, v)$ and obtain $\widehat{f}(x, y)$.

3.3.2 The Back-projection filtering (BPF) algorithm

As presented previously, the back-projection is the adjoint operator of the Radon transform :

$$b(x, y) = \frac{1}{2\pi} \int_0^{\pi} \int_{-\infty}^{+\infty} g(r, \phi) \delta(x \cos \phi + y \sin \phi - r) \, \mathrm{d}r \, \mathrm{d}\phi. \tag{3.10}$$

Consider that :

$$g(r, \phi) = \mathcal{F}_{1D(\rho)}^{-1} \{F(\rho, \phi)\} = \int_{-\infty}^{+\infty} F(\rho, \phi) \exp \{i 2\pi \rho r\} \, \mathrm{d}\rho, \tag{3.11}$$

we have :

$$\begin{aligned} b(x, y) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_0^{\pi} \int_{-\infty}^{+\infty} F(\rho, \phi) \delta(x \cos \phi + y \sin \phi - r) \exp \{i 2\pi \rho r\} \, \mathrm{d}r \, \mathrm{d}\phi \, \mathrm{d}\rho \\ &= \frac{1}{2\pi} \int_0^{\pi} \int_{-\infty}^{+\infty} F(\rho, \phi) \exp \{i 2\pi \rho (x \cos \phi + y \sin \phi)\} \, \mathrm{d}\phi \, \mathrm{d}\rho. \end{aligned} \tag{3.12}$$

From the definition of FT we have $F(-\rho, \phi) = F(\rho, \phi + \pi)$, so we obtain :

$$\begin{aligned} b(x, y) &= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{+\infty} \frac{F(\rho, \phi)}{\rho} \exp \{i 2\pi \rho (x \cos \phi + y \sin \phi)\} \, \rho \, \mathrm{d}\phi \, \mathrm{d}\rho \\ &= \frac{1}{2\pi} \mathcal{F}_{2D}^{-1} \left\{ \frac{F(\rho, \phi)}{\rho} \right\}. \end{aligned} \tag{3.13}$$

From the obtained relation we get : $f_{BPF}(x, y) = \mathcal{F}_{2D}^{-1} \left\{ \rho \mathcal{F}_{2D} \left\{ \widehat{f}(x, y) \right\} \right\}$, and the Back-projection filtering method implies the following steps :

1. Back-projection from $g(r, \phi)$ to get $b(x, y)$.
2. Apply 2D FT for $b(x, y)$ and we get $\frac{F(\rho, \phi)}{\rho} = \mathcal{F}_{2D} \{b(x, y)\}$.
3. Filter with ρ and obtain $F(\rho, \phi)$.
4. Apply inverse 2D FT to get $\widehat{f}_{BPF}(x, y)$.

The BPF algorithm is illustrated in Figure 3.6.

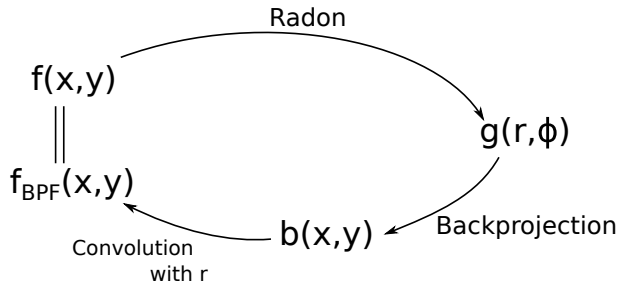


FIGURE 3.6 – The Back-projection Filtering reconstruction algorithm.

3.3.3 The Filtered Back-projection

The Filtered Back-Projection (FBP), illustrated in Figure 3.7, is a conventional analytical technique for the reconstruction, and it is the most commonly used technique in real medical and industrial applications. In FBP, the projection data is convoluted by a filter function in the Fourier space, and then a back-projection is applied. Mathematically it is presented by $\widehat{f}_{FBP} = \boldsymbol{B}\mathcal{F}_{1D}^{-1}\left|\rho\right|\mathcal{F}_{1D}g$ where $\boldsymbol{B}$ represents the back-projection. Comparing with back-projection, FBP compensates the high frequency component of the phantom, and derives results with much clearer contours than the back-projection.

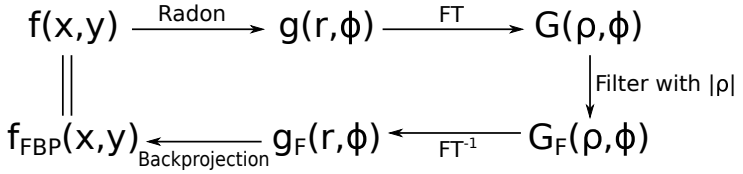


FIGURE 3.7 – The Filtered Back-projection reconstruction algorithm.

The algorithm of the FBP method implies the following two steps :

1. Filter the projection with $|\rho|$ in the Fourier domain or by a convolution method.
2. Apply the back-projection.

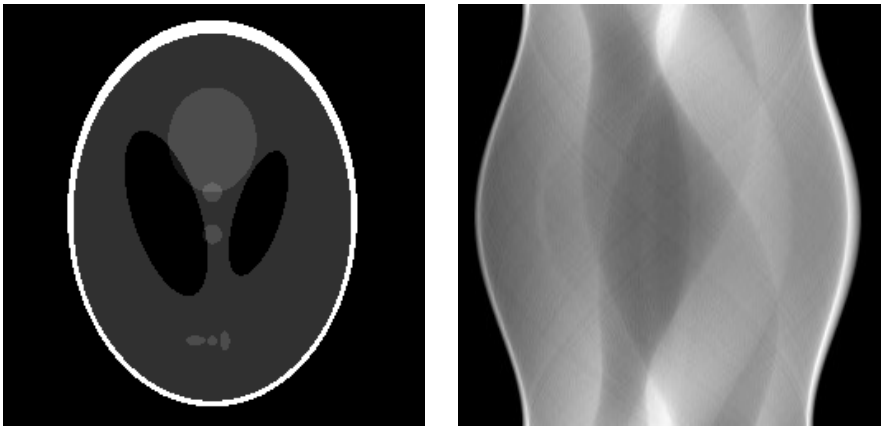
With these two steps we have :

$$\begin{aligned}
\widehat{f}_{FBP}(x,y) &= \int_0^{\pi}\int_{-\infty}^{+\infty}\mathcal{F}_{1D}^{-1}\left\{\left|\rho\right|\mathcal{F}_{1D}\left\{g(r,\phi)\right\}\right\}\delta\left(x\cos\phi+y\sin\phi-r\right)\,\mathrm{d}r\,\mathrm{d}\phi \\
&= \int_0^{\pi}\int_{-\infty}^{+\infty}\mathcal{F}_{1D}^{-1}\left\{\left|\rho\right|F(\rho,\phi)\right\}\delta\left(x\cos\phi+y\sin\phi-r\right)\,\mathrm{d}r\,\mathrm{d}\phi \\
&= \int_{-\infty}^{+\infty}\int_0^{\pi}\int_{-\infty}^{+\infty}\left|\rho\right|F(\rho,\phi)\exp\left\{i2\pi\rho r\right\}\delta\left(x\cos\phi+y\sin\phi-r\right)\,\mathrm{d}\rho\,\mathrm{d}r\,\mathrm{d}\phi \\
&= \int_0^{\pi}\int_{-\infty}^{+\infty}F(\rho,\phi)\exp\left\{i2\pi\rho\left(x\cos\phi+y\sin\phi\right)\right\}\left|\rho\right|\,\mathrm{d}\rho\,\mathrm{d}\phi \\
&= \int_0^{2\pi}\int_0^{+\infty}F(\rho,\phi)\exp\left\{i2\pi\rho\left(x\cos\phi+y\sin\phi\right)\right\}\rho\,\mathrm{d}\rho\,\mathrm{d}\phi \\
&= \mathcal{F}_{2D}^{-1}\left\{F(\rho,\phi)\right\} = f(x,y).
\end{aligned} \tag{3.14}$$

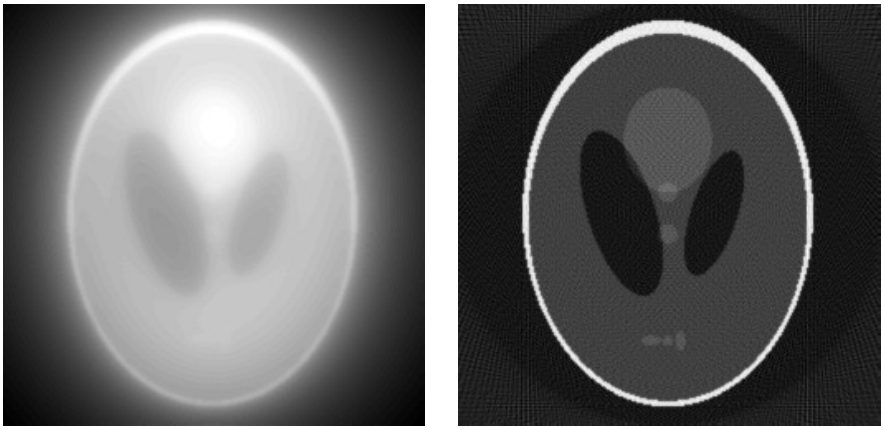
In the real applications, a low-pass filter, for example the Hanning filter, is used during the filter for the projection in order to reduce the noise.

However, the presented analytical reconstruction methods, for example the most commonly used FBP method, have a good performance only when there is sufficient number of projections and that the projections are distributed from 0 to π. We can see from the results that, when there are 64 projections distributed from 0 to π, the reconstructed figure is clear and we can distinguish the details. When there are only 32 projections, distributed from 0 to π, some obvious artifacts appears in the reconstructed figure. In the case where the angle of projection is limited, the reconstruction has a poor quality.

Figure 3.8 shows the original Shepp Logan figure of size 256^2 and the projection sinogram with 64 projections evenly distributed in $[0, \pi]$. The reconstruction results by using the conventional Back-projection (BP) and Filtered Back-projection (FBP) methods are also shown.



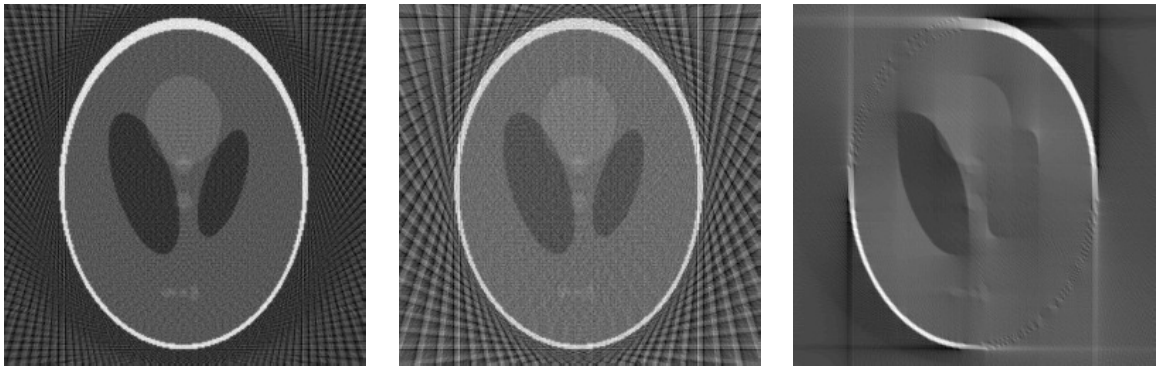
Original $f(x, y)$ Sinogram $g(r, \phi)$



BP $\widehat{f}_{BP}(x, y)$ FBP $\widehat{f}_{FBP}(x, y)$

FIGURE 3.8 – The Original Shepp Logan figure, the projection sinogram without noise by using 64 projections and the reconstruction results by using the Back-projection and Filtered Back-projection methods.

In Figure 3.9, the reconstruction of the Shepp Logan figure by using the FBP algorithm is presented. As a conventional analytical method, we can see that when the number of projections is sufficient and the projections are obtained from 0 to π, the reconstruction result is clear. However when the number of projections or the angle of projections are insufficient, the reconstruction results by using the FBP method are blurred and some obvious artifacts appear in the results. Another shortcoming of the FBP method is its sensitiveness to the noise, because of that the high frequency parameters corresponding to the noise are not filtered in the projection



$\widehat{f}$ with 64 proj $\in [0, \pi]$ $\widehat{f}$ with 32 proj $\in [0, \pi]$ $\widehat{f}$ with 128 proj $\in \left[0, \frac{\pi}{2}\right]$

FIGURE 3.9 – The reconstructed figure by using 64 projections distributed in $[0, \pi]$, 32 projections distributed in $[0, \pi]$ and 64 projections distributed in $[0, \frac{\pi}{2}]$.

domain.

3.4 The algebraic methods

In imaging problems, the function $f(x, y)$ is typically considered in a discrete form, and is presented by a vector $\boldsymbol{f}$. The projection is presented by vector $\boldsymbol{g}$. In a discrete system, the integral operation becomes the sum. Figure 3.10 illustrates the Radon transform of a discretized 2D image. After discretization, the image is represented by pixels. The value of each pixel is the average of the continuous function $f(x, y)$ in the pixel area.

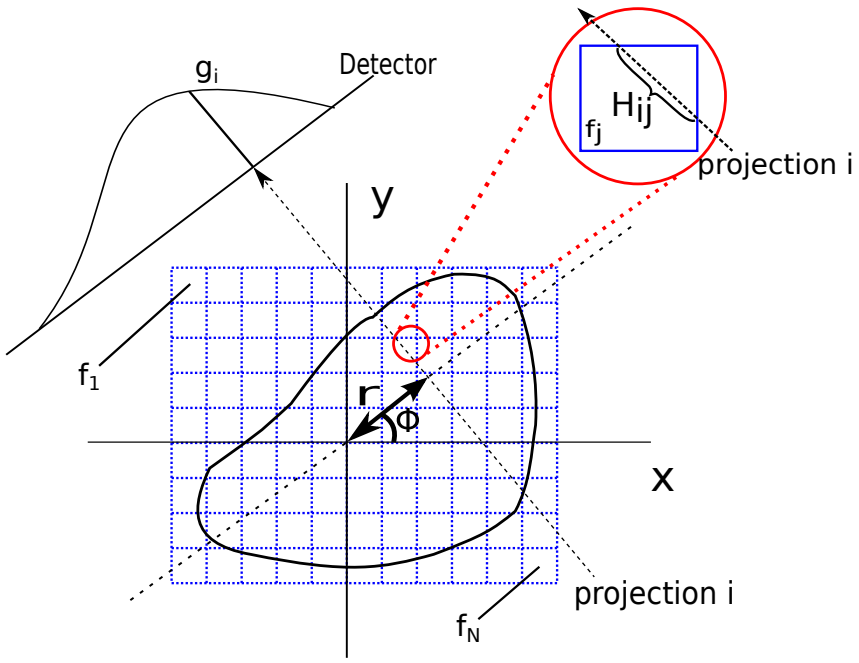


FIGURE 3.10 – The Radon transform of discrete system.

From the figure, we can see that the projection line passes the pixels, with different length of segment in different pixels. We use H_{ij} to represent the length of the ith projection line in

the pixel j. According to the definition, we have :

$$g_i = \sum_{j=1}^{N} H_{ij} f_j, \tag{3.15}$$

where $j \in [1, N]$ and N is the size of the image $\boldsymbol{f}$. It can also be expressed as :

$$g_i = [H_{i1}\ H_{i2}\ \cdots\ H_{iN}] \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{bmatrix}. \tag{3.16}$$

We use the vector $\boldsymbol{g} \in \mathbb{R}^{M\times 1}$ to represent the dataset of the projection, $\boldsymbol{g} = [g_1,\ g_2,\ \cdots,\ g_M]^T$, and we can obtain :

$$\begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_M \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & \cdots & H_{1N} \\ H_{21} & H_{22} & \cdots & H_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ H_{M1} & H_{M2} & \cdots & H_{MN} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{bmatrix}, \tag{3.17}$$

and it can be expressed as :

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}. \tag{3.18}$$

This expression is used as the forward model of the projection system. Each row of matrix $\boldsymbol{H}$ corresponds to a projection line, considering every pixel of the image. Each column of matrix $\boldsymbol{H}$ corresponds to all the projection lines in all angles for one pixel f_j.

Consequently, the transpose of the matrix $\boldsymbol{H}$ is used as the back-projection operator :

$$\widehat{\boldsymbol{f}}_{BP} = \boldsymbol{H}^T \boldsymbol{g}. \tag{3.19}$$

Based on the discrete forward model of the projection system in Eq.(3.18), the algebraic methods are used, such as the algebraic reconstruction technique (ART) [Gor74, GBH70], the related simultaneous ART (SART) [AK84] and the Simultaneous Iterative Reconstruction Technique (SIRT) [TL90, BYL+00]. The update step for a voxel f_j in ART, SART and SIRT are respectively expressed as :

$$\text{ART} : f_j^{(k+1)} = f_j^{(k)} + \lambda \frac{g_i - \sum_{n=1}^{N} H_{in} f_n^{(k)}}{\sum_{n=1}^{N} H_{in}^2} H_{ij}, \tag{3.20}$$

$$\text{SART} : f_j^{(k+1)} = f_j^{(k)} + \lambda \frac{\sum_{g_i \in G_\phi} \left(\frac{g_i - \sum_{n=1}^{N} H_{in} f_n^{(k)}}{\sum_{n=1}^{N} H_{in}} \right) H_{ij}}{\sum_{g_i \in G_\phi} H_{ij}}, \tag{3.21}$$

$$\text{SIRT} : f_j^{(k+1)} = f_j^{(k)} + \lambda \frac{\sum_{g_i \in G} \left(\frac{g_i - \sum_{n=1}^{N} H_{in} f_n^{(k)}}{\sum_{n=1}^{N} H_{in}} \right) H_{ij}}{\sum_{g_i \in G} H_{ij}}, \tag{3.22}$$

where λ is a relaxation parameter and H_{ij} is the weight with which the voxel j contributes its value to the projection i. In ART, only one ray in one projection direction is considered for one iteration of voxel j. In SART, all rays in one projection G_ϕ in direction ϕ is considered for one

iteration. In SIRT, all rays in all projections are considered for one iteration. A Discrete Algebraic Reconstruction Technique (DART) is presented in [BS11]. The DART technique considers the reconstruction of the strictly piecewise constant objects, which consists of several different homogeneous materials, and the gray level of each different material is known. For this special types of problems, the DART method has a very good performance with insufficient number of projections and is robust to the noise. However, in our work, we consider a more general case, where we have no information on the number of materials and the gray level of each material.

In Figure 3.11, we show the reconstructed Shepp Logan figure of size 256^2 by using the SIRT method. From the results we can see that, when there are sufficient number of projections, the shape of the reconstructed phantom is clear, but the contours in the phantom is a little blurring. When there is insufficient number of projections, there will be many artifacts in the reconstructed figure.

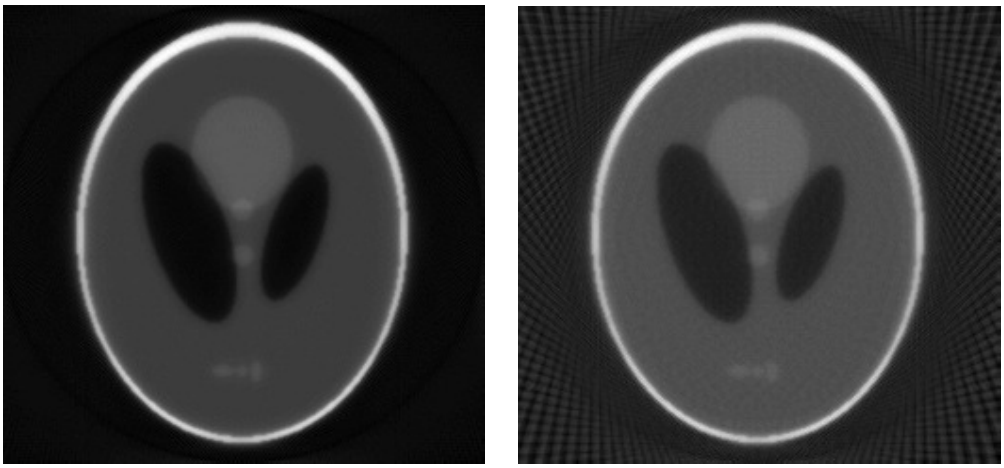


FIGURE 3.11 – Reconstruction of 2D Shepp Logan phantom of size 256^2 by using the SIRT method, with respectively 180 projections (left) and 60 projections (right).

From the definition and the simulation results, we can see that the limitation of the algebraic reconstruction techniques is the very costly computation for the calculation of the weighted sum. What's more, the size of detected data $\boldsymbol{g}$ will have a big influence on the reconstruction result. In the ill-posed reconstruction problems, where the number of observed data is much less than the unknown data, the prior information need to be considered during the reconstructions.

In the following section, I am going to present the regularization methods, which are fundamental and widely used methods concerning the prior information.

3.5 The Regularization methods

In X-ray CT, generally the object under testing needs to be reconstructed with precise details and the dimension of voxel is very small compared with the size of object. With a large size of unknown variable, the acquisition of sufficient dataset is too expensive. In most of the cases we get an under-determined system of equations ($N > M$). The following direct model is used to account for the modeling error and measurement noise :

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}, \tag{3.23}$$

where $\boldsymbol{\epsilon} \in \mathbb{R}^{M\times 1}$ corresponds to all the additive noise and uncertainties of the projection system.

The regularization methods are derived from this model. In regularization methods, we optimize a criterion which consists of generally two terms : the data consistency $Q(\boldsymbol{g}, \boldsymbol{f})$ and the regularization term $R(\boldsymbol{f})$:

$$J(\boldsymbol{f}) = Q(\boldsymbol{g}, \boldsymbol{f}) + \lambda R(\boldsymbol{f}), \tag{3.24}$$

where λ is called the regularization parameter.

The data consistency term $Q(\boldsymbol{g}, \boldsymbol{f})$ describes the discrepancy between the observation and estimation. Based on direct model defined in Eq.(3.24), it is commonly under form of $Q(|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}|)$. The l_2 norm is mostly used in the regularization methods, with :

$$Q(\boldsymbol{g}, \boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|_2^2 . \tag{3.25}$$

The L_q norm is a general form, with $Q(\boldsymbol{g}, \boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|_q^q$, for example $q = 1$.

The diversification of the criterion of regularization methods is normally on the regularization term $R(\boldsymbol{f})$. This term is fixed by considering the prior information that we known about the unknown variable $\boldsymbol{f}$. We can see that when $R(\boldsymbol{f}) = 0$, we obtain a Least Square (LS) method [YAT+97, SC00].

3.5.1 Different regularization criterion

The regularization term $R(\boldsymbol{f})$ is typically a penalty on the complexity criterion of $\boldsymbol{f}$. By choosing different regularization function $R(\boldsymbol{f})$, different regularization methods are referred to.

When $R(\boldsymbol{f}) = \|\Phi(\boldsymbol{f})\|_2^2$, the ridge penalty, or the Quadratic Regularization (QR) [RSD09] is derived, where the linear operator $\Phi(\cdot)$ can be the multiplication by an identical matrix which lead to $R(\boldsymbol{f}) = \|\boldsymbol{f}\|_2^2$. $\Phi(\cdot)$ can also be the gradient operator. The L_2 form of the regularization term will result in the smoothness in the final results.

The Total Variation (TV) method [CL97, GO09, PTS+15] is defined by $R(\boldsymbol{f}) = \|\boldsymbol{D}_{TV}\boldsymbol{f}\|_{TV}$ where $\boldsymbol{D}_{TV}$ is the gradient operator. In the anisotropic form, $R(\boldsymbol{f}) = \|\boldsymbol{D}_x\boldsymbol{f}\|_1 + \|\boldsymbol{D}_y\boldsymbol{f}\|_1$ for 2D image and $R(\boldsymbol{f}) = \|\boldsymbol{D}_x\boldsymbol{f}\|_1 + \|\boldsymbol{D}_y\boldsymbol{f}\|_1 + \|\boldsymbol{D}_z\boldsymbol{f}\|_1$ for 3D object, where $\boldsymbol{D}_x$, $\boldsymbol{D}_y$ and $\boldsymbol{D}_z$ are respectively the gradient in the x, y and z direction of the Cartesian coordinate system. L_1 norm is used for sparse estimations, which enforces the sparsity of the $\boldsymbol{D}_{TV}\boldsymbol{f}$ term.

Sometimes there are more than one regularization terms in the criterion. For example in the Elastic-net regularization method [WVVH10], the regularization term is $R(\boldsymbol{f}) = \lambda_1 \|\boldsymbol{f}\|_1 + \lambda_2 \|\boldsymbol{f}\|_2^2$. The elastic net penalty is a convex combination of the lasso and ridge penalty. When $\lambda_1 = 0$, it becomes a simple ridge regression.

More general regularization methods are developed basing on constrained and dual-variable regularization method :

$$J(\boldsymbol{f}, \boldsymbol{z}) = Q_1(\boldsymbol{g}, \boldsymbol{f}) + \eta Q_2(\boldsymbol{f}, \boldsymbol{z}) + \lambda R(\boldsymbol{z}). \tag{3.26}$$

In such a model, the penalty regularization term is set on $\boldsymbol{z}$, which is associated with $\boldsymbol{f}$ via a linear transformation. The loss function $Q_1(\boldsymbol{g}, \boldsymbol{f})$ and $Q_2(\boldsymbol{f}, \boldsymbol{z})$ are generally quadratic, i.e. $Q_1(\boldsymbol{g}, \boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|_2^2$ and $Q(\boldsymbol{f}, \boldsymbol{z}) = \|\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}\|_2^2$. They describe respectively the forward model consistency and the prior model consistency. $R(\boldsymbol{z})$ is the regularization term on $\boldsymbol{z}$.

3.5.2 Optimization algorithms

In the problem of optimizing large data size variable, the optimization result sometimes contains the inversion of big size matrix, which is impossible to compute. In these cases, the optimization algorithms are often used. The optimization of the variables are derived by minimizing the loss function $J(\boldsymbol{f})$ with respect to the unknowns.

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{J(\boldsymbol{f})\}. \tag{3.27}$$

The optimization algorithms fall in two major categories : the first order optimization algorithms and the second order optimization algorithms.

The first order optimization algorithms minimize or maximize the loss functions by considering their gradient. The Gradient Descent (GD) algorithm [Bat92, VCR97], which is the most important technique and the foundation of the optimization algorithms, belongs to the first order optimization category. The GD algorithm takes steps proportional to the negative of the gradient of the loss function at the current point and iteratively reaches the local minimum or maximum of the loss function :

$$\boldsymbol{f}^{(k+1)} = \boldsymbol{f}^{(k)} - \gamma \nabla J(\boldsymbol{f}), \tag{3.28}$$

where $\nabla J(\boldsymbol{f})$ is the derivative of loss function and γ is called the descent step length.

Though the most widely used algorithm, the GD algorithm has some drawbacks, for example the convergence is slow and the final result might be the local minimum. Many other first order algorithms are developed to overcome these problems, for example the conjugate gradient algorithm [Bat92, Sca87, FB99], the stochastic gradient descent algorithm [Bot10, NWS14, KPSV09], etc.

The second order optimization algorithms are less frequently used, as they use the second order derivative, also known as the Hessian, to minimize or maximize the loss function, and is more costly than the first order derivative. However, in some cases the form of the second order derivative is known, they can be used for the big data size optimization problems. The most famous second order optimization algorithm is the Newton's method [Kel99, Fis92].

In gradient based optimization algorithms, the derivative of the criterion function always need to be calculated. Note that the L_1 norm is not derivable everywhere, the conventional optimization algorithms can not be used. Many methods have been proposed to solve this L_1 norm optimization problem, for example the Primal Dual method [CGM99], the Split Bregman method [GO09] and so on.

Other methods are also used for the large scale regularization methods, for example the augmented Lagragian method [TW09, ABDF11], the Alternating Direction Method of Multipliers (ADMM) [WBAW12], etc. These methods are often applied to solve the constrained optimization problems. The ADMM method considers to minimize $\Phi(\boldsymbol{f}) + \Psi(\boldsymbol{z})$ subject to $\boldsymbol{A}\boldsymbol{f} + \boldsymbol{B}\boldsymbol{z} = \boldsymbol{C}$, and it covers a large amount of estimation forms. An example is when $\Phi(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|_2^2$, $\Psi(\boldsymbol{z}) = R(\boldsymbol{z})$, $\boldsymbol{A} = \boldsymbol{I}$, $\boldsymbol{B} = -\boldsymbol{D}$ and $\boldsymbol{C} = 0$, referring to the above mentioned bi-variable regularization method corresponding to Eq.(3.26).

3.5.3 The regularization parameter

In the above mentioned regularization methods, there is always a regularization parameter λ to be fixed. This parameter compromise between not having enough regularization to reduce the noise and too much regularization that the solution becomes too smooth. Sometimes there are more than one regularization terms in the minimization criterion, for example the Elastic-net method. Hence two or even more regularization parameters are introduced. These parameters tune the terms of regularization. The value of these parameters will effect on the optimization result.

Figure 3.12 and Figure 3.13 demonstrate respectively the influence of the regularization parameter λ in the QR and TV method for the reconstruction of the Shepp Logan phantom. From the figures we can see that the optimal regularization parameter value is different for the cases with different number of projections or different SNR.

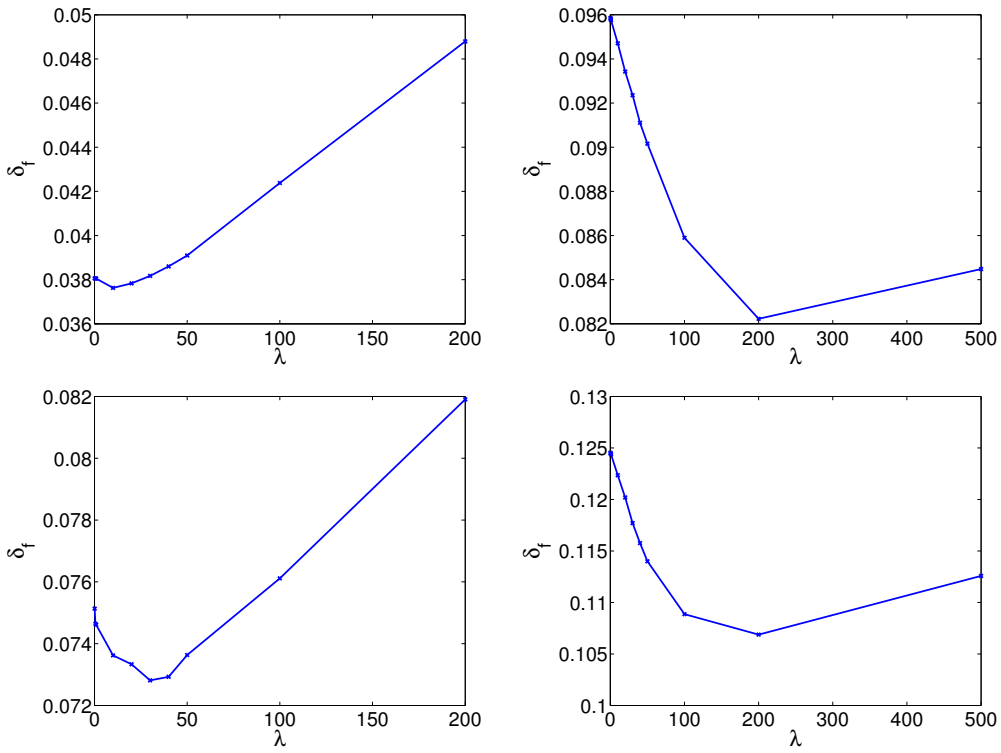


FIGURE 3.12 – The influence of the regularization parameter on the relative mean square error of the reconstruction results in the **Quadratic Regularization** method with 128 projections (top) or 64 projections (bottom) with a noise of SNR=40dB (left) or SNR=20dB (right).

In order to fix the regularization parameter, many method have been proposed. A most intuitive strategy is to fix them manually. In [KVS+06], the authors proposed to fix the parameters with training dataset with similar distribution structure as the unknown variable. They demonstrated that a fixed parameter usually performs well for similar imaging data. The reconstruction program should be done many times for the training object in order to find the suitable parameter. What's more, the parameter is not ensured to be suitable for the same structured data with different noise level.

The Generalized Cross Validation (GCV) [GHW79, RLR+12] and L-curve [CMRS00, HO93] methods are also proposed to determine suitable values for this parameter. The prin-

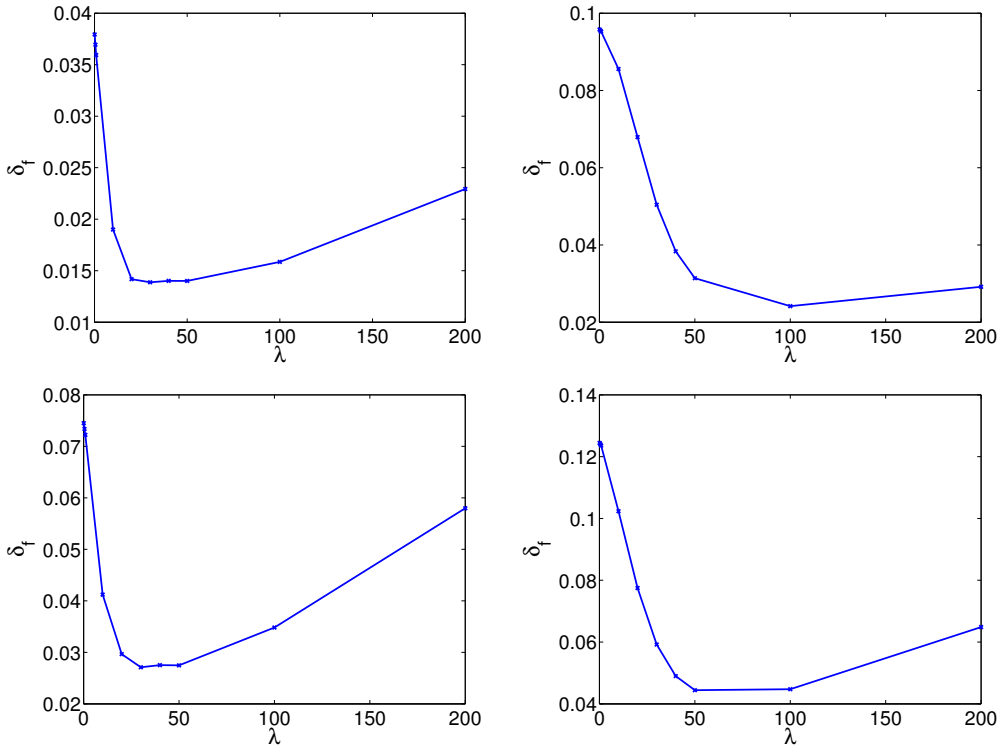


FIGURE 3.13 – The influence of the regularization parameter on the relative mean square error of the reconstruction results in the **Total Variation** method with 128 projections (top) or 64 projections (bottom) with a noise of SNR=40dB (left) or SNR=20dB (right).

cipal strategy of the L-curve method is to set the data consistency term and the regularization term of the criterion as the abscissa and the ordinate respectively, and by setting different value of the parameter λ, we can get a curve which has a 'L' shape. The point which has the biggest curvature corresponds to the suitable parameter.

In the GCV method, a GCV function is defined depending on parameter λ. For example for the Tikhonov regularization $J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|_2^2 + \lambda \|\boldsymbol{D}\boldsymbol{f}\|_2^2$, the GCV function is :

$$\mathcal{G}(\lambda) = \frac{\left\| \left(\boldsymbol{H} \left(\boldsymbol{H}^T \boldsymbol{H} + \lambda \boldsymbol{D}^T \boldsymbol{D} \right)^{-1} - \boldsymbol{I} \right) \boldsymbol{g} \right\|_2^2}{\left(\boldsymbol{T}r \left(\boldsymbol{I} - \boldsymbol{H} \left(\boldsymbol{H}^T \boldsymbol{H} + \lambda \boldsymbol{D}^T \boldsymbol{D} \right)^{-1} \boldsymbol{H}^T \right) \right)^2}. \tag{3.29}$$

Then the optimal parameter value is estimated by minimizing the GCV function. However, in many cases the GCV function does not have an explicit minimum point and the parameter obtained by the minimization is not accurate. For the reconstruction problem of large data size object, this method is too costly.

3.6 The statistical methods

When data are corrupted by noise, statistical methods are often used. With the development of Computer Science, the statistical inference [WSQ09] are used in X ray tomography reconstruction. For the inverse problems which consider a large amount of data, the statistical

methods set a probability model. Instead of estimating all the parameters, the parametric model is examined.

In the forward model presented in Eq.(3.23), the noise $\boldsymbol{\epsilon}$ is modeled by a probabilistic density function (pdf). In tomography, the acquisition of photon count is obtained by detecting the photons arrived at the detectors, which is modeled by a Poisson distribution :

$$p(\boldsymbol{g}|\boldsymbol{f}) = \prod_{i=1}^{M} \frac{[\boldsymbol{H}\boldsymbol{f}]_i^{g_i}}{g_i!} e^{-[\boldsymbol{H}\boldsymbol{f}]_i}. \tag{3.30}$$

where $\boldsymbol{g}_i$ is the detected photon rate, as shown in Figure 3.14.

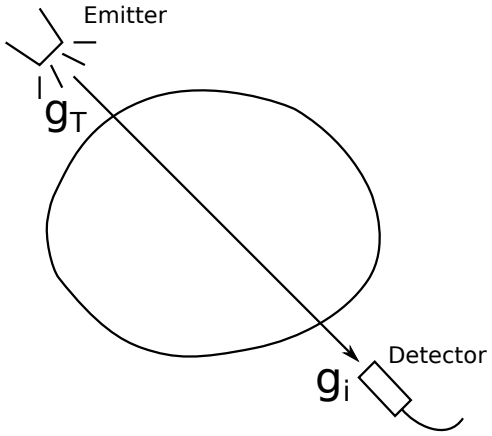


FIGURE 3.14 – One X ray projection from emitter to detector.

In X-ray CT, the amount of photon is very big. With the large amount of photon, the Poisson distribution can be approximate by a Gaussian distribution [FDK84].

The likelihood plays an important role in the statistical inference. It describes the probability density function of parameters of the model knowing data. For a forward model shown in Eq.(3.23), the likelihood is $p(\boldsymbol{g}|\boldsymbol{f})$. In X-ray CT, it is modeled by a Gaussian distribution :

$$p(\boldsymbol{g}|\boldsymbol{f}) = \mathcal{N}(\boldsymbol{g}|\boldsymbol{H}\boldsymbol{f}, \boldsymbol{V}_\epsilon), \tag{3.31}$$

where $\boldsymbol{V}_\epsilon = \text{diag}\,[\boldsymbol{v}_\epsilon]$, $\boldsymbol{v}_\epsilon = [v_{\epsilon_1}, \cdots, v_{\epsilon_i}, \cdots, v_{\epsilon_M}]$ and v_{ϵ_i} is the variance of noise of projection i.

When $v_{\epsilon_i} = v_\epsilon, \ \forall i$, it refers to a stationary noise. However for the X-ray CT applications, the noise depends on the projection system and also the object. The shape and the material structure of the object will have an influence of the detection noise. Hence a non-stationary noise is considered.

By using the Maximum Likelihood (ML) methods [RW84], the statistical parameters are estimated by maximizing the likelihood of making the observations given the parameters :

$$\widehat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \{p(\boldsymbol{g}|\boldsymbol{f})\}. \tag{3.32}$$

In the Emission Computed Tomography, the detection of the photons is modeled by the Poisson distribution, and so the logarithm of the likelihood is :

$$\ln p(\boldsymbol{g}|\boldsymbol{f}) = \sum_{i=1}^{M} \left(-\sum_{n=1}^{N} H_{in} f_n + g_i \ln \left(\sum_{n=1}^{N} H_{in} f_n \right) - \ln g_i! \right). \tag{3.33}$$

As demonstrated in [VSK85], the EM algorithm of this ML method can be written as :

$$\begin{aligned} f_j^{(k+1)} &= f_j^{(k)} \left(1 + \frac{\sum_{i \in \mathcal{I}_j} \left(\frac{g_i - \sum_{n=1}^{N} H_{in} f_n^{(k)}}{\sum_{n=1}^{N} H_{in} f_n^{(k)}} \right) H_{ij}}{\sum_{i \in \mathcal{I}_j} H_{ij}} \right) \\ &= f_j^{(k)} + f_j^{(k)} \frac{\sum_{i \in \mathcal{I}_j} \left(\frac{g_i - \sum_{n=1}^{N} H_{in} f_n^{(k)}}{\sum_{n=1}^{N} H_{in} f_n^{(k)}} \right) H_{ij}}{\sum_{i \in \mathcal{I}_j} H_{ij}}, \end{aligned} \tag{3.34}$$

where $\mathcal{I}_j$ is the set of projections to which the pixel j contributes. We can see that this result has a very similar form as the SIRT method shown in Eq.(3.22).

As mentioned before, the Poisson noise is approximated by a Gaussian noise in X-ray CT. By using the Gaussian model, the maximization of the likelihood is equivalent to the weighted Least Square regression method :

$$\begin{aligned} \widehat{\boldsymbol{f}} &= \arg\max_{\boldsymbol{f}} \left\{ \mathcal{N}(\boldsymbol{g} | \boldsymbol{f}, \boldsymbol{V}_\epsilon) \right\} \\ &= \arg\max_{\boldsymbol{f}} \left\{ |\boldsymbol{V}_\epsilon|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}) \boldsymbol{V}_\epsilon^{-1} (\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}) \right\} \right\} \\ &= \arg\min_{\boldsymbol{f}} \left\{ (\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}) \boldsymbol{V}_\epsilon^{-1} (\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}) \right\} \\ &= \arg\min_{\boldsymbol{f}} \left\{ \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|_{\boldsymbol{V}_\epsilon^{-1}} \right\}, \end{aligned} \tag{3.35}$$

where $\boldsymbol{V}_\epsilon^{-1}$ is the weight of the Least Square method.

The Expectation Maximization (EM) [Moo96] algorithms are used to solve the Maximum Likelihood estimations when the equations are difficult to solve directly. Typically the models depends on some other unobserved hidden variables. The EM iteration alternates between performing an expectation (E) step, and a maximization (M) step.

Some of its extensions, for example the stochastic EM (SEM) [TET+04] or the Ordered subsets-EM (OS-EM) [HL94] are also used for the tomography reconstruction problems.

The shortcomings of the Maximum Likelihood methods are that, the numerical estimation is usually non-trivial, except for a few cases where the maximum likelihood formulas are simple, and that they do not consider any prior information. Consequently, they are not robust in the ill-posed reconstruction systems where the observations are much less the unknowns. Therefore, we present the Bayesian methods, which combine the model and the prior information together to estimate the unknowns.

3.7 The Bayesian inference

The Bayesian inference [SB92, GRL+93, MD96, BMD08, GCSR14, ZFR15] is a type of statistical methods, which is widely used in the inverse problems. In Bayesian inference, Bayes' theorem is used to update the probability for a hypothesis as more evidence or information becomes available. The Bayes' rule derives the posterior distribution as a consequence of the

prior law and the likelihood, with the basic expression :

$$p(\boldsymbol{f}|\boldsymbol{g}) = \frac{p(\boldsymbol{g}|\boldsymbol{f})p(\boldsymbol{f})}{p(\boldsymbol{g})}, \tag{3.36}$$

with

$$p(\boldsymbol{g}) = \int p(\boldsymbol{g}|\boldsymbol{f})p(\boldsymbol{f})\,\mathrm{d}\boldsymbol{f}, \tag{3.37}$$

where $p(\boldsymbol{g}|\boldsymbol{f})$ is the likelihood, $p(\boldsymbol{f})$ is the prior law and $p(\boldsymbol{f}|\boldsymbol{g})$ is the posterior law. $p(\boldsymbol{g})$ is called the marginal likelihood or the evidence. As with the marginalization the variable $\boldsymbol{f}$ doesn't appear in this term, so it is normally not considered in the determination of the variables. Figure 3.15 shows the generative graph of the basic Bayesian model. In this model, the parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are fixed and are considered as known constant in the model.

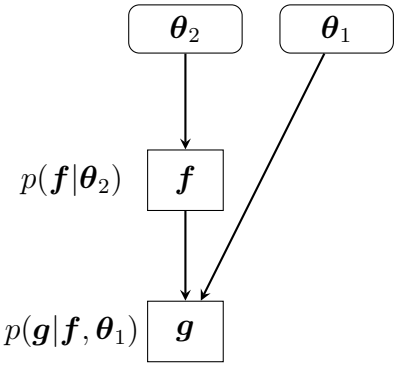


FIGURE 3.15 – Generative graph of the basic Bayesian model.

In the supervised or semi-supervised systems, some parameters are also estimated. The unknown variables contains not only $\boldsymbol{f}$ but also the parameters $\boldsymbol{\theta}$. In this case, Bayes' rule derives the joint posterior probabilistic law :

$$p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g}) = \frac{p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta}_1)p(\boldsymbol{f}|\boldsymbol{\theta}_2)p(\boldsymbol{\theta})}{p(\boldsymbol{g})}, \tag{3.38}$$

where vector $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2]$ contains all the parameters to be estimated. This joint posterior law can then be used with different Bayesian point estimation methods to infer on $\boldsymbol{f}$ and $\boldsymbol{\theta}$. Figure 3.16 illustrate the basic Bayesian model corresponding to Eq.(3.38). In this model, the parameters $\boldsymbol{\theta}$ are unknown and depends on the fixed hyper-parameters $\boldsymbol{\kappa} = [\boldsymbol{\kappa}_1,\ \boldsymbol{\kappa}_2]$.

For a hierarchical structured model where a hidden variable $\boldsymbol{z}$ appears, a posterior law with all the unknown variables is obtained via the Bayes' rule :

$$p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}|\boldsymbol{g}) = \frac{p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta}_1)p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{\theta}_2)p(\boldsymbol{z}|\boldsymbol{\theta}_3)p(\boldsymbol{\theta})}{p(\boldsymbol{g})}, \tag{3.39}$$

where $\boldsymbol{\theta} = [\boldsymbol{\theta}_1,\ \boldsymbol{\theta}_2,\ \boldsymbol{\theta}_3]$ corresponds to the parameters which are estimated during reconstruction, and $\boldsymbol{\kappa} = [\boldsymbol{\kappa}_1,\ \boldsymbol{\kappa}_2,\ \boldsymbol{\kappa}_3]$ are the fixed hyper-parameters. This model is illustrated by the generative graph shown in Figure 3.17.

3.7.1 Bayesian point estimators

In Bayesian inference, the posterior distribution is typically considered for the estimation of parameters and variables. The Bayesian point estimators aims at choosing the "best value"

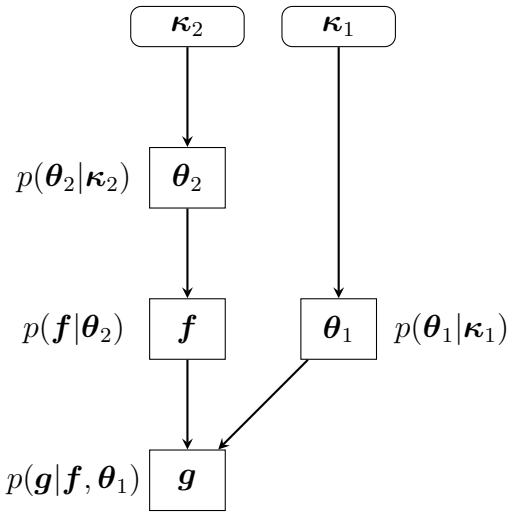


FIGURE 3.16 – Generative graph of the semi-supervised Bayesian model.

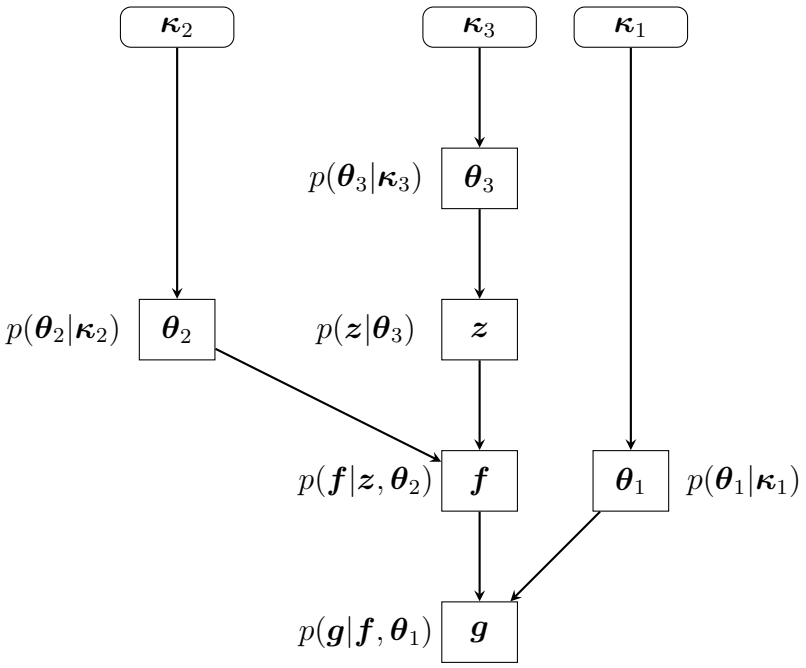


FIGURE 3.17 – Generative graph of the model with hidden variable $\boldsymbol{z}$.

of the unknown variable from its posterior distribution. They imply the posterior distribution's statistics of central tendency, for example the mean, the median, or the mode.

Consequently, three types of point estimators are derived : Posterior Mean (PM) [Eri69], Posterior Median [BS05] and the Maximum a Posterior (MAP) [QL00] estimation. For some distributions, there is no explicit form for the median value. Hence typically only PM and MAP estimation are used for the determination of variables is Bayesian inference.

With the posterior distribution obtained from an unsupervised Bayesian inference [MD96] as in Eq.(3.38), we distinguish three estimation methods. The first is to integrate out θ from $p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})$ to obtain $p(\boldsymbol{f}|\boldsymbol{g})$ and then use it to infer on $\boldsymbol{f}$. Unfortunately, this approach does not often give explicit expression for $p(\boldsymbol{f}|\boldsymbol{g})$. The second approach is that one can first marginalize it with respect to $\boldsymbol{f}$ to obtain $p(\boldsymbol{\theta}|\boldsymbol{g}) = \int p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})\, \mathrm{d}\boldsymbol{f}$ and estimate $\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \{p(\boldsymbol{\theta}|\boldsymbol{g})\}$, then use it as it was known. The third and easiest algorithm to implement is the joint optimization, which estimates the variable $\boldsymbol{f}$ and parameter $\boldsymbol{\theta}$ iteratively and alternately.

In MAP estimation, the unknown variables are estimated by maximizing the posterior distribution. When there are several variables, they are optimized jointly :

$$(\boldsymbol{f}, \boldsymbol{\theta}) = \arg\max_{\boldsymbol{f},\boldsymbol{\theta}} \{p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})\}\,. \tag{3.40}$$

The posterior distribution has normally an negative exponential form, and hence the maximization of the posterior distribution are normally considered as a minimization of the negative logarithm of posterior distribution :

$$(\boldsymbol{f}, \boldsymbol{\theta}) = \arg\min_{\boldsymbol{f},\boldsymbol{\theta}} \{-\ln p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})\}\,. \tag{3.41}$$

In Posterior Mean estimation, when there are several related variables and the posterior distribution is not separable for the variables, the Variational Bayesian Approximation (VBA) [FR12, TLG08, AMD10] is always used. The VBA method approximates the posterior distribution $p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})$ by a separable distribution $q(\boldsymbol{f}, \boldsymbol{\theta})$ by minimizing the Kullback–Leibler divergence [BMD08, CB01], $p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}) \approx q(\boldsymbol{f}, \boldsymbol{\theta}) = q(\boldsymbol{f})q(\boldsymbol{\theta})$. The conjugate prior distributions are normally used to define the prior model in order to simplify the computation of the posterior model and the parameters.

3.7.2 Link between Bayesian and Regularization methods

By using the MAP estimation for the posterior distribution, the links between the Bayesian inference and the regularization methods are derived. For example we consider the following MAP problem :

$$p(\boldsymbol{f}) = \exp\{-R(\boldsymbol{f})\}\,, \tag{3.42}$$
$$p(\boldsymbol{g}|\boldsymbol{f}) = \exp\{-Q(\boldsymbol{g}, \boldsymbol{f})\}\,. \tag{3.43}$$

The posterior distribution obtained from the likelihood and prior model is :

$$p(\boldsymbol{f}|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f})p(\boldsymbol{f}) = \exp\{-Q(\boldsymbol{g}, \boldsymbol{f}) - R(\boldsymbol{f})\}\,. \tag{3.44}$$

By minimizing the negative logarithm of the posterior distribution, the variable is estimated via :

$$\widehat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \{p(\boldsymbol{f}|\boldsymbol{g})\} = \arg\min_{\boldsymbol{f}} \{-\ln p(\boldsymbol{f}|\boldsymbol{g})\} = \arg\min_{\boldsymbol{f}} \{Q(\boldsymbol{g},\boldsymbol{f}) + R(\boldsymbol{f})\}, \tag{3.45}$$

which is exactly the general model of the regularization method presented in Eq.(3.24).

As in X-ray CT we the noise is modeled by a Gaussian distribution, we have

$$p(\boldsymbol{g}|\boldsymbol{f}) = \mathcal{N}(\boldsymbol{g}|\boldsymbol{H}\boldsymbol{f}, \boldsymbol{V}_\epsilon) \propto |\boldsymbol{V}_\epsilon|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f})^T \boldsymbol{V}_\epsilon^{-1}(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f})\right\}, \tag{3.46}$$

and we get :

$$p(\boldsymbol{f}|\boldsymbol{g}) \propto \exp\left\{-\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|^2_{\boldsymbol{V}_\epsilon} - R(\boldsymbol{f})\right\}, \tag{3.47}$$

and

$$\widehat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \left\{\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|^2_{\boldsymbol{V}_\epsilon} + R(\boldsymbol{f})\right\}. \tag{3.48}$$

When the prior distribution of $\boldsymbol{f}$ is a Gaussian model, with zero mean and variance $\boldsymbol{v}_f$:

$$p(\boldsymbol{f}) = \mathcal{N}(\boldsymbol{f}|0, \boldsymbol{v}_f) \propto \exp\left\{-\frac{1}{2}\|\boldsymbol{f}\|^2_{\boldsymbol{V}_f}\right\}, \tag{3.49}$$

we get $\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \left\{\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|^2_{\boldsymbol{V}_\epsilon} + \|\boldsymbol{f}\|^2_{\boldsymbol{V}_f}\right\}$. We can see that when $\boldsymbol{V}_\epsilon = v_\epsilon \boldsymbol{I}_M$ and $\boldsymbol{V}_f = v_f \boldsymbol{I}_N$ where $\boldsymbol{I}$ is an identity matrix, we obtain the Quadratic Regularization method.

When the prior distribution of $\boldsymbol{f}$ is a Laplacian model :

$$p(\boldsymbol{f}) = \mathcal{L}(\boldsymbol{f}|0, \boldsymbol{V}_f) \propto \exp\left\{-\frac{1}{2}\|\boldsymbol{f}\|_{1\boldsymbol{V}_f}\right\}, \tag{3.50}$$

we get $\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \left\{\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|^2_{\boldsymbol{V}_\epsilon} + \|\boldsymbol{f}\|_{1\boldsymbol{V}_f}\right\}$. When $\boldsymbol{V}_\epsilon = v_\epsilon^{-1}\boldsymbol{I}_M$ and $\boldsymbol{V}_f = v_f^{-1}\boldsymbol{I}_N$, we obtain the LASSO regularization method [Tib96]. If $p(\boldsymbol{f}) = \mathcal{L}(\boldsymbol{f}|0, \left(\boldsymbol{D}^T\boldsymbol{V}_f\boldsymbol{D}\right)^{-1})$, we get $\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \left\{\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|^2_{\boldsymbol{V}_\epsilon} + \|\boldsymbol{D}\boldsymbol{f}\|_{1\boldsymbol{V}_f}\right\}$ and it corresponds to the TV method.

3.8 Conclusions

In this chapter, a framework of the X-ray CT reconstruction methods is presented. Among these conventional methods, the analytical and algebraic ones considering the forward model without noise : $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$, are not robust to the noise, and no additional information is considered except for the biased data.

The regularization methods and the statistical methods consider an additive noise in the forward model $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$. These methods, especially the Bayesian ones, take into consideration the prior information as a supplement criterion for the ill-posed reconstruction problems. In the Bayesian methods, the information on the object is modeled by some appropriate probability density function (pdf). An important objective in the Bayesian method is to choose suitable pdfs for the unknown variables.

In the next chapter, I am going to give a brief review of the Bayesian methods, and present some conventional used prior models for the piecewise continuous variables in the X-ray CT problems.

4

Bayesian method and sparse enforcing prior distributions

4.1 Introduction

In this chapter, we present the Bayesian inference and its utilization in the CT reconstruction problem. As the most important part of the Bayesian methods, the definition of the prior models are discussed in this chapter. The prior model for defining the piecewise continuous images or objects are typically realized by using a Markovian model or by introducing a sparse hidden variable. The definition of the sparse prior model can be realized by using plenty of distributions. In this chapter we present a generalization of the Student-t distribution and discuss its property.

Bayesian inference has provided a useful tool for the inverse problems in different domains, especially in the ill-posed problems where the number of detected data is insufficient. This advantage of the Bayesian inference comes from the fact that the prior information is considered in the method in terms of the prior model of the variables. The prior information are combined with the likelihood and the detected data in the reconstruction.

We recall the Bayes rule for the unsupervised system model :

$$\begin{aligned} p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}) =& \frac{p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1)p(\boldsymbol{f}|\boldsymbol{\theta}_2)p(\boldsymbol{\theta}_2)}{p(\boldsymbol{g})} \\ \propto& p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1)p(\boldsymbol{f}|\boldsymbol{\theta}_2)p(\boldsymbol{\theta}_2). \end{aligned} \tag{4.1}$$

In this formula, $p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})$ is the posterior distribution, $p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1)$ is the likelihood, $p(\boldsymbol{f}|\boldsymbol{\theta}_2)$ and $p(\boldsymbol{\theta}_2)$ are the prior distributions.

We can see that the posterior distribution depends on the likelihood and the prior distributions. In the Bayesian methods, an essential work is to define the prior models. The prior information of the variable $\boldsymbol{f}$ and parameter $\boldsymbol{\theta} = [\boldsymbol{\theta}_1,\ \boldsymbol{\theta}_2]$ are considered and described by some suitable distribution. Some examples of the prior information are the ”positivity”, ”continuity”,

"sparsity", etc.

In X-ray CT, the objects under consideration are for example the brain or organs in medical applications, or components of industrial objects in Non Destructive Testing (NDT).

The industrial components normally consist of several different homogeneous materials, for example the steel, iron, air, etc. In medical CT, the brains, bones, teeth or other organs also consist of some homogeneous substances. We can also see from the examples shown in Figure 3.1 and Figure 3.2 that the objects under examination are **piecewise continuous**.

In this thesis, the piecewise continuous property of the object is particularly considered in the Bayesian methods when choosing the prior models. In the following sections I'm going to present the choice of the prior models adapting to the piecewise continuous prior information.

4.2 Gradient and Markovian models

The "homogeneous" property implies that the voxels in the object are not independent. Each voxel depends on other voxels. But the coherence between two voxels decreases rapidly when the distance between them increases. Consequently, only the nearest neighbor voxels are related to the current voxel.

Noted by $\boldsymbol{r}$ the current voxel position, and $f(\boldsymbol{r})$ the value of the voxel, we use $\boldsymbol{r}'$ to represent the neighbor voxels of $\boldsymbol{r}$, $\boldsymbol{r}' \in \boldsymbol{N}(\boldsymbol{r})$.

Four or eight neighbors are the most commonly used cases. In order to consider the coherent relations between voxels, the Markovian method is generally used.

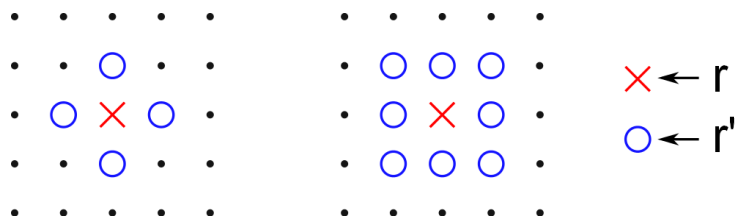
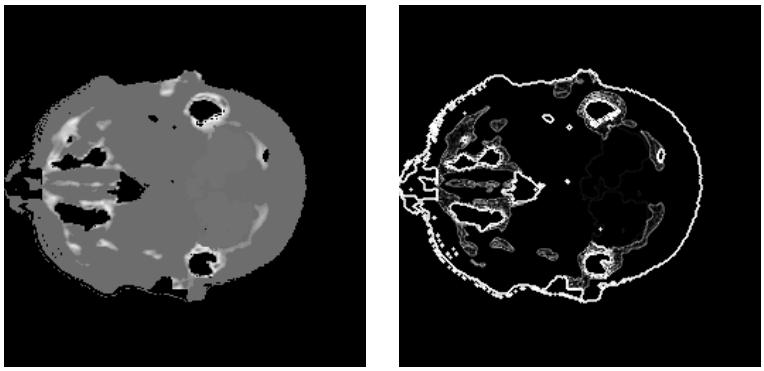


FIGURE 4.1 – The positions of current pixel $\boldsymbol{r}$ and neighbor pixels $\boldsymbol{r}'$ by considering left : 4 neighbor pixels or right : 8 neighbor pixels.

In many methods, the gradient of the piecewise continuous variable is used as a sparse hidden variable. In the gradient-based methods, the difference between the neighbour pixels are considered. As shown in Figure 4.2, for a piece-wise constant object, the gradient is zero in most of parts except for the contours areas.

In the Markovian models, the strategy is also to consider the neighbour pixels in the prior models of the piecewise continuous variables. There are mainly two types of Markovian models, known as the homogeneous Markovian models and the inhomogeneous Markovian models. In the homogeneous Markovian models, the big variation of the pixel values is penalized for all the pixels, thus the noise is limited. However, the contours in the image are also obscured. The inhomogeneous Markovian models solves this problem by considering separately the homogeneous areas and the contours.

Shepp Logan phantom | Contours of Shepp Logan



Head object | Contours of Head

FIGURE 4.2 – The piecewise constant objects and their gradient.

4.2.1 Homogeneous Markovian model

The homogeneous Markovian model defines the same Markovian model for every voxel, with the following prior model expression :

$$p(\boldsymbol{f}) = \Psi\left(\Phi\left(f(\boldsymbol{r}), f(\boldsymbol{r}')\right)\right). \tag{4.2}$$

The function $\Phi\left(f(\boldsymbol{r}), f(\boldsymbol{r}')\right)$ describes the influence of the neighbor pixels on the current pixel. In the Figure 4.1, we show the example of considering 4 and 8 neighbor pixels, and the corresponding $\Phi(f(\boldsymbol{r}), f(\boldsymbol{r}'))$ of these two cases are :

$$\Phi\left(f(\boldsymbol{r})\right) = f(\boldsymbol{r}) - \frac{1}{4} \sum_{\boldsymbol{r}' \in \boldsymbol{N}_4(\boldsymbol{r})} f(\boldsymbol{r}'), \tag{4.3}$$

$$\Phi\left(f(\boldsymbol{r})\right) = f(\boldsymbol{r}) - \frac{1}{8} \sum_{\boldsymbol{r}' \in \boldsymbol{N}_8(\boldsymbol{r})} f(\boldsymbol{r}'), \tag{4.4}$$

where $\boldsymbol{N}_4(\cdot)$ and $\boldsymbol{N}_8(\cdot)$ represent respectively the set of 4 and 8 neighbor pixels.

The prior distribution Ψ with exponential form are commonly used :

$$p(\boldsymbol{f}) = \Psi\left(\Phi\left(f(\boldsymbol{r}), f(\boldsymbol{r}')\right)\right) \propto \exp\left\{-\lambda \sum_{r} \left\|\Phi\left(f(\boldsymbol{r}), f(\boldsymbol{r}')\right)\right\|_2^2\right\}, \tag{4.5}$$

where λ is a parameter of this prior distribution.

The Laplacian Markovian model is also used in some methods :

$$p(f) = \Psi\left(\Phi\left(f(\boldsymbol{r}), f(\boldsymbol{r}')\right)\right) \propto \exp\left\{-\lambda \sum_{r} \left\|\Phi\left(f(\boldsymbol{r}), f(\boldsymbol{r}')\right)\right\|_1\right\}. \tag{4.6}$$

By using these two prior models in Bayesian method and optimize via the MAP algorithm, we obtain respectively the conventional Quadratic Regularization method (QR) and the Total Variation method (TV) :

$$\begin{aligned} \text{QR}: \ J_1(\boldsymbol{f}) &= \|\boldsymbol{g} - \boldsymbol{f}\|_2^2 + \lambda \|\boldsymbol{D}\boldsymbol{f}\|_2^2, \\ \text{TV}: \ J_2(\boldsymbol{f}) &= \|\boldsymbol{g} - \boldsymbol{f}\|_2^2 + \lambda \|\boldsymbol{D}\boldsymbol{f}\|_1, \end{aligned} \tag{4.7}$$

where $\boldsymbol{D}$ is the gradient operator, corresponding to $\Phi\left(f(\boldsymbol{r}), f(\boldsymbol{r}')\right)$.

4.2.2 Inhomogeneous Markovian models

The homogeneous Markovian models are frequently used, but a shortcoming is that these models do not distinguish the homogeneous areas and the contours in the object, thus the contours in the reconstructed image will not be preserved.

The inhomogeneous Markovian model solves this problem. In these models, one or several hidden variables are defined to identify different areas. For example, the label of different homogeneous zones, or the label of contours.

The word "inhomogeneous" implies that the probability model for the pixels are not the same for all the pixels. As shown in Figure 4.3, the pixels f_1 and f_4 have 2 neighbour pixels on the contours, f_2 has only 1 neighbour pixel on the contours, and f_3 has no neighbour pixel on the contours. In the inhomogeneous Markovian model, these four pixels have different prior distributions, depending on the neighbour pixels in the homogeneous areas (circles) while ignoring the neighbour pixels on the contours (triangles).

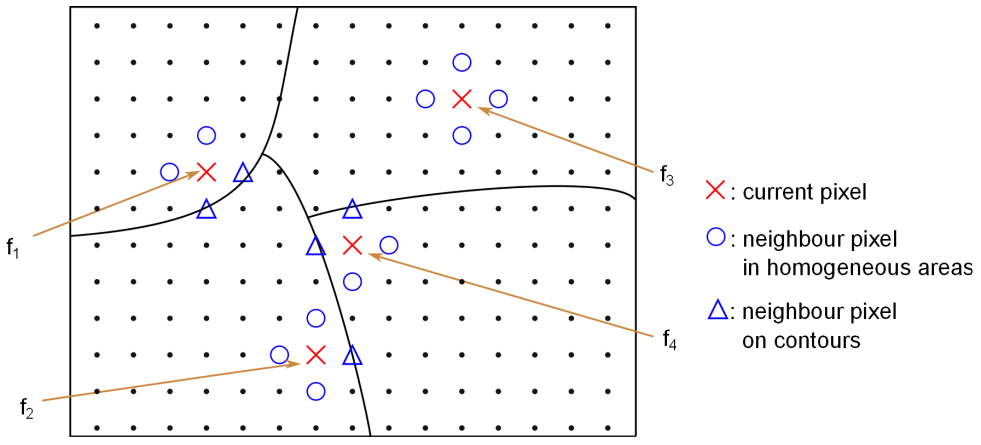


FIGURE 4.3 – Example of inhomogeneous Markovian model.

The inhomogeneous Markovian method depending on the label of different blocks is presented in [AMD10], and the prior model for the object is :

$$p(\boldsymbol{f}|\boldsymbol{z}) = \prod_{R} \mathcal{N}(\mu_z(R), v_z(R)), \tag{4.8}$$

where $R \in \mathcal{R}$ is the label of different blocks, $\mu_z(R)$ and $v_z(R)$ are respectively the mean and variance of the pixel z in label R. From the model we see that in different blocks, $\boldsymbol{f}$ follows a Gaussian distribution with different parameters, $p(\boldsymbol{f}|\boldsymbol{z}) = \prod_R \mathcal{N}(\mu_z(R), v_z(R))$.

The gradient can also be used to control the Markovian model, separating the model for voxels into homogeneous areas and contours areas. A method using the gradient value in the prior model in proposed and presented in Chapter 6.

4.3 Sparse transformation base coefficients

The piecewise continuous variables can be represented by a optimally sparse base. A simple example is the wavelet transformation. The wavelet transformation has been used in many works of the X-ray CT reconstruction [BKW96, CKL01, RVJ+06] where the variable of interest is piecewise continuous.

In Chapter 5, a Bayesian method basing on a prior model concerning the Haar transform coefficients is presented.

With the wavelet transformation, or dictionary base transform, the piecewise continuous prior property of the object become the sparse structured property of the coefficient of the transformation. A prior model for enforcing the sparse structure is then used.

As it was presented in the previous section, the piecewise continuous variable can be represented by another sparse variable, via some linear or non linear transformation operations. We present here some commonly used sparse representations.

The linear wavelet transformations include for example the Haar transformation [SF03], the contourlet transformation [DV05], the curvelet transformation [SCD02], etc. We present here the definition of an orthogonal transformation.

A linear transformation can be expressed as a multiplication by a matrix. Here we define a linear transformation :

$$\boldsymbol{z} = \boldsymbol{D}\boldsymbol{f}, \tag{4.9}$$

where $\boldsymbol{D} \in \mathbb{R}^{M \times N}$ represents the transformation matrix, $\boldsymbol{f} \in \mathbb{R}^{N \times 1}$ is a vector and $\boldsymbol{z} \in \mathbb{R}^{M \times 1}$ is the transform coefficient.

An orthogonal matrix is square, thus $\boldsymbol{D} \in \mathbb{R}^{N \times N}$, and $\boldsymbol{z} \in \mathbb{R}^{N \times 1}$. In linear algebra, an orthogonal matrix is a square matrix with real entries whose columns and rows are orthogonal unit vectors (i.e., orthonormal vectors), and thus

$$\boldsymbol{D}^T\boldsymbol{D} = \boldsymbol{I}, \tag{4.10}$$

where $\boldsymbol{I}$ is an identity matrix.

The equivalent characterization of the orthogonal matrix is that : its transpose is equal to its inverse :

$$\boldsymbol{D}^T = \boldsymbol{D}^{-1}. \tag{4.11}$$

One of the frequently used orthogonal transformation is the wavelet transformation. For example the Discrete Haar transformation. The $N \times N$ Haar transformation matrix is :

$$\boldsymbol{D} = \left[\boldsymbol{d}_0^T,\ \boldsymbol{d}_1^T,\ \cdots,\ \boldsymbol{d}_{N-1}^T\right]^T, \tag{4.12}$$

and

$$\boldsymbol{D}^T = [\boldsymbol{d}_0,\ \boldsymbol{d}_1,\ \cdots,\ \boldsymbol{d}_{N-1}]\,. \tag{4.13}$$

$\boldsymbol{d}_k$ is called the Haar function, and is defined as :

$$\boldsymbol{d}_k(t) = \frac{1}{\sqrt{N}} \begin{cases} 2^{p/2} & (q-1)/2^p \le t < (q-0.5)/2^p \\ -2^{p/2} & (q-0.5)/2^p \le t < q/2^p \\ 0 & \text{otherwise} \end{cases}, \tag{4.14}$$

and $\boldsymbol{d}_0(t) = 1/\sqrt{N}$. p and q are defined as :

$$k = 2^p + q - 1, \tag{4.15}$$

and 2^p is the largest power of 2 that is smaller than k.

when $N = 2$:

$$\boldsymbol{D}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{4.16}$$

When $N = 4$, we have :

$$\boldsymbol{D}_4 = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ \sqrt{2} & -\sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & -\sqrt{2} \end{bmatrix}. \tag{4.17}$$

In Figure 4.4, the wavelet shapes show the corresponding Haar functions. We can see that the function has a shape as the step function and the period of the wavelet reduces by half when level of p augments by 1.

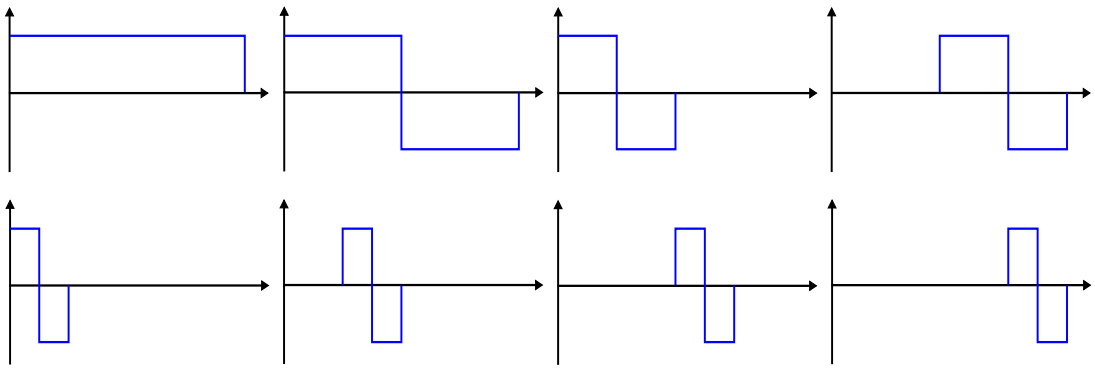


FIGURE 4.4 – The wavelet shape of the Haar functions.

We can see that according to the definition, we have :

$$\boldsymbol{d}_i^T * \boldsymbol{d}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{4.18}$$

Therefore we obtain :

$$\boldsymbol{D}\boldsymbol{D}^T = \begin{bmatrix} \boldsymbol{d}_0^T \\ \boldsymbol{d}_1^T \\ \vdots \\ \boldsymbol{d}_{N-1}^T \end{bmatrix} [\boldsymbol{d}_0\ \boldsymbol{d}_1\ \cdots\ \boldsymbol{d}_{N-1}] = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}, \tag{4.19}$$

and we obtain :

$$\boldsymbol{D}\boldsymbol{D}^T = \boldsymbol{I}. \tag{4.20}$$

Consequently $\boldsymbol{D}^T = \boldsymbol{D}^{-1}$.

In Figure 4.5, the Shepp Logan figure of size 256^2 and its multilevel Haar transformation coefficients are shown. The scale of the transformation coefficients are adapted such that they are visually clearly.

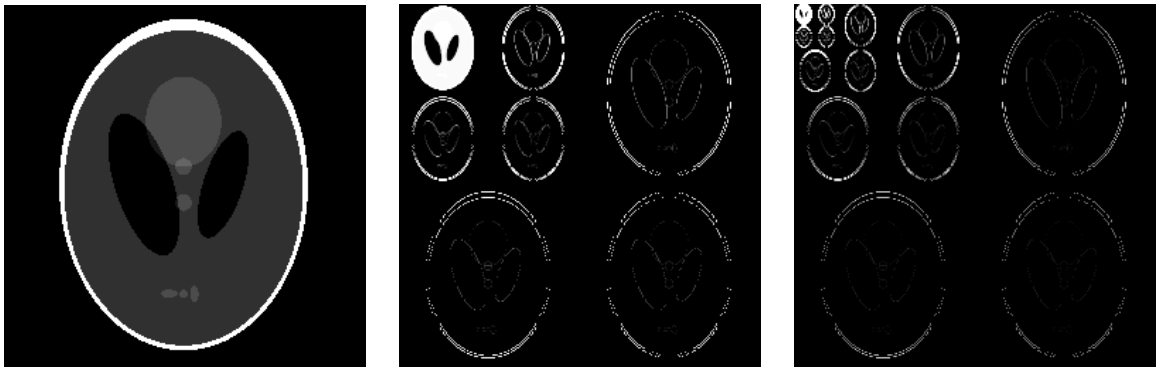


FIGURE 4.5 – The Shepp Logan figure of size 256^2 (left), its 2-level Haar transformation coefficient (middle) and its 4-level Haar transformation coefficient (right).

There exist many orthogonal transformations, for example the wavelet transformation [RVJ+06, BKW96], the curvelet transformation [SCD02, DS07], etc. By using the orthogonal transformations in the Bayesian method, the transformation of the operator can be replaced by the inversion of the operator, which is quick and convenient.

4.4 Sparsity enforcing prior distributions

As it was mentioned previously, the sparse transformation coefficients of a piecewise continuous variable is used as a hidden variable in many Bayesian methods. In a sparse vector, most of the elements are zero. The sparse transformations are widely used in compressed sensing problems [Don06, Bar07].

4.4.1 Distributions for sparse variables

In Bayesian inference, a sparse structured variable can be defined by some sparsity enforcing prior distributions. Generally, three types of distributions are used to define the sparse variables :

1. The Generalized Gaussian distributions :

$$p(\boldsymbol{z}) \propto \exp\left\{-\frac{1}{2}\alpha \left\|\boldsymbol{z}\right\|_\beta^\beta\right\}. \tag{4.21}$$

   For example when $\beta = 1$, it becomes Lapacian distribution.

2. The Mixture distributions :

$$p(\boldsymbol{z}) \propto \prod_k \exp\left\{-\frac{1}{2}\alpha_k \left\|\boldsymbol{z}\right\|_2^2\right\}. \tag{4.22}$$

3. The heavy tailed distributions [Kle03], for example the Student-t distribution. The tails of this type of distributions are not exponentially bounded, with the definition :

$$\lim_{x\to\infty} e^{\lambda x} p(X > x) = \infty \ \ \forall \ \ \lambda > 0. \tag{4.23}$$

The generalized Gaussian distribution or the Mixture distributions which can be used to define the sparse variables are also heavy tailed.

4.4.2 Generalized Student-t distribution

Among all the sparsity enforcing distributions, the Laplacian distribution is frequently considered. However, the shortcoming is that the computation is complex and slow because of the $L1$ norm optimization. The Student-t distribution is another well known sparsity enforcing distribution. In my work, I use a two-parameter version of the Student-t distribution, benefiting the convenience that it can be derived via a hierarchical model and this model leads to a semi-supervised system model for the Bayesian method.

As a heavy tailed distribution, the standard Student-t distribution has a probability density function as :

$$\mathcal{S}t(x|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \ \ \nu > 0, \tag{4.24}$$

where ν is the number of degrees of freedom and Γ represents the Gamma function.

The standard Student-t distribution can also be expressed as an Infinite Gaussian Scale Mixture (IGSM) distribution [Tre01, PSWS03] :

$$\mathcal{S}t(x|\nu) = \int_0^\infty \mathcal{N}\left(x|0, \frac{1}{z}\right) \mathcal{G}\left(z|\frac{\nu}{2}, \frac{\nu}{2}\right) \ \mathrm{d}z, \tag{4.25}$$

where $\mathcal{N}(\cdot)$ is the Gaussian distribution and $\mathcal{G}(\cdot)$ is the Gamma distribution. A Normal-Inverse Gamma marginalization can also be used to define the Student-t distribution :

$$\mathcal{S}t(x|\nu) = \int_0^\infty \mathcal{N}\left(x|0, z\right) \mathcal{IG}\left(z|\frac{\nu}{2}, \frac{\nu}{2}\right). \tag{4.26}$$

Comparing with many other heavy tailed distributions, for example the Laplacian distribution and the Generalized Gaussian distribution, the Student-t distribution is particularly interesting thanks to the IGSM expression. What's more, the IGSM expression is composed by a Normal and a Gamma or Inverse Gamma distribution, which are conjugate likelihood and prior pairs. This property facilitates the computation in Bayesian methods.

In our work, we replace the one-parameter Inverse Gamma distribution $\mathcal{IG}(z|\frac{\nu}{2}, \frac{\nu}{2})$ by the two-parameter Inverse Gamma distribution $\mathcal{IG}(z|\alpha, \beta)$, and we call it the generalized Student-t distribution. It is expressed as :

$$\mathcal{S}t_g\left(x|\alpha, \beta\right) = \int_0^\infty \mathcal{N}\left(x|0, z\right) \mathcal{IG}\left(z|\alpha, \beta\right) \ \mathrm{d}z. \tag{4.27}$$

And the explicit expression of the generalized Student-t distribution is :

$$\mathcal{S}t_g\left(x|\alpha, \beta\right) = \frac{\Gamma\left(\alpha + \frac{1}{2}\right)}{\sqrt{2\beta\pi}\Gamma(\alpha)} \left(1 + \frac{x^2}{2\beta}\right)^{-\left(\alpha+\frac{1}{2}\right)}, \ \alpha > 0, \ \beta > 0. \tag{4.28}$$

When $\alpha = \beta = \frac{\nu}{2}$, we obtain the standard Student-t distribution.

In Figure 4.6, the form of the pdf of Gaussian distribution, Student-t distribution and the generalized Student-t distribution are compared. From the comparison of the standard Student-t distribution and the generalized Student-t distribution, we can see that the standard Student-t distribution has a limitation : the shape of the pdf is either heavy tailed or with narrow peak. These two properties can not be realized simultaneously. The generalized Student-t distribution, however, can have a pdf with both narrow peak and a heavy tail, by choosing suitable values for the parameters α and β.

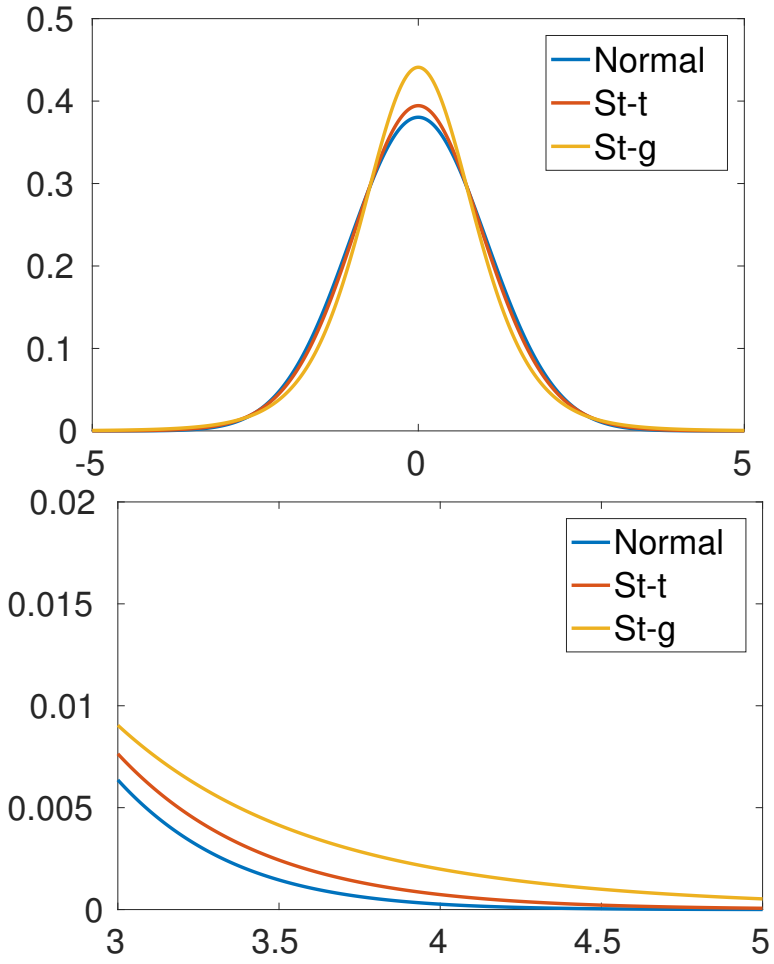


FIGURE 4.6 – The comparison of the pdfs of the Normal distribution, the Student-t distribution and the generalized Student-t distribution.

In Bayesian inference, the sparse variables are defined belonging to a generalized Student-t distribution. With the Normal-Inverse Gamma marginalization property, a sparse variable x is defined as a Gaussian distributed variable, with mean equal to zero and variance z, and the variance z is defined belonging to an Inverse Gamma distribution with parameters α and β :

$$\begin{aligned} p(x|z) &= \mathcal{N}(x|0, z), \\ p(z|\alpha, \beta) &= \mathcal{IG}(z|\alpha, \beta). \end{aligned} \tag{4.29}$$

By choosing different values for the (α, β) pair, we get different shape of the probability distributed function. The standard Student-t distribution can be approached to a Gaussian distribution when $\nu \to \infty$.

Demonstration.

$$\lim_{\nu\to\infty} p(x|\nu) = \lim_{\nu\to\infty} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}. \tag{4.30}$$

By using the theorem :

$$\lim_{\nu\to\infty} \frac{\Gamma(\nu+\alpha)}{\Gamma(\nu)\nu^{\alpha}} = 1, \tag{4.31}$$

we have :

$$\lim_{\nu\to\infty} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} = \lim_{\nu\to\infty} \frac{\Gamma\left(\frac{\nu}{2}+\frac{1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\left(\frac{\nu}{2}\right)^{\frac{1}{2}}\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi}}. \tag{4.32}$$

And with the Taylor series we have :

$$\begin{aligned}
\lim_{\nu\to\infty} \left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} &= \lim_{\nu\to\infty} \exp\left\{-\frac{\nu+1}{2}\ln\left(1+\frac{x^2}{\nu}\right)\right\} \\
&= \lim_{\nu\to\infty} \exp\left\{-\frac{\nu+1}{2}\left(\frac{x^2}{\nu}-\frac{1}{2}\frac{x^4}{\nu^2}+\frac{1}{3}\frac{x^6}{\nu^3}+\cdots\right)\right\} \\
&= \lim_{\nu\to\infty} \exp\left\{-\frac{1}{2}\frac{\nu+1}{\nu}x^2\right\} = \exp\left\{-\frac{1}{2}x^2\right\}.
\end{aligned} \tag{4.33}$$

And we get a Gaussian distribution. For the $\mathcal{S}t_g$ distribution, we can also use the same strategy :

$$\begin{aligned}
&\lim_{\alpha,\beta\to\infty} \frac{\Gamma\left(\alpha+\frac{1}{2}\right)}{\sqrt{2\pi\beta}\Gamma(\alpha)} \left(1+\frac{x^2}{2\beta}\right)^{-\left(\alpha+\frac{1}{2}\right)} \\
&\qquad = \lim_{\alpha,\beta\to\infty} \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\alpha}{\beta}} \exp\left\{-(\alpha+\frac{1}{2})\left(\frac{x^2}{2\beta}-\frac{1}{2}\frac{x^4}{4\beta^2}+\cdots\right)\right\} \\
&\qquad = \lim_{\alpha,\beta\to\infty} \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\alpha}{\beta}} \exp\left\{-\frac{1}{2}\frac{\alpha+\frac{1}{2}}{\beta}x^2\right\}.
\end{aligned} \tag{4.34}$$

When α and β are both big enough and they are defined on the same scale, the approximation is also a Gaussian distribution :

$$\lim_{\alpha,\beta\to\infty} \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\alpha}{\beta}} \exp\left\{-(\alpha+\frac{1}{2})\left(\frac{x^2}{2\beta}-\frac{1}{2}\frac{x^4}{4\beta^2}+\cdots\right)\right\} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\}. \tag{4.35}$$

□

4.5 Conclusion

We have introduced in this chapter the context of the Bayesian inference and discussed the choice of the prior model for the piecewise continuous variable. The generalized Student-t distribution is presented in this chapter. From the discussion, we can see that the $\mathcal{S}t_g$ distribution could be used as the prior law for the sparse hidden variable in the Bayesian method. The orthogonal transformation is convenient as the inversion of operator and the transpose of operator are identical.

In the next two chapters, we present the proposed Bayesian methods basing on two principal strategies.

The first strategy is to define a hidden variable, z, which is a sparse transformation coefficient of the object, f, in the prior model. A hierarchical model is proposed. We call it the Hierarchical Haar based Bayesian Method (HHBM). In this method, both the coefficient z and the object f are considered as unknown variables in the model, and the generalized Student-t distribution is used as the prior distribution for the sparse coefficient z. This proposed method and some extensions are presented in Chapter 5.

The second strategy (ROCC method), presented in Chapter 6, uses an inhomogeneous Markovian model. In this proposed method, the Laplacian of the object is considered as a hidden variable, and is modeled by the generalized Student-t distribution because of its sparse structure. The optimized Laplacian is used as a hyperparameter in the Markovian model for the object, and the optimized object then is used in the optimization of Laplacian. Iteratively and alternately, the object and the Laplacian are optimized.

5

Sparsity Enforcing Unsupervised Hierarchical Model

5.1 Introduction

As we have presented previously, the objects that we take into consideration in Non Destructive Testing (NDT) are commonly piecewise continuous. As presented in the previous chapter, the prior models for the piecewise continuous variables are normally a Markovian model or a hierarchical model with a hidden variable. This hidden variable is for example a sparse transformation coefficient of the piecewise continuous variable.

In this chapter, a Bayesian method with a hierarchical structured prior model is proposed. In this model, the coefficient of a sparse orthogonal transformation is considered as a hidden variable. Meanwhile, the semi-supervised system is realized such that the parameters are estimated simultaneously with the unknown variables. The initialization for the hyperparameters is discussed and demonstrated in this chapter. The initialization of hyperparameters is insensible to the final results, which is convenient comparing with the state-of-the-art methods.

5.2 The choice of the sparse transformation

In the big data size applications, the transformation is normally realized by a convolution operator, as the multiplication matrix will have a huge size. In this case, the inverse of a transformation is normally not expensive to compute. We use an orthonormal transformation in our work. Thanks to the fact that the inversion and transpose of an orthonomal transformation are identical, the computation with big data size can be simplified, with the transpose being replaced by the inversion.

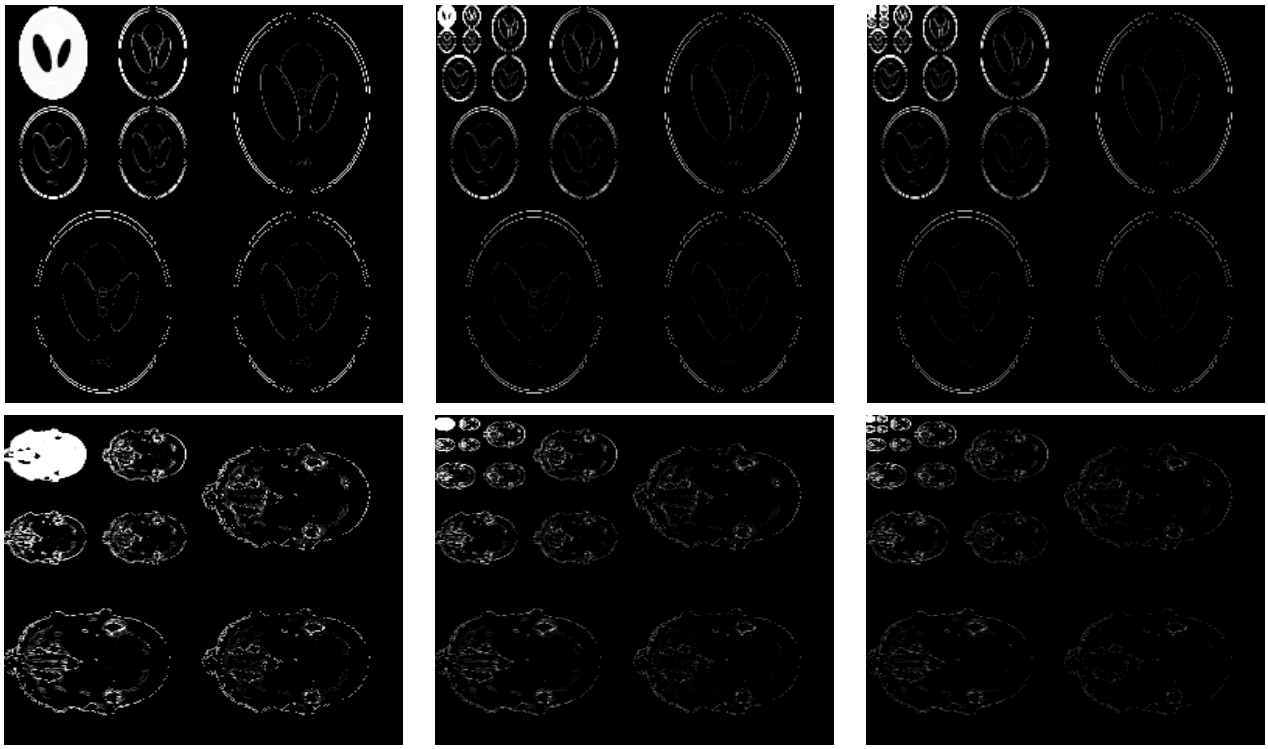
Wavelets and multiscale analysis have proved to be useful in a wide range of applications

[FBS02, DB95, RVJ+06, LNDD07]. Some commonly used transforms are for example the discrete wavelet transform (DWT) [AU96, TSP+08], the discrete cosinus transformation (DCT) [RY14, SK13], the discrete Haar transformation [SF03], the discrete curvlet transformation [CDDY06], the dual-tree Complex Wavelet Transform (DT-CWT) [SBK05, Kin98].

Among these transforms, the discrete Haar transformation is practical in the discrete phantoms, and especially for the piecewise constant or continous images. The Haar transform is one of the earliest transformation. As one of the simplest and oldest orthonormal wavelet, the Haar transformation is used in many applications as for example image coding [ABMD92, CDSB03], edge extraction [GLPS07], and recently image restoration [PP10].

In the Bayesian method proposed in this chapter, the Haar transformation coefficient is used as a sparse hidden variable. In this section, the details of the choice of the transformation level are presented.

Figure 5.1 shows the $2-level$, $4-level$ and $5-level$ Haar transformation coefficient of Shepp Logan phantom and the Head object, both of size 256×256. The scale of these figures have been adapted to be visually more clearly. In general the degree of sparsity increases when the transformation level augments.



Haar coefficient with $l = 2$ | Haar coefficient with $l = 4$ | Haar coefficient with $l = 5$

FIGURE 5.1 – The multilevel discrete Haar transformation coefficients of the 2D Shepp Logan phantom (top) and the Head object (bottom). The transformation level are respectively 2 levels (left), 4 levels (middle) and 5 levels (right).

In order to analyze the properties of HT of different levels, we present in Figure 5.2 and Figure 5.3 the histograms of the transformation coefficients for the Shepp Logan object and the Head object respectively. From the histogram we can see that, when l is small, the coefficients are less compressed with bigger values. When l increases, the sparsity rate increases. We can see from the figures that, when the transformation level is equal to or bigger than 4, the histograms

are identical and hence the sparsity rate are on the same scale. Thus in our work, we choose to use a 5-level HT in order to get a stable sparse structure.

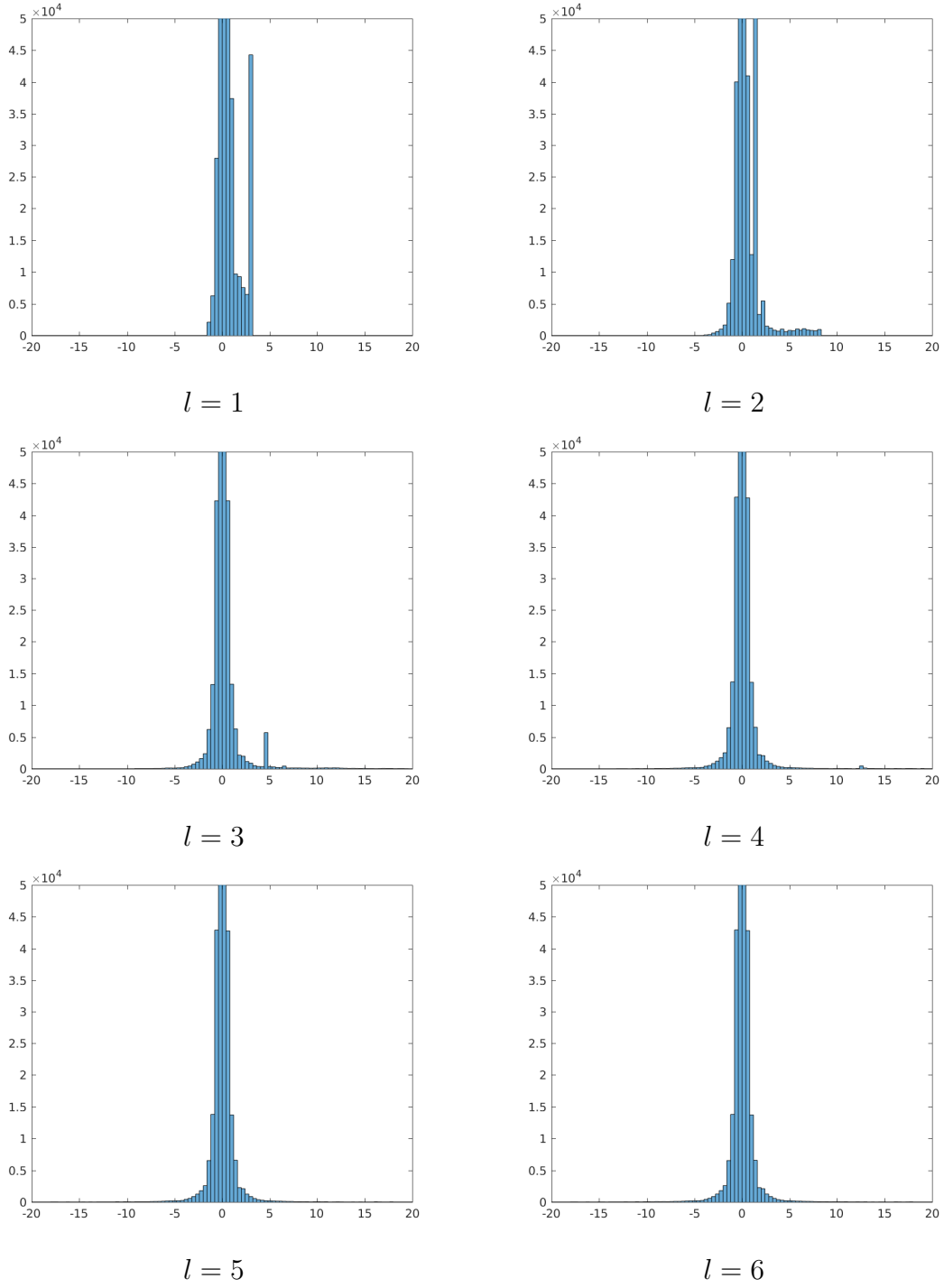


FIGURE 5.2 – The histogram of the multilevel Haar transformation of the **Shepp Logan phantom**. Comparison of the histogram of different transformation levels.

As we have presented previously, the sparse variable can be modeled principally by three types of distributions : the generalized Gaussian distributions for example the Laplacian distribution, the Gaussian mixture distributions for example a mixture of two Gaussian distributions and the heavy tailed distributions for example the Student-t distribution. As presented in

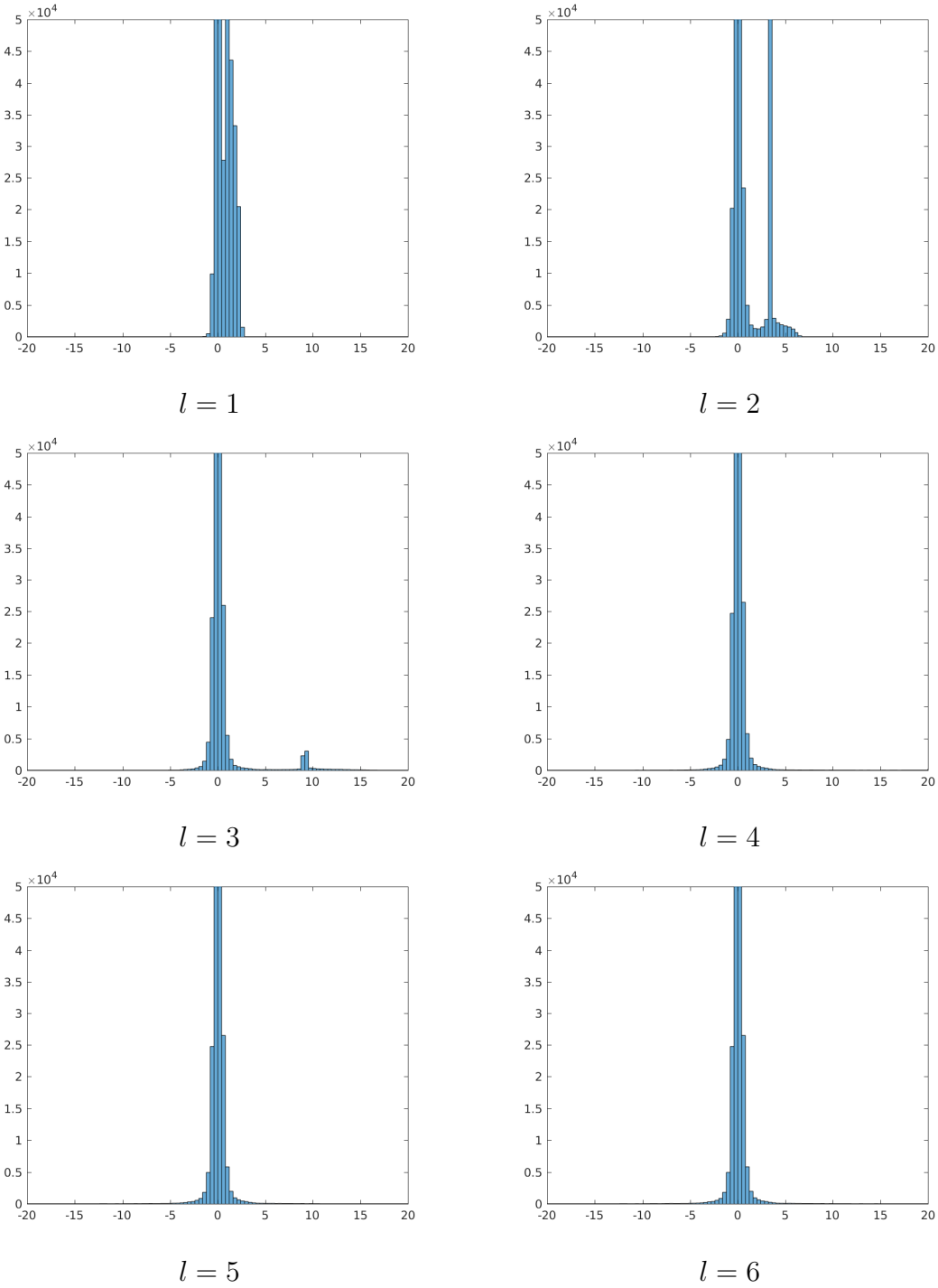


FIGURE 5.3 – The histogram of the multilevel Haar transformation of the **Head object**. Comparison of the histogram of different transformation levels.

Section 4.4, the generalized Student-t distribution is used to model the sparse transformation coefficients.

5.3 A synthesis hierarchical Bayesian model

In the published works, the prior model types can be classified into two types : the analysis type and the synthesis type [EMR07]. The analysis based approach derives the likelihood of the target signal from a forward transform applied to it. An example is the gradient in TV method. This analysis type commonly appears as a regularization term in optimization :

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{J(\boldsymbol{f})\} \quad \text{with} \quad J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|_2^2 + \lambda R(\boldsymbol{D}\boldsymbol{f}) \,. \tag{5.1}$$

The synthesis type of prior, on the other hand, changes the signal of consideration into the coefficient of a forward transform which will be estimated. The coefficient then will be used in a post processing to obtain the final result of the target signal. An example is as following : with the forward model $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$ and the forward transform $\boldsymbol{f} = \boldsymbol{D}\boldsymbol{z}$, $\boldsymbol{z}$ is estimated by optimizing :

$$\widehat{\boldsymbol{f}} = \boldsymbol{D}\widehat{\boldsymbol{z}} \;\text{ with }\; \widehat{\boldsymbol{z}} = \arg\min_{\boldsymbol{z}} \{J(\boldsymbol{z})\} \;\text{ with }\; J(\boldsymbol{z}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{D}\boldsymbol{z}\|_2^2 + \lambda R(\boldsymbol{z}), \tag{5.2}$$

and $\widehat{\boldsymbol{f}} = \boldsymbol{D}\widehat{\boldsymbol{z}}$ is derived as the final result of the reconstruction.

First of all, we consider a synthesis model with which we obtain directly the coefficient of the transformation.

5.3.1 The Hierarchical Model.

Again, we mention here the forward system model :

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}, \tag{5.3}$$

where $\boldsymbol{\epsilon}$ is an additive noise representing the uncertainties of the projection system. In the X-ray CT models, it is modeled by a Gaussian distribution, with mean equal to zero :

$$p(\boldsymbol{\epsilon}|v_\epsilon) = \mathcal{N}(\boldsymbol{\epsilon}|0, v_\epsilon \boldsymbol{I}). \tag{5.4}$$

The variance of the distribution, v_ϵ, is a fixed scalar in a basic model, with the hypothesis of a stationary noise.

In most of the cases, the variance of the additive noise in the forward model, Eq.(5.3), is unknown. In that case, we consider it as an unknown variable and a prior probability distribution is assigned to it.

To choose a prior model for the variance of a Gaussian likelihood, a conjugate distribution is always preferred. Taking into consideration of the positivity of the variance, the Gamma or Inverse Gamma distributions are commonly used. In our work, we use the Inverse Gamma distribution to take advantage of the Normal-Inverse Gamma bivariate distribution property, which will be presented later.

Définition. *The Inverse Gamma (IG) distribution's probability density function over the support* $x > 0$ *is :*

$$p(x|\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\beta}{\alpha}\right), \tag{5.5}$$

with shape parameter α *and rate parameter* β. $\Gamma(\cdot)$ *denotes the gamma function :*

$$\Gamma(x) = \int_0^{\infty} t^x e^{-t}\, dt. \tag{5.6}$$

As we can see in Figure 5.4 which presents the pdf of the inverse gamma distribution, the variable is always positive, and most of the values are centralized at small values near zero, which exactly adapt to the property of the variance. Indeed, the inverse gamma distribution is a conjugate distribution for a Gaussian likelihood with unknown variance, the Bayesian inference methods are therefore simplified with a simple and regular posterior distribution.

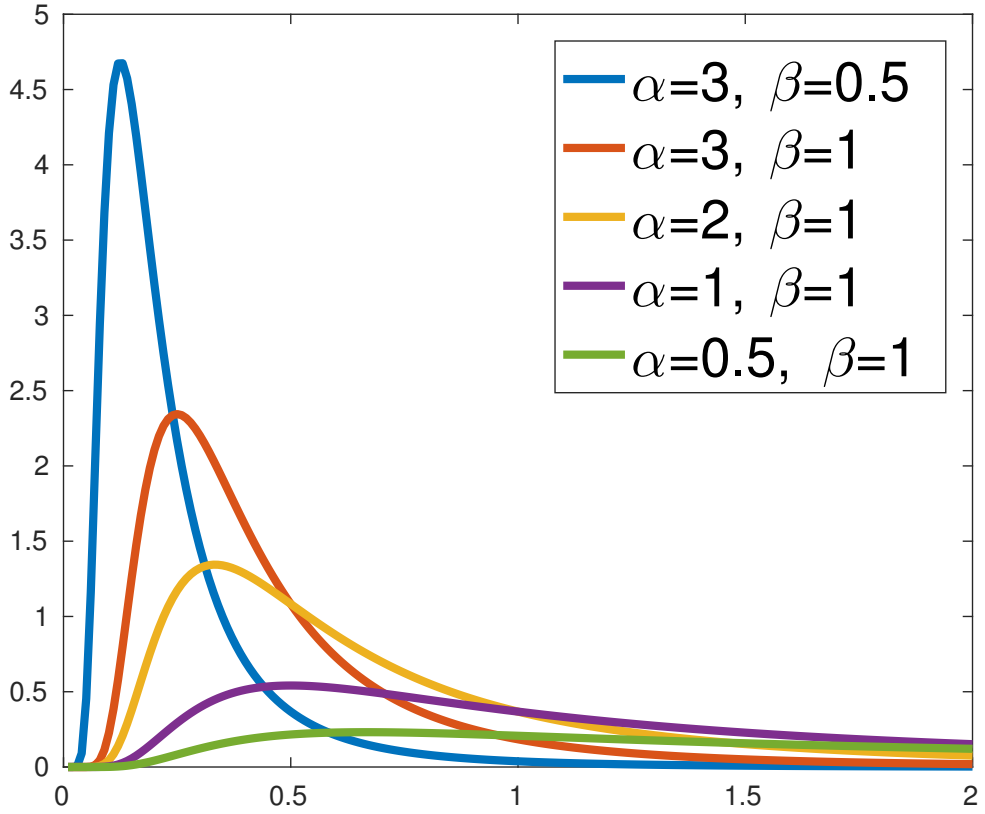


FIGURE 5.4 – The pdfs of Inverse Gamma distribution with different parameter values.

The inverse gamma model for v_ϵ is denoted by :

$$p(v_\epsilon|\alpha_{\epsilon_0},\beta_{\epsilon_0}) = \mathcal{IG}(v_\epsilon|\alpha_{\epsilon_0},\beta_{\epsilon_0}). \tag{5.7}$$

As is presented in Chapter 4, a Gaussian distribution, Eq.(5.4), with the variance belonging to an inverse gamma distribution, Eq.(5.7), leads to a generalized Student-t distribution, which can either approach a Gaussian distribution or tending to a heavy tailed sparse enforcing distribution depending on the choice of the two hyperparameters α_{ϵ_0} and β_{ϵ_0}.

For a more general case, we consider a non-stationary noise, thus for each measurement the variance of the noise is different. This appears for example in X ray CT for a slab material.

The noise will be more important in the projection direction along the longer side of the object, as there will be more absorption and scattering of photons when the radiations pass through the object. In these cases, the variance of noise is defined as a vector, each element of which corresponds to one projection direction. Therefore the prior system model and the noise model are :

$$p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{v}_\epsilon) = \mathcal{N}(\boldsymbol{g}|\boldsymbol{H}\boldsymbol{f},\boldsymbol{V}_\epsilon), \text{ where } \boldsymbol{V}_\epsilon = \text{diag}\left[\boldsymbol{v}_\epsilon\right], \tag{5.8}$$

$$p(\boldsymbol{v}_\epsilon|\alpha_{\epsilon_0},\beta_{\epsilon_0}) = \mathcal{IG}(\boldsymbol{v}_\epsilon|\alpha_{\epsilon_0},\beta_{\epsilon_0}), \tag{5.9}$$

where $\boldsymbol{v}_\epsilon = [v_{\epsilon_1},\cdots,v_{\epsilon_M}]$. By supposing the independence of $v_{\epsilon_i},\ \forall i$, we obtain :

$$p(\boldsymbol{v}_\epsilon|\alpha_{\epsilon_0},\beta_{\epsilon_0}) = \prod_{i=1}^{M} \mathcal{IG}(v_{\epsilon_i}|\alpha_{\epsilon_0},\beta_{\epsilon_0}). \tag{5.10}$$

Each element v_{ϵ_i} belongs to an inverse gamma distribution.

Now, a prior model is defined for the additive noise $\boldsymbol{\epsilon}$. For the reconstruction of $\boldsymbol{f}$, the most important model to be defined is the prior model of the object $\boldsymbol{f}$. As it is presented previously, in the X ray CT the phantoms under consideration are generally piecewise continuous, and the discrete multilevel Haar transform coefficient is a suitable representation of the phantom. We take into consideration the sparsity of the Haar transform coefficient with the relation :

$$\boldsymbol{f} = \boldsymbol{D}\boldsymbol{z}, \tag{5.11}$$

where $\boldsymbol{z}$ is the multilevel Haar transform coefficient of $\boldsymbol{f}$ and $\boldsymbol{D}^{-1}$ is the transformation operator, with $\boldsymbol{z} = \boldsymbol{D}^{-1}\boldsymbol{f}$. As this transformation operator is orthogonal with $\boldsymbol{D}^{-1} = \boldsymbol{D}^T$, and we have $\boldsymbol{z} = \boldsymbol{D}^T\boldsymbol{f}$.

With this relation, we could replace the variable $\boldsymbol{f}$ in Eq.(5.3) by the transformation coefficient $\boldsymbol{z}$:

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{D}\boldsymbol{z} + \boldsymbol{\epsilon}. \tag{5.12}$$

By doing so, using the MAP estimation will lead to exactly the same criterion as the synthesis method, shown in Eq. (5.2).

As the Haar transformation coefficient, the variable $\boldsymbol{z}$ is sparse. Consequently, we use the generalized Student-t distribution, presented in Chapter 4, to define each element z_j :

$$p(z_j|\alpha_{z_0},\beta_{z_0}) = \frac{\Gamma\left(\alpha_{z_0}+\frac{1}{2}\right)}{\sqrt{2\pi\beta_{z_0}}\Gamma\left(\alpha_{z_0}\right)}\left(1+\frac{z_j^2}{2\beta_{z_0}}\right)^{-\left(\alpha_{z_0}+\frac{1}{2}\right)}. \tag{5.13}$$

Thanks to the Gaussian-Inverse Gamma marginal structure of the $\mathcal{S}t_g$ distribution, we could use a hierarchical model to define the hidden variable :

$$\begin{aligned} p(\boldsymbol{z}|\boldsymbol{v}_z) &= \mathcal{N}(\boldsymbol{z}|0,\boldsymbol{V}_z), \text{ where } \boldsymbol{V}_z = \text{diag}\left[\boldsymbol{v}_z\right], \\ p(\boldsymbol{v}_z|\alpha_{z_0},\beta_{z_0}) &= \prod_{j=1}^{N} \mathcal{IG}(v_{z_j}|\alpha_{z_0},\beta_{z_0}), \end{aligned} \tag{5.14}$$

where $\boldsymbol{v}_z = [v_{z_1},\cdots,v_{z_N}]$, and its elements are supposed to be a priori independent and identically distributed.

Now we have all the prior models with the hyper-parameters $(\alpha_{\epsilon_0}, \beta_{\epsilon_0})$ and $(\alpha_{z_0}, \beta_{z_0})$:

$$
\begin{aligned}
p(\boldsymbol{g}|\boldsymbol{z}, \boldsymbol{v}_\epsilon) &= \mathcal{N}(\boldsymbol{g}|\boldsymbol{HDz}, \boldsymbol{V}_\epsilon) \propto |\boldsymbol{V}_\epsilon|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{g}-\boldsymbol{HDz})^T \boldsymbol{V}_\epsilon^{-1}(\boldsymbol{g}-\boldsymbol{HDz})\right\}, \\
p(\boldsymbol{z}|\boldsymbol{v}_z) &= \mathcal{N}(\boldsymbol{z}|0, \boldsymbol{V}_z) \propto |\boldsymbol{V}_z|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{z}^T \boldsymbol{V}_z^{-1}\boldsymbol{z}\right\}, \\
p(\boldsymbol{v}_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) &= \prod_{i=1}^{M} \mathcal{IG}(v_{\epsilon_i}|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \propto \prod_{i=1}^{M} \left\{v_{\epsilon_i}^{-(\alpha_{\epsilon_0}+1)} \exp\left\{-\beta_{\epsilon_0} v_{\epsilon_i}^{-1}\right\}\right\}, \\
p(\boldsymbol{v}_z|\alpha_{z_0}, \beta_{z_0}) &= \prod_{i=1}^{N} \mathcal{IG}(v_{z_j}|\alpha_{z_0}, \beta_{z_0}) \propto \prod_{j=1}^{N} \left\{v_{z_j}^{-(\alpha_{z_0}+1)} \exp\left\{-\beta_{z_0} v_{z_j}^{-1}\right\}\right\}.
\end{aligned} \tag{5.15}
$$

Figure 5.5 present the generative graph of the hierarchical system. In this figure, the circles stand for the unknown variables and hidden variables, and the squares stand for the fixed parameters and hyperparameters.

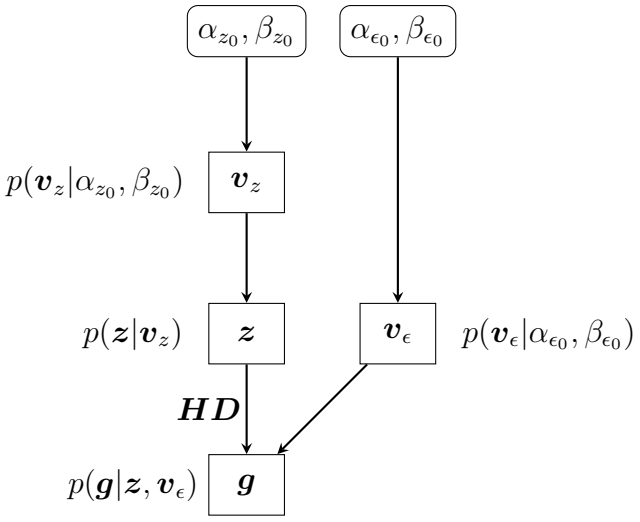


FIGURE 5.5 – Generative graph of the synthesis model.

Then we use the Bayes rule to obtain the posterior distribution :

$$
p(\boldsymbol{f}, \boldsymbol{v}_\epsilon, \boldsymbol{v}_z|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{z}, \boldsymbol{v}_\epsilon) p(\boldsymbol{z}|\boldsymbol{v}_z) p(\boldsymbol{v}_z) p(\boldsymbol{v}_\epsilon) \tag{5.16}
$$

The posterior distribution is obtained by taking into consideration Equations (5.15) :

$$\begin{aligned}
p(\boldsymbol{z}, \boldsymbol{v}_z, \boldsymbol{v}_\epsilon|\boldsymbol{g}) = & |\boldsymbol{V}_\epsilon|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{g}-\boldsymbol{HDz}\right)^T \boldsymbol{V}_\epsilon^{-1}\left(\boldsymbol{g}-\boldsymbol{HDz}\right)\right\}\cdot \\
& |\boldsymbol{V}_z|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{z}^T\boldsymbol{V}_z^{-1}\boldsymbol{z}\right\}\left(\prod_{i=1}^{M} v_{\epsilon_i}^{-(\alpha_{\epsilon_0}+1)} \exp\left\{-\beta_{\epsilon_0} v_{\epsilon_i}^{-1}\right\}\right)\cdot \\
& \prod_{j=1}^{N}\left(v_{z_j}^{-(\alpha_{z_0}+1)} \exp\left\{-\beta_{z_0} v_{z_j}^{-1}\right\}\right) \\
= & \prod_{i=1}^{M} v_{\epsilon_i}^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{M} v_{\epsilon_i}^{-1}\left[\boldsymbol{g}-\boldsymbol{HDz}\right]_i^2\right\}\cdot \\
& \prod_{j=1}^{N} v_{z_j}^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{N} v_{z_j}^{-1}\boldsymbol{z}_j^2\right\}\prod_{i=1}^{M} v_{\epsilon_i}^{-(\alpha_{\epsilon_0}+1)}\prod_{i=1}^{M}\exp\left\{-\beta_{\epsilon_0} v_{\epsilon_i}^{-1}\right\}\cdot \\
& \prod_{j=1}^{N} v_{z_j}^{-(\alpha_{z_0}+1)}\prod_{j=1}^{N}\exp\left\{-\beta_{z_0} v_{z_j}^{-1}\right\}.
\end{aligned} \tag{5.17}$$

From the posterior distribution we can apply a Bayesian point estimation to solve the reconstruction problem.

5.3.2 Bayesian point estimation

The commonly used Bayesian point estimations are the Joint Maximum A Posterior (JMAP) estimation and the Posterior Mean (PM) estimation via Variational Bayesian Approach (VBA). JMAP estimation maximizes the posterior distribution with respect to each variable alternately. This algorithm is convenient for the big data size problems. Meanwhile, the PM algorithm is more complicated. By using a VBA method, the computation could be simplified, but the computational costs are still very expensive for the simulations of 3D phantom. In my work, the JMAP method is used.

In 3D phantom reconstruction, the size of data is so big that the elements of the matrix $\boldsymbol{H}$ are not accessible. The only things that we have access to are the geometric operator of projection $\boldsymbol{Hf}$ and the geometric operator of back-projection $\boldsymbol{H}^T\boldsymbol{g}$. MCMC method is theoretically feasible but is too costly because of the huge data size and the huge amount of variables.

5.3.3 The Joint MAP estimation.

In the JMAP estimation, the maximization of the posterior distribution is changed to the minimization of the negative logarithm of the posterior distribution. This comes from the fact that the distributions are always in the form of an exponential of a negative polynomial, and that the logarithm operation is monotonous :

$$\begin{aligned}
(\widehat{\boldsymbol{z}}, \widehat{\boldsymbol{v}}_z, \widehat{\boldsymbol{v}}_\epsilon) &= \arg\max_{\boldsymbol{z},\boldsymbol{v}_z,\boldsymbol{v}_\epsilon}\left\{p\left(\boldsymbol{z}, \boldsymbol{v}_z, \boldsymbol{v}_\epsilon|\boldsymbol{g}\right)\right\} \\
&= \arg\min_{\boldsymbol{z},\boldsymbol{v}_z,\boldsymbol{v}_\epsilon}\left\{-\ln p\left(\boldsymbol{z}, \boldsymbol{v}_z, \boldsymbol{v}_\epsilon|\boldsymbol{g}\right)\right\}.
\end{aligned} \tag{5.18}$$

From Eq.(5.17), we calculate the criterion of the optimization :

$$\begin{aligned} J(\boldsymbol{z},\boldsymbol{v}_z,\boldsymbol{v}_\epsilon) =& -\ln p\left(\boldsymbol{z},\boldsymbol{v}_z,\boldsymbol{v}_\epsilon|\boldsymbol{g}\right) \\ =& \frac{1}{2}\sum_{i=1}^{M}\ln v_{\epsilon_i} + \frac{1}{2}\sum_{i=1}^{M} v_{\epsilon_i}^{-1}\left[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{D}\boldsymbol{z}\right]_i^2 + \frac{1}{2}\sum_{j=1}^{N}\ln v_{z_j} + \frac{1}{2}\sum_{j=1}^{N} v_{z_j}^{-1} z_j^2 \\ &+ (\alpha_{\epsilon_0}+1)\sum_{i=1}^{M}\ln v_{\epsilon_i} + \beta_{\epsilon_0}\sum_{i=1}^{M}\ln v_{\epsilon_i} + (\alpha_{z_0}+1)\sum_{j=1}^{N}\ln v_{z_j} + \beta_{z_0}\sum_{j=1}^{N}\ln v_{z_j}. \end{aligned} \tag{5.19}$$

In order to estimate three unknown variables from one joint criterion, we use an alternate optimization method, which estimates each of the unknown variable alternately. When optimizing one unknown variable, the others are fixed. It is called the Joint Maximum A Posteriori (JMAP) estimation in this thesis.

The optimization criterion for variables are respectively :

$$J(\boldsymbol{z}) = \frac{1}{2}\sum_{i=1}^{M} v_{\epsilon_i}^{-1}\left[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{D}\boldsymbol{z}\right]_i^2 + \frac{1}{2}\sum_{j=1}^{N} v_{z_j}^{-1} z_j^2, \tag{5.20}$$

$$J(\boldsymbol{v}_z) = \frac{1}{2}\sum_{j=1}^{N}\ln v_{z_j} + \frac{1}{2}\sum_{j=1}^{N} v_{z_j}^{-1} z_j^2 + (\alpha_{z_0}+1)\sum_{j=1}^{N}\ln v_{z_j} + \beta_{z_0}\sum_{j=1}^{N}\ln v_{z_j}^{-1}, \tag{5.21}$$

$$J(\boldsymbol{v}_\epsilon) = \frac{1}{2}\sum_{i=1}^{M}\ln v_{\epsilon_i} + \frac{1}{2}\sum_{i=1}^{M} v_{\epsilon_i}^{-1}\left[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{D}\boldsymbol{z}\right]_i^2 + (\alpha_{\epsilon_0}+1)\sum_{i=1}^{M}\ln v_{\epsilon_i} + \beta_{\epsilon_0}\sum_{i=1}^{M}\ln v_{\epsilon_i}^{-1}. \tag{5.22}$$

By considering $v_{z_j},\ \forall j \in [1,N]$ and $v_{\epsilon_i},\ \forall i \in [1,M]$ separately accounting for the hypothesis that each of them are independent, we obtain :

$$J(\boldsymbol{z}) = \frac{1}{2}\sum_{i=1}^{M} v_{\epsilon_i}^{-1}\left[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{D}\boldsymbol{z}\right]_i^2 + \frac{1}{2}\sum_{j=1}^{N} v_{z_j}^{-1} z_j^2, \tag{5.23}$$

$$J(v_{z_j}) = \left(\alpha_{z_0}+\frac{3}{2}\right)\ln v_{z_j} + \left(\beta_{z_0}+\frac{1}{2}z_j^2\right) v_{z_j}^{-1}, \tag{5.24}$$

$$J(v_{\epsilon_i}) = \left(\alpha_{\epsilon_0}+\frac{3}{2}\right)\ln v_{\epsilon_i} + \left(\beta_{\epsilon_0}+\frac{1}{2}\left[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{D}\boldsymbol{z}\right]_i^2\right) v_{\epsilon_i}^{-1}. \tag{5.25}$$

The gradient of $\nabla J(\boldsymbol{z})$ is :

$$\begin{aligned} \nabla J(\boldsymbol{z}) =& \frac{\mathrm{d}\left(\frac{1}{2}\left(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{D}\boldsymbol{z}\right)^T\boldsymbol{V}_\epsilon^{-1}\left(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{D}\boldsymbol{z}\right)+\frac{1}{2}\boldsymbol{z}^T\boldsymbol{V}_z^{-1}\boldsymbol{z}\right)}{\mathrm{d}\boldsymbol{z}} \\ =& \frac{1}{2}\left(-\boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{g} - \boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{g} + 2\boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{H}\boldsymbol{D}\boldsymbol{z}\right) + \frac{1}{2}\cdot 2\boldsymbol{V}_z^{-1}\boldsymbol{z} \\ =& -\boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{g} + \boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{H}\boldsymbol{D}\boldsymbol{z} + \boldsymbol{V}_z^{-1}\boldsymbol{z} \\ =& -\boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{g} + \left(\boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{H}\boldsymbol{D} + \boldsymbol{V}_z^{-1}\right)\boldsymbol{z}. \end{aligned} \tag{5.26}$$

A local minimum value $\widehat{\boldsymbol{z}}$ appears when $\nabla J(\boldsymbol{z})|_{\widehat{\boldsymbol{z}}} = 0$:

$$-\boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{g} + \left(\boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{H}\boldsymbol{D} + \boldsymbol{V}_z^{-1}\right)\widehat{\boldsymbol{z}} = 0. \tag{5.27}$$

And so :

$$\widehat{\boldsymbol{z}} = \left(\boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{H}\boldsymbol{D} + \boldsymbol{V}_z^{-1}\right)^{-1}\boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{g}. \tag{5.28}$$

However, this doesn't work in the 3D X ray CT problems. As in the huge data size case, it is impossible to calculate the inversion of the matrix $\left(\boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{H}\boldsymbol{D} + \boldsymbol{V}_z^{-1}\right)$, needless to say to do this computation in each iteration.

Some algorithms are proposed and used to solve this problem, for example the gradient descent algorithm [Bat92, VCR97], the conjugate gradient algorithm [FB99, Wu01], etc. In our work we used the gradient descent method to optimize the variable, by iteratively computing :

$$\widehat{\boldsymbol{z}}^{(k+1)} = \widehat{\boldsymbol{z}}^{(k)} - \gamma_z^{(k)}\nabla J(\widehat{\boldsymbol{z}}^{(k)}), \tag{5.29}$$

where $\nabla J(\widehat{\boldsymbol{z}}^{(k)}) = \frac{\mathrm{d}J(\boldsymbol{z})}{\mathrm{d}\boldsymbol{z}}$ is the gradient of the criterion and $\gamma_z^{(k)}$ is the descent step length. $\gamma_z^{(k)}$ can either be fixed as a constant or change in every iteration adapting the gradient. The last case is called the optimized step length strategy [BV04], in which $\gamma_z^{(k)}$ is calculated in each iteration by minimizing the criterion :

$$\gamma_z^{(k)} = \arg\min_{\gamma_z}\left\{J(\widehat{\boldsymbol{z}}^{(k)} - \gamma_z\nabla J(\widehat{\boldsymbol{z}}^{(k)}))\right\}, \tag{5.30}$$

where $J(\boldsymbol{z} - \gamma_z\nabla J_z)$ is simplified as :

$$\begin{aligned}
&J(\boldsymbol{z} - \gamma_z\nabla J(\boldsymbol{z}))\\
=&\frac{1}{2}\left(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{D}\left(\boldsymbol{z} - \gamma_z\nabla J(\boldsymbol{z})\right)\right)^T\boldsymbol{V}_\epsilon^{-1}\left(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{D}\left(\boldsymbol{z} - \gamma_z\nabla J(\boldsymbol{z})\right)\right)\\
&+\frac{1}{2}\left(\boldsymbol{z} - \gamma_z\nabla J(\boldsymbol{z})\right)^T\boldsymbol{V}_z^{-1}\left(\boldsymbol{z} - \gamma_z\nabla J(\boldsymbol{z})\right)\\
=&\frac{1}{2}\left(\boldsymbol{g}^T - \boldsymbol{z}^T\boldsymbol{D}^T\boldsymbol{H}^T + \gamma_z\nabla J(\boldsymbol{z})^T\boldsymbol{D}^T\boldsymbol{H}^T\right)\boldsymbol{V}_\epsilon^{-1}\left(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{D}\boldsymbol{z} + \gamma_z\boldsymbol{H}\boldsymbol{D}\nabla J(\boldsymbol{z})\right)\\
&+\frac{1}{2}\left(\boldsymbol{z}^T - \gamma_z\nabla J(\boldsymbol{z})^T\right)\boldsymbol{V}_z^{-1}\left(\boldsymbol{z} - \gamma_z\nabla J(\boldsymbol{z})\right)\\
=&\frac{1}{2}\left(\gamma_z\boldsymbol{g}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{H}\boldsymbol{D}\nabla J(\boldsymbol{z}) - \gamma_z\boldsymbol{z}^T\boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{H}\boldsymbol{D}\nabla J(\boldsymbol{z}) + \gamma_z\nabla J(\boldsymbol{z})^T\boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{g}\right.\\
&\left.-\gamma_z\nabla J(\boldsymbol{z})^T\boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{H}\boldsymbol{D}\boldsymbol{z} + \gamma_z^2\nabla J(\boldsymbol{z})^T\boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{H}\boldsymbol{D}\nabla J(\boldsymbol{z})\right)\\
&+\frac{1}{2}\left(-\gamma_z\boldsymbol{z}^T\boldsymbol{V}_z^{-1}\nabla J(\boldsymbol{z}) - \gamma_z\nabla J(\boldsymbol{z})^T\boldsymbol{V}_z^{-1}\boldsymbol{z} + \gamma_z^2\nabla J(\boldsymbol{z})^T\boldsymbol{V}_z^{-1}\nabla J(\boldsymbol{z})\right) + \mathrm{Cst}\\
=&\frac{1}{2}\gamma_z\left(2\nabla J(\boldsymbol{z})^T\boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{g} - 2\nabla J(\boldsymbol{z})^T\boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{H}\boldsymbol{D}\boldsymbol{z} - 2\nabla J(\boldsymbol{z})^T\boldsymbol{V}_z^{-1}\boldsymbol{z}\right)\\
&+\frac{1}{2}\gamma_z^2\left(\nabla J(\boldsymbol{z})^T\boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{H}\boldsymbol{D}\nabla J(\boldsymbol{z}) + \nabla J(\boldsymbol{z})^T\boldsymbol{V}_z^{-1}\nabla J(\boldsymbol{z})\right) + \mathrm{Cst}\\
=&\gamma_z\nabla J(\boldsymbol{z})^T\left[\boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{g} - \boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{H}\boldsymbol{D}\boldsymbol{z} - \boldsymbol{V}_z^{-1}\boldsymbol{z}\right]\\
&+\frac{1}{2}\gamma_z^2\left[\|\boldsymbol{Y}_\epsilon\boldsymbol{H}\boldsymbol{D}\nabla J(\boldsymbol{z})\|_2^2 + \|\boldsymbol{Y}_z\nabla J(\boldsymbol{z})\|_2^2\right] + \mathrm{Cst},
\end{aligned}$$

where $\boldsymbol{Y}_\epsilon = \boldsymbol{V}_\epsilon^{-\frac{1}{2}}$ and $\boldsymbol{Y}_z = \boldsymbol{V}_z^{-\frac{1}{2}}$. From Eq.(5.26) we have :

$$\begin{aligned}
J(\boldsymbol{z}) &= -\boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\left(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{D}\boldsymbol{z}\right) + \boldsymbol{V}_z^{-1}\boldsymbol{z}\\
&= -\boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{g} + \boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{H}\boldsymbol{D}\boldsymbol{z} + \boldsymbol{V}_z^{-1}\boldsymbol{z}.
\end{aligned}$$

So that :

$$J(\boldsymbol{z}-\gamma_z\nabla J(\boldsymbol{z})) = -\gamma_z\nabla J(\boldsymbol{z})^T\nabla J(\boldsymbol{z}) + \frac{1}{2}\gamma_z^2\left[\|\boldsymbol{Y}_\epsilon\boldsymbol{H}\boldsymbol{D}\nabla J(\boldsymbol{z})\|_2^2 + \|\boldsymbol{Y}_z\nabla J(\boldsymbol{z})\|_2^2\right] + \text{Cst}$$
$$= -\gamma_z\|\nabla J(\boldsymbol{z})\|_2^2 + \frac{1}{2}\gamma_z^2\left[\|\boldsymbol{Y}_\epsilon\boldsymbol{H}\boldsymbol{D}\nabla J(\boldsymbol{z})\|_2^2 + \|\boldsymbol{Y}_z\nabla J(\boldsymbol{z})\|_2^2\right].$$

The gradient of the criterion is :

$$\frac{\mathrm{d}J(\boldsymbol{z}-\gamma_z\nabla J(\boldsymbol{z}))}{\mathrm{d}\gamma_z} = -\|\nabla J(\boldsymbol{z})\|_2^2 + \gamma_z\left[\|\boldsymbol{Y}_\epsilon\boldsymbol{H}\boldsymbol{D}\nabla J(\boldsymbol{z})\|_2^2 + \|\boldsymbol{Y}_z\nabla J(\boldsymbol{z})\|_2^2\right].$$

By setting

$$\frac{\mathrm{d}J(\boldsymbol{z}-\gamma_z\nabla J(\boldsymbol{z}))}{\mathrm{d}\gamma_z}\Big|_{\gamma_z^{(k)}} = 0,$$

we get :

$$\gamma_z^{(k)} = \frac{\|\nabla J(\boldsymbol{z})\|_2^2}{\|\boldsymbol{Y}_\epsilon\boldsymbol{H}\boldsymbol{D}\nabla J(\boldsymbol{z})\|_2^2 + \|\boldsymbol{Y}_z\nabla J(\boldsymbol{z})\|_2^2}.$$

Therefore, the gradient descent optimization for variable $\boldsymbol{z}$ is summarized as following :

$$\widehat{\boldsymbol{z}}^{(k+1)} = \widehat{\boldsymbol{z}}^{(k)} - \gamma_z^{(k)}\nabla J(\widehat{\boldsymbol{z}}^{(k)}), \tag{5.31}$$
$$\gamma_z^{(k)} = \frac{\|\nabla J(\boldsymbol{z})\|_2^2}{\|\boldsymbol{Y}_\epsilon\boldsymbol{H}\boldsymbol{D}\nabla J(\boldsymbol{z})\|_2^2 + \|\boldsymbol{Y}_z\nabla J(\boldsymbol{z})\|_2^2}, \tag{5.32}$$

with

$$J(\boldsymbol{z}) = \frac{1}{2}\sum_{i=1}^{M} v_{\epsilon_i}^{-1}\left[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{D}\boldsymbol{z}\right]_i^2 + \frac{1}{2}\sum_{j=1}^{N} v_{z_j}^{-1}z_j^2, \tag{5.33}$$
$$\nabla J(\boldsymbol{z}) = -\boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{g} + \boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{H}\boldsymbol{D}\boldsymbol{z} + \boldsymbol{V}_z^{-1}\boldsymbol{z}. \tag{5.34}$$

Back to the Eq.(5.21) and Eq.(5.22), we calculate the gradient of these two criterions :

$$\nabla J(v_{z_j}) = \left(\alpha_{z_0} + \frac{3}{2}\right)v_{z_j}^{-1} - \left(\beta_{z_0} + \frac{1}{2}z_j^2\right)v_{z_j}^{-2}, \tag{5.35}$$
$$\nabla J(v_{\epsilon_i}) = \left(\alpha_{\epsilon_0} + \frac{3}{2}\right)v_{\epsilon_i}^{-1} - \left(\beta_{\epsilon_0} + \frac{1}{2}\left[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{D}\boldsymbol{z}\right]_i^2\right)v_{\epsilon_i}^{-2}. \tag{5.36}$$

Now we discuss if the local minimum is the global minimum. In order to consider both of these two variables, we consider a general case, for the criterion of the form

$$J(x) = a\ln x + bx^{-1},$$

where $a > 0$, $b > 0$ and $x \geq 0$. First of all, we look for the point where the gradient of the criterion equals to zero :

$$\nabla J(x) = \frac{a}{x} - \frac{b}{x^2} = \frac{1}{x}\left(a - \frac{b}{x}\right),$$

where $x = 0$ is not defined in this case, so the only zero point is when $x_0 = \frac{b}{a}$.

For $x > x_0 = b/a$, we have :

$$\frac{1}{x} < \frac{a}{b} \implies \frac{b}{x} < a \implies -\frac{b}{x} > -a \implies a - \frac{b}{x} > 0 \implies \frac{1}{x}\left(a - \frac{b}{x}\right) > 0$$
$$\implies \nabla J(x) > 0.$$

For $0 < x < x_0 = b/a$

$$\frac{1}{x} > \frac{a}{b} \implies \frac{b}{x} > a \implies -\frac{b}{x} > -a \implies a - \frac{b}{x} < 0 \implies \frac{1}{x}\left(a - \frac{b}{x}\right) < 0$$
$$\implies \nabla J(x) < 0.$$

And when $x \to 0_+ \implies \nabla J(x) \to -\infty$, where 0_+ is a positive value infinitely approaching zero.

We have demonstrated that when $x < x_0$ the gradient is always negative and when $x > x_0$ the gradient is always positive. So that the critical value $x_0 = b/a$ is exactly the global minimum of the criterion. We get :

$$\widehat{v}_{zj} = \frac{\beta_{z_0} + \frac{1}{2} z_j^2}{\alpha_{z_0} + \frac{3}{2}}, \tag{5.37}$$

$$\widehat{v}_{\epsilon_i} = \frac{\beta_{\epsilon_0} + \frac{1}{2}\left[\boldsymbol{g} - \boldsymbol{H}\boldsymbol{D}\boldsymbol{z}\right]_i^2}{\alpha_{\epsilon_0} + \frac{3}{2}}. \tag{5.38}$$

5.3.4 Algorithm and implementation

The variables $\widehat{\boldsymbol{z}}$, $\widehat{v}_{zj}$ and $\widehat{v}_{\epsilon_i}$ are optimized alternately and iteratively by using Eq.(5.31)-Eq.(5.34), Eq.(5.37) and Eq.(5.38). The computation algorithm for updating the variable $\boldsymbol{z}$ is shown in Algorithm 1.

Algorithm 1 Gradient descent algorithm for $\widehat{\boldsymbol{z}}$

Input : $\boldsymbol{H}$, $\boldsymbol{D}$, $\boldsymbol{g}$
Output : $\widehat{\boldsymbol{z}}^{(k+1)}$
Initialization : $\widehat{\boldsymbol{z}}^{(0)}$
$k = 0$
repeat
    $k = k + 1$
    Calculate $\nabla J(\widehat{\boldsymbol{z}}^{(k)})$ according to Eq.(5.34)
    Update $\gamma_z^{(k)}$ according to Eq.(5.32)
    Update $\widehat{\boldsymbol{z}}^{(k+1)} = \widehat{\boldsymbol{z}}^{(k)} - \gamma_z^{(k)} \nabla J(\widehat{\boldsymbol{z}}^{(k)})$
until convergence or max iteration

The entire optimization algorithm which estimate all the variables are shown in Algorithm 2.

After the estimation of the variable $\boldsymbol{z}$, a post processing is applied in order to obtain the reconstruction of phantom :

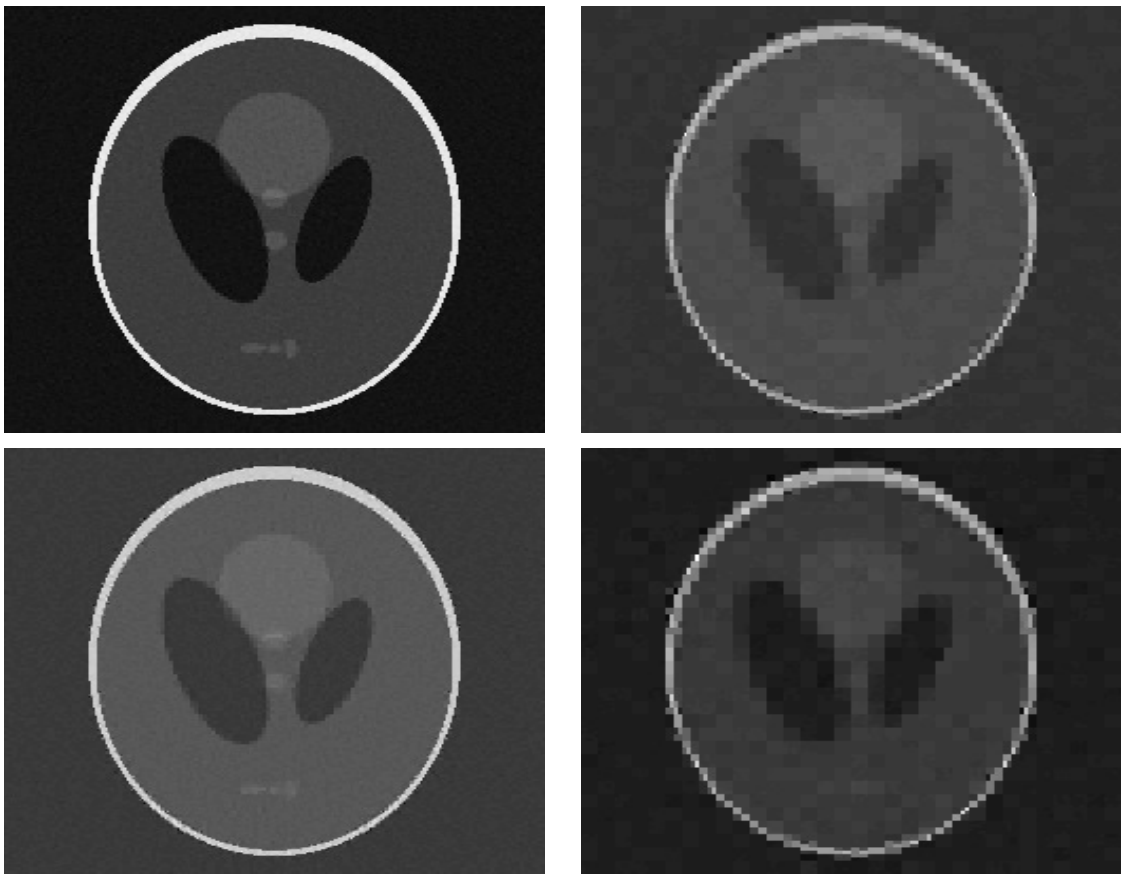
$$\widehat{\boldsymbol{f}} = \boldsymbol{D}\widehat{\boldsymbol{z}}. \tag{5.39}$$

Algorithm 2 Summary of the JMAP algorithm for the synthesis Bayesian hierarchical method

1: Fix parameters α_{z_0}, β_{z_0}, α_{ϵ_0}, β_{ϵ_0}
2: **Input :** $\boldsymbol{H}$, $\boldsymbol{D}$, $\boldsymbol{g}$
3: **Output :** $\widehat{\boldsymbol{z}}$, $\widehat{\boldsymbol{v}}_z$, $\widehat{\boldsymbol{v}}_\epsilon$
4: **Initialization :** $\widehat{\boldsymbol{z}}^{(0)} = \boldsymbol{D}^{-1}\widehat{\boldsymbol{f}}_{FBP}$, where $\widehat{\boldsymbol{f}}_{FBP}$ is the result of the FBP method.
5: **repeat**
6: Update $\widehat{\boldsymbol{z}}$ according to Algorithm 1
7: Compute $\boldsymbol{v}_z$ according to Eq.(5.37)
8: Compute $\boldsymbol{v}_\epsilon$ according to Eq.(5.38)
9: **until** convergence or max iteration

5.3.5 Simulation results

We use the projections dataset of the Shepp Logan phantom of size 256^3 to reconstruct the phantom. The dataset is obtained respectively by using : 128 projections with a high SNR=40dB, 128 projections with a low SNR=20dB, 64 projections with a high SNR=40dB and 64 projections with a low SNR=20dB.



40dB 20dB

FIGURE 5.6 – Middle slice of reconstructed Shepp Logan phantom of size 256^3 from projection sinogram with 128 projections (top) or 64 projections (bottom) and a noise of SNR=40dB (left) or SNR=20dB (right).

From the results, we can see that from the reconstructed coefficient, we obtain a reconstruc-

ted phantom with block artifacts. This phenomena comes from the fact that, the inverse wavelet transformation can only reconstruct to a perfect object when the coefficient is not contaminated. We can see that the artifacts has square shapes as the Haar transformation considers only two wavelet direction.

Obviously, this result with the square artifacts is unsatisfactory. We propose another Bayesian method basing on this method. The new Bayesian method estimates the object $\boldsymbol{f}$ and the transform coefficient $\boldsymbol{z}$ simultaneously. The estimated coefficient is used during the reconstruction of the object in each iteration. This proposed method is named the HHBM (Haar transformation based Hierarchical Bayesian Method) in our work.

5.4 HHBM method

From the simulation results of the previously presented method, we find that the final result of the reconstruction contains a lot of artifacts. This phenomenon comes from the last step of the process : the transform from $\widehat{\boldsymbol{z}}$ to $\widehat{\boldsymbol{f}}$. A small error on some of the coefficients, z_j, affects the image $\boldsymbol{f}$ greatly, in particular when z_j is in the upper level. So we propose another Bayesian method, basing on the previous method, but reconstruct the phantom $\widehat{\boldsymbol{f}}$ simultaneously while estimating the coefficient $\boldsymbol{z}$ and other variables.

The system forward model is the same :

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}. \tag{5.40}$$

However, instead of using an exact discrete multilevel Haar transform from $\boldsymbol{f}$ to $\boldsymbol{z}$, we consider a relaxed relation :

$$\boldsymbol{f} = \boldsymbol{D}\boldsymbol{z} + \boldsymbol{\xi}, \tag{5.41}$$

where $\boldsymbol{\xi}$ is an additive noise representing uncertainties of the transformation. While choosing the prior model for $\boldsymbol{\xi}$, the objective is that $\boldsymbol{f}$ and $\boldsymbol{D}\boldsymbol{z}$ are as approximate as possible. Therefore we define $\boldsymbol{\xi}$ as a sparse noise with most of the values approaching to zero. The generalized Student-t distribution is used to model $\boldsymbol{\xi}$, enforcing its sparse structure. The $\mathcal{S}t_g$ distribution for $\boldsymbol{\xi}$ is defined via the hierarchical model :

$$p(\boldsymbol{\xi}|\boldsymbol{v}_\xi) = \mathcal{N}(\boldsymbol{\xi}|0, \boldsymbol{V}_\xi) \propto |\boldsymbol{V}_\xi|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{\xi}^T \boldsymbol{V}_\xi^{-1}\boldsymbol{\xi}\right\}, \tag{5.42}$$

$$p(\boldsymbol{v}_\xi|\alpha_{\xi_0}, \beta_{\xi_0}) = \prod_{j=1}^{N} \mathcal{IG}\left(v_{\xi_j}|\alpha_{\xi_0}, \beta_{\xi_0}\right) \propto \prod_{j=1}^{N} \left(v_{\xi_j}^{-(\alpha_{\xi_0}+1)} \exp\left\{-\beta_{\xi_0} v_{\xi_j}^{-1}\right\}\right), \tag{5.43}$$

where $\boldsymbol{V}_\xi = \text{diag}\left[\boldsymbol{v}_\xi\right]$, $\boldsymbol{v}_\xi = \left[v_{\xi_1}, \cdots, v_{\xi_j}, \cdots, v_{\xi_N}\right]$, and v_{ξ_j} are iid. In this model, both $\boldsymbol{z}$ and $\boldsymbol{\xi}$ are sparse and belong to a $\mathcal{S}t_g$ distribution.

With this forward transform model and the prior model of $\boldsymbol{\xi}$, we have :

$$\begin{aligned} p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{v}_\xi) =& \mathcal{N}(\boldsymbol{f}|\boldsymbol{D}\boldsymbol{z}, \boldsymbol{V}_\xi) \\ &\propto |\boldsymbol{V}_\xi|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}\right)^T \boldsymbol{V}_\xi^{-1}\left(\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}\right)\right\}. \end{aligned} \tag{5.44}$$

Vector $\boldsymbol{z} = [z_1, z_2, \cdots, z_N]$ represents the discrete multi-level Haar transform coefficient of piece-wise continuous phantom $\boldsymbol{f}$. We use the generalized Student-t distribution ($\mathcal{S}t_g$) [Dum16] to model $\boldsymbol{z}$, enforcing the sparsity property :

$$p(\boldsymbol{z}|\boldsymbol{v}_z) = \mathcal{N}(\boldsymbol{z}|0, \boldsymbol{V}_z), \tag{5.45}$$

$$p(\boldsymbol{v}_z|\alpha_{z_0}, \beta_{z_0}) = \prod_{j=1}^{N} \mathcal{IG}(v_{z_j}|\alpha_{z_0}, \beta_{z_0}). \tag{5.46}$$

Thanks to the fact that the Inverse Gamma distribution is conjugate for a Gaussian likelihood with known experience and unknown variance, using this $\mathcal{S}t_g$ distribution simplifies the calculations when using the Bayesian point optimization methods, for example the Posterior Mean via Variational Bayesian Approximation (VBA) method [AMD10].

We list out all the proposed likelihood and prior models depending on the prior property of each variable :

$$p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{v}_\epsilon) \propto |\boldsymbol{V}_\epsilon|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f})^T \boldsymbol{V}_\epsilon^{-1}(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f})\right\} \tag{5.47}$$

$$p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{v}_\xi) \propto |\boldsymbol{V}_\xi|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z})^T \boldsymbol{V}_\xi^{-1}(\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z})\right\}, \tag{5.48}$$

$$p(\boldsymbol{z}|\boldsymbol{v}_z) \propto |\boldsymbol{V}_z|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{z}^T \boldsymbol{V}_z^{-1}\boldsymbol{z}\right\}, \tag{5.49}$$

$$p(\boldsymbol{v}_z|\alpha_{z_0}, \beta_{z_0}) \propto \prod_{j=1}^{N} \left(v_{z_j}^{-(\alpha_{z_0}+1)} \exp\left\{-\beta_{z_0} v_{z_j}^{-1}\right\}\right), \tag{5.50}$$

$$p(\boldsymbol{v}_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \propto \prod_{i=1}^{M} \left(v_{\epsilon_i}^{-(\alpha_{\epsilon_0}+1)} \exp\left\{-\beta_{\epsilon_0} v_{\epsilon_i}^{-1}\right\}\right), \tag{5.51}$$

$$p(\boldsymbol{v}_\xi|\alpha_{\xi_0}, \beta_{\xi_0}) \propto \prod_{j=1}^{N} \left(v_{\xi_j}^{-(\alpha_{\xi_0}+1)} \exp\left\{-\beta_{\xi_0} v_{\xi_j}^{-1}\right\}\right). \tag{5.52}$$

In this model, v_{z_j}, v_{ϵ_i} and v_{ξ_j}, $\forall i \in [1, M]$, $\forall j \in [1, N]$, are supposed to be independent and identically distributed (iid). The generative graph of this hierarchical structured model is given in Figure 5.7.

By using the Bayes rule, Eq.(3.39), the posterior distribution is obtained :

$$\begin{aligned} p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_\epsilon, \boldsymbol{v}_\xi, \boldsymbol{v}_z|\boldsymbol{g}) &= \frac{p(\boldsymbol{g}, \boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_\epsilon, \boldsymbol{v}_\xi, \boldsymbol{v}_z)}{p(\boldsymbol{g})} \\ &= \frac{p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{v}_\epsilon)p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{v}_\xi)p(\boldsymbol{z}|\boldsymbol{v}_z)p(\boldsymbol{v}_z)p(\boldsymbol{v}_\epsilon)p(\boldsymbol{v}_\xi)}{p(\boldsymbol{g})} \\ &\propto p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{v}_\epsilon)p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{v}_\xi)p(\boldsymbol{z}|\boldsymbol{v}_z)p(\boldsymbol{v}_z)p(\boldsymbol{v}_\epsilon)p(\boldsymbol{v}_\xi). \end{aligned} \tag{5.53}$$

Substituting the prior distribution models in Eq.(5.47)-Eq.(5.52), we get the posterior distri-

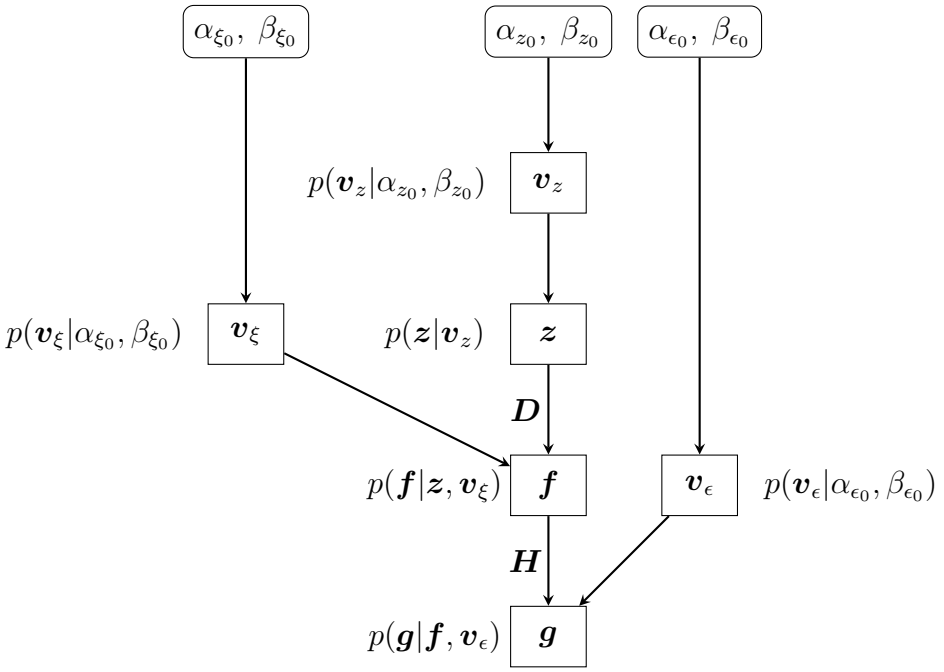


FIGURE 5.7 – Generative graph of the HHBM method.

bution :

$$
\begin{aligned}
&p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{v}_z,\boldsymbol{v}_\epsilon,\boldsymbol{v}_\xi|\boldsymbol{g}) \\
&\propto |\boldsymbol{V}_\epsilon|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f})^T \boldsymbol{V}_\epsilon^{-1}(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f})\right\} \\
&\quad |\boldsymbol{V}_\xi|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z})^T \boldsymbol{V}_\xi^{-1}(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z})\right\} \\
&\quad |\boldsymbol{V}_z|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{z}^T\boldsymbol{V}_z^{-1}\boldsymbol{z}\right\} \prod_{j=1}^{N}\left(v_{z_j}^{-(\alpha_{z_0}+1)} \exp\left\{-\beta_{z_0}v_{z_j}^{-1}\right\}\right) \\
&\quad \prod_{i=1}^{M}\left(v_{\epsilon_i}^{-(\alpha_{\epsilon_0}+1)} \exp\left\{-\beta_{\epsilon_0}v_{\epsilon_i}^{-1}\right\}\right) \prod_{j=1}^{N}\left(v_{\xi_j}^{-(\alpha_{\xi_0}+1)} \exp\left\{-\beta_{\xi_0}v_{\xi_j}^{-1}\right\}\right) \\
&= \prod_{i=1}^{M} v_{\epsilon_i}^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f})^T \boldsymbol{V}_\epsilon^{-1}(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f})\right\} \\
&\quad \prod_{j=1}^{N} v_{\xi_j}^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z})^T \boldsymbol{V}_\xi^{-1}(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z})\right\} \\
&\quad \prod_{j=1}^{N} v_{z_j}^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{z}^T\boldsymbol{V}_z^{-1}\boldsymbol{z}\right\} \prod_{j=1}^{N} v_{z_j}^{-(\alpha_{z_0}+1)} \prod_{j=1}^{N} \exp\left\{-\beta_{z_0}v_{z_j}^{-1}\right\} \\
&\quad \prod_{i=1}^{M} v_{\epsilon_i}^{-(\alpha_{\epsilon_0}+1)} \prod_{i=1}^{M} \exp\left\{-\beta_{\epsilon_0}v_{\epsilon_i}^{-1}\right\} \prod_{j=1}^{N} v_{\xi_j}^{-(\alpha_{\xi_0}+1)} \prod_{j=1}^{N} \exp\left\{-\beta_{\xi_0}v_{\xi_j}^{-1}\right\}.
\end{aligned}
\tag{5.54}
$$

As the reasons presented in the previous section, the JMAP estimation algorithm is used to optimize all the unknown variables.

5.4.1 Joint Maximum A Posterior estimation

The Joint Maximum A Posterior (JMAP) method estimate all the unknown variables by maximizing the posterior distribution obtained below :

$$\left(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{z}}, \widehat{\boldsymbol{v}}_z, \widehat{\boldsymbol{v}}_\epsilon, \widehat{\boldsymbol{v}}_\xi\right) = \arg \max_{\boldsymbol{f},\boldsymbol{z},\boldsymbol{v}_z,\boldsymbol{v}_\epsilon,\boldsymbol{v}_\xi} \left\{p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_z, \boldsymbol{v}_\epsilon, \boldsymbol{v}_\xi|\boldsymbol{g})\right\}. \tag{5.55}$$

As the posterior distribution has the form of exponential distribution, it becomes a minimization of the negative logarithm of the posterior distribution :

$$\begin{aligned}\left(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{z}}, \widehat{\boldsymbol{v}}_z, \widehat{\boldsymbol{v}}_\epsilon, \widehat{\boldsymbol{v}}_\xi\right) &= \arg \min_{\boldsymbol{f},\boldsymbol{z},\boldsymbol{v}_z,\boldsymbol{v}_\epsilon,\boldsymbol{v}_\xi} \left\{-\ln p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_z, \boldsymbol{v}_\epsilon, \boldsymbol{v}_\xi|\boldsymbol{g})\right\} \\ &= \arg \min_{\boldsymbol{f},\boldsymbol{z},\boldsymbol{v}_z,\boldsymbol{v}_\epsilon,\boldsymbol{v}_\xi} \left\{J(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_z, \boldsymbol{v}_\epsilon, \boldsymbol{v}_\xi)\right\}.\end{aligned} \tag{5.56}$$

So, the criterion of optimization is :

$$\begin{aligned}J(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_z, \boldsymbol{v}_\epsilon, \boldsymbol{v}_\xi) =& -\ln p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_z, \boldsymbol{v}_\epsilon, \boldsymbol{v}_\xi|\boldsymbol{g}) \\ =& \frac{1}{2}\sum_{i=1}^{M} \ln v_{\epsilon_i} + \frac{1}{2}(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f})^T \boldsymbol{V}_\epsilon^{-1} (\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}) + \frac{1}{2}\sum_{j=1}^{N} \ln v_{\xi_j} \\ &+ \frac{1}{2}(\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z})^T \boldsymbol{V}_\xi^{-1} (\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}) + \frac{1}{2}\sum_{j=1}^{N} \ln v_{z_j} + \frac{1}{2}\boldsymbol{z}^T \boldsymbol{V}_z^{-1} \boldsymbol{z} \\ &+ (\alpha_{z_0} + 1)\sum_{j=1}^{N} \ln v_{z_j} + \beta_{z_0}\sum_{j=1}^{N} v_{z_j}^{-1} + (\alpha_{\epsilon_0} + 1)\sum_{i=1}^{M} \ln v_{\epsilon_i} \\ &+ \beta_{\epsilon_0}\sum_{i=1}^{M} v_{\epsilon_i}^{-1} + (\alpha_{\xi_0} + 1)\sum_{j=1}^{N} \ln v_{\xi_j} + \beta_{\xi_0}\sum_{j=1}^{N} v_{\xi_j}^{-1}.\end{aligned} \tag{5.57}$$

The variables are estimated alternately and iteratively. The optimization criterion for each variable consist of the terms containing the corresponding variable :

$$\begin{aligned}J(\boldsymbol{f}) &= \frac{1}{2}(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f})^T \boldsymbol{V}_\epsilon^{-1} (\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}) + \frac{1}{2}(\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z})^T \boldsymbol{V}_\xi^{-1} (\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}), \\ J(\boldsymbol{z}) &= \frac{1}{2}(\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z})^T \boldsymbol{V}_\xi^{-1} (\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}) + \frac{1}{2}\boldsymbol{z}^T \boldsymbol{V}_z^{-1} \boldsymbol{z}, \\ J(\boldsymbol{v}_z) &= \frac{1}{2}\sum_{j=1}^{N} \ln v_{z_j} + \frac{1}{2}\boldsymbol{z}^T \boldsymbol{V}_z^{-1} \boldsymbol{z} + (\alpha_{z_0} + 1)\sum_{j=1}^{N} \ln v_{z_j} + \beta_{z_0}\sum_{j=1}^{N} v_{z_j}^{-1}, \\ J(\boldsymbol{v}_\epsilon) &= \frac{1}{2}\sum_{i=1}^{M} \ln v_{\epsilon_i} + \frac{1}{2}(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f})^T \boldsymbol{V}_\epsilon^{-1} (\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}) + (\alpha_{\epsilon_0} + 1)\sum_{i=1}^{M} \ln v_{\epsilon_i} + \beta_{\epsilon_0}\sum_{i=1}^{M} v_{\epsilon_i}^{-1}, \\ J(\boldsymbol{v}_\xi) &= \frac{1}{2}\sum_{j=1}^{N} \ln v_{\xi_j} + \frac{1}{2}(\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z})^T \boldsymbol{V}_\xi^{-1} (\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}) + (\alpha_{\xi_0} + 1)\sum_{j=1}^{N} \ln v_{\xi_j} + \beta_{\xi_0}\sum_{j=1}^{N} v_{\xi_j}^{-1}.\end{aligned} \tag{5.58}$$

And the optimization results are obtained by minimizing the criterions :

$$\begin{aligned}
\widehat{\boldsymbol{f}} &= \arg\min_{\boldsymbol{f}} \left\{ J(\boldsymbol{f}) \right\}, \\
\widehat{\boldsymbol{z}} &= \arg\min_{\boldsymbol{z}} \left\{ J(\boldsymbol{z}) \right\}, \\
\widehat{\boldsymbol{v}}_z &= \arg\min_{\boldsymbol{v}_z} \left\{ J(\boldsymbol{v}_z) \right\}, \\
\widehat{\boldsymbol{v}}_\epsilon &= \arg\min_{\boldsymbol{v}_\epsilon} \left\{ J(\boldsymbol{v}_\epsilon) \right\}, \\
\widehat{\boldsymbol{v}}_\xi &= \arg\min_{\boldsymbol{v}_\xi} \left\{ J(\boldsymbol{v}_\xi) \right\}.
\end{aligned} \tag{5.59}$$

The minimization locate at the critical points, where the gradient of the criterion equal to zero. In order to obtain these critical values, the gradient of the criterion are calculated :

$$\nabla J(\boldsymbol{f}) = -\boldsymbol{H}^T \boldsymbol{V}_\epsilon^{-1} (\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}) + \boldsymbol{V}_\xi^{-1} (\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}), \tag{5.60}$$

$$\nabla J(\boldsymbol{z}) = -\boldsymbol{D}^T \boldsymbol{V}_\xi^{-1} (\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}) + \boldsymbol{V}_z^{-1} \boldsymbol{z}. \tag{5.61}$$

And for vectors $\boldsymbol{v}_z$, $\boldsymbol{v}_\epsilon$ and $\boldsymbol{v}_\xi$, each element of them are considered separately because of the independence hypothesis :

$$\nabla J(v_{z_j}) = \frac{1}{2} v_{z_j}^{-1} - \frac{1}{2} z_j^2 v_{z_j}^{-2} + (\alpha_{z_0} + 1) v_{z_j}^{-1} - \beta_{z_0} v_{z_j}^{-2}, \tag{5.62}$$

$$\nabla J(v_{\epsilon_i}) = \frac{1}{2} v_{\epsilon_i}^{-1} - \frac{1}{2} \left[\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f} \right]_i^2 v_{\epsilon_i}^{-2} + (\alpha_{\epsilon_0} + 1) v_{\epsilon_i}^{-1} - \beta_{\epsilon_0} v_{\epsilon_i}^{-2}, \tag{5.63}$$

$$\nabla J(v_{\xi_j}) = \frac{1}{2} v_{\xi_j}^{-1} - \frac{1}{2} \left[\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z} \right]_j^2 v_{\xi_j}^{-2} + (\alpha_{\xi_0} + 1) v_{\xi_j}^{-1} - \beta_{\xi_0} v_{\xi_j}^{-2}. \tag{5.64}$$

From Eq.(5.60) and Eq.(5.61) :

$$\nabla J(\boldsymbol{f}) \Big|_{\widehat{\boldsymbol{f}}} = 0 \qquad \Longrightarrow \widehat{\boldsymbol{f}} = \left(\boldsymbol{H}^T \boldsymbol{V}_\epsilon^{-1} \boldsymbol{H} + \boldsymbol{V}_\xi^{-1} \right)^{-1} \left(\boldsymbol{H}^T \boldsymbol{V}_\epsilon^{-1} \boldsymbol{g} + \boldsymbol{V}_\xi^{-1} \boldsymbol{D}\boldsymbol{z} \right), \tag{5.65}$$

$$\nabla J(\boldsymbol{z}) \,|_{\widehat{\boldsymbol{z}}} = 0 \qquad \Longrightarrow \widehat{\boldsymbol{z}} = \left(\boldsymbol{D}^T \boldsymbol{V}_\xi^{-1} \boldsymbol{D} + \boldsymbol{V}_z^{-1} \right)^{-1} \boldsymbol{D}^T \boldsymbol{V}_\xi^{-1} \boldsymbol{f}. \tag{5.66}$$

As we have mentioned previously, the inversion of the matrix $\left(\boldsymbol{H}^T \boldsymbol{V}_\epsilon^{-1} \boldsymbol{H} + \boldsymbol{V}_\xi^{-1} \right)$ and matrix $\left(\boldsymbol{D}^T \boldsymbol{V}_\xi^{-1} \boldsymbol{D} + \boldsymbol{V}_z^{-1} \right)$ are impossible in the reconstruction of 3D objects because of the huge data size. The gradient descent algorithm is used to realize the optimization for these two variables by computing :

$$\begin{aligned}
\widehat{\boldsymbol{f}}^{(k+1)} &= \widehat{\boldsymbol{f}}^{(k)} - \gamma_f^{(k)} \nabla J(\widehat{\boldsymbol{f}}^{(k)}), \\
\widehat{\boldsymbol{z}}^{(k+1)} &= \widehat{\boldsymbol{z}}^{(k)} - \gamma_z^{(k)} \nabla J(\widehat{\boldsymbol{z}}^{(k)}),
\end{aligned} \tag{5.67}$$

where $\gamma_f^{(k)}$ and $\gamma_z^{(k)}$ are the gradient descent step length and are obtained by minimizing $J(\widehat{\boldsymbol{f}}^{(k)} - \gamma_f \nabla J(\widehat{\boldsymbol{f}}^{(k)}))$ and $J(\widehat{\boldsymbol{z}}^{(k)} - \gamma_z \nabla J(\widehat{\boldsymbol{z}}^{(k)}))$ respectively :

$$\begin{aligned}
& J(\boldsymbol{f} - \gamma_f \nabla J(\boldsymbol{f})) \\
=& \gamma_f \nabla J(\boldsymbol{f})^T \left[\boldsymbol{H}^T \boldsymbol{V}_\epsilon^{-1} (\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}) - \boldsymbol{V}_\xi^{-1} (\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}) \right] \\
& + \frac{1}{2} \gamma_f^2 \left[\| \boldsymbol{Y}_\epsilon \boldsymbol{H} \nabla J(\boldsymbol{f}) \|_2^2 + \| \boldsymbol{Y}_\xi \nabla J(\boldsymbol{f}) \|_2^2 \right] + \text{Constant, where } \boldsymbol{Y}_\epsilon = \boldsymbol{V}_\epsilon^{-\frac{1}{2}} \text{ and } \boldsymbol{Y}_\xi = \boldsymbol{V}_\xi^{-\frac{1}{2}}.
\end{aligned} \tag{5.68}$$

Substituting Eq.(5.60) we obtain :

$$\begin{aligned} & J(\boldsymbol{f} - \gamma_f \nabla J(\boldsymbol{f})) \\ &= \arg\min_{\gamma_f} \left\{ -\gamma_f \nabla J(\boldsymbol{f})^T \nabla J(\boldsymbol{f}) + \frac{1}{2}\gamma_f^2 \left[\|\boldsymbol{Y}_\epsilon \boldsymbol{H} \nabla J(\boldsymbol{f})\|_2^2 + \|\boldsymbol{Y}_\xi \nabla J(\boldsymbol{f})\|_2^2 \right] + \text{Cst} \right\}. \end{aligned} \tag{5.69}$$

The gradient of the criterion is :

$$\nabla J(\boldsymbol{f} - \gamma_f \nabla J(\boldsymbol{f})) = - \|\nabla J(\boldsymbol{f})\|_2^2 + \gamma_f \left[\|\boldsymbol{Y}_\epsilon \boldsymbol{H} \nabla J(\boldsymbol{f})\|_2^2 + \|\boldsymbol{Y}_\xi \nabla J(\boldsymbol{f})\|_2^2 \right].$$

By setting

$$\nabla J(\boldsymbol{f} - \gamma_f \nabla J(\boldsymbol{f})) \Big|_{\gamma_f^{(k)}} = 0,$$

we get :

$$\gamma_f^{(k)} = \frac{\|\nabla J(\boldsymbol{f})\|_2^2}{\|\boldsymbol{Y}_\epsilon \boldsymbol{H} \nabla J(\boldsymbol{f})\|_2^2 + \|\boldsymbol{Y}_\xi \nabla J(\boldsymbol{f})\|_2^2}, \text{ where } \boldsymbol{Y}_\epsilon = \boldsymbol{V}_\epsilon^{-\frac{1}{2}} \text{ and } \boldsymbol{Y}_\xi = \boldsymbol{V}_\xi^{-\frac{1}{2}}. \tag{5.70}$$

For $\gamma_z^{(k)}$ the same algorithm is used :

$$\begin{aligned} J(\boldsymbol{z} - \gamma_z \nabla J(\boldsymbol{z})) =& \gamma_z \nabla J(\boldsymbol{z})^T \left[\boldsymbol{D}^T \boldsymbol{V}_\xi^{-1} (\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}) - \boldsymbol{V}_z^{-1} \boldsymbol{z} \right] \\ &+ \frac{1}{2}\gamma_z^2 \left[\|\boldsymbol{Y}_\xi \boldsymbol{D} \nabla J(\boldsymbol{z})\|_2^2 + \|\boldsymbol{Y}_z \nabla J(\boldsymbol{z})\|_2^2 \right] + \text{Cst}, \text{ where } \boldsymbol{Y}_z = \boldsymbol{V}_z^{-\frac{1}{2}}. \end{aligned} \tag{5.71}$$

Substituting Eq.(5.61) we have :

$$\begin{aligned} & J(\boldsymbol{z} - \gamma_z \nabla J(\boldsymbol{z})) \\ &= \arg\min_{\gamma_z} \left\{ -\gamma_z \|\nabla J(\boldsymbol{z})\|_2^2 + \frac{1}{2}\gamma_z^2 \left[\|\boldsymbol{Y}_\xi \boldsymbol{D} \nabla J(\boldsymbol{z})\|_2^2 + \|\boldsymbol{Y}_z \nabla J(\boldsymbol{z})\|_2^2 \right] + \text{Cst} \right\}. \end{aligned} \tag{5.72}$$

The gradient of the criterion is :

$$\nabla J(\boldsymbol{z} - \gamma_z \nabla J(\boldsymbol{z})) = - \|\nabla J(\boldsymbol{z})\|_2^2 + \gamma_z \left[\|\boldsymbol{Y}_\xi \boldsymbol{D} \nabla J(\boldsymbol{z})\|_2^2 + \|\boldsymbol{Y}_z \nabla J(\boldsymbol{z})\|_2^2 \right]. \tag{5.73}$$

By setting

$$\nabla J(\boldsymbol{z} - \gamma_z \nabla J(\boldsymbol{z})) \Big|_{\gamma_z^{(k)}} = 0,$$

we get :

$$\gamma_z^{(k)} = \frac{\|\nabla J(\boldsymbol{z})\|_2^2}{\|\boldsymbol{Y}_\xi \boldsymbol{D} \nabla J(\boldsymbol{z})\|_2^2 + \|\boldsymbol{Y}_z \nabla J(\boldsymbol{z})\|_2^2}, \text{ where } \boldsymbol{Y}_z = \boldsymbol{V}_z^{-\frac{1}{2}}. \tag{5.74}$$

The optimizations for $\boldsymbol{f}$ and $\boldsymbol{z}$ by using gradient descent algorithm are :

$$\text{for } k = 1 \to I_2 : \ \widehat{\boldsymbol{f}}^{(k+1)} = \widehat{\boldsymbol{f}}^{(k)} - \gamma_f^{(k)} \nabla J(\widehat{\boldsymbol{f}}^{(k)}), \tag{5.75}$$

$$\text{for } k = 1 \to I_2 : \ \widehat{\boldsymbol{z}}^{(k+1)} = \widehat{\boldsymbol{z}}^{(k)} - \gamma_z^{(k)} \nabla J(\widehat{\boldsymbol{z}}^{(k)}), \tag{5.76}$$

where I_2 is the number of iterations for the gradient descent algorithm.

The optimization for the other variables : v_{z_j}, v_{ϵ_i} and v_{ξ_j} are obtained by minimizing the corresponding criterion. From Eq.(5.62)-Eq.(5.64), and considering the discussion of the global

minimum point for the criterion of form $a \ln x + bx^{-1}$ in Section 5.3.3, the critical point of each criterion is the minimum :

$$\nabla J(v_{z_j}) \big|_{\widehat{v}_{zj}} = 0 \qquad \Longrightarrow \widehat{v}_{zj} = \frac{\beta_{z_0} + \frac{1}{2} v_{z_j}^{-1}}{\alpha_{z_0} + \frac{3}{2}}, \tag{5.77}$$

$$\nabla J(v_{\epsilon_i}) \big|_{\widehat{v}_{\epsilon_i}} = 0 \qquad \Longrightarrow \widehat{v}_{\epsilon_i} = \frac{\beta_{\epsilon_0} + \frac{1}{2} [\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}]_i^2}{\alpha_{\epsilon_0} + \frac{3}{2}}, \tag{5.78}$$

$$\nabla J(v_{\xi_j}) \Big|_{\widehat{v}_{\xi_j}} = 0 \qquad \Longrightarrow \widehat{v}_{\xi_j} = \frac{\beta_{\xi_0} + \frac{1}{2} [\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}]_j^2}{\alpha_{\xi_0} + \frac{3}{2}}, \tag{5.79}$$

$\forall i \in [1, M]$ and $\forall j \in [1, N]$.

The algorithms of the implementation of estimation for $\widehat{\boldsymbol{f}}$ and $\widehat{\boldsymbol{z}}$ are shown in Algorithm 3 and Algorithm 4.

Algorithm 3 Gradient descent algorithm for $\widehat{\boldsymbol{f}}$

Input : $\boldsymbol{H},\ \boldsymbol{D},\ \boldsymbol{g},\ \widehat{\boldsymbol{z}},\ \boldsymbol{f}_0$
Output : $\widehat{\boldsymbol{f}}$
Initialization : $\widehat{\boldsymbol{f}}_{ini} = \boldsymbol{f}_0$
$k = 0$
repeat
    $k = k + 1$
    Calculate $\nabla J(\widehat{\boldsymbol{f}}^{(k-1)})$ according to Eq.(5.65)
    Update $\widehat{\gamma}_f^{(k)}$ according to Eq.(5.70)
    Update $\widehat{\boldsymbol{f}}^{(k)} = \widehat{\boldsymbol{f}}^{(k-1)} - \widehat{\gamma}_f^{(k)} \nabla J(\widehat{\boldsymbol{f}}^{(k-1)})$
until convergence or maximum iteration

Algorithm 4 Gradient descent algorithm for $\widehat{\boldsymbol{z}}$

Input : $\boldsymbol{H},\ \boldsymbol{D},\ \boldsymbol{g},\ \widehat{\boldsymbol{f}},\ \boldsymbol{z}_0$
Output : $\widehat{\boldsymbol{z}}$
Initialization : $\widehat{\boldsymbol{z}}_{ini} = \boldsymbol{z}_0$
$k = 0$
repeat
    $k = k + 1$
    Calculate $\nabla J(\widehat{\boldsymbol{z}}^{(k-1)})$ according to Eq.(5.66)
    Update $\widehat{\gamma}_z^{(k)}$ according to Eq.(5.74)
    Update $\widehat{\boldsymbol{z}}^{(k)} = \widehat{\boldsymbol{z}}^{(k-1)} - \widehat{\gamma}_z^{(k)} \nabla J(\widehat{\boldsymbol{z}}^{(k-1)})$
until convergence or maximum iteration

An integrated algorithm concerning the estimation of all the unknown variables is given in Algorithm 5.

5.4.2 Analysis of hyper-parameters and initialization

In the regularization methods, there are always one or several regularization parameters. One of the shortcoming of the regularization method is that the choice of these parameters is

Algorithm 5 Summary of the JMAP algorithm for HHBM method

Fix parameters α_{z_0}, β_{z_0}, α_{ϵ_0}, β_{ϵ_0}, α_{ξ_0}, β_{ξ_0}, l
Input : $\boldsymbol{H}$, $\boldsymbol{D}$, $\boldsymbol{g}$
Output : $\widehat{\boldsymbol{f}}$, $\widehat{\boldsymbol{z}}$, $\widehat{\boldsymbol{v}}_z$, $\widehat{\boldsymbol{v}}_\epsilon$, $\widehat{\boldsymbol{v}}_\xi$
Initialization :
$\widehat{\boldsymbol{f}}^{(0)} \leftarrow$ normalized FDK
$\widehat{\boldsymbol{z}}^{(0)} \leftarrow l-$level Haar transformation of $\widehat{\boldsymbol{f}}^{(0)}$
$\boldsymbol{V}_z$, $\boldsymbol{V}_\epsilon$ and $\boldsymbol{V}_\xi$ according to Eq.(5.77)-Eq.(5.79)
repeat
    Update $\widehat{\boldsymbol{f}}$ according to Algorithm 3
    Update $\widehat{\boldsymbol{z}}$ according to Algorithm 4
    Compute $v_{z_j}, \forall j \in [1:N]$ according to Eq.(5.77)
    Compute $v_{\epsilon_i}, \forall i \in [1:M]$ according to Eq.(5.78)
    Compute $v_{\xi_j}, \forall j \in [1:N]$ according to Eq.(5.79)
until convergence or maximum iteration

crucial in the reconstruction, meanwhile, for different size and different quality of the measured dataset, the optimal value of this parameter normally changes. Thus, for each different case, one should always consider to initialize with suitable values for the parameters.

In the proposed Bayesian method, however, the case is different. Even though we should always initialize the hyperparameters, we get a theoretical basis to fix the initialization values of hyperparameters. The method is therefore semi-supervised, which is a priority property in the ill-posed reconstruction problems.

The initialization for the hyperparameters is considered based on the hierarchical prior model chosen for the variables. For $\boldsymbol{\epsilon}$ and $\boldsymbol{\xi}$, both of them are modeled by a Gaussian distribution with an unknown variance belonging to an Inverse Gamma distribution. $\boldsymbol{\epsilon}$ is supposed to be a Gaussian noise and $\boldsymbol{\xi}$ is supposed to be sparse such that $\boldsymbol{f}$ and $\boldsymbol{z}$ are satisfying the transformation relationship for most of elements. Hyperparameters α_{ϵ_0}, β_{ϵ_0}, α_{ξ_0}, β_{ξ_0} are initialized according to these properties. The hidden variable $\boldsymbol{z}$ is sparse, and is defined by the $\mathcal{S}t_g$ distribution. The hyperparameters α_{z_0} and β_{z_0} are initialized based on this sparse enforcing prior model.

Fixing α_{z_0} and β_{z_0}

As we can see in Figure 5.8, there are several different ranks of transformation coefficient in the multilevel discrete Haar transformation coefficient. We use $r \in [1, l+1]$ to represent the rank of the coefficient of the l-level Haar transformation (shown in Figure 5.8). The low frequency part of the coefficient $\boldsymbol{z}$, in the top-left of the coefficient Figure 5.8 with $r = 1$, is less sparse than the other coefficients. The sparsity rate augments when when r is bigger. Consequently, we set a different $(\alpha_{z_0}, \beta_{z_0})$ pairs for coefficients in different coefficient ranks. For $(\alpha_{z_0}(j), \beta_{z_0}(j))$ in rank r, we set (α_r, β_r). The strategy of the initialization is to fix one hyperparameter, α_{z_0}, and use different values for hyperparameter β_{z_0}, to realize the definition of different sparsity rate for different ranks. We set $\alpha_{z_0} = 2.01$, so that the variance of the $\mathcal{S}t_g$ distribution will be approximate to β_{z_0} according to the definition $\mathrm{Var}[z|\alpha_{z_0}, \beta_{z_0}] = \beta_{z_0}/(\alpha_{z_0} - 1)$. When α_{z_0} is fixed, we get a higher sparsity rate with a smaller value for $\beta_{z_0} = \beta_r$. We set

$\boldsymbol{\beta}_r = [\beta_1, \beta_2, \cdots, \beta_{l+1}]$ where $\beta_r = 10^{-r+1}$. When $r = 1$ we have $\beta = 1$. In the simulations, we set $l = 5$, hence we have $r = [1, 2, 3, 4, 5, 6]$, and the hyperparameters $\beta_{z_0}(j)$ are initialized with 6 ranks : $\beta_r = [10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}]$.

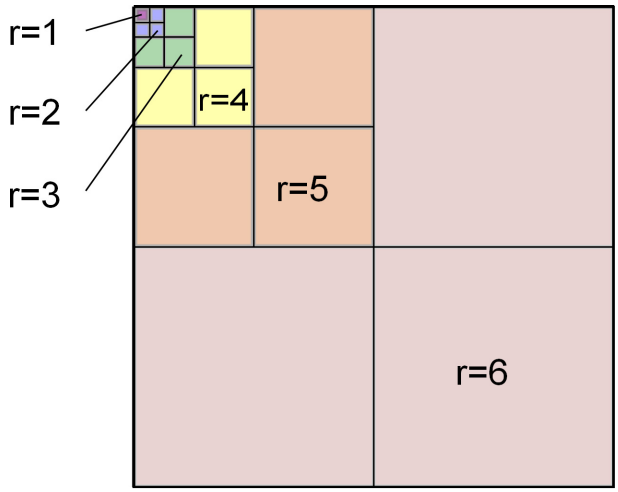


FIGURE 5.8 – The illustration of rank r of a 5-level discrete Haar transformation coefficient.

Fixing α_{ϵ_0} and β_{ϵ_0}

The noise $\boldsymbol{\epsilon}$ depends on the SNR of the dataset. The biased dataset is expressed as the sum of uncontaminated dataset and the additive noise :

$$\boldsymbol{g} = \boldsymbol{g}_0 + \boldsymbol{\epsilon}. \tag{5.80}$$

As the noise $\boldsymbol{\epsilon}$ and the uncontaminated data $\boldsymbol{g}_0$ are supposed to be independent, we have :

$$\|\boldsymbol{g}\|_2^2 = \|\boldsymbol{g}_0\|_2^2 + \|\boldsymbol{\epsilon}\|_2^2 . \tag{5.81}$$

The SNR of the dataset is :

$$\mathrm{SNR} = 10 \log \frac{\|\boldsymbol{g}_0\|_2^2}{\|\boldsymbol{\epsilon}\|_2^2} = 10 \log \frac{\|\boldsymbol{g}\|_2^2 - \|\boldsymbol{\epsilon}\|_2^2}{\|\boldsymbol{\epsilon}\|_2^2}. \tag{5.82}$$

As the expectation of $\boldsymbol{\epsilon}$ is $\mathbb{E}[\boldsymbol{\epsilon}] = 0$,

$$v_\epsilon = \mathbb{E}\left[\boldsymbol{\epsilon}^2\right] \approx \frac{\|\boldsymbol{\epsilon}\|_2^2}{M} = \frac{\|\boldsymbol{g}\|_2^2}{M} \times \frac{1}{1 + 10^{SNR/10}}. \tag{5.83}$$

Belonging to an Inverse Gamma distribution, the expectation of the variance $\mathbb{E}[v_{\epsilon_i} | \alpha_{\epsilon_0}, \beta_{\epsilon_0}] = \frac{\beta_{\epsilon_0}}{\alpha_{\epsilon_0} - 1}$, therefore :

$$\beta_{\epsilon_0} = \frac{\|\boldsymbol{g}\|_2^2}{M} \times \frac{1}{1 + 10^{SNR/10}} \times (\alpha_{\epsilon_0} - 1) . \tag{5.84}$$

The two hyperparameters α_{ϵ_0} and β_{ϵ_0} are combined according to Eq.(5.84) and hence initialization for one of them is sufficient. Figure 5.9 shows the influence of the value of α_{ϵ_0} on the reconstruction. According to the results, a bigger value for α_{ϵ_0} results to a smaller value on RMSE, for different number of projections and SNR of dataset. For the case with a low SNR

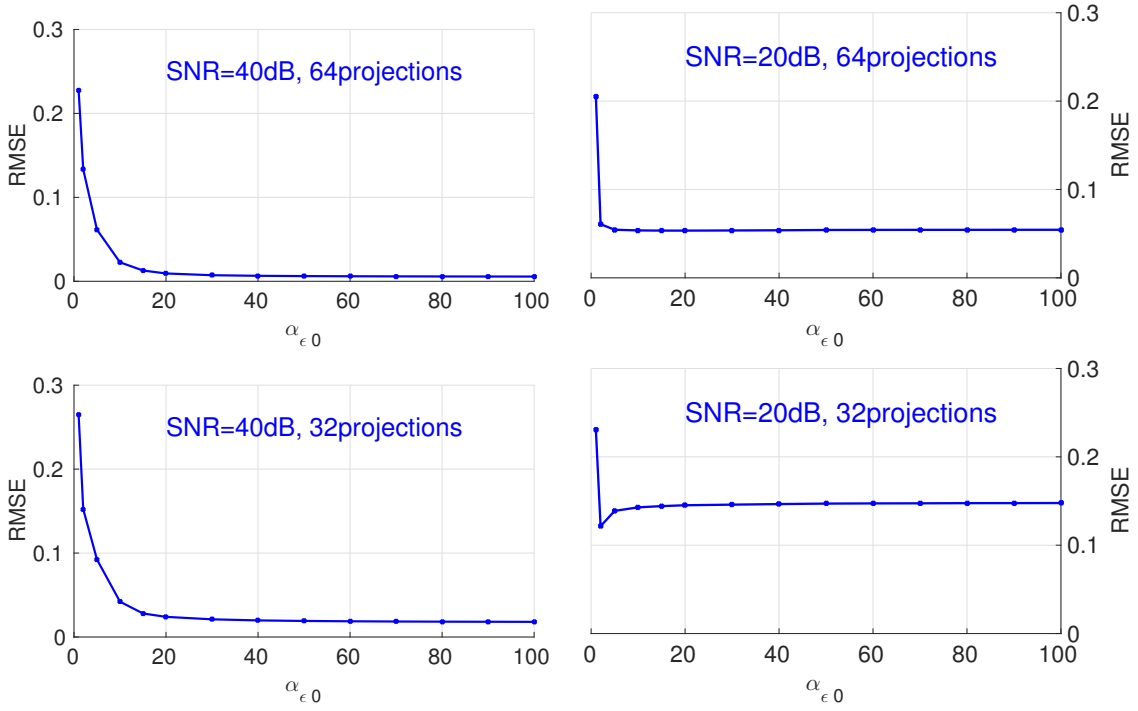


FIGURE 5.9 – Influence of hyperparameter α_{ϵ_0} on RMSE of final reconstruction results for different number of projections and noise.

and less projections numbers, there is a slight augment of RMSE when the hyperparameter increase, however, it soon arrives at a platform and stay robust. This monotonous property solves the problem of initialization of this hyperparameter, as a big value for α_{ϵ_0} satisfies all cases, since when α_{ϵ_0} is greater than a certain value, the RMSE does not change by using different initialization value α_{ϵ_0}.

In the reconstruction of a real data, we do not know the SNR of the data set. But we can estimate the noise $\boldsymbol{\epsilon}$ by some experiments. One could estimate the detected data of in two cases : the first is when there is no source of X-ray, and the second is to emit X-rays without any object. By using these two data sets, we could obtain a rough additive noise in order to get the information of $\boldsymbol{v}_\epsilon$.

Fixing α_{ξ_0} and β_{ξ_0}

For $\boldsymbol{\xi}$, we do not have evident information on it, therefore the initialization for both α_{ξ_0} and β_{ξ_0} is necessary. Figure 5.10 and Figure 5.11 show respectively the influence of the initialization of α_{ξ_0} and β_{ξ_0} on the reconstruction. For α_{ξ_0}, Figure 5.10, the influence is monotone and the reconstructed phantom has a lower RMSE when α_{ξ_0} is smaller.

Figure 5.11 shows the influence of the hyperparameter β_{ξ_0} on RMSE of reconstructed phantom. From the curves we find that the minimum RMSE value appears when β_{ξ_0} is approaching 0.01. Consequently we use $\beta_{\xi_0} = 0.01$. There is an augment of the RMSE value when β_{ξ_0} increases, but it soon arrives at a platform which is slightly bigger than the minimum RMSE value. So we conclude that the value of β_{ξ_0} is relatively insensitive to the reconstruction quality.

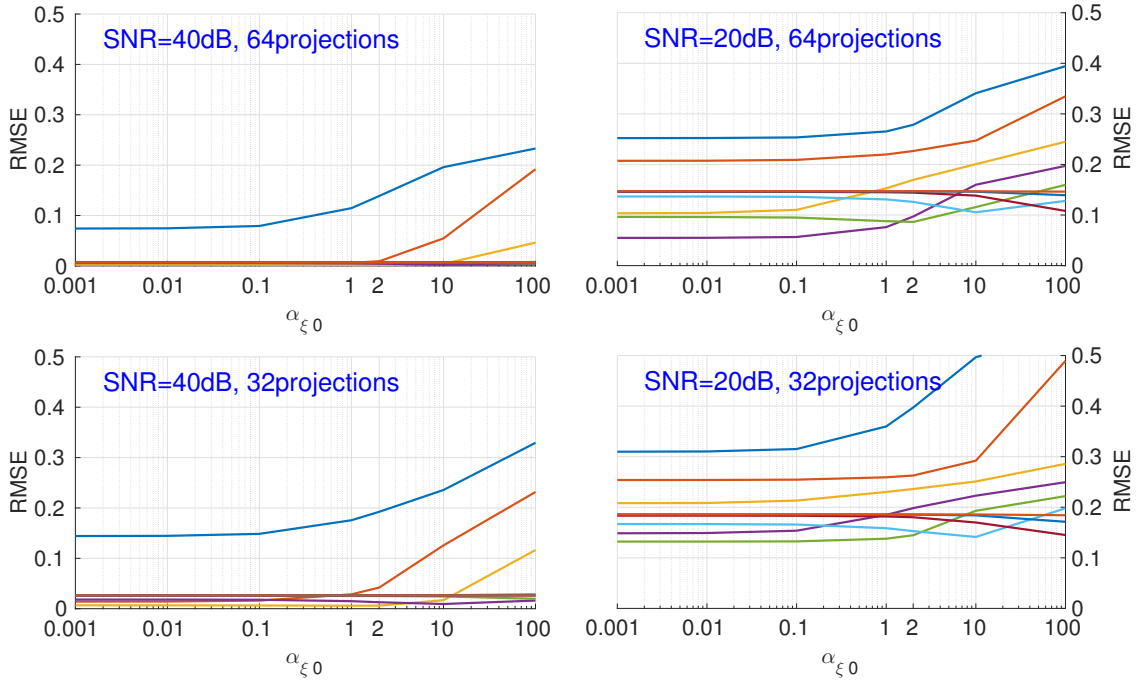


FIGURE 5.10 – Influence of hyperparameter α_{ξ_0}, for different values of β_{ξ_0}, on RMSE of reconstruction results for different number of projections and noise. Each different color corresponds to a different value of β_{ξ_0}.

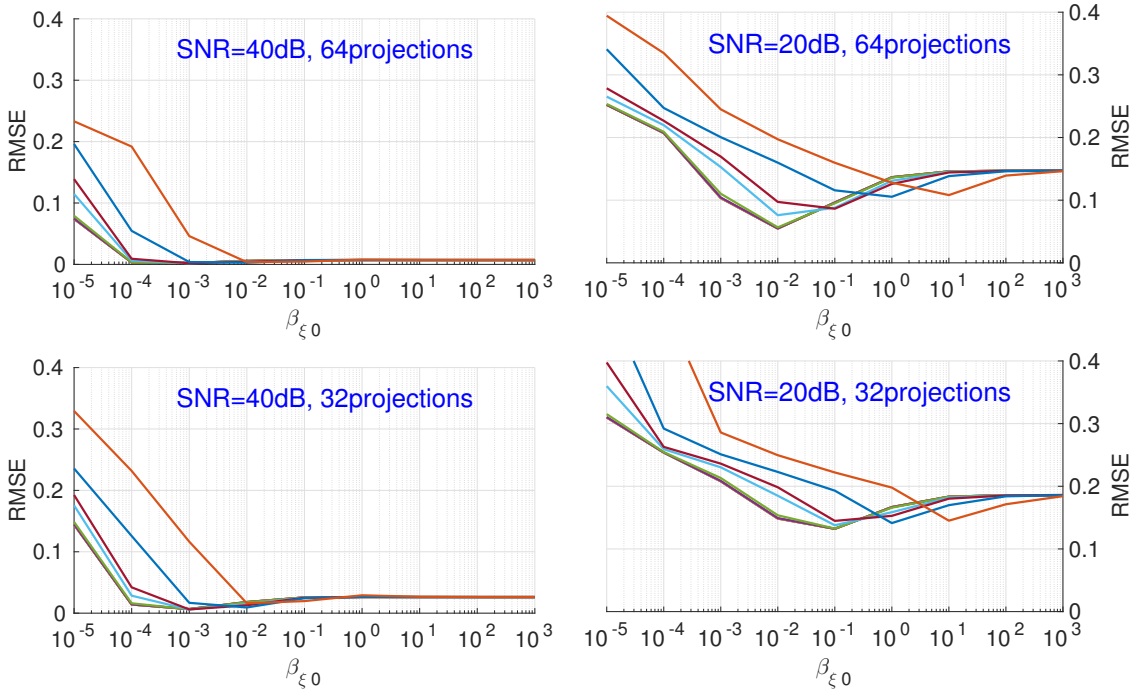


FIGURE 5.11 – Influence of hyperparameter β_{ξ_0}, for different values of α_{ξ_0}, on RMSE of reconstruction results for different number of projections and noise. Each different color corresponds to a different value of α_{ξ_0}.

In [Dum16], it is pointed out that when α and β of the $\mathcal{S}t_g$ distribution are both big, it approaches a Gaussian distribution, which is the case for the additive noise $\boldsymbol{\epsilon}$; if α and β are small (approaching to 0), it is an non-informative distribution (Jeffreys distribution); while α and β are both small but not zero, then the $\mathcal{S}t_g$ has the sparsity enforcing property, which is the case for the additive noise $\boldsymbol{\xi}$.

5.4.3 Simulation results with limited number of projections

We apply respectively $180,\ 90,\ 60,\ 45,\ 36$ and 18 projections evenly distributed in $[0°, 180°]$ for reconstruction of the 3D Shepp Logan phantom of size 256^3, each projection contains 256×256 detectors. The number of projections are chosen such that there are respectively one projection every $1°,\ 2°,\ 3°,\ 4°,\ 5°$ and $10°$. The iterative methods have all been initialized with the FBP reconstruction which has a high frequency artifacts when limited number of projections are used.

TABLE 5.1 – Comparison of relative error of reconstructed phantom with 50 global iterations (10 iterations of gradient descent in each global iteration) and the computation time of each iteration of different methods. The value of regularization parameter are respectively $\lambda_{QR} = 10$ for SNR=40dB, $\lambda_{QR} = 600$ for SNR=20dB, $\lambda_{TV} = 50$ for SNR=40dB, $\lambda_{TV} = 100$ for SNR=20dB.

	256*256*256											
	180 projections						90 projections					
	40dB			20dB			40dB			20dB		
	QR	TV	HHBM	QR	TV	HHBM	QR	TV	HHBM	QR	TV	HHBM
RMSE	0.0236	0.0114	**0.0069**	0.1309	**0.0209**	0.0755	0.0401	0.0212	**0.0092**	0.1558	**0.0491**	0.1117
ISNR	5.5584	8.7217	**10.9346**	7.2024	**15.1775**	10.2162	6.6136	9.3832	**12.9973**	8.4583	**13.4765**	9.9056
PSNR	30.0675	33.2308	**35.4437**	22.6318	**30.6069**	25.0209	27.7743	30.5439	**34.1579**	21.8754	**26.8937**	23.3227
SSIM	0.9999	0.9999	**1.0000**	0.9992	**0.9999**	0.9995	0.9997	0.9999	**0.9999**	0.9990	**0.9997**	0.9993
	60 projections						45 projections					
	40dB			20dB			40dB			20dB		
	QR	TV	HHBM	QR	TV	HHBM	QR	TV	HHBM	QR	TV	HHBM
RMSE	0.0636	0.0321	**0.0107**	0.1656	**0.0753**	0.1293	0.0904	0.0474	**0.0132**	0.1854	0.0901	**0.1414**
ISNR	9.3826	12.3480	**17.1346**	9.1492	**12.5701**	10.2226	10.3301	13.1308	**18.6839**	10.0137	13.1476	**11.1916**
PSNR	25.7693	28.7347	**33.5214**	21.6116	**25.0325**	22.6849	24.2404	27.0412	**32.5942**	21.1195	24.2535	**22.2974**
SSIM	0.9996	0.9995	**0.9999**	0.9990	**0.9995**	0.9992	0.9994	0.9997	**0.9999**	0.9988	0.9994	**0.9991**
	36 projections						18 projections					
	40dB			20dB			40dB			20dB		
	QR	TV	HHBM	QR	TV	HHBM	QR	TV	HHBM	QR	TV	HHBM
RMSE	0.1177	0.0680	**0.0169**	0.1957	**0.1116**	0.1500	0.2581	0.2104	**0.0574**	0.2907	0.2313	**0.2014**
ISNR	10.6591	13.0424	**19.0933**	10.8633	**13.3032**	12.0187	10.7122	11.5992	**17.2373**	10.8088	11.8022	**12.4036**
PSNR	23.0949	25.4783	**31.5292**	20.8865	**23.3264**	22.0420	19.6263	20.5133	**26.1514**	19.1085	20.1020	**20.7033**
SSIM	0.9993	0.9996	**0.9999**	0.9988	**0.9993**	0.9990	0.9983	0.9987	**0.9996**	0.9981	0.9985	**0.9987**

In Table 5.1, different metrics of evaluation of the reconstructed 256^3 Shepp Logan phantom are compared. It is shown that the HHBM method performs not always better than the TV method, especially when there are sufficient number of projections. But when there are insufficient detected data, for example in the case of 18 projections, the HHBM method stay more robust than TV. On the other hand, as we all know that the choice of regularization parameter plays an important role in the regularization methods like QR or TV, and the value for the regularization parameter should be selected for each different case, the HHBM method solves this problem. As we can see from Figure 5.9, Figure 5.10 and Figure 5.11, once we have chosen the hyperparameters in a certain interval which is not difficult to fix according to the theoretical base, we can get the appropriate reconstruction results. What is more important, in the Bayesian approach, the prior model can be chosen by many other suitable distributions, which gives more

possibilities for the models than the conventional regularization methods. We may also choose different point estimators from the posterior distribution, for example the Posterior Mean, etc.

Figure 5.12 shows the reconstructed middle slice of "Shepp Logan" phantom and "Head" object by using TV and HHBM methods with 36 projections and SNR=40dB. The red curve illustrate the profile of the blue line position. In the reconstructed Shepp Logan phantom by using TV method, the three small circles on the top of the slice are not evident. By using the HHBM method, we can distinguish these three small circles. By comparing the profiles of the slice of reconstructed Shepp Logan phantom, we can see that by using HHBM method, the contour positions on the profile is closer to the original profile than the TV method. In the reconstructed Head object, there are more details than the simulated Shepp Logan phantom, especially in the zoom area in the second line in Figure 5.12. By comparing the results, we can see that for the type of object which contains some small details, the TV method derives a results with smoother homogeneous areas, but less details at the contour areas than the HHBM method. For the Head object, we can also see that HHBM gives more details than TV, and is closer to the original one. Some of the white material which is dispersed into discontinuous small blocks in Head object are connected in result of TV method. From these images we conclude that in the case of insufficient number of projections the proposed method gives results with clearer contours and details.

Figure 5.13 shows the reconstructed Shepp Logan phantom from 18 projections and SNR=40dB. In this very underdetermined case, the HHBM method can still get a result which is clear enough to distinguish the primary zones and contours of the object.

Figure 5.14 and Figure 5.15 show the comparison between the QR, the TV and the HHBM method with a high SNR=40dB and a low SNR=20dB dataset respectively. The abscissa corresponds to the number of projections evenly distributed from $0°$ to $180°$, and the ordinate is the RMSE after 50 iterations. When the number of projections becomes less and less sufficient, the HHBM method outperforms the other two conventional methods.

5.4.4 Simulation results with limited angle of projections

In medical X-ray CT, a commonly faced problem is the limit of projection angles. In this part of the simulation, we treat the dataset of projection evenly distributed in a limited range of angles for the simulated 3D "Shepp Logan" phantom and the 3D "Head" object, both of which have a size of 256^3.

Figure 5.16 shows the middle slice of the reconstructed Shepp Logan phantom and the Head object, from 90 projections distributed between $0°$ and $90°$. By using TV method, the reconstructed object is blurry along the the diagonal direction where there is no projection data estimated, and there is an obvious square corner while the object should have a round shape. By using the HHBM method, we get results with closer shape and clearer contours.

Figure 5.17 and Figure 5.18 show the comparison of the performance in terms of RMSE of different method with a high SNR=40dB and a low SNR=20dB. In this comparison, four cases of limited angle of projections are considered. They are respectively : 45, 90, 135 and 180 projections evenly distributed in respectively $[0°, 45°]$, $[0°, 90°]$, $[0°, 135°]$ and $[0°, 180°]$. Thus, there is one projection every $1°$. From these two figures, we conclude that the proposed HHBM method stay more robust than the other two conventional methods when there is limited angle of projections.

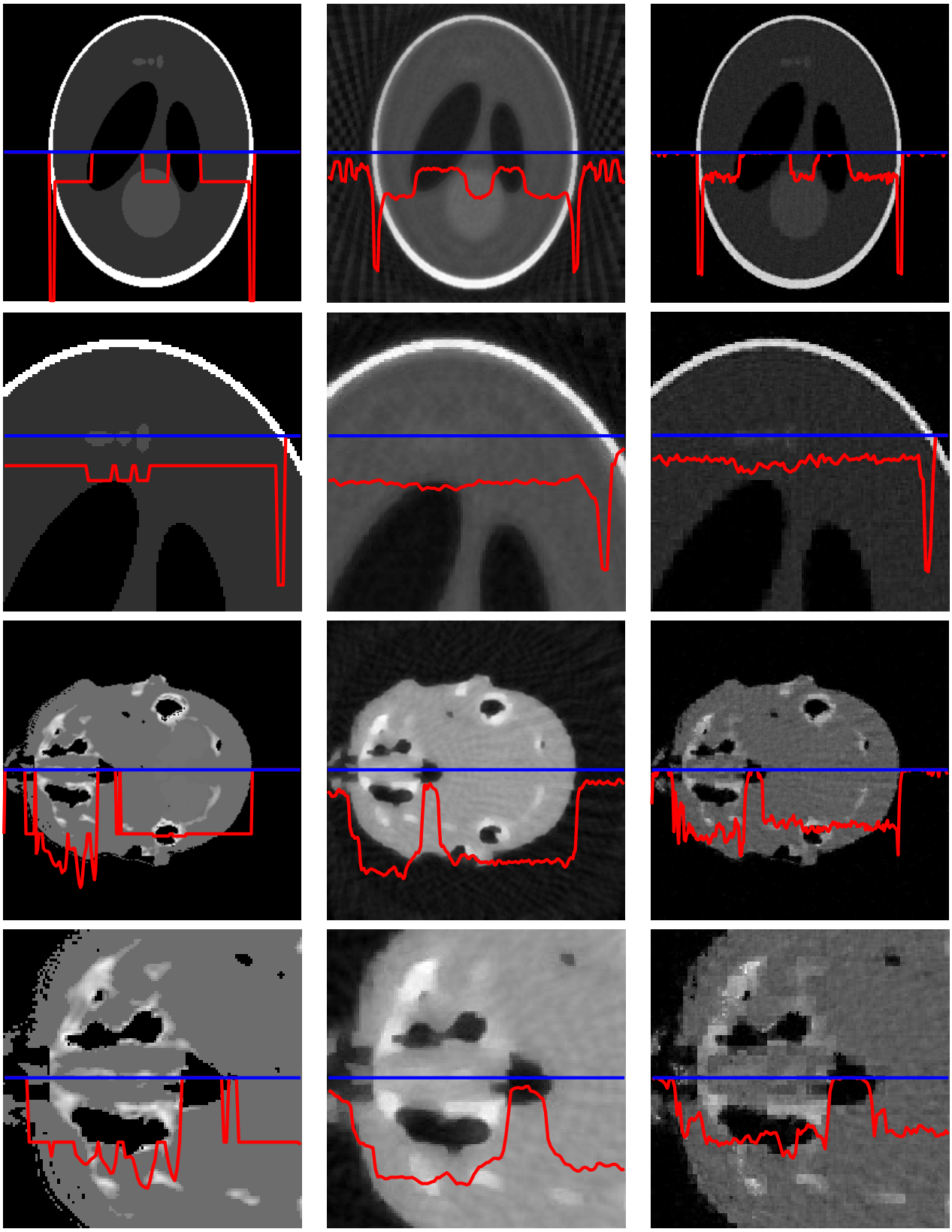


FIGURE 5.12 – Slice of reconstructed ”Shepp Logan” and ”Head” phantom of size 256^3, with dataset of 36 projections and SNR=40 dB, by using TV and HHBM methods respectively. Bottom figures are part of the corresponding top figures.

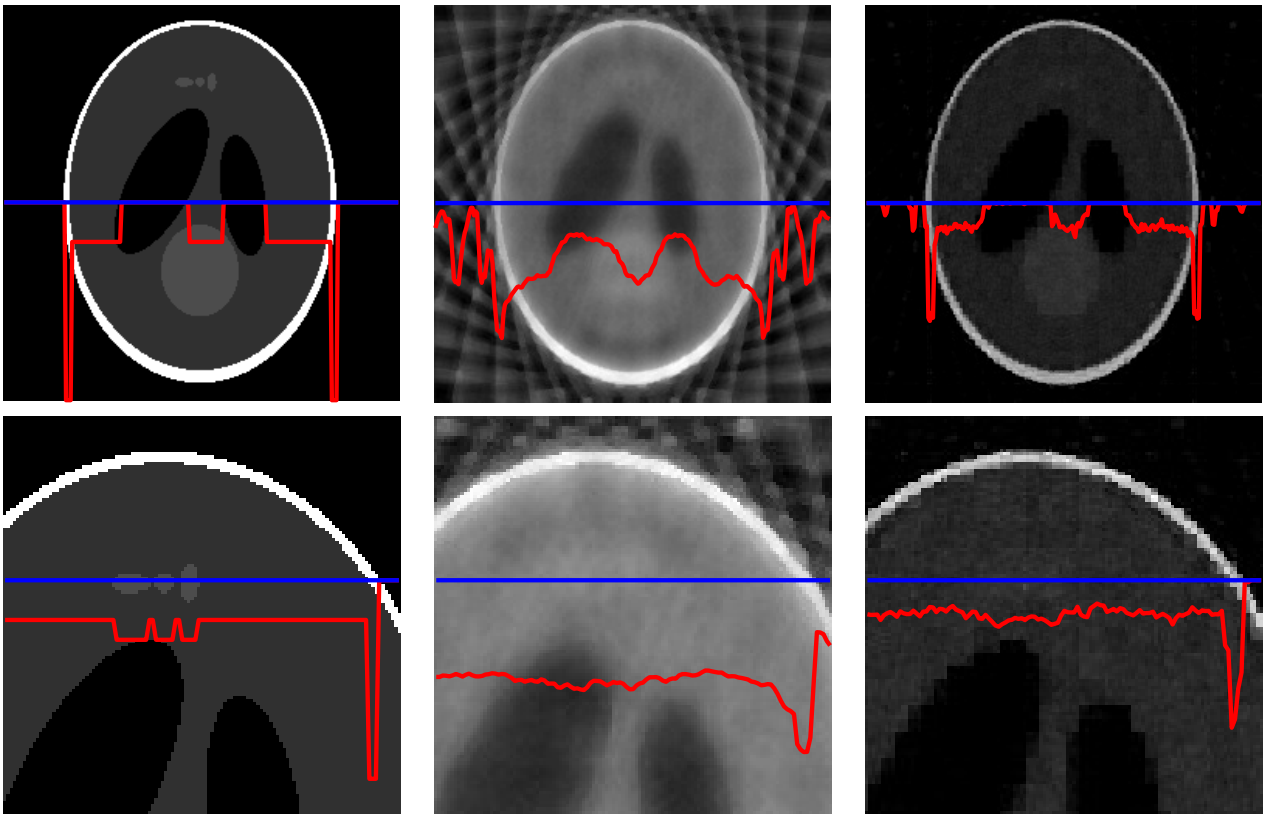


FIGURE 5.13 – Slice of reconstructed Shepp Logan phantom of size 256^3, with dataset of 18 projections and SNR=40 dB, by using TV (left) and HHBM (right) methods respectively. The red curves are the profile at the position of the corresponding blue lines.

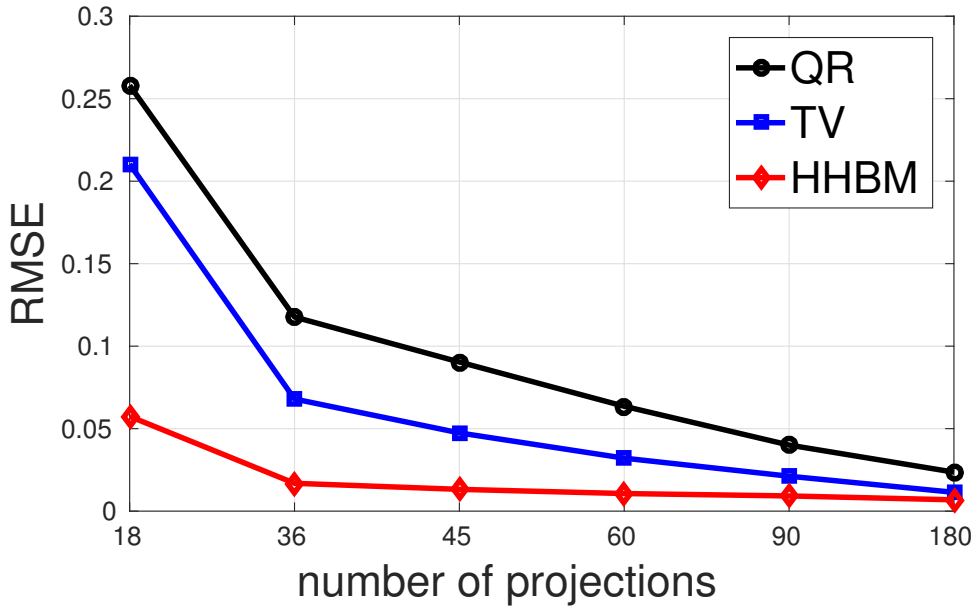


FIGURE 5.14 – The performance of different methods for reconstructing Shepp Logan phantom in terms of RMSE with different number of projections evenly distributed in $[0°, 180°]$ and a high SNR=40dB.

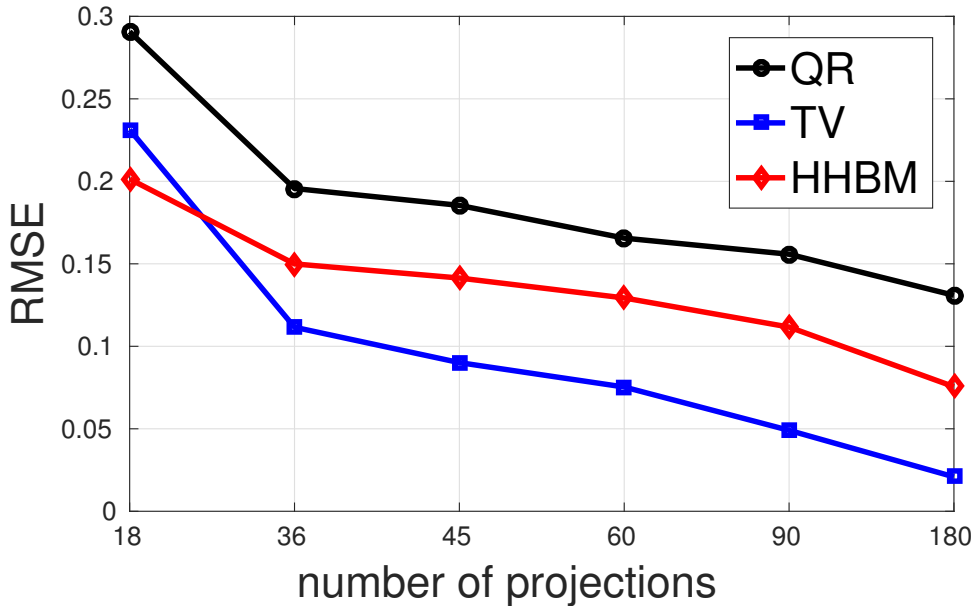


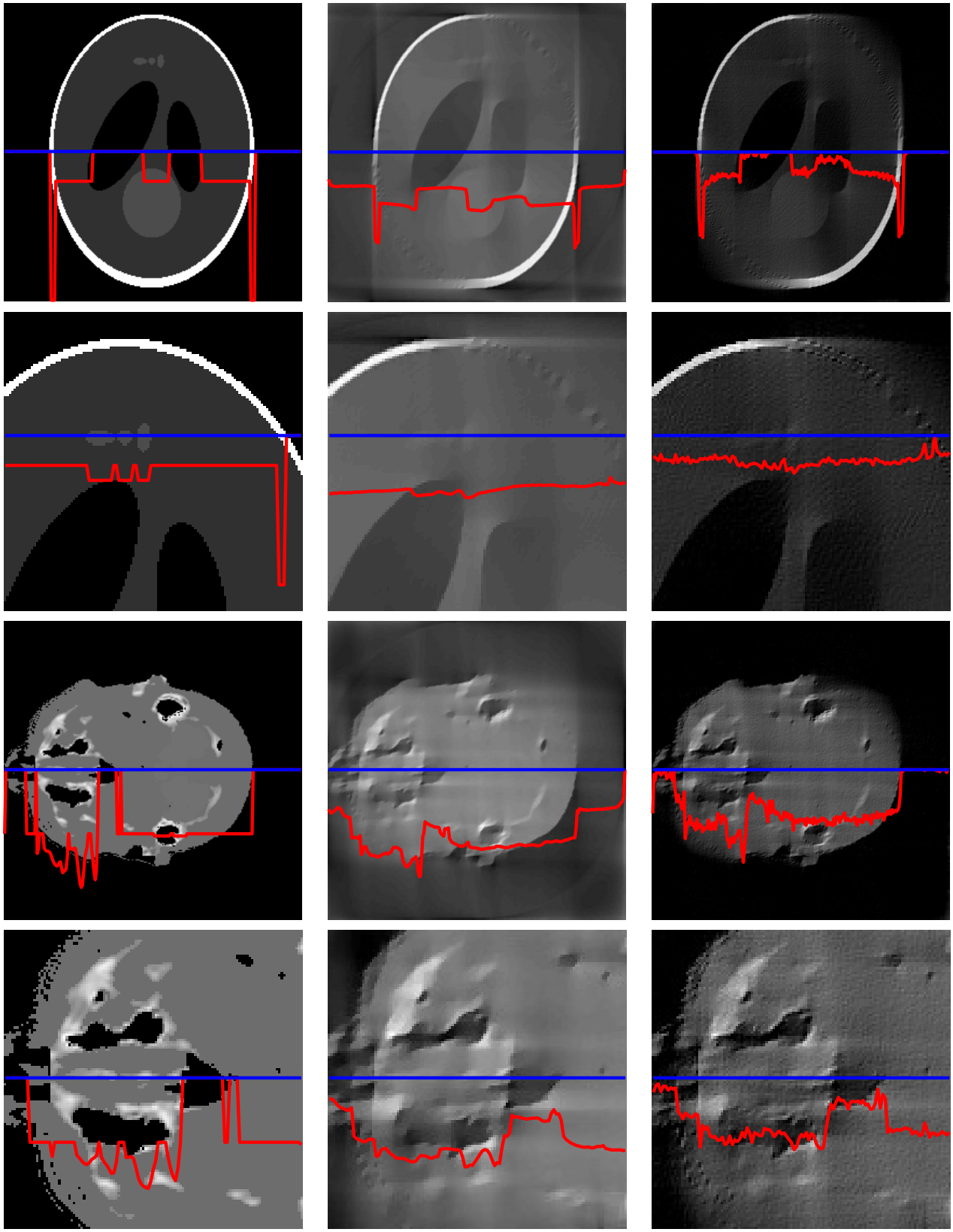
FIGURE 5.15 – The performance of different methods for reconstructing Shepp Logan phantom in terms of RMSE with different number of projections evenly distributed in $[0°, 180°]$ and a low SNR=20dB.

5.5 Dual-tree complex wavelet transform (DT-CWT) Based Hierarchical Bayesian Method

In the proposed method, the multilevel Haar transformation coefficient of the phantom is considered as a hidden variable and is estimated during iterations. The reason of choosing Haar transformation is that the phantom under consideration is normally discrete and strictly piecewise continuous. Haar transformation is one of the simpliest and basic transformation among all the orthonormal transformations. Its computation can be efficiently accelerated thanks to a parallelization on GPUs. What's more, it is orthonormal and therefore it simplifies the computation in the Bayesian methods. Recently, there are more and more new transforms proposed with very good properties, for example the curvelet [CDDY06, SCD02], the contourlet [DV05], etc.

When the wavelet transformations are used, some disadvantages should be considered [SBK05] :

- Oscillations. In the wavelet transformation, the coefficients tend to oscillate positive and negative around the singularities. These oscillations around the singularities complicate the wavelet-based processing, and makes the singularity extraction and signal modeling more complicated.
- Shift variance. In the wavelet transformation, a small shift of the signal will perturb the wavelet coefficient oscillation pattern around the singularities because of the positive-negative oscillation property of the coefficients. This property also complicates the wavelet domain processing.
- Lack of directionality. While Fourier sinusoids in higher dimensions correspond to highly directional plane waves, the standard tensor product construction of M-D wavelets produces a checkerboard pattern that is simultaneously oriented along only several directions. This lack of directional selectivity complicates the modeling and processing



Original TV HHBM

FIGURE 5.16 – Slice of reconstructed 3D Shepp Logan phantom and 3D Head object, with 90 projections evenly distributed in $[0°, 90°]$.

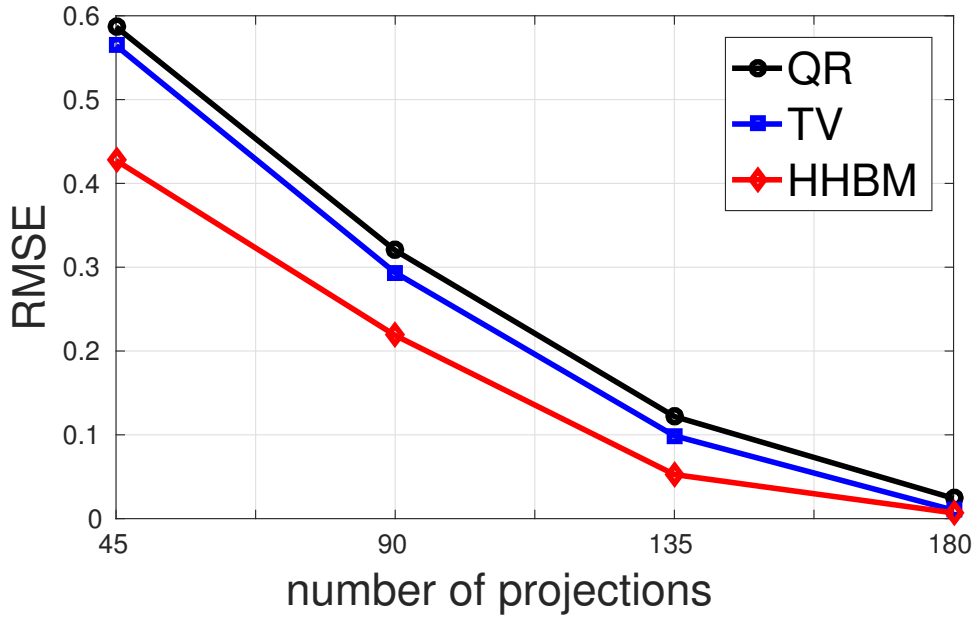


FIGURE 5.17 – The performance of different methods for reconstructing Shepp Logan phantom in terms of RMSE with different limited angles of projections and a high SNR=40dB.

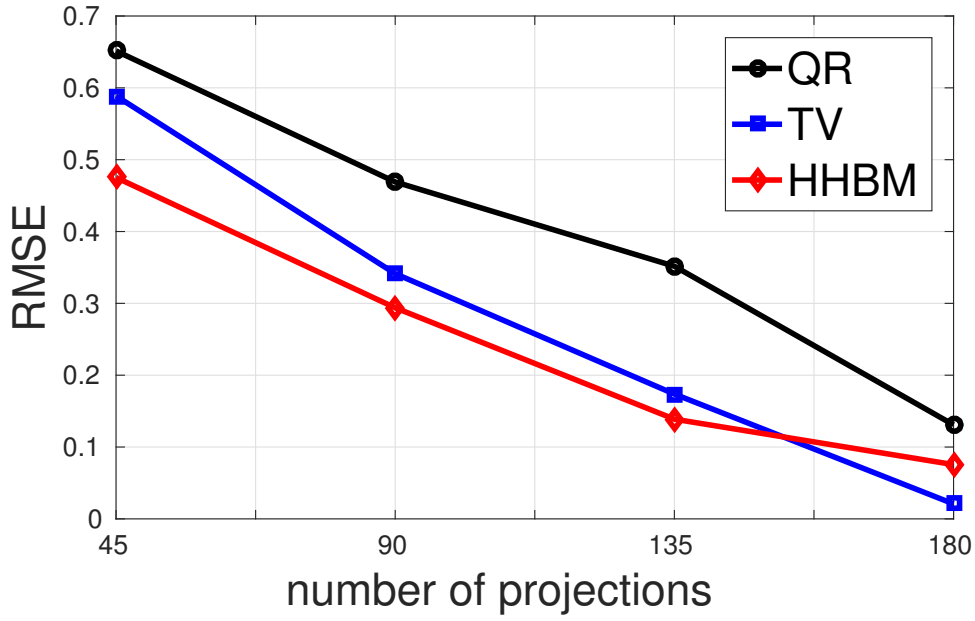


FIGURE 5.18 – The performance of different methods for reconstructing Shepp Logan phantom in terms of RMSE with different limited angles of projections and a low SNR=20dB.

of geometric image features, for example the ridges and edges.

To overcome these problems, the Dual Tree Complex Wavelet Transform (DT-CWT), first introduced by Kingsbury in 1998 [Kin98], is proposed. The dual-tree structure of wavelet filters are used to obtain the real and imaginary parts of complex wavelet coefficients. Figure 5.19 shows the filter bank analysis of the DT-CWT. The DT-CWT employs two real DWTs. The first DWT gives the real part of the transform while the second real DWT gives the imaginary part. The two real DWTs use two different sets of filters. $h_0(n)$, $h_1(n)$ denote the low-pass/high-pass filter pair for the upper filter bank, and $g_0(n)$, $g_1(n)$ denote the low-pass/high-pass filter pair for the lower filter bank. These two filter banks are realized by a typical DWT with two pairs of adapted filters. The inverse of DT-CWT is also a synthesis of two real DWTs. The real part and the imaginary part are each inverted to obtain two real signals. These two real signals are then averaged to obtain the final signal. If the two real DWTs are orthonormal transforms, then so does the DT-CWT. It means that the inverse of the DT-CWT can be performed using its transpose, so that the operator $\boldsymbol{D}$ is orthogonal.

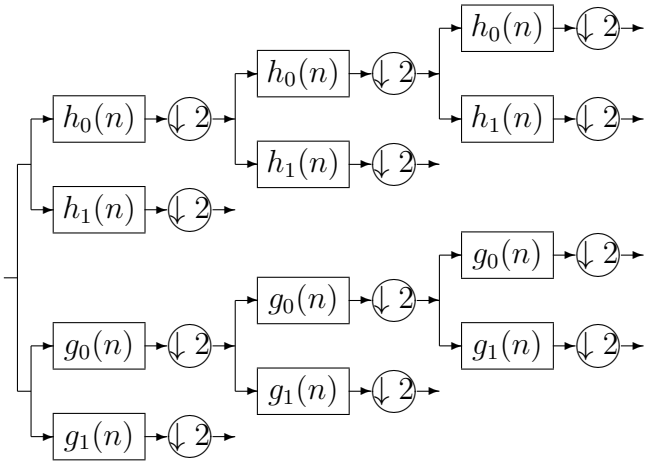


FIGURE 5.19 – Analysis Filter Bank for the dual-tree CWT with a different set of filters at each stage.

Figure 5.20 illustrates the corresponding synthesis filter bank of the DT-CWT transformation.

Also, it was pointed out in [XYM+12] that the dictionary redundancy improves the sparsity of representation. In the discrete Haar transform, the size of the coefficient is the same as the size of image. In DT-CWT, the size of coefficients is two times the size of image in 2D cases (see Figure 5.21) and 4 times the size of phantom in 3D cases. According to the definition, dual-tree of wavelet filters is used to obtain the real and imaginary parts of complex wavelet coefficients. So the coefficients can be divided into 2 groups, which correspond to the real and imaginary parts, respectively.

In this section, I am going to show the comparison of the DT-CWT with the HT. Then DT-CWT will be used in the proposed hierarchical Bayesian method and compared with the HHBM method.

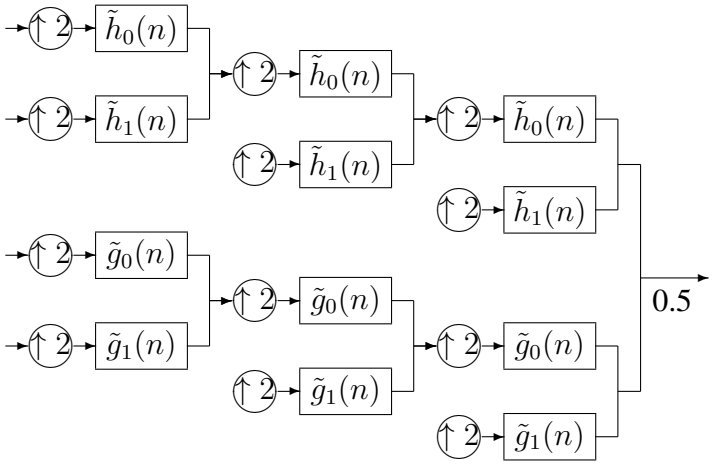
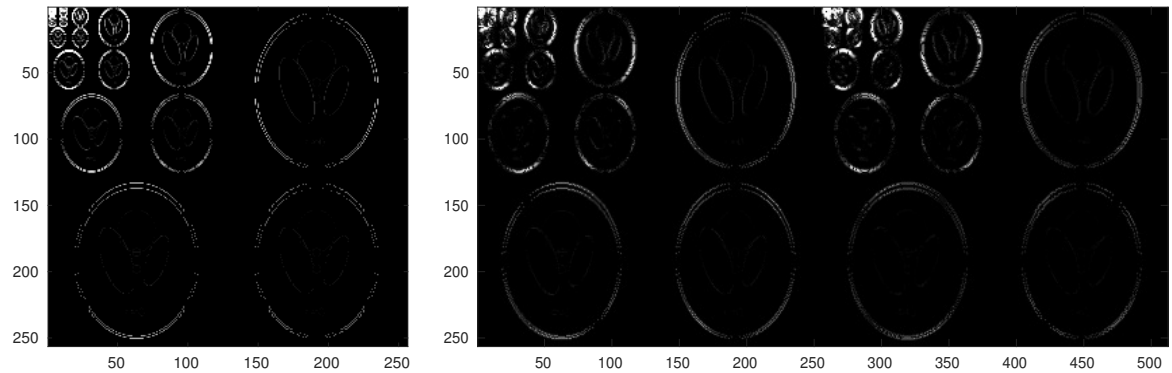


FIGURE 5.20 – Synthesis Filter Bank for the dual-tree CWT with a different set of filters at each stage.



Haar DT-CWT

FIGURE 5.21 – The 5-level Haar transformation coefficient and 5-level DT-CWT transformation coefficient of the 2D Shepp Logan phantom of size 256×256. The scales are adapted in order to be visually clearer.

5.5.1 Comparison with Haar transform

The DT-CWT transform and the Haar transform is compared by analysing several different cases of biased transform coefficient. Error of the coefficient will have an influence on the inverse transformation results. We search for a transformation which is more robust against the noise of coefficient, thus the corresponding inverse-transformed data depends less on the noise of the coefficient.

In Figure 5.21, the HT coefficient and the DT-CWT coefficient are presented. It shows that the DT-CWT transformation import a redundancy to the coefficient. Although the size of the coefficient is two times the HT coefficient, thanks to the development of the computational capacity, the computation can be fulfilled.

In order to compare the HT and DT-CWT, we consider two types of biased transformation coefficients. First of all we truncate the coefficient and compare the recontructed image. Then, a Gaussian noise is added to the coefficient and we compare the reconstructed results of these two transformations.

5.5.2 Truncating parts of coefficients z

In image processing, the wavelet transformations are frequently used in many applications, for example the denoising problem, the restoration, etc. In some of these applications, the coefficient of transformations are truncated. So here we analyse the influence of the truncation of coefficient to the image.

In this simulation, the transformation coefficients are truncated. Noting that for Haar transformation the coefficients matrix has size $\boldsymbol{N} = \boldsymbol{N}_x \times \boldsymbol{N}_y$, and for DT-CWT is $2\boldsymbol{N} = \boldsymbol{N}_x \times 2\boldsymbol{N}_y$. For both of the cases, the 80%, 90%, 95% and 99% smallest coefficients are removed. The image of the inverse transformation is obtained and compared.

The first comparison is given by doing the following steps :

1. computing the transform of image $\boldsymbol{z} = \boldsymbol{D}^{-1}\boldsymbol{f}$,
2. truncating parts of the minimum coefficients in $\boldsymbol{z}$ (80%, 90%, 95% and 99%),
3. using the truncated coefficients to reconstruct the image $\widehat{\boldsymbol{f}}$.

The Relative Mean Square Error (RMSE) of $\widehat{\boldsymbol{f}}$ are given below the corresponding images which are given in Figure 5.22. From the results, we can see that when 80% of the smallest coefficients are truncated, the reconstructed image is still clear by using both transformation. When 90% coefficients are truncated, the reconstructed figures is still clear by using the Haar transformation, but the figure becomes blurry when the DT-CWT is used. When more coefficients are truncated, for example 95% and 99%, both of the two transformations will reconstruct a blurry image. The discrete Haar transformation generates square textures in the reconstructed image, and the DT-CWT transformation gives a smoother image at the contours. From this simulation we can conclude that, the coefficients of HT are more compressed than DT-CWT. The results demonstrate that the proportion of the non-null coefficient is less than 10% while in DT-CWT it is more than 10%. However, when there are more truncated coefficients, the DT-CWT outperforms the HT. Visually, there is also less artifacts in the results of the inverse DT-CWT transformation.

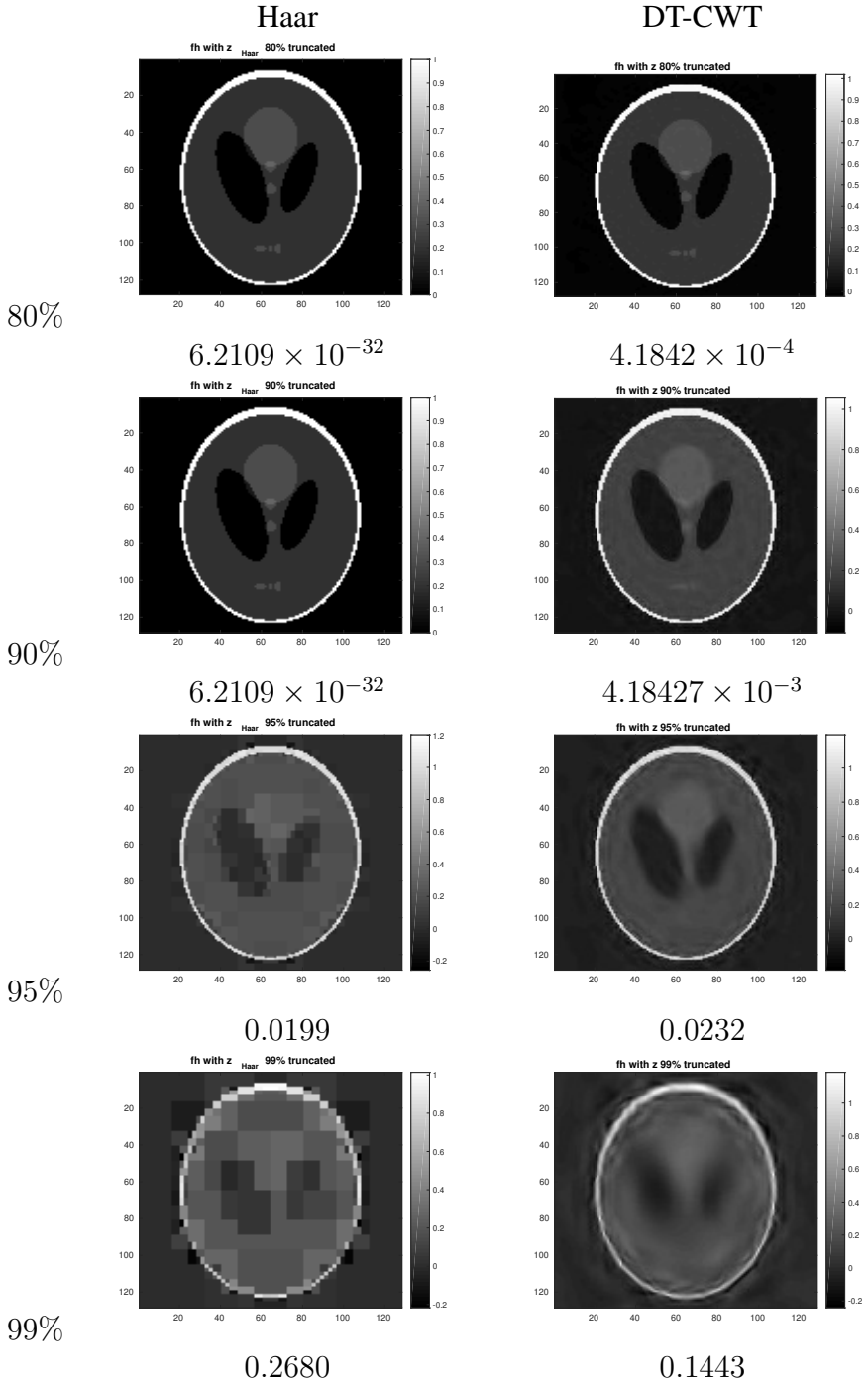


FIGURE 5.22 – Inverse transformed images with different amount of truncated coefficients by using different transformations, with the Relative Mean Square Error (RMSE) of image shown below.

5.5.3 Adding noise to coefficients z

When the coefficients are treated as a variable in the algorithms, there will be appearence of noise in the coefficient. So in the second simulation, I am going to add noise to the coefficient with different SNR values ($30dB$, $20dB$, $10dB$ and $5dB$) and compare the reconstructed image $\widehat{f}$ and the RMSE of them. The comparison is shown in Figure 5.23.

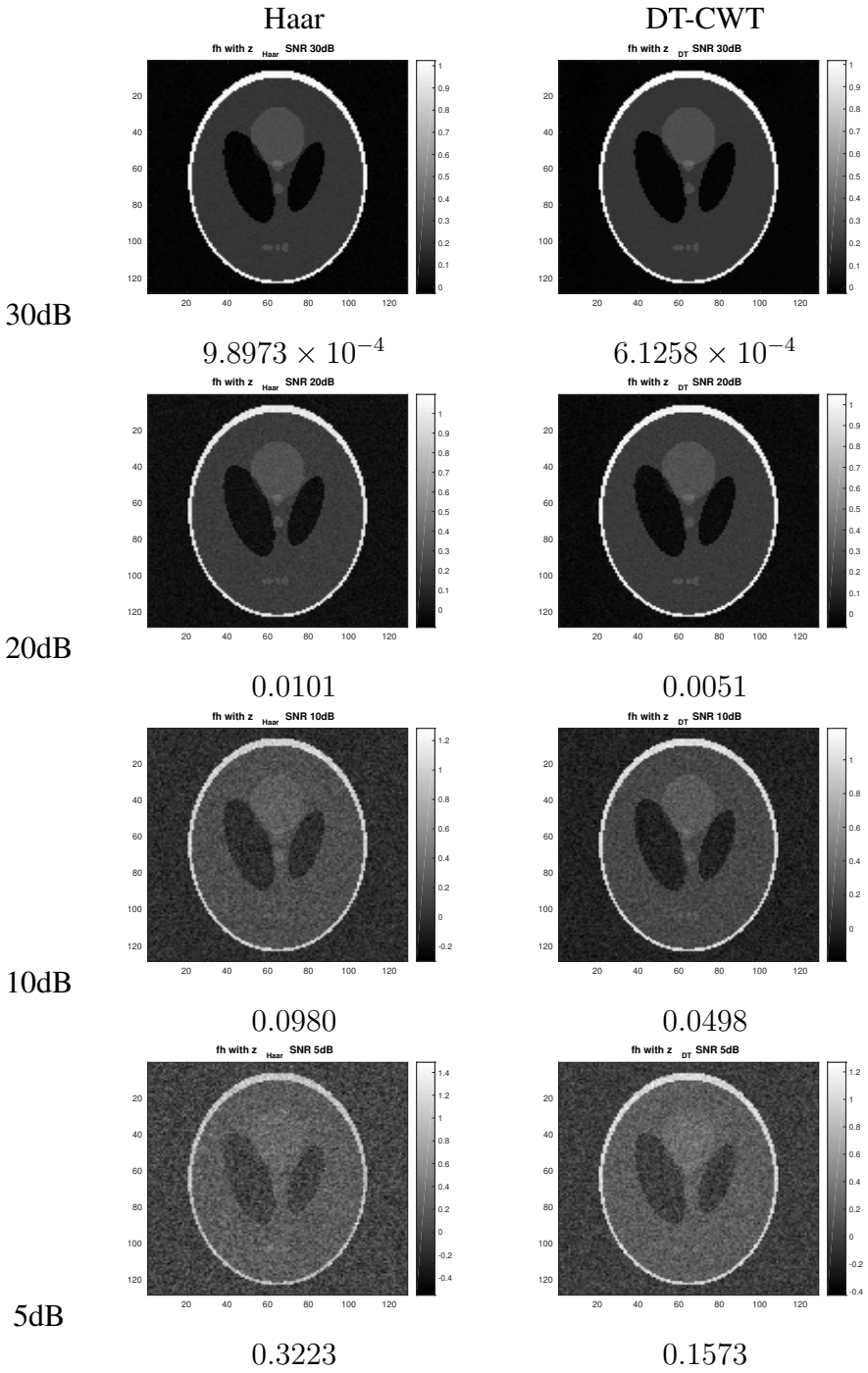
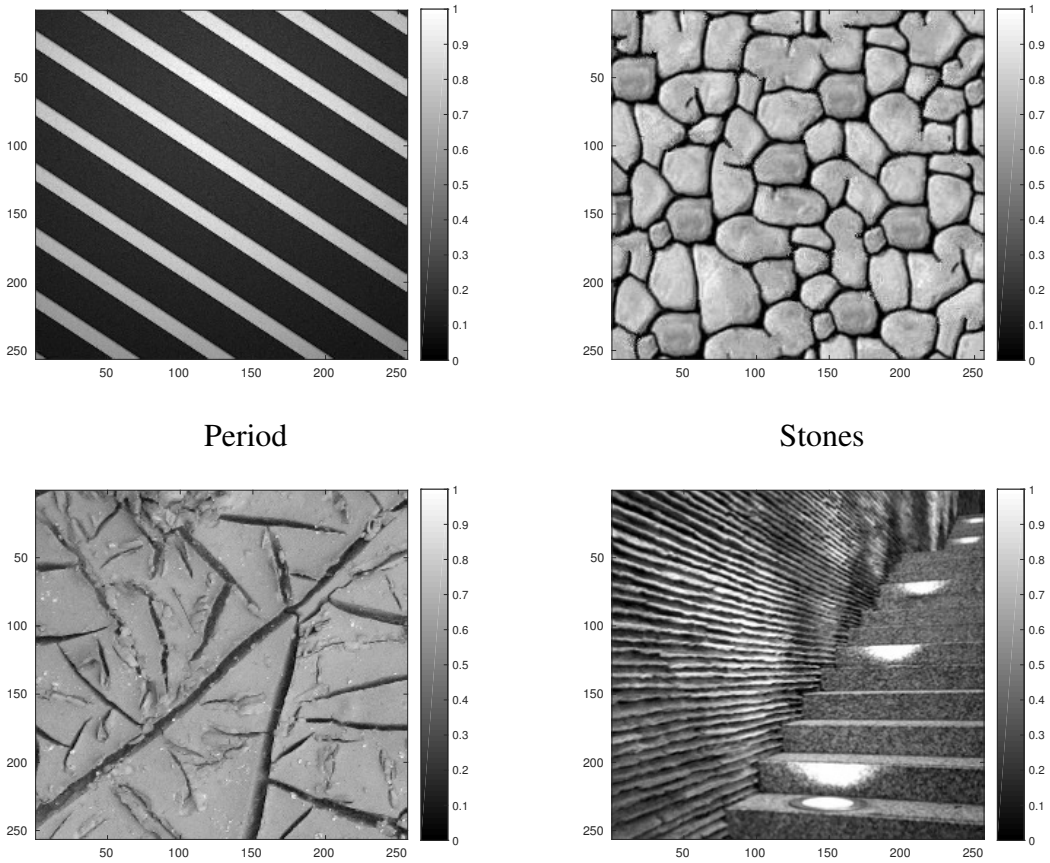


FIGURE 5.23 – Inverse transformed images with different noise added to coefficients by using different transformations, with the Relative Mean Square Error (RMSE) of image shown below.

From the comparison, we can see that when there is a noise on the coefficient, the inverse transformed image will be biased. By using the DT-CWT, the reconstructed image is better than the HT with a small RMSE value. We can conclude from the comparison that DT-CWT is more robust than HT against the noise added to the transformation coefficients.

5.6 Test on different images

The Shepp Logan phantom is artificially created and is strictly piecewise constant. In this case, the Haar transformation is appropriate. The same comparisons have been done by using some real images of different types, for example the image with textures, the piecewise-continuous image and the image with blocs and textures. The images used are the ”Period”, ”Stones”, ”Ground” and ”Stairs”, shown in Figure 5.24.



Period Stones

Ground Stairs

FIGURE 5.24 – The images with different shape of textures.

5.6.1 Test 1 : Truncation in the transform coefficients z

Table 5.2 shows the Relative Mean Square Error, $\mathbf{RMSE} = \left\|\widehat{\boldsymbol{f}} - \boldsymbol{f}\right\|_2^2 / \|\boldsymbol{f}\|_2^2$, of the images obtained by inverse transformation from the truncated coefficient of the Haar and the

DT-CWT transformation. The 80%, 90%, 95% and 99% smallest part of the coefficients of Haar transform and the DT-CWT transform are removed respectively. The results demonstrate that when the coefficient is truncated in the algorithms, the DT-CWT transformation has a better performance for the reconstruction from the transform coefficients.

TABLE 5.2 – Truncating the Transform Coefficients $\boldsymbol{z}$

	Period		Stones		Ground		Stairs	
truncate	Haar	DT-CWT	Haar	DT-CWT	Haar	DT-CWT	Haar	DT-CWT
80%	0.0008	$4.1652e^{-5}$	0.0029	$3.0084e^{-5}$	0.0017	$4.5542e^{-5}$	0.0109	$1.3460e^{-4}$
90%	0.0054	$6.5086e^{-5}$	0.0094	$4.8588e^{-5}$	0.0046	$6.9940e^{-5}$	0.0258	$2.0881e^{-4}$
95%	0.0242	$8.0747e^{-5}$	0.0205	$6.1184e^{-5}$	0.0089	$8.5687e^{-5}$	0.0458	$2.5838e^{-4}$
99%	0.1127	$9.5165e^{-5}$	0.0658	$7.2906e^{-5}$	0.0231	$1.0029e^{-4}$	0.1011	$3.0478e^{-4}$

5.6.2 Test 2 : Adding noise to the transform coefficients $\boldsymbol{z}$

Table 5.3 shows the RMSE of the images obtained by inverse transformation from the biased coefficient of the Haar and the DT-CWT transformation. A Gaussian noise of $5dB$, $10dB$, $15dB$, $20dB$ and $30dB$ are added respectively. The same conclusion is drawn : by using the DT-CWT transformation, the reconstructed images are more robust against the noise added to the transform coefficient.

TABLE 5.3 – Adding noise to the Transform Coefficients $\boldsymbol{z}$

	Period		Stones		Ground		Stairs	
noise	Haar	DT-CWT	Haar	DT-CWT	Haar	DT-CWT	Haar	DT-CWT
5dB	0.3166	0.1596	0.3180	0.1586	0.3142	0.1592	0.3156	0.1563
10dB	0.0990	0.0495	0.1008	0.0501	0.1004	0.0497	0.0993	0.0498
15dB	0.0316	0.0158	0.0313	0.0159	0.0315	0.0160	0.0315	0.0159
20dB	0.0100	0.0050	0.0100	0.0050	0.0100	0.0050	0.0100	0.0050
30dB	0.0010	0.0005	0.0010	0.0005	0.0010	0.0005	0.0110	0.0005

5.7 The use of DT-CWT in Bayesian method

In the previously proposed Bayesian method, the Haar transformation is used and its transformation coefficient is considered as a hidden variable. By comparing the DT-CWT with HT, we demonstrated some advantages of the DT-CWT. In this section, we replace the HT coefficient in the HHBM method by the DT-CWT coefficient, and compare the new Bayesian method with the HHBM method.

The system model depends always on the following two forward models :

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}, \tag{5.85}$$

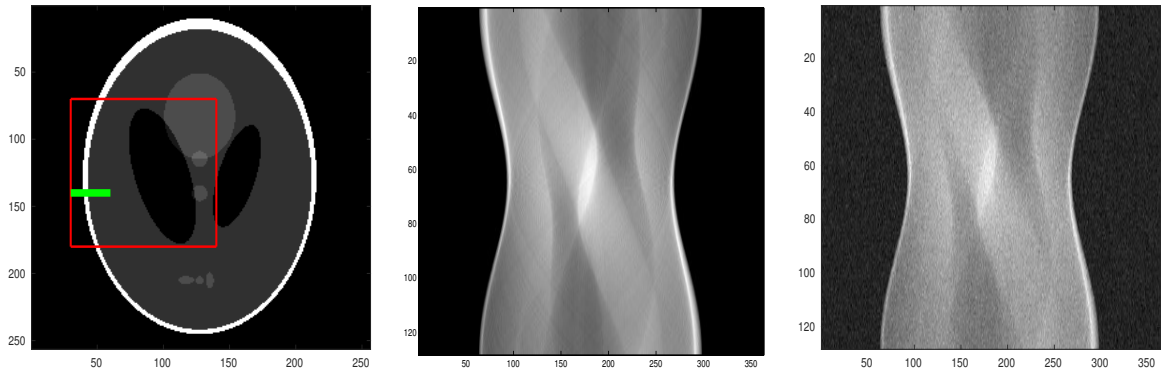
$$\boldsymbol{f} = \boldsymbol{D}\boldsymbol{z} + \boldsymbol{\xi}, \tag{5.86}$$

where $\boldsymbol{D}$ is the DT-CWT and $\boldsymbol{z}$ the DT-CWT coefficient in this case.

5.7.1 Simulation results

In the HHBM method, the discrete multilevel Haar transformation coefficient is considered as the hidden variable. In this section, we replace the Haar transformation with the DT-CWT transformation. The transformation coefficient, $\boldsymbol{z}$, is the DT-CWT coefficient of the object $\boldsymbol{f}$. The generalized Student-t distribution is always used to model the coefficient $\boldsymbol{z}$ to enforce its sparse structure. The method of reconstruction by considering the DT-CWT transformation coefficient is named the **BH-DTCWT** method.

In the simulation, the 2D Shepp Logan phantom of size 256^2 is considered as the original object. The projections are evenly distributed from $0°$ to $180°$. In Figure 5.25, the Original Shepp Logan figure and its projection sinogram are shown. For the illustration of reconstruction phantom, we show the zone of the phantom which is in the red block areas shown in Figure 5.25.



(a) $\boldsymbol{f}$ (b) $\boldsymbol{g}_0 = \boldsymbol{H}\boldsymbol{f}$ (c) $\boldsymbol{g} = \boldsymbol{g}_0 + \boldsymbol{\epsilon}$

FIGURE 5.25 – Original Shepp Logan image of size 256^2 and its (b) sinogram with 128 projections and (c) the same data with additive noise with SNR=20dB.

Figure 5.26 shows the comparison of the middle slice of the reconstructed Shepp Logan phantom by using different reconstruction methods. From the comparison, we can see that, by using the hierarchical structured Bayesian method, the reconstructed phantom is clearer than the TV method when the number of projections is insufficient. Visually, there is no big difference between the results of the HHBM method and the BH-DTCWT method.

Figure 5.27 shows the comparison of a segment of profile around the contour in the Shepp Logan phantom. From the comparison, we see that the BH-DTCWT outperforms the TV method at the contour areas, as it is closer to the original profile.

In order to compare the HHBM and the BH-DTCWT methods, we show in Figure 5.28, Figure 5.29 and Figure 5.30 the evolution of the RMSE during reconstruction of different methods. The comparison shows that the convergence of BH-DTCWT is faster than the HHBM method, and this advantage reduces when there is less number of projections. While comparing with the TV method, the advantage of the HHBM and BH-DTCWT methods is obvious when the projection number is limited.

5.7.2 Conclusions and perspectives

By comparing the Haar transformation and the DT-CWT transformation, we see that the later is more robust when the coefficient is truncated or contaminated. By using DT-CWT trans-

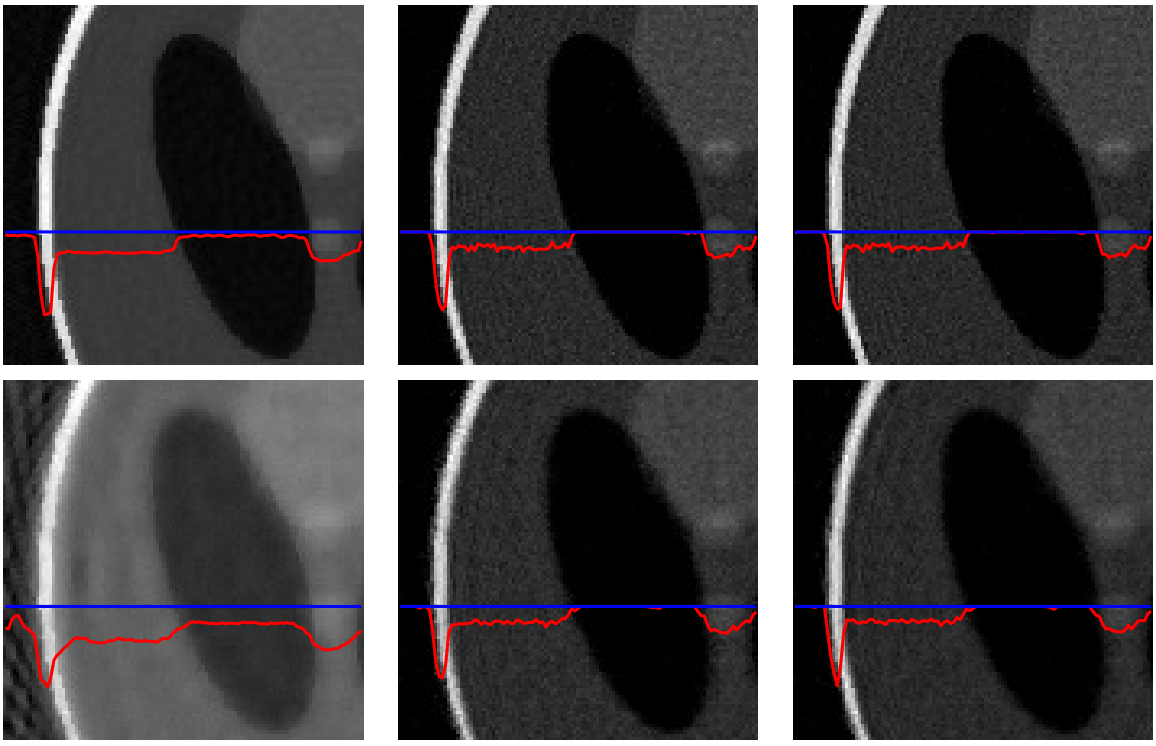


FIGURE 5.26 – The middle slice of reconstructed Shepp Logan phantom by using different reconstruction methods. The reconstructed form projection dataset with top : 64 projections and bottom : 32 projections respectively.

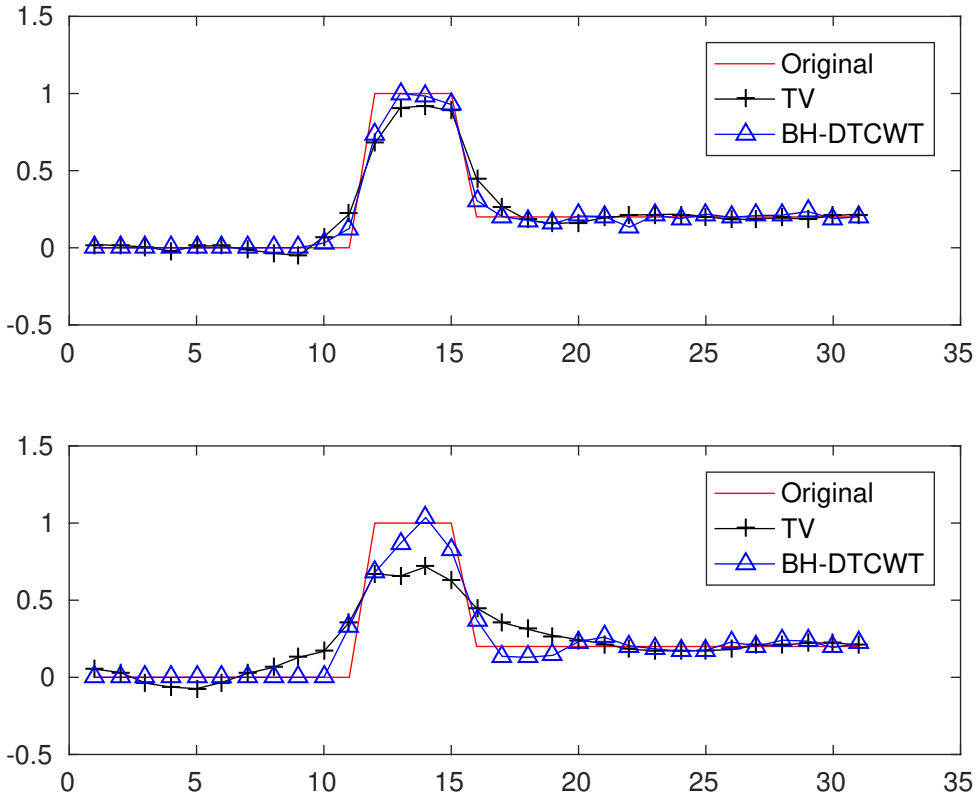


FIGURE 5.27 – Comparison of a zone of profile with TV method, from dataset of 64 projections on the top, and 32 projections on the bottom.

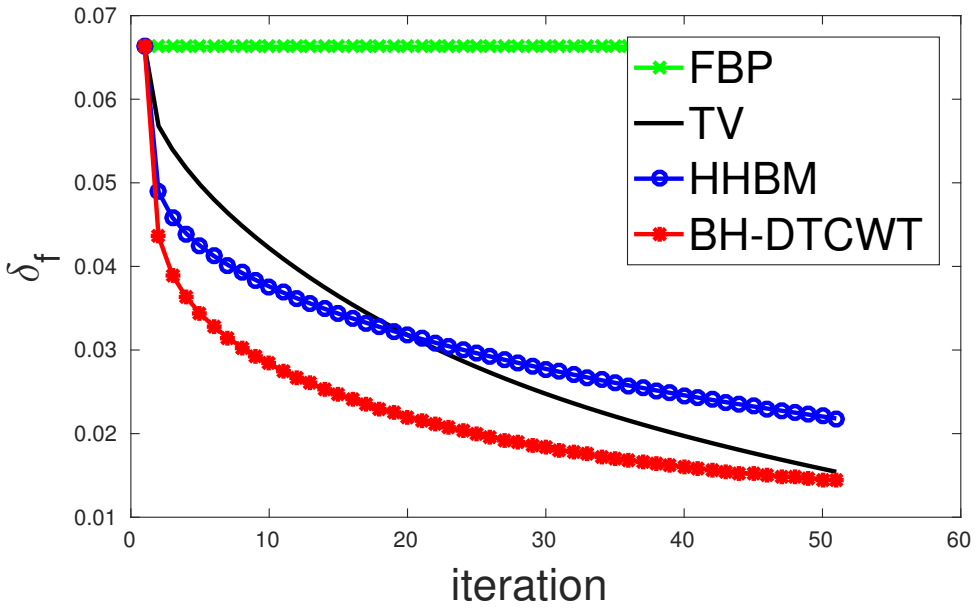


FIGURE 5.28 – NMSE of reconstructed image along iterations from sinogram with 128 projections and a high SNR=40dB.

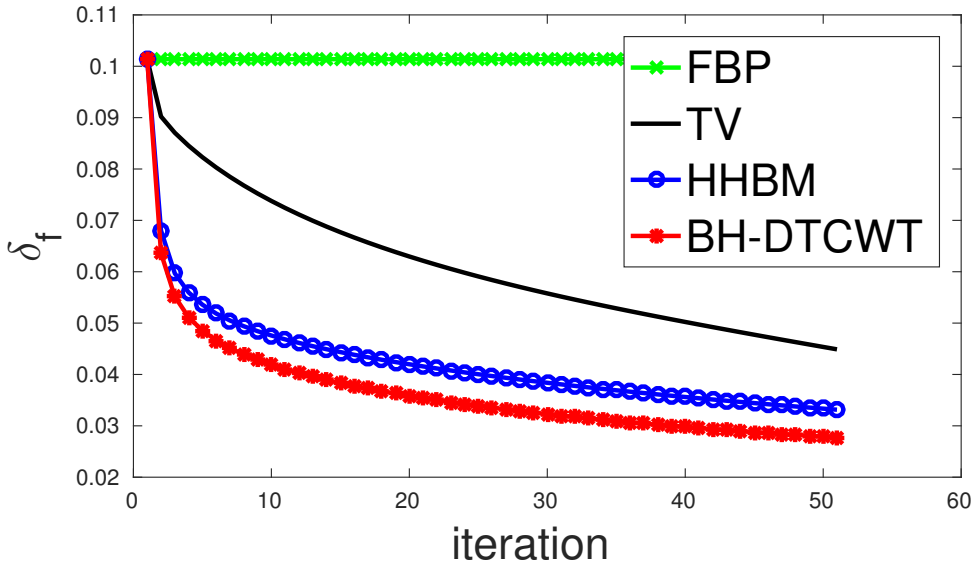


FIGURE 5.29 – NMSE of reconstructed image along iterations from sinogram with 64 projections and a high SNR=40dB.

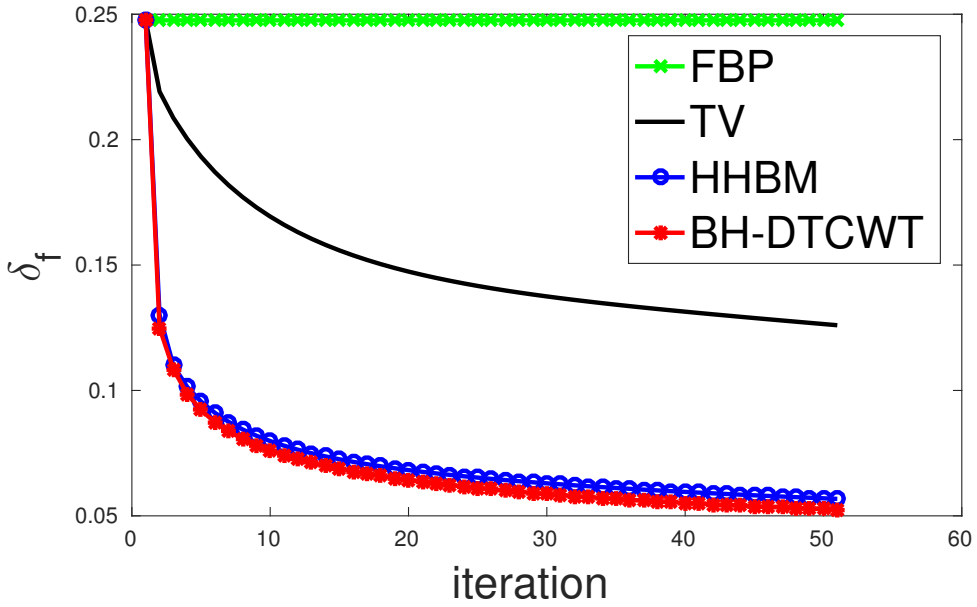


FIGURE 5.30 – NMSE of reconstructed image along iterations from sinogram with 32 projections and a high SNR=40dB.

formation in the Bayesian method, the reconstructed phantom has clearer contours than the TV method. Visually, the reconstructed phantom by using the Bayesian method with Haar transformation or DT-CWT transformation does not have much difference, but the BH-DTCWT converges faster than HHBM.

In future work we will compare some other wavelet transformations which consider the property of the geometric figures and the direction of the contours, for example the curvelet transformation [CDDY06, SCD02], the contourlet transformation [DV05], etc. Meanwhile, the Bayesian method using DT-CWT transformation is being implemented in 3D simulations. The computation of the DT-CWT transformation is expensive, particularly in 3D implementation. So we will implement the transformation by using GPU processor in order to accelerate the computation.

5.8 Variable Splitting Algorithm for HHBM

From the regularization point of view, variable splitting refers to the process of introducing auxiliary constraint variables to separate coupled components in the cost function. This procedure transforms the original minimization problem into an equivalent constrained optimization problem that can be effectively solved using classical constrained optimization schemes. The variable splitting approach has become popular recently for solving reconstruction problems in image processing [ABDF10, RF12, NVL+13].

5.8.1 Variable Splitting (VS)

Consider an unconstrained optimization problem in which the objective function is the sum of two functions, one of which is written as the composition of two functions :

$$\min_{\boldsymbol{f}} R_1(\boldsymbol{f}) + R_2(\Phi(\boldsymbol{f})), \tag{5.87}$$

where Φ is a linear function. Variable splitting is a very simple procedure that consists in creating a new variable, $\boldsymbol{z}$, to serve as the argument of R_2, under the constraint that $\Phi(\boldsymbol{f}) = \boldsymbol{z}$. This leads to the constrained problem :

$$\begin{aligned} &\min_{\boldsymbol{f},\boldsymbol{z}} R_1(\boldsymbol{f}) + R_2(\boldsymbol{z}), \\ &\text{subject to } \ \boldsymbol{z} = \Phi(\boldsymbol{f}). \end{aligned} \tag{5.88}$$

The convenience of the variable splitting methods is that it may be easier to solve the constrained problem in Eq.(5.88) than that of Eq.(5.87). The splitting variable has been recently used in many inverse problem methods, for example in [WYYZ08].

The constrained problem in Eq.(5.88) can be derived into a quadratic penalty approach, by solving :

$$\min_{\boldsymbol{f},\boldsymbol{z}} R_1(\boldsymbol{f}) + R_2(\boldsymbol{z}) + \frac{\lambda}{2} \|\Phi(\boldsymbol{f}) - \boldsymbol{z}\|_2^2 \tag{5.89}$$

by alternately minimizing $\boldsymbol{f}$ and $\boldsymbol{z}$ respectively.

While splitting the variable $\boldsymbol{f}$, we get the criterion similar to the form of the MAP optimization criterion of the HHBM method proposed previously, with one penalty term on $\boldsymbol{f}$, one on $\boldsymbol{z}$ and one penalize the disparity between them.

5.8.2 VS in Bayesian inference

In the regularization methods, the splitting of variables normally leads to the splitting of regularization terms. In the Bayesian inference, on the other hand, it is considered as a splitting of the variables. By using MAP optimization, we can see the links between the regularization methods and a Bayesian inference. However, the Bayesian approach provide the possibilities to estimate the hyperparameters of the inverse problem which are often crucial in real applications.

In X-ray CT models, typically only one variable is used to represent all the additive noise. Consequently, the system noise and the uncertainties of the forward model are all modeled by the same prior distribution, for example the Gaussian distribution or the Student-t distribution that we used previously. In our work, we propose a noise splitting model. This model considers the system noise in different aspects and use different distributions for the modeling of prior variables.

In a real projection system of X-ray CT, the acquisition of the sinogram dataset is obtained mainly by two steps : the projection and the detection. The projection step is the attenuation of the intensity of radiations when passing through the object, and the detection step is the acquisition of the sinogram by the detectors.

Supposing that the true projection data before detection is $\boldsymbol{g}_0$, and the detected projection data is $\boldsymbol{g}$, and theoretically we have $\boldsymbol{g} = \boldsymbol{g}_0$ if the detection noise is not considered. The noise of

detection in tomography is typically modeled by a Poison distribution. In X-ray CT, the number of the photon is big and the Poison distribution is approximated to a Gaussian distribution [SB93]. So the detection of the attenuated radiation leads to an additive noise $\boldsymbol{\epsilon}$:

$$\boldsymbol{g} = \boldsymbol{g}_0 + \boldsymbol{\epsilon}, \tag{5.90}$$

where $\boldsymbol{\epsilon}$ is a Gaussian additive noise, with mean equal to zero and variance $\boldsymbol{v}_\epsilon$.

Before the detection of the radiation, these rays are attenuated by the object. The exact attenuated data $\boldsymbol{g}_0$ is obtained according to the Beer's law, Eq.(3.3), and the linear system is model by the direct model $\boldsymbol{g}_0 = \boldsymbol{H}\boldsymbol{f}$.

Nevertheless, the physical phenomenons during the attenuation of X radiation exist, for example the scattering of some photons, and sometimes the photons with low energies would be absorbed by some materials, for example the metals. These kinds of physical phenomenons will cause uncertainties for the linear model $\boldsymbol{g}_0 = \boldsymbol{H}\boldsymbol{f}$, and hence an additive noise should be added for the direct model :

$$\boldsymbol{g}_0 = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\rho}. \tag{5.91}$$

The uncertainties caused by the scattering or the absorption appears only for few of the photons among a large amount of photons, and hence $\boldsymbol{\rho}$ is sparse. It is modeled by the generalized Student-t distribution in our work. Notice that in our work we consider only the additive noise model.

By splitting the noise, the following model is considered :

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\rho} + \boldsymbol{\epsilon}, \tag{5.92}$$

where $\boldsymbol{\epsilon}$ represents the measurement noise and $\boldsymbol{\rho}$ can represent the modelling errors. It can be expressed as :

$$\begin{aligned} \boldsymbol{g} &= \boldsymbol{g}_0 + \boldsymbol{\epsilon}, \\ \boldsymbol{g}_0 &= \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\rho}, \end{aligned} \tag{5.93}$$

where $\boldsymbol{g}_0$ is the hidden variable of the system. Therefore, the additive noise $\boldsymbol{\epsilon}$ stands for the noise of the detectors, and $\boldsymbol{\rho}$ for the sparse noise.

With this model, by using the classical regularization methods, the solution is defined as the optimizer of :

$$J(\boldsymbol{f}, \boldsymbol{g}_0) = \|\boldsymbol{g} - \boldsymbol{g}_0\|_2^2 + \lambda_1 \|\boldsymbol{g}_0 - \boldsymbol{H}\boldsymbol{f}\|_2^2 + \lambda_2 R(\boldsymbol{f}), \tag{5.94}$$

where parameter λ_1 and λ_2 are chosen beforehand, manually or using an ad-hoc algorithm. An alternate optimization algorithm is :

$$\begin{aligned} \boldsymbol{f}^{(k+1)} &= \boldsymbol{f}^{(k)} + \alpha_1 \left[2\lambda_1 \boldsymbol{H}^T \left(\boldsymbol{g}_0 - \boldsymbol{H}\boldsymbol{f}^{(k)} \right) - \lambda_2 \nabla R(\boldsymbol{f}^{(k)}) \right], \\ \boldsymbol{g}_0^{(k+1)} &= \boldsymbol{g}_0^{(k)} + \alpha_2 \lambda_1 \left(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}^{(k)} \right). \end{aligned} \tag{5.95}$$

We can compare it with the ADMM method presented in Chapter 3 where $\boldsymbol{g}_0$ plays the role of $\boldsymbol{\mu}$.

5.8.3 Gaussian-$\mathcal{S}t_g$ prior model for the variable splitting method (HHBM-GS)

In Bayesian inference, prior models are defined according to the prior information. The noise of detector, $\boldsymbol{\epsilon}$, is modeled by a Gaussian distribution, with a fixed variance $\boldsymbol{v}_\epsilon$, wile the uncertainties $\boldsymbol{\rho}$ is modeled by a heavy tailed distribution, enforcing its sparse structure.

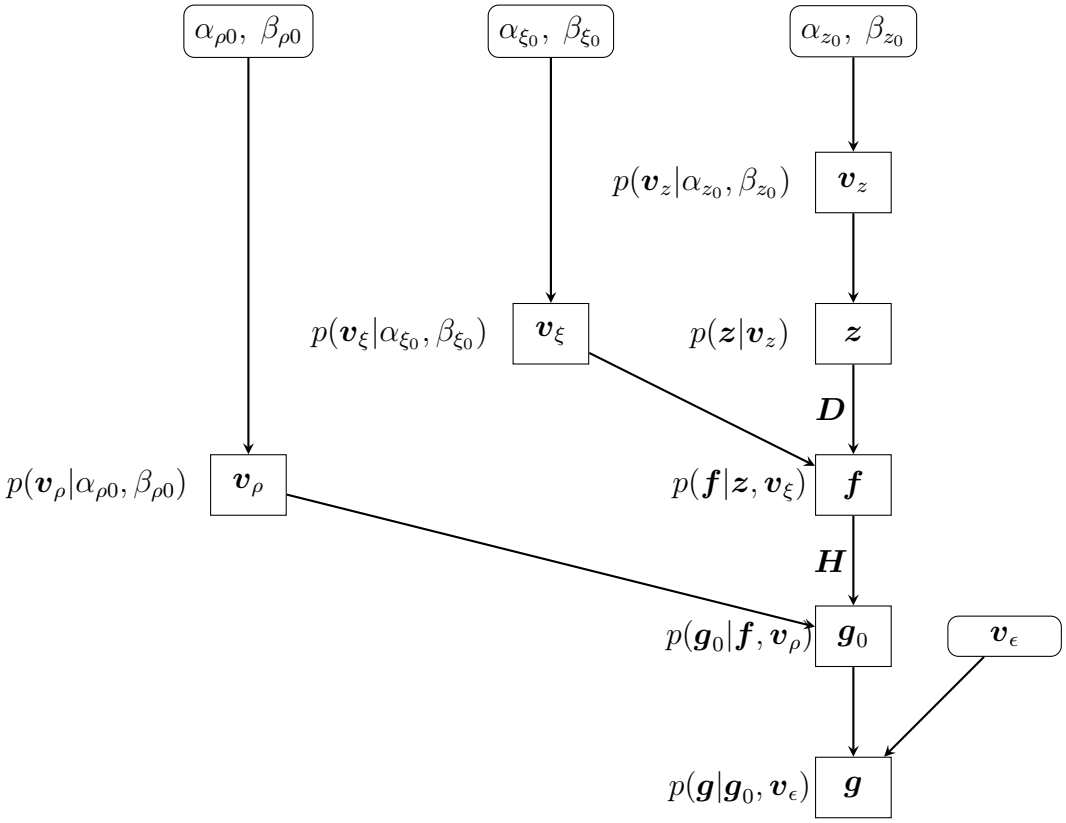


FIGURE 5.31 – Generative graph of the Gaussian-$\mathcal{S}t_g$ Variable Splitting method.

The statistical model is expressed via the forward model, Eq.(5.93), and the generative graph of this system model is illustrated in Figure 5.31.

$$
\begin{aligned}
&p(\boldsymbol{g}|\boldsymbol{g}_0,\boldsymbol{v}_\epsilon)=\mathcal{N}(\boldsymbol{g}|\boldsymbol{g}_0,\boldsymbol{V}_\epsilon)\propto\exp\left\{-\frac{1}{2}\left(\boldsymbol{g}-\boldsymbol{g}_0\right)^T\boldsymbol{V}_\epsilon^{-1}\left(\boldsymbol{g}-\boldsymbol{g}_0\right)\right\},\\
&p(\boldsymbol{g}_0|\boldsymbol{f},\boldsymbol{v}_\rho)=\mathcal{N}(\boldsymbol{g}_0|\boldsymbol{H}\boldsymbol{f},\boldsymbol{V}_\rho)\propto|\boldsymbol{V}_\rho|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left(\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f}\right)^T\boldsymbol{V}_\rho^{-1}\left(\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f}\right)\right\},\\
&p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{v}_\xi)=\mathcal{N}(\boldsymbol{f}|\boldsymbol{D}\boldsymbol{z},\boldsymbol{V}_\xi)\propto|\boldsymbol{V}_\xi|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\right)^T\boldsymbol{V}_\xi^{-1}\left(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\right)\right\},\\
&p(\boldsymbol{z}|\boldsymbol{v}_z)=\mathcal{N}(\boldsymbol{z}|\boldsymbol{0},\boldsymbol{V}_z)\propto|\boldsymbol{V}_z|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\boldsymbol{z}^T\boldsymbol{V}_z^{-1}\boldsymbol{z}\right\}\ \text{ where }\ \boldsymbol{V}_z=\operatorname{diag}\left[\boldsymbol{v}_z\right],\\
&p(\boldsymbol{v}_\rho|\alpha_{\rho0},\beta_{\rho0})=\mathcal{IG}(\boldsymbol{v}_\rho|\alpha_{\rho0},\beta_{\rho0})=\prod_{i=1}^{M}\mathcal{IG}(v_{\rho_i}|\alpha_{\rho0},\beta_{\rho0})\propto\prod_{i=1}^{M}\left\{v_{\rho_i}^{-(\alpha_{\rho0}+1)}\exp\left\{-\beta_{\rho0}v_{\rho_i}^{-1}\right\}\right\},\\
&p(\boldsymbol{v}_\xi|\alpha_{\xi_0},\beta_{\xi_0})=\mathcal{IG}(\boldsymbol{v}_\xi|\alpha_{\xi_0},\beta_{\xi_0})=\prod_{j=1}^{N}\mathcal{IG}(v_{\xi_j}|\alpha_{\xi_0},\beta_{\xi_0})\propto\prod_{j=1}^{N}\left\{v_{\xi_j}^{-(\alpha_{\xi_0}+1)}\exp\left\{-\beta_{\xi_0}v_{\xi_j}^{-1}\right\}\right\},\\
&p(\boldsymbol{v}_z|\alpha_{z_0},\beta_{z_0})=\mathcal{IG}(\boldsymbol{v}_z|\alpha_{z_0},\beta_{z_0})=\prod_{j=1}^{N}\mathcal{IG}(v_{z_j}|\alpha_{z_0},\beta_{z_0})\propto\prod_{j=1}^{N}\left\{v_{z_j}^{-(\alpha_{z_0}+1)}\exp\left\{-\beta_{z_0}v_{z_j}^{-1}\right\}\right\}.
\end{aligned}
\tag{5.96}
$$

As previous, the posterior distribution is obtained from these models via the Bayes rule :

$$
\begin{aligned}
&p(\boldsymbol{g}_0,\boldsymbol{f},\boldsymbol{z},\boldsymbol{v}_\rho,\boldsymbol{v}_\xi,\boldsymbol{v}_z|\boldsymbol{g})\\
&\quad\propto p(\boldsymbol{g}|\boldsymbol{g}_0,\boldsymbol{v}_\epsilon)p(\boldsymbol{g}_0|\boldsymbol{f},\boldsymbol{v}_\rho)p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{v}_\xi)p(\boldsymbol{z}|\boldsymbol{v}_z)p(\boldsymbol{v}_\rho|\alpha_{\rho0},\beta_{\rho0})p(\boldsymbol{v}_\xi|\alpha_{\xi_0},\beta_{\xi_0})p(\boldsymbol{v}_z|\alpha_{z_0},\beta_{z_0})\\
&\quad=\exp\left\{-\frac{1}{2}\left(\boldsymbol{g}-\boldsymbol{g}_0\right)^T\boldsymbol{V}_\epsilon^{-1}\left(\boldsymbol{g}-\boldsymbol{g}_0\right)\right\}|\boldsymbol{V}_\rho|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left(\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f}\right)^T\boldsymbol{V}_\rho^{-1}\left(\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f}\right)\right\}\\
&\qquad|\boldsymbol{V}_\xi|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\right)^T\boldsymbol{V}_\xi^{-1}\left(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\right)\right\}|\boldsymbol{V}_z|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\boldsymbol{z}^T\boldsymbol{V}_z^{-1}\boldsymbol{z}\right\}\\
&\qquad\prod_{i=0}^{M}v_{\rho_i}^{-(\alpha_{\rho0}+1)}\prod_{i=0}^{M}\exp\left\{-\beta_{\rho0}v_{\rho_i}^{-1}\right\}\prod_{j=0}^{N}v_{\xi_j}^{-(\alpha_{\xi_0}+1)}\prod_{j=0}^{N}\exp\left\{-\beta_{\xi_0}v_{\xi_j}^{-1}\right\}\\
&\qquad\prod_{j=0}^{N}v_{z_j}^{-(\alpha_{z_0}+1)}\prod_{j=0}^{N}\exp\left\{-\beta_{z_0}v_{z_j}^{-1}\right\}.
\end{aligned}
\tag{5.97}
$$

The negative logarithm of the posterior distribution is calculated to simplify the optimiza-

tion :

$$\begin{aligned}
J(\boldsymbol{g}_0,\boldsymbol{f},\boldsymbol{z},\boldsymbol{v}_\rho,\boldsymbol{v}_\xi,\boldsymbol{v}_z) &= -\ln p(\boldsymbol{g}_0,\boldsymbol{f},\boldsymbol{z},\boldsymbol{v}_\rho,\boldsymbol{v}_\xi,\boldsymbol{v}_z|\boldsymbol{g}) \\
&= \frac{1}{2}(\boldsymbol{g}-\boldsymbol{g}_0)^T\boldsymbol{V}_\epsilon^{-1}(\boldsymbol{g}-\boldsymbol{g}_0) + \frac{1}{2}\sum_{i=1}^{M}\ln v_{\rho_i} + \frac{1}{2}(\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f})^T\boldsymbol{V}_\rho^{-1}(\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f}) \\
&\quad + \frac{1}{2}\sum_{j=1}^{N}\ln v_{\xi_j} + \frac{1}{2}(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z})^T\boldsymbol{V}_\xi^{-1}(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}) + \frac{1}{2}\sum_{j=1}^{N}\ln v_{z_j} + \frac{1}{2}\boldsymbol{z}^T\boldsymbol{V}_z^{-1}\boldsymbol{z} \\
&\quad + (\alpha_{\rho 0}+1)\sum_{i=1}^{M}\ln v_{\rho_i} + \beta_{\rho 0}\sum_{i=1}^{M}v_{\rho_i}^{-1} + (\alpha_{\xi_0}+1)\sum_{j=1}^{N}\ln v_{\xi_j} + \beta_{\xi_0}\sum_{j=1}^{N}v_{\xi_j}^{-1} \\
&\quad + (\alpha_{z_0}+1)\sum_{j=1}^{N}\ln v_{z_j} + \beta_{z_0}\sum_{j=1}^{N}v_{z_j}^{-1}.
\end{aligned} \tag{5.98}$$

Then the unknowns are optimized alternately by minimizing the criterions :

$$\widehat{\boldsymbol{g}}_0 = \arg\min_{\boldsymbol{g}_0}\left\{\frac{1}{2}(\boldsymbol{g}-\boldsymbol{g}_0)^T\boldsymbol{V}_\epsilon^{-1}(\boldsymbol{g}-\boldsymbol{g}_0) + \frac{1}{2}(\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f})^T\boldsymbol{V}_\rho^{-1}(\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f})\right\}, \tag{5.99}$$

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}}\left\{\frac{1}{2}(\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f})^T\boldsymbol{V}_\rho^{-1}(\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f}) + \frac{1}{2}(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z})^T\boldsymbol{V}_\xi^{-1}(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z})\right\}, \tag{5.100}$$

$$\widehat{\boldsymbol{z}} = \arg\min_{\boldsymbol{z}}\left\{\frac{1}{2}(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z})^T\boldsymbol{V}_\xi^{-1}(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}) + \frac{1}{2}\boldsymbol{z}^T\boldsymbol{V}_z^{-1}\boldsymbol{z}\right\}, \tag{5.101}$$

$$\widehat{v_{\rho_i}} = \arg\min_{v_{\rho_i}}\left\{\frac{1}{2}\ln v_{\rho_i} + \frac{1}{2}[\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f}]_i^2 v_{\rho_i}^{-1} + (\alpha_{\rho 0}+1)\ln v_{\rho_i} + \beta_{\rho 0}v_{\rho_i}^{-1}\right\}, \tag{5.102}$$

$$\widehat{v_{\xi_j}} = \arg\min_{v_{\xi_j}}\left\{\frac{1}{2}\ln v_{\xi_j} + \frac{1}{2}[\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}]_j^2 v_{\xi_j}^{-1} + (\alpha_{\xi_0}+1)v_{\xi_j}^{-1} + \beta_{\xi_0}v_{\xi_j}^{-1}\right\}, \tag{5.103}$$

$$\widehat{v_{zj}} = \arg\min_{v_{z_j}}\left\{\frac{1}{2}\ln v_{z_j} + \frac{1}{2}[\boldsymbol{z}]_j^2 v_{z_j}^{-1} + (\alpha_{z_0}+1)\ln v_{z_j} + \beta_{z_0}v_{z_j}^{-1}\right\}. \tag{5.104}$$

For the optimization of $\boldsymbol{g}_0$, $\boldsymbol{f}$ and $\boldsymbol{z}$, the analytical expressions are :

$$\widehat{\boldsymbol{g}}_0 = \left(\boldsymbol{V}_\epsilon^{-1}+\boldsymbol{V}_\rho^{-1}\right)^{-1}\left(\boldsymbol{V}_\epsilon^{-1}\boldsymbol{g}+\boldsymbol{V}_\rho^{-1}\boldsymbol{H}\boldsymbol{f}\right), \tag{5.105}$$

$$\widehat{\boldsymbol{f}} = \left(\boldsymbol{H}^T\boldsymbol{V}_\rho^{-1}\boldsymbol{H}+\boldsymbol{V}_\xi^{-1}\right)^{-1}\left(\boldsymbol{H}^T\boldsymbol{V}_\rho^{-1}\boldsymbol{g}_0+\boldsymbol{V}_\xi^{-1}\boldsymbol{D}\boldsymbol{z}\right), \tag{5.106}$$

$$\widehat{\boldsymbol{z}} = \left(\boldsymbol{D}^T\boldsymbol{V}_z^{-1}\boldsymbol{D}+\boldsymbol{V}_z^{-1}\right)^{-1}\boldsymbol{D}^T\boldsymbol{V}_z^{-1}\boldsymbol{f}. \tag{5.107}$$

In these three expressions, the analytical expression for $\widehat{\boldsymbol{g}}_0$ is realizable because the matrix $\left(\boldsymbol{V}_\epsilon^{-1}+\boldsymbol{V}_\rho^{-1}\right)$ is a diagonal matrix. The inversion of a diagonal matrix is another diagonal matrix with each element equal to inversion of corresponding element of the original matrix. Hence :

$$\widehat{\boldsymbol{g}}_0 = \text{diag}\left[\frac{1}{v_{\epsilon_i}^{-1}+v_{\rho_i}^{-1}}\right]\left(\boldsymbol{V}_\epsilon^{-1}\boldsymbol{g}+\boldsymbol{V}_\rho^{-1}\boldsymbol{H}\boldsymbol{f}\right) \tag{5.108}$$

But for $\widehat{\boldsymbol{f}}$ and $\widehat{\boldsymbol{z}}$, the terms of matrix inversion is not realizable in the background of big data size images reconstruction. So as we have presented previously, the gradient descent optimization algorithm is used.

The analytical expression for optimization of $\widehat{v}_{\rho_i}$, $\widehat{v}_{\xi_j}$ and $\widehat{v}_{zj}$ are also obtained by calculating the critical point of the corresponding criterion.

Consequently, the iterative optimization expressions of all the unknown variables are given bellow :

$$\begin{aligned}
&\widehat{\boldsymbol{g}}_0 = \left(\boldsymbol{V}_{\epsilon}^{-1} + \boldsymbol{V}_{\rho}^{-1}\right)^{-1}\left(\boldsymbol{V}_{\epsilon}^{-1}\boldsymbol{g} + \boldsymbol{V}_{\rho}^{-1}\boldsymbol{H}\boldsymbol{f}\right),\\
&\text{for } i = 1 : I_2 \ \widehat{\boldsymbol{f}}^{(k+1)} = \widehat{\boldsymbol{f}}^{(k)} - \gamma_f^{(k)}\nabla J(\widehat{\boldsymbol{f}}^{(k)}),\\
&\text{for } i = 1 : I_2 \ \widehat{\boldsymbol{z}}^{(k+1)} = \widehat{\boldsymbol{z}}^{(k)} - \gamma_z^{(k)}\nabla J(\widehat{\boldsymbol{z}}^{(k)}),\\
&\widehat{v}_{\rho_i} = \frac{\beta_{\rho 0} + \frac{1}{2}\left[\boldsymbol{g}_0 - \boldsymbol{H}\boldsymbol{f}\right]_i^2}{\alpha_{\rho 0} + \frac{3}{2}},\\
&\widehat{v}_{\xi_j} = \frac{\beta_{\xi_0} + \frac{1}{2}\left[\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}\right]_j^2}{\alpha_{z_0} + \frac{3}{2}},\\
&\widehat{v}_{zj} = \frac{\beta_{z_0} + \frac{1}{2}z_j^2}{\alpha_{z_0} + \frac{3}{2}},
\end{aligned} \tag{5.109}$$

where

$$\nabla J(\widehat{\boldsymbol{f}}^{(k)}) = -\boldsymbol{H}^T\boldsymbol{V}_{\rho}^{-1}\left(\boldsymbol{g}_0 - \boldsymbol{H}\boldsymbol{f}\right) + \boldsymbol{V}_{\xi}^{-1}\left(\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}\right), \tag{5.110}$$

$$\nabla J(\widehat{\boldsymbol{z}}^{(k)}) = -\boldsymbol{D}^T\boldsymbol{V}_z^{-1}\left(\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}\right) + \boldsymbol{V}_z^{-1}\boldsymbol{z}, \tag{5.111}$$

$$\gamma_f^{(k)} = \frac{\left\|\nabla J(\boldsymbol{f})\right\|_2^2}{\left\|\boldsymbol{Y}_{\rho}\boldsymbol{H}\nabla J(\boldsymbol{f})\right\|_2^2 + \left\|\boldsymbol{Y}_{\xi}\nabla J(\boldsymbol{f})\right\|_2^2}, \text{ where } \boldsymbol{Y}_{\rho} = \boldsymbol{V}_{\rho}^{-\frac{1}{2}} \text{ and } \boldsymbol{Y}_{\xi} = \boldsymbol{V}_{\xi}^{-\frac{1}{2}}, \tag{5.112}$$

$$\gamma_z^{(k)} = \frac{\left\|\nabla J(\boldsymbol{z})\right\|_2^2}{\left\|\boldsymbol{Y}_{\xi}\boldsymbol{D}\nabla J(\boldsymbol{z})\right\|_2^2 + \left\|\boldsymbol{Y}_{z}\nabla J(\boldsymbol{z})\right\|_2^2}, \text{ where } \boldsymbol{Y}_{z} = \boldsymbol{V}_{z}^{-\frac{1}{2}}. \tag{5.113}$$

As we have demonstrated, the Normal distribution can be approximated by a $\mathcal{S}t_g$ distribution. So in the variable splitting method, we also consider to use the $\mathcal{S}t_g$ distribution to model the noise $\boldsymbol{\epsilon}$, presented in the next section.

5.8.4 $\mathcal{S}t_g$-$\mathcal{S}t_g$ prior model for the variable splitting method (HHBM-SS)

Another prior model for the variable splitting method is to model both the noise $\boldsymbol{\rho}$ and $\boldsymbol{\epsilon}$ by a generalized Student-t distribution. As we have mentioned previously, the $\mathcal{S}t_g$ distribution can be either heavy-tailed for modeling a sparse variable or approach to a Gaussian distribution for non-sparse variable by choosing different pair of hyper-parameters α and β.

In this model, both two additive noises, $\boldsymbol{\epsilon}$ and $\boldsymbol{\rho}$ are model by a generalized Student-t distribution. Comparing with the previous model, $\boldsymbol{\epsilon}$ is modeled with a Gaussian distribution with an unknown variance $\boldsymbol{v}_{\epsilon}$ belonging to an Inverse Gamma distribution :

$$\begin{aligned}
&p(\boldsymbol{\epsilon}|\boldsymbol{v}_{\epsilon}) = \mathcal{N}(\boldsymbol{\epsilon}|0, \boldsymbol{V}_{\epsilon}), \text{ where } \boldsymbol{V}_{\epsilon} = \text{diag}\left[\boldsymbol{v}_{\epsilon}\right],\\
&p(\boldsymbol{v}_{\epsilon}|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) = \mathcal{IG}(\boldsymbol{v}_{\epsilon}|\alpha_{\epsilon_0}, \beta_{\epsilon_0}).
\end{aligned}$$

Figure 5.32 illustrates the Generative graph of this Variable Splitting model.

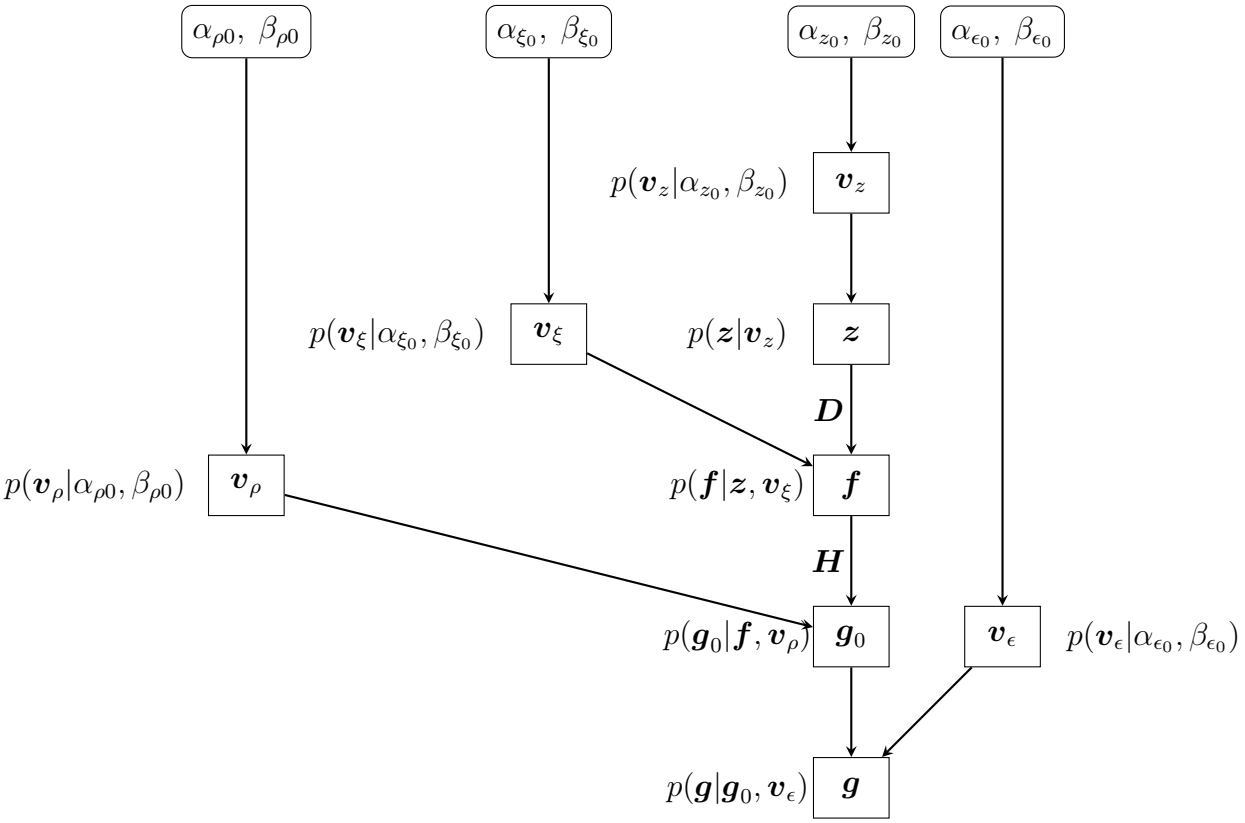


FIGURE 5.32 – Generative graph of the $\mathcal{S}t_g$-$\mathcal{S}t_g$ Variable Splitting method.

Consequently the posterior distribution is :

$$
\begin{aligned}
&p(\boldsymbol{g}_0,\boldsymbol{f},\boldsymbol{z},\boldsymbol{v}_\epsilon,\boldsymbol{v}_\rho,\boldsymbol{v}_\xi,\boldsymbol{v}_z|\boldsymbol{g}) \\
&\quad\propto p(\boldsymbol{g}|\boldsymbol{g}_0,\boldsymbol{v}_\epsilon)p(\boldsymbol{g}_0|\boldsymbol{f},\boldsymbol{v}_\rho)p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{v}_\xi)p(\boldsymbol{z}|\boldsymbol{v}_z)p(\boldsymbol{v}_\epsilon|\alpha_{\epsilon_0},\beta_{\epsilon_0})p(\boldsymbol{v}_\rho|\alpha_{\rho 0},\beta_{\rho 0})p(\boldsymbol{v}_\xi|\alpha_{\xi_0},\beta_{\xi_0})p(\boldsymbol{v}_z|\alpha_{z_0},\beta_{z_0}) \\
&\quad= |\boldsymbol{V}_\epsilon|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{g}-\boldsymbol{g}_0\right)^T \boldsymbol{V}_\epsilon^{-1}\left(\boldsymbol{g}-\boldsymbol{g}_0\right)\right\} |\boldsymbol{V}_\rho|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f}\right)^T \boldsymbol{V}_\rho^{-1}\left(\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f}\right)\right\} \\
&\qquad |\boldsymbol{V}_\xi|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\right)^T \boldsymbol{V}_\xi^{-1}\left(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\right)\right\} |\boldsymbol{V}_z|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{z}^T\boldsymbol{V}_z^{-1}\boldsymbol{z}\right\} \\
&\qquad \prod_{i=0}^{M} v_{\epsilon_i}^{-(\alpha_{\epsilon_0}+1)} \prod_{i=0}^{M} \exp\left\{-\beta_{\epsilon_0} v_{\epsilon_i}^{-1}\right\} \prod_{i=0}^{M} v_{\rho_i}^{-(\alpha_{\rho 0}+1)} \prod_{i=0}^{M} \exp\left\{-\beta_{\rho 0} v_{\rho_i}^{-1}\right\} \\
&\qquad \prod_{j=0}^{N} v_{\xi_j}^{-(\alpha_{\xi_0}+1)} \prod_{j=0}^{N} \exp\left\{-\beta_{\xi_0} v_{\xi_j}^{-1}\right\} \prod_{j=0}^{N} v_{z_j}^{-(\alpha_{z_0}+1)} \prod_{j=0}^{N} \exp\left\{-\beta_{z_0} v_{z_j}^{-1}\right\},
\end{aligned}
\tag{5.114}
$$

and the negative logarithm of the posterior distribution is :

$$\begin{aligned}
J(\boldsymbol{g}_0,\boldsymbol{f},\boldsymbol{z},\boldsymbol{v}_\rho,\boldsymbol{v}_\xi,\boldsymbol{v}_z) &= -\ln p(\boldsymbol{g}_0,\boldsymbol{f},\boldsymbol{z},\boldsymbol{v}_\rho,\boldsymbol{v}_\xi,\boldsymbol{v}_z|\boldsymbol{g}) \\
&= \frac{1}{2}\left(\boldsymbol{g}-\boldsymbol{g}_0\right)^T \boldsymbol{V}_\epsilon^{-1}\left(\boldsymbol{g}-\boldsymbol{g}_0\right) + \frac{1}{2}\sum_{i=1}^{M}\ln v_{\rho_i} + \frac{1}{2}\left(\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f}\right)^T \boldsymbol{V}_\rho^{-1}\left(\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f}\right) \\
&\quad + \frac{1}{2}\sum_{j=1}^{N}\ln v_{\xi_j} + \frac{1}{2}\left(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\right)^T \boldsymbol{V}_\xi^{-1}\left(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\right) + \frac{1}{2}\sum_{j=1}^{N}\ln v_{z_j} + \frac{1}{2}\boldsymbol{z}^T\boldsymbol{V}_z^{-1}\boldsymbol{z} \\
&\quad + (\alpha_{\rho 0}+1)\sum_{i=1}^{M}\ln v_{\rho_i} + \beta_{\rho 0}\sum_{i=1}^{M} v_{\rho_i}^{-1} + (\alpha_{\xi_0}+1)\sum_{j=1}^{N}\ln v_{\xi_j} + \beta_{\xi_0}\sum_{j=1}^{N} v_{\xi_j}^{-1} \\
&\quad + (\alpha_{z_0}+1)\sum_{j=1}^{N}\ln v_{z_j} + \beta_{z_0}\sum_{j=1}^{N} v_{z_j}^{-1}.
\end{aligned} \tag{5.115}$$

The unknowns are optimized alternately :

$$\widehat{\boldsymbol{g}}_0 = \arg\min_{\boldsymbol{g}_0}\left\{\frac{1}{2}\left(\boldsymbol{g}-\boldsymbol{g}_0\right)^T \boldsymbol{V}_\epsilon^{-1}\left(\boldsymbol{g}-\boldsymbol{g}_0\right) + \frac{1}{2}\left(\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f}\right)^T \boldsymbol{V}_\rho^{-1}\left(\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f}\right)\right\}, \tag{5.116}$$

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}}\left\{\frac{1}{2}\left(\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f}\right)^T \boldsymbol{V}_\rho^{-1}\left(\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f}\right) + \frac{1}{2}\left(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\right)^T \boldsymbol{V}_\xi^{-1}\left(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\right)\right\}, \tag{5.117}$$

$$\widehat{\boldsymbol{z}} = \arg\min_{\boldsymbol{z}}\left\{\frac{1}{2}\left(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\right)^T \boldsymbol{V}_\xi^{-1}\left(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\right) + \frac{1}{2}\boldsymbol{z}^T\boldsymbol{V}_z^{-1}\boldsymbol{z}\right\}, \tag{5.118}$$

$$\widehat{v}_{\epsilon_i} = \arg\min_{v_{\epsilon_i}}\left\{\frac{1}{2}\ln v_{\epsilon_i} + \frac{1}{2}\left[\boldsymbol{g}-\boldsymbol{g}_0\right]_i^2 v_{\epsilon_i}^{-1} + (\alpha_{\epsilon_0}+1)\ln v_{\epsilon_i} + \beta_{\epsilon_0} v_{\epsilon_i}^{-1}\right\}, \tag{5.119}$$

$$\widehat{v}_{\rho_i} = \arg\min_{v_{\rho_i}}\left\{\frac{1}{2}\ln v_{\rho_i} + \frac{1}{2}\left[\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f}\right]_i^2 v_{\rho_i}^{-1} + (\alpha_{\rho 0}+1)\ln v_{\rho_i} + \beta_{\rho 0} v_{\rho_i}^{-1}\right\}, \tag{5.120}$$

$$v_{\xi_j} = \arg\min_{v_{\xi_j}}\left\{\frac{1}{2}\ln v_{\xi_j} + \frac{1}{2}\left[\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\right]_j^2 v_{\xi_j}^{-1} + (\alpha_{\xi_0}+1)\, v_{\xi_j}^{-1} + \beta_{\xi_0} v_{\xi_j}^{-1}\right\}, \tag{5.121}$$

$$v_{z_j} = \arg\min_{v_{z_j}}\left\{\frac{1}{2}\ln v_{z_j} + \frac{1}{2}z_j^2 v_{z_j}^{-1} + (\alpha_{z_0}+1)\ln v_{z_j} + \beta_{z_0} v_{z_j}^{-1}\right\}. \tag{5.122}$$

As indicated previously, the estimation of variables $\boldsymbol{f}$ and $\boldsymbol{z}$ are realized by using the descent

gradient algorithm. The iterative optimization expressions of all the unknown variables are :

$$\begin{aligned}
&\widehat{\boldsymbol{g}}_0 = \left(\boldsymbol{V}_{\epsilon}^{-1} + \boldsymbol{V}_{\rho}^{-1}\right)^{-1}\left(\boldsymbol{V}_{\epsilon}^{-1}\boldsymbol{g} + \boldsymbol{V}_{\rho}^{-1}\boldsymbol{H}\boldsymbol{f}\right),\\
&\text{for } i = 1 : I_2 \; \widehat{\boldsymbol{f}}^{(k+1)} = \widehat{\boldsymbol{f}}^{(k)} - \gamma_f^{(k)}\nabla J(\widehat{\boldsymbol{f}}^{(k)}),\\
&\text{for } i = 1 : I_2 \; \widehat{\boldsymbol{z}}^{(k+1)} = \widehat{\boldsymbol{z}}^{(k)} - \gamma_z^{(k)}\nabla J(\widehat{\boldsymbol{z}}^{(k)}),\\
&\widehat{v}_{\epsilon_i} = \frac{\beta_{\epsilon_0} + \frac{1}{2}\left[\boldsymbol{g} - \boldsymbol{g}_0\right]_i^2}{\alpha_{\epsilon_0} + \frac{3}{2}},\\
&\widehat{v}_{\rho_i} = \frac{\beta_{\rho 0} + \frac{1}{2}\left[\boldsymbol{g}_0 - \boldsymbol{H}\boldsymbol{f}\right]_i^2}{\alpha_{\rho 0} + \frac{3}{2}},\\
&\widehat{v}_{\xi_j} = \frac{\beta_{\xi_0} + \frac{1}{2}\left[\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}\right]_j^2}{\alpha_{z_0} + \frac{3}{2}},\\
&\widehat{v}_{zj} = \frac{\beta_{z_0} + \frac{1}{2}z_j^2}{\alpha_{z_0} + \frac{3}{2}},
\end{aligned} \tag{5.123}$$

where

$$\nabla J(\widehat{\boldsymbol{f}}^{(k)}) = -\boldsymbol{H}^T\boldsymbol{V}_{\rho}^{-1}\left(\boldsymbol{g}_0 - \boldsymbol{H}\boldsymbol{f}\right) + \boldsymbol{V}_{\xi}^{-1}\left(\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}\right), \tag{5.124}$$

$$\nabla J(\widehat{\boldsymbol{z}}^{(k)}) = -\boldsymbol{D}^T\boldsymbol{V}_{z}^{-1}\left(\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}\right) + \boldsymbol{V}_{z}^{-1}\boldsymbol{z}, \tag{5.125}$$

$$\gamma_f^{(k)} = \frac{\|\nabla J(\boldsymbol{f})\|_2^2}{\|\boldsymbol{Y}_{\rho}\boldsymbol{H}\nabla J(\boldsymbol{f})\|_2^2 + \|\boldsymbol{Y}_{\xi}\nabla J(\boldsymbol{f})\|_2^2}, \tag{5.126}$$

$$\gamma_z^{(k)} = \frac{\|\nabla J(\boldsymbol{z})\|_2^2}{\|\boldsymbol{Y}_{\xi}\boldsymbol{D}\nabla J(\boldsymbol{z})\|_2^2 + \|\boldsymbol{Y}_{z}\nabla J(\boldsymbol{z})\|_2^2}. \tag{5.127}$$

5.8.5 Initialization of hyperparameters for HHBM-GS and HHBM-SS methods

The hyperparameters, α_{ϵ_0}, β_{ϵ_0}, $\alpha_{\rho 0}$ and $\beta_{\rho 0}$, of the $\mathcal{S}t_g$ distribution are initialized by considering the prior property of the two noises $\boldsymbol{\epsilon}$ and $\boldsymbol{\rho}$. According to the system model of X-ray CT, the hyperparameters are initialized such that $\boldsymbol{\epsilon}$ is sparse and $\boldsymbol{\rho}$ approach to a Gaussian noise.

As presented in Section 5.4.2, the initialization of the hyperparameters considers the SNR of the dataset. The variance of noise v_n can be estimated beforehand with the projection data without any object. According to Eq.(5.83), we have :

$$v_n \approx \frac{\|\boldsymbol{g}\|_2^2}{M} \times \frac{1}{1 + 10^{SNR/10}}. \tag{5.128}$$

In our work, without any further information about the two different noise of system $\boldsymbol{\epsilon}$ and $\boldsymbol{\rho}$, we suppose that $v_{\epsilon} = v_{\rho}$, and each of them equal to half of the total noise variance :

$$v_{\epsilon} = v_{\rho} = \frac{1}{2}v_n. \tag{5.129}$$

- For **HHBM-GS** method, we fix $v_{\epsilon} = v_n$, and the relation $\beta_{\rho 0} = v_{\rho}\left(\alpha_{\rho 0} - 1\right)$ is considered. We fix $\alpha_{\rho 0} = 2.01$ in order to have a sparse structure for $\boldsymbol{\rho}$.

- For **HHBM-SS** method, the following relations are considered :

$$\beta_{\epsilon_0} = v_\epsilon \left(\alpha_{\epsilon_0} - 1\right), \tag{5.130}$$
$$\beta_{\rho 0} = v_\rho \left(\alpha_{\rho 0} - 1\right). \tag{5.131}$$

  In our work, we initialized different values for α_{ϵ_0} and $\alpha_{\rho 0}$ in order to realize different prior distributions. We set $\alpha_{\rho 0} = 2.01$ in order to have a sparse structure for $\boldsymbol{\rho}$, and we set $\alpha_{\epsilon_0} = 100$ in order to have a Gaussian structured noise $\boldsymbol{\epsilon}$.

Next we use these two different noise splitting methods for the reconstruction of Shepp Logan figure and compare the simulation results.

5.8.6 Simulation results of the noise splitting method

In the simulations, we compared the proposed HHBM method with the variable splitting method. In our work, we considered to split the noise into two parts, and two strategies of the prior modeling are proposed. In the first strategy, called **HHBM-GS**, the projection noise $\boldsymbol{\rho}$ is modeled by a $\mathcal{S}t_g$ distribution while the detection noise is modeled by a Gaussian distribution. In the second strategy, called **HHBM-SS**, two $\mathcal{S}t_g$ distributions are used to model these two noises, $\boldsymbol{\epsilon}$ and $\boldsymbol{\rho}$, with different values for the initialization of the corresponding hyperparameters.

In the simulations, the Shepp Logan phantom of size 256^2 is used as the original figure. The projections are evenly distributed from $0°$ to $180°$. The initialization for the figure $\widehat{\boldsymbol{f}}^{(0)}$ is the reconstructed figure of the FBP method.

Figure 5.33 shows the comparison of the convergence of the RMSE of the reconstruction by using the HHBM method, the variable splitting method with two $\mathcal{S}t_g$ models (**HHBM-SS**) and the variable splitting method with a Gaussian and a $\mathcal{S}t_g$ model (**HHBM-GS**), with 128 projections and 64 projections, respectively.

From the comparisons, we can see that, by considering two $\mathcal{S}t_g$ noises, the variable splitting method (**HHBM-SS**) has a slightly better reconstruction result than the HHBM method in terms of RMSE value. By modeling one of the noise by a Gaussian distribution and the other by the $\mathcal{S}t_g$ distribution, the reconstruction results surpasses that of the HHBM method and the HHBM-SS method, with a faster convergence than the other methods.

5.9 Conclusions and perspectives

In this chapter, the HHBM method is presented. In this method, a hierarchical Bayesian model is defined, in which a hidden variable is introduced. The posterior distribution in the Bayesian method is obtained from the hierarchical structured prior model. From the posterior distribution, the unknown variables and parameters are estimated via MAP algorithm. Compared with the regularization approaching methods, the proposed Bayesian method estimates the parameters and variables simultaneously. The hyperparameters which defines the prior distribution of the parameters are fixed during initialization. A proposition of the initialization for the hyperparameters is presented in this chapter, and the simulation results proved a relative sensitiveness of the hyperparameters' values. Simulations with both limited number of projections and limited angle of projections are done, and it is proved that the proposed Bayesian method

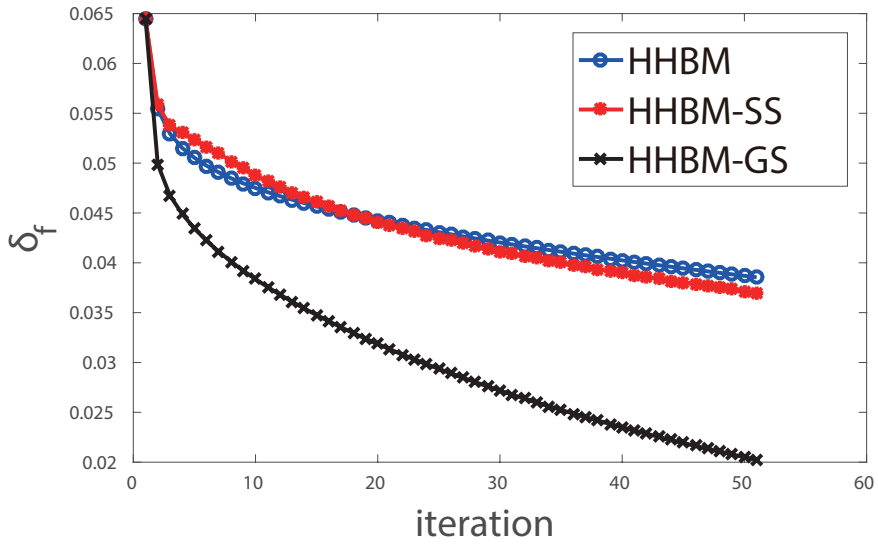


FIGURE 5.33 – Comparison of the HHBM method and the variational splitting methods in terms of the evolution of the RMSE during reconstruction from 128 projections and SNR=40dB.

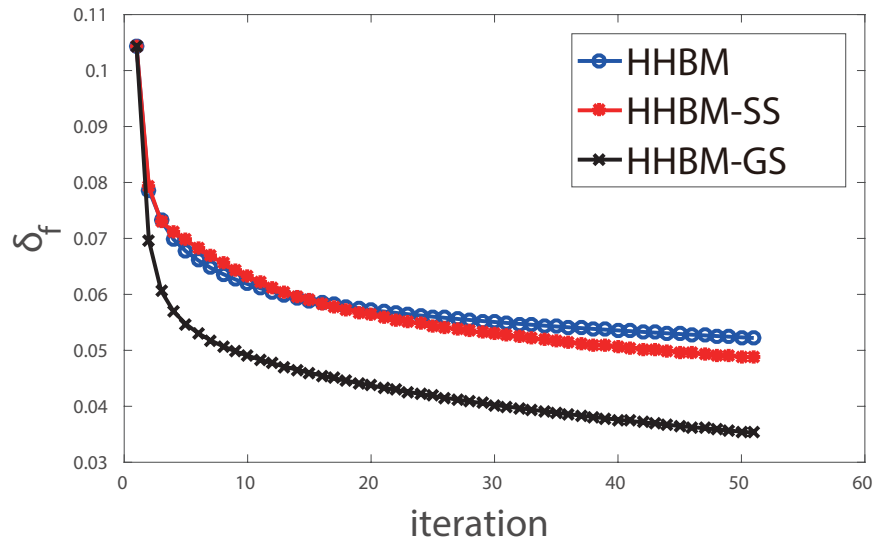


FIGURE 5.34 – Comparison of the HHBM method and the variational splitting methods in terms of the evolution of the RMSE during reconstruction from 64 projections and SNR=40dB.

is more robust than the conventional quadratic regularization method or total variation method when the number or angle of projection is limited.

We also compared the Dual-Tree complex wavelet transformation (DT-CWT) with the Haar transformation. From the comparison, we observed that the DT-CWT transformation is more robust to the contamination than the Haar transformaion. Thanks to that, the Bayesian method BH-DTCWT based on this wavelet transformation outperforms the HHBM method using Haar wavelet transformation. The short term future work devote to find some other transformation which is computationally faster and considers more wavelet directions instead of only two directions in Haar transformation. The DT-CWT could be implemented in GPU for 3D simulations in the future in order to accelerate the computation.

Based on the proposed HHBM method, a noise splitting method is proposed and presented in this chapter. The splitting of noise considers the noise in two parts : a Gaussian noise and a $\mathcal{S}t_g$ noise. By considering these two noise, the reconstruction surpasses the HHBM in terms of convergence speed and the quality of reconstruction.

6

Simultaneous Object-Contour Reconstruction Model

6.1 Introduction

In the previous chapter, a hierarchical Bayesian method which uses a heavy tailed generalized Student-t distribution to enforce the sparse structure of the discrete wavelet transformation coefficient is used. As it was pointed out, the advantage of the wavelet transformation comparing with the gradient operator, or the Laplacian operator, is that, while the inversion of gradient operator is inconvenient, the wavelet transformation operator is reversible and orthogonal.

However, the most accurate transformation to obtain a sparse structured coefficient of a piecewise continuous image is the gradient operator or the Laplacian operator, because in the wavelet transformation coefficient, the part corresponding to the low-pass filtered coefficients is not sparse.

In this chapter, we present a Bayesian method which combines the reconstruction of the object and the reconstruction of the Laplacian operator of object. For a piecewise continuous object, the Laplacian is sparse. The $\mathcal{S}t_g$ distribution is used as the prior model for the sparse variable. The prior distribution for the variable $\boldsymbol{f}$ is a non-homogeneous Markovian model. In this non-homogeneous Markovian model, the distribution of each voxel f_i depends not only on its six neighbor voxels, but also the Laplacian value of the neighbor voxels. By taking into consideration the Laplacian values, we can distinguish whether the element is in a homogeneous area or on a contour, and the weight of impact of this neighbor voxel on the current voxel will depend on this Laplacian value.

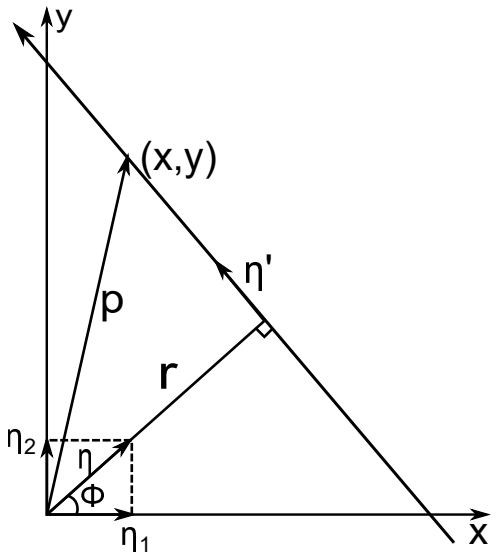


FIGURE 6.1 – The Radon Transform with the direction of radiation defined by vector $\boldsymbol{\eta}$.

6.2 Forward model : Radon Transform and notations

As presented in Chapter 3, the Radon transform is analytically expressed as :

$$g(r,\phi) = \mathcal{R}f(x,y) = \int_{L_{r,\phi}} f(x,y)\mathrm{d}l, \tag{6.1}$$

where $\mathcal{R}$ represents the operator which transforms f to Radon space, $f(x,y)$ represents the image, g the detected projection data, r the perpendicular length from center point of coordinate and ϕ the considered X ray angle. $L_{r,\phi}$ is the length of ray (r,ϕ) passing through the image.

We can also use a vector to replace ϕ to define the direction of the radiation. As shown in Figure 6.1, the direction of the radiation can be defined by the unit vector $\boldsymbol{\eta}$, and r is always the vertical distance from the origin to the radiation. The RT can be represented by the following expression :

$$g(r,\boldsymbol{\eta}) = \int_{-\infty}^{+\infty} f(x,y)\delta\left(r - \mathbf{p}\boldsymbol{\eta}\right)\,\mathrm{d}t \tag{6.2}$$

where $p\boldsymbol{\eta} = x\boldsymbol{\eta}_1 + y\boldsymbol{\eta}_2 = x\cos\phi + y\sin\phi$, so that :

$$g(r,\boldsymbol{\eta}) \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\delta(r - x\boldsymbol{\eta}_1 - y\boldsymbol{\eta}_2)\,\mathrm{d}x\,\mathrm{d}y \tag{6.3}$$

6.3 Basic properties of Radon transformation

Lemma 6.3.1. The shifting property :

$$\mathcal{R}f(x-a,y-b) = g(r - a\boldsymbol{\eta}_1 - b\boldsymbol{\eta}_2, \boldsymbol{\eta}) \tag{6.4}$$

Demonstration. By using the definition given in Eq.(6.3), we have :

$$\mathcal{R}f(x-a,y-b) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x-a,y-b)\delta\left(r - x\boldsymbol{\eta}_1 - y\boldsymbol{\eta}_2\right)\,\mathrm{d}x\,\mathrm{d}y \tag{6.5}$$

By changing the variable : $x' = x - a$ and $y' = y - b$, we get :

$$\begin{aligned}\mathcal{R}f(x', y') = &\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x', y')\delta\left((r - a\boldsymbol{\eta}_1 - b\boldsymbol{\eta}_2) - x'\boldsymbol{\eta}_1 - y'\boldsymbol{\eta}_2\right)\,\mathrm{d}x'\,\mathrm{d}y' \\ =&g\left(r - a\boldsymbol{\eta}_1 - b\boldsymbol{\eta}_2, \boldsymbol{\eta}\right)\end{aligned} \tag{6.6}$$

Hence, the following properties are derived :

$$\begin{aligned}\mathcal{R}f(x - a, y) &= g(r - a\boldsymbol{\eta}_1, \boldsymbol{\eta}) \\ \mathcal{R}f(x, y - b) &= g(r - b\boldsymbol{\eta}_2, \boldsymbol{\eta})\end{aligned} \tag{6.7}$$

□

Lemma 6.3.2. Derivation :

$$\mathcal{R}\left[\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right] = \frac{\partial^2 g}{\partial r^2} \tag{6.8}$$

Demonstration. The definition of the second order derivative is :

$$\begin{aligned}\frac{\partial^2 f(x, y)}{\partial x^2} &= \lim_{\varepsilon\to 0}\frac{f(x + \varepsilon, y) + f(x - \varepsilon, y) - 2f(x, y)}{\varepsilon^2} \\ \frac{\partial^2 f(x, y)}{\partial y^2} &= \lim_{\varepsilon\to 0}\frac{f(x, y + \varepsilon) + f(x, y - \varepsilon) - 2f(x, y)}{\varepsilon^2}\end{aligned} \tag{6.9}$$

The Radon transform of the derivatives are respectively :

$$\begin{aligned}\mathcal{R}\frac{\partial^2 f(x, y)}{\partial x^2} &= \lim_{\varepsilon\to 0}\frac{\mathcal{R}f(x + \varepsilon, y) + \mathcal{R}f(x - \varepsilon, y) - 2\mathcal{R}f(x, y)}{\varepsilon^2} \\ \mathcal{R}\frac{\partial^2 f(x, y)}{\partial y^2} &= \lim_{\varepsilon\to 0}\frac{\mathcal{R}f(x, y + \varepsilon) + \mathcal{R}f(x, y - \varepsilon) - 2\mathcal{R}f(x, y)}{\varepsilon^2}\end{aligned} \tag{6.10}$$

According to the shifting property, Lemma 6.3.1, we get :

$$\begin{aligned}\mathcal{R}\frac{\partial^2 f(x, y)}{\partial x^2} =&\lim_{\varepsilon\to 0}\frac{g\left(r + \varepsilon\boldsymbol{\eta}_1, \boldsymbol{\eta}\right) + g\left(r - \varepsilon\boldsymbol{\eta}_1, \boldsymbol{\eta}\right) - 2g(r, \boldsymbol{\eta})}{\varepsilon^2} \\ =&\boldsymbol{\eta}_1^2\lim_{(\boldsymbol{\eta}_1\varepsilon)\to 0}\frac{g\left(r + \varepsilon\boldsymbol{\eta}_1, \boldsymbol{\eta}\right) + g\left(r - \varepsilon\boldsymbol{\eta}_1, \boldsymbol{\eta}\right) - 2g(r, \boldsymbol{\eta})}{(\varepsilon\boldsymbol{\eta}_1)^2} \\ =&\boldsymbol{\eta}_1^2\frac{\partial^2 g(r, \boldsymbol{\eta})}{\partial r^2}\end{aligned} \tag{6.11}$$

$$\begin{aligned}\mathcal{R}\frac{\partial^2 f(x, y)}{\partial y^2} =&\lim_{\varepsilon\to 0}\frac{g\left(r + \varepsilon\boldsymbol{\eta}_2, \boldsymbol{\eta}\right) + g\left(r - \varepsilon\boldsymbol{\eta}_2, \boldsymbol{\eta}\right) - 2g(r, \boldsymbol{\eta})}{\varepsilon^2} \\ =&\boldsymbol{\eta}_1^2\lim_{(\boldsymbol{\eta}_2\varepsilon)\to 0}\frac{g\left(r + \varepsilon\boldsymbol{\eta}_2, \boldsymbol{\eta}\right) + g\left(r - \varepsilon\boldsymbol{\eta}_2, \boldsymbol{\eta}\right) - 2g(r, \boldsymbol{\eta})}{(\varepsilon\boldsymbol{\eta}_2)^2} \\ =&\boldsymbol{\eta}_2^2\frac{\partial^2 g(r, \boldsymbol{\eta})}{\partial r^2}\end{aligned} \tag{6.12}$$

As $\boldsymbol{\eta}_1^2 + \boldsymbol{\eta}_2^2 = 1$, we obtain :

$$\begin{aligned}\mathcal{R}\left[\frac{\partial^2 f(x, y)}{\partial x^2} + \frac{\partial^2 f(x, y)}{\partial y^2}\right] &= \left(\boldsymbol{\eta}_1^2 + \boldsymbol{\eta}_2^2\right)\frac{\partial^2 g(r, \boldsymbol{\eta})}{\partial r^2} \\ &= \frac{\partial^2 g(r, \boldsymbol{\eta})}{\partial r^2}\end{aligned} \tag{6.13}$$

□

According to the previous demonstrated property, we obtain the following relationships between the objection $\boldsymbol{f}$ and projection $\boldsymbol{g}$ and between the Laplacian of object $\ddot{\boldsymbol{f}}$ and the Laplacian of projection $\ddot{\boldsymbol{g}}$:

$$f(x,y) \rightarrow \mathrm{RT} \rightarrow g(r,\phi), \tag{6.14}$$

$$\ddot{f}(x,y) \rightarrow \mathrm{RT} \rightarrow \ddot{g}(r,\phi), \tag{6.15}$$

where

$$\begin{aligned} \ddot{f}(x,y) &= \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) f(x,y), \\ \ddot{g}(r,\phi) &= \frac{\partial^2}{\partial r^2} g(r,\phi). \end{aligned} \tag{6.16}$$

The same property can also be demonstrated in 3D projection.

Figure 6.2 shows, from left to right and top to bottom, the Shepp Logan phantom of size 256^2, the projection of Shepp Logan phantom with 256 projections, the Laplacian of the Shepp Logan phantom and the Laplacian of the projection.

In the statistical methods, the discretized data are considered. In the discretized Radon transform system, the transformation operation is represented via the multiplier of a matrix $\boldsymbol{H}$. So that :

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}. \tag{6.17}$$

where an additive noise is presented by vector $\boldsymbol{\epsilon}$, representing the error of detection and uncertainties of the projection system.

Meanwhile, the Radon transform from Laplacian of object to the derivation of projection, shown in Eq.(6.15), is expressed as :

$$\ddot{\boldsymbol{g}} = \boldsymbol{H}\ddot{\boldsymbol{f}} + \boldsymbol{\tau} \tag{6.18}$$

where the additive error represents the uncertainties of the transform system and other errors of the Laplacian operation.

Noted by $\ddot{\boldsymbol{f}}$ the discretized $\ddot{f}(x,y)$ and by $\ddot{\boldsymbol{g}}$ discretized $\ddot{g}(r,\phi)$, they are approximated by convolution of the corresponding variable with kernels :

$$\ddot{\boldsymbol{g}} = [-1 \;\; 2 \;\; -1] * \boldsymbol{g} = \boldsymbol{d}_1 * \boldsymbol{g} \tag{6.19}$$

$$\ddot{\boldsymbol{f}} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} * \boldsymbol{f} = \boldsymbol{d}_2 * \boldsymbol{f} \tag{6.20}$$

As the convolution $\boldsymbol{d}_1 * \boldsymbol{g}$ and $\boldsymbol{d}_2 * \boldsymbol{f}$ are linear operations, they are expressed as a linear multiplier with the corresponding matrix $\boldsymbol{D}_1 \in \mathbb{R}^{M \times M}$ and $\boldsymbol{D}_2 \in \mathbb{R}^{N \times N}$:

$$\begin{aligned} \ddot{\boldsymbol{g}} &= \boldsymbol{D}_1 \boldsymbol{g}, \\ \ddot{\boldsymbol{f}} &= \boldsymbol{D}_2 \boldsymbol{f}. \end{aligned} \tag{6.21}$$

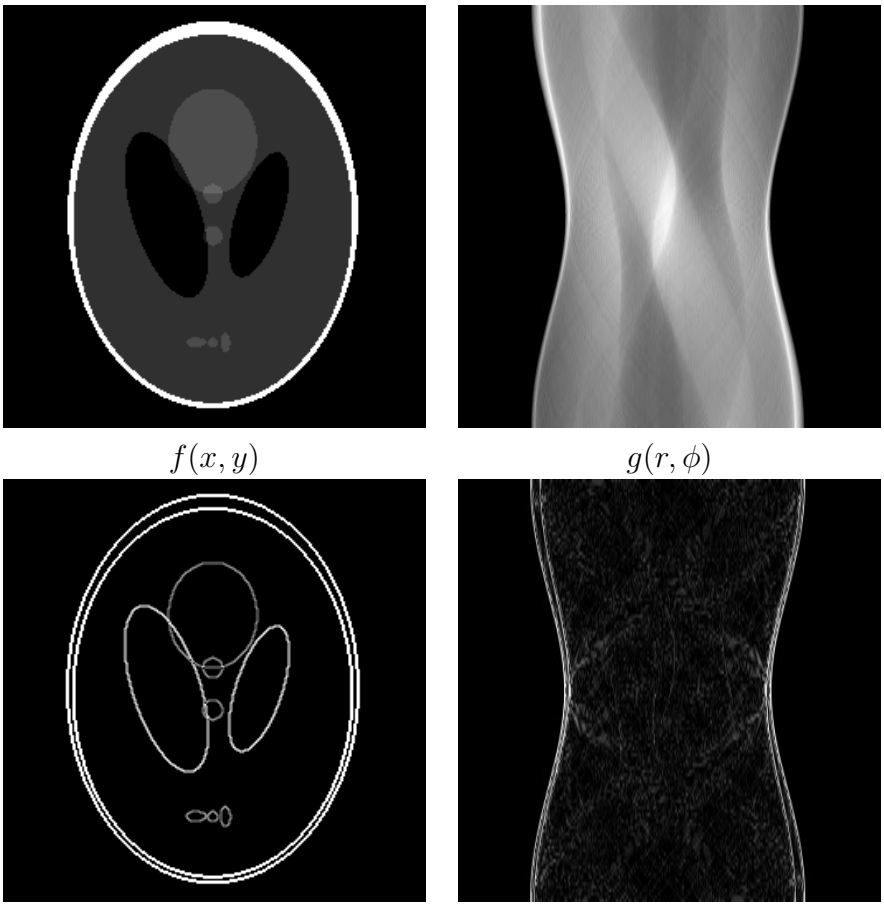


FIGURE 6.2 – The original Shepp Logan figure of size 256^2 (top-left) $f(x,y)$, its Laplacian (bottom-left) $\ddot{f}(x,y)$, the projection sinogram (top-right) $g(r,\phi)$ and the Laplacian of sinogram (bottom-right) $\ddot{g}(r,\phi)$.

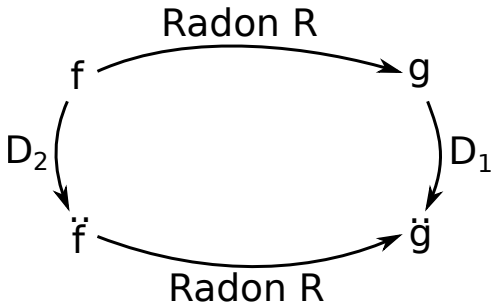


FIGURE 6.3 – The relations between the original figure, the projection, the Laplacian of the object and the Laplacian of the projection.

And the relations between the object, the projection, the Laplacian of object and the Laplacian of projections are shown in Figure 6.3.

In the conventional image reconstruction methods, for example the TV method, the gradient of phantom $\boldsymbol{f}$ is used as a criterion to preserve the edges of image. In the proposed Reconstruction of Object and Contour Combined method (ROCC), the Laplacian of the object $\ddot{\boldsymbol{f}}$ is considered as an unknown variable and is estimated during the reconstruction of the object.

6.4 Proposed Method of Reconstruction of Object while Considering Contours (ROCC)

In this proposed ROCC method, we propose to use the forward relations in Eq.(6.21) to simultaneously reconstruct the image $\boldsymbol{f}$ and detect the contours $\ddot{\boldsymbol{f}}$ by using a Bayesian approach. A hierarchical inhomogeneous Markovian model is defined for the object $\boldsymbol{f}$. A hidden variables $\boldsymbol{q}$, which is the normalized Laplacian, $\left(\left|\ddot{\boldsymbol{f}}\right| - \min(\left|\ddot{\boldsymbol{f}}\right|)\right) / \left(\max(\left|\ddot{\boldsymbol{f}}\right|) - \min(\left|\ddot{\boldsymbol{f}}\right|)\right)$, is used as a parameter of this prior model. A sparsity enforcing Generalized Student-t prior distribution is used to enforce the sparse structure of $\ddot{\boldsymbol{f}}$.

6.4.1 Noise model

Starting from the forward model in Eq.(6.17) and Eq.(6.18), by assuming that the additive noise $\boldsymbol{\epsilon}$ and $\boldsymbol{\tau}$ are iid Gaussian distributed with variance v_ϵ and v_τ, we obtain the likelihood :

$$\begin{aligned} p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{v}_\epsilon) &= \mathcal{N}(\boldsymbol{g}|\boldsymbol{H}\boldsymbol{f}, \boldsymbol{V}_\epsilon), \\ p(\ddot{\boldsymbol{g}}|\ddot{\boldsymbol{f}}, \boldsymbol{v}_\tau) &= \mathcal{N}(\ddot{\boldsymbol{g}}|\boldsymbol{H}\ddot{\boldsymbol{f}}, \boldsymbol{V}_\tau), \end{aligned} \tag{6.22}$$

where $\boldsymbol{V}_\epsilon = \text{diag}\,[\boldsymbol{v}_\epsilon]$ and $\boldsymbol{V}_\tau = \text{diag}\,[\boldsymbol{v}_\tau]$ are diagonal matrix. Considering that the variances v_{ϵ_i} and v_{τ_i} are positive and are not very big, they are defined by Inverse Gamma distributions :

$$\begin{aligned} p(\boldsymbol{v}_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) &= \mathcal{IG}(\boldsymbol{v}_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}), \\ p(\boldsymbol{v}_\tau|\alpha_{\tau_0}, \beta_{\tau_0}) &= \mathcal{IG}(\boldsymbol{v}_\tau|\alpha_{\tau_0}, \beta_{\tau_0}), \end{aligned} \tag{6.23}$$

where α_{ϵ_0}, β_{ϵ_0}, α_{τ_0} and β_{τ_0} are hyper parameters.

Another convenience of choosing the Inverse Gamma distribution for the variance is that, when the likelihood is Gaussian distributed, they are therefore conjugate distributions, and the form of the posterior distribution will still be Gaussian and Inverse Gamma distribution.

6.4.2 Sparsity enforcing Student-t prior model for $\ddot{\boldsymbol{f}}$

As we have shown in Chapter 4, in X-ray CT, the object under consideration is commonly piecewise continuous. The Laplacian of this type of object is sparse, as shown in Figure 6.2, with most of the part approaching to zero except the contours.

As $\ddot{\boldsymbol{f}}$ is sparse, we use a sparsity enforcing prior to define it. As presented in the previous chapter, the generalized Student-t distribution is a suitable distribution and it has many appro-

priate properties in the Bayesian method. By considering its Normal-Inverse Gamma marginalization form, we model the hidden variable $\ddot{\boldsymbol{f}}$ by :

$$\begin{aligned} p(\ddot{\boldsymbol{f}}|\boldsymbol{v}_c) &= \mathcal{N}(\ddot{\boldsymbol{f}}|0, \boldsymbol{V}_c), \\ p(\boldsymbol{v}_c|\alpha_{c_0}, \beta_{c_0}) &= \mathcal{IG}(\boldsymbol{v}_c|\alpha_{c_0}, \beta_{c_0}), \end{aligned} \tag{6.24}$$

where $\boldsymbol{V}_c = \text{diag}\,[\boldsymbol{v}_c]$, and α_{c_0} and β_{c_0} are two hyper-parameters of generalized Student-t distribution.

6.4.3 The prior model for object

In the ROCC method, $\boldsymbol{f}$ is defined by an inhomogeneous Markovian distribution basing on the Laplacian $\ddot{\boldsymbol{f}}$. In this Markovian model, the neighbor pixels are considered. For 2D image, four neighbor pixels are considered, and for 3D object, six neighbor voxels are considered, as shown in Figure 6.4.

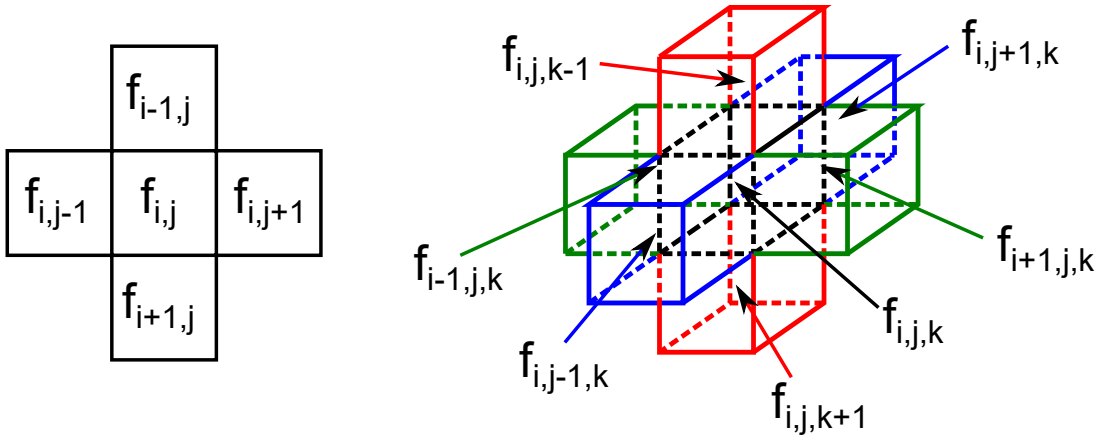


FIGURE 6.4 – The neighbour pixels of current pixel in 2D image (left) and the neighbour voxels of current voxel in 3D object (right).

In the conventional homogeneous Markovian model, every neighbor pixel will have the same weight on the current pixel. It works well for the pixels in the homogeneous areas. As shown in Figure 6.5, when the pixel is around the contours between different homogeneous areas, the neighbor pixels should have different weight of influence on the current pixel. Here we use f_r to represent the current pixel value, and $f_{r'}$ to represent its neighbor pixels values. When $f_{r'}$ is not on the contour, it means that $f_{r'}$ is in the same homogeneous zone as f_r, and that they are approach to each other. When $f_{r'}$ is on the contour, on the other hand, $f_{r'}$ should not have much influence on f_r.

By considering these relations, we propose to use a normalized Laplacian value $\boldsymbol{q}$:

$$q_r = \frac{\left|\ddot{\boldsymbol{f}}_r\right| - \min\left(\left|\ddot{\boldsymbol{f}}_r\right|\right)}{\max\left(\left|\ddot{\boldsymbol{f}}_r\right|\right) - \min\left(\left|\ddot{\boldsymbol{f}}_r\right|\right)}. \tag{6.25}$$

So $q_{r'}$ is between 0 and 1, $\forall r' \in \boldsymbol{N}r(r)$ and $\forall r \in [1, N]$, where $\boldsymbol{N}r(r)$ means the neighbor pixels of r. When $q_{r'} = 0$, it corresponds to a minimum Laplacian value, and it means that the neighbor pixels $f_{r'}$ is in the same homogeneous areas as f_r and f_r should depend on the value

$f_{r'}$. When $q_{r'} = 1$, it corresponds to a maximum Laplacian value, representing that the neighbor pixel $f_{r'}$ is on the contour and should not have influence on f_r. Consequently, the influence of the neighbors pixels in the Markovian model is inversely proportional to the value of $q_{r'}$.

We define the variable $\boldsymbol{f}$ by the following Gaussian distribution :

$$p(\boldsymbol{f}|\boldsymbol{m}_f, \boldsymbol{v}_f) = \mathcal{N}(\boldsymbol{f}|\boldsymbol{m}_f, \boldsymbol{V}_f) \tag{6.26}$$

where the variance $\boldsymbol{v}_f$ are supposed belonging to an Inverse Gamma distribution, in order to realize a semi-supervised model. The mean of this Gaussian distribution, $\boldsymbol{m}_f$, depends on the neighbors of each pixel. The definition of the mean of the pixel r is :

$$m_r = \frac{\sum_{r' \in \boldsymbol{N}_{r(r)}} (1 - q_{r'}) f_{r'}}{\sum_{r' \in \boldsymbol{N}_{r(r)}} (1 - q_{r'})}. \tag{6.27}$$

For each neighbor pixel, r', the pixel value $f_{r'}$ is considered with a weight $(1 - q_{r'})$. $q_{r'}$ is the normalized Laplacian value of the neighbor pixel r', defined in Eq.(6.25). Term $\frac{1}{\sum_{r' \in \boldsymbol{N}_{r(r)}} (1 - q_{r'})}$ is added for the normalization.

Therefore, the expression of the prior distribution for each pixel f_r is :

$$\begin{aligned} p(f_r|\boldsymbol{f}_{-r}, \boldsymbol{q}, v_{f_r}) =& \mathcal{N}(f_r|m_r, v_{f_r}) \\ \propto& v_{f_r}^{-1} \exp\left\{ -\frac{1}{2} v_{f_r}^{-1} \left(f_r - \frac{\sum_{r' \in \boldsymbol{N}_{r(r)}} (1 - q_{r'}) f_{r'}}{\sum_{r' \in \boldsymbol{N}_{r(r)}} (1 - q_{r'})} \right)^2 \right\}, \end{aligned} \tag{6.28}$$

where $\boldsymbol{f}_{-r}$ represents all the elements in vector $\boldsymbol{f}$ except f_r.

We show an example in Figure 6.5 as an explanation of this prior model.

In this example, f_r is the value of the current pixel, q_r is the normalized Laplacian of pixel r. r'_1, r'_2, r'_3 and r'_4 are the four neighbor pixel of r. Among them, r'_1 and r'_4 are on the contour with normalized Laplacian values $q_{r'}$ equal to 1, while r'_2 and r'_3 are in the homogeneous areas with $q_{r'}$ equal to 0. Thus, $q_{r'_1} = q_{r'_4} = 1$ and $q_{r'_2} = q_{r'_3} = 0$. As defined in Eq.(6.27), the mean of prior distribution for f_r is :

$$\begin{aligned} m_r =& \frac{\left(1 - q_{r'_1}\right) f_{r'_1} + \left(1 - q_{r'_2}\right) f_{r'_2} + \left(1 - q_{r'_3}\right) f_{r'_3} + \left(1 - q_{r'_4}\right) f_{r'_4}}{\left(1 - q_{r'_1}\right) + \left(1 - q_{r'_2}\right) + \left(1 - q_{r'_3}\right) + \left(1 - q_{r'_4}\right)} \\ =& \frac{f_{r'_2} + f_{r'_3}}{2}. \end{aligned} \tag{6.29}$$

So that f_r belong to a Gaussian distribution :

$$p(f_r) = \mathcal{N}\left(f_r | \frac{f_{r'_2} + f_{r'_3}}{2}, v_{f_r} \right). \tag{6.30}$$

In a more complicated case, the value q_r is distributed in $[0, 1]$. Consequently, a neighbor pixel with a bigger value of $q_{r'}$ will have less influence on the current pixel.

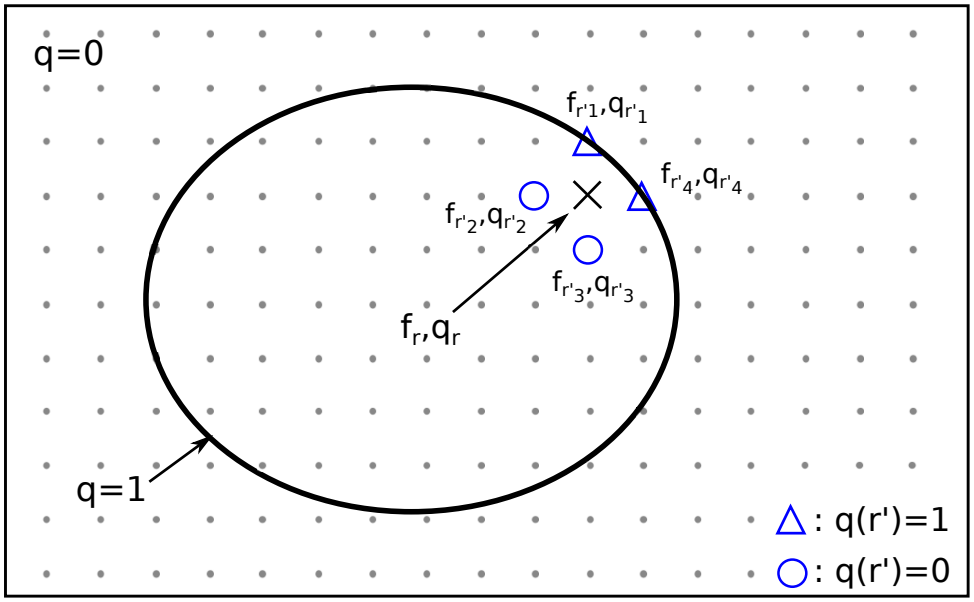


FIGURE 6.5 – Example of the inhomogeneous Markovian prior model of $\boldsymbol{f}$

6.5 Semi-supervised ROCC method

In most of the cases, initializing the variance of phantom $\boldsymbol{v}_f$ manually demand some prior information, which is typically not accessible in real applications. A semi-supervised model is considered in our work, in which $\boldsymbol{v}_f$ is an unknown vector and the Inverse Gamma distribution is defined for it :

$$p(\boldsymbol{v}_f|\alpha_{f_0},\beta_{f_0}) = \mathcal{IG}(\boldsymbol{v}_f|\alpha_{f_0},\beta_{f_0}) \propto \prod_{j=1}^{N} v_{f_j}^{-(\alpha_{f_0}+1)} \exp\left\{-\beta_{f_0} v_{f_j}^{-1}\right\}. \tag{6.31}$$

6.5.1 The system model

Basing on the previously presented prior model for the object $\boldsymbol{f}$, and by considering $\boldsymbol{v}_f$ and an unknown variable, belonging to an Inverse Gamma distribution, we get the following system

model :

$$
\begin{aligned}
&p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{v}_\epsilon)=\mathcal{N}(\boldsymbol{g}|\boldsymbol{H}\boldsymbol{f},\boldsymbol{V}_\epsilon)\propto|\boldsymbol{V}_\epsilon|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\right)^T\boldsymbol{V}_\epsilon^{-1}\left(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\right)\right\},\\
&p(\ddot{\boldsymbol{g}}|\ddot{\boldsymbol{f}},\boldsymbol{v}_\tau)=\mathcal{N}(\ddot{\boldsymbol{g}}|\boldsymbol{H}\ddot{\boldsymbol{f}},\boldsymbol{V}_\tau)\propto|\boldsymbol{V}_\tau|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left(\ddot{\boldsymbol{g}}-\boldsymbol{H}\ddot{\boldsymbol{f}}\right)^T\boldsymbol{V}_\tau^{-1}\left(\ddot{\boldsymbol{g}}-\boldsymbol{H}\ddot{\boldsymbol{f}}\right)\right\},\\
&p(\boldsymbol{f}|\boldsymbol{m}_f,\boldsymbol{v}_f)=\boldsymbol{N}(\boldsymbol{f}|\boldsymbol{m}_f,\boldsymbol{V}_f)\propto|\boldsymbol{V}_f|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left(\boldsymbol{f}-\boldsymbol{m}_f\right)^T\boldsymbol{V}_f^{-1}\left(\boldsymbol{f}-\boldsymbol{m}_f\right)\right\},\\
&p(\ddot{\boldsymbol{f}}|\boldsymbol{v}_c)=\mathcal{N}(\ddot{\boldsymbol{f}}|\boldsymbol{0},\boldsymbol{V}_c)\propto|\boldsymbol{V}_c|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\ddot{\boldsymbol{f}}^T\boldsymbol{V}_c^{-1}\ddot{\boldsymbol{f}}\right\},\\
&p(\boldsymbol{v}_\epsilon|\alpha_{\epsilon_0},\beta_{\epsilon_0})=\mathcal{IG}(\boldsymbol{v}_\epsilon|\alpha_{\epsilon_0},\beta_{\epsilon_0})\propto\prod_{i=1}^{M}v_{\epsilon_i}^{-(\alpha_{\epsilon_0}+1)}\prod_{i=1}^{M}\exp\left\{-\beta_{\epsilon_0}v_{\epsilon_i}^{-1}\right\},\\
&p(\boldsymbol{v}_\tau|\alpha_{\tau_0},\beta_{\tau_0})=\mathcal{IG}(\boldsymbol{v}_\tau|\alpha_{\tau_0},\beta_{\tau_0})\propto\prod_{i=1}^{M}v_{\tau_i}^{-(\alpha_{\tau_0}+1)}\prod_{i=1}^{M}\exp\left\{-\beta_{\tau_0}v_{\tau_i}^{-1}\right\},\\
&p(\boldsymbol{v}_f|\alpha_{f_0},\beta_{f_0})=\mathcal{IG}(\boldsymbol{v}_f|\alpha_{f_0},\beta_{f_0})\propto\prod_{j=1}^{N}v_{f_j}^{-(\alpha_{f_0}+1)}\prod_{j=1}^{N}\exp\left\{-\beta_{f_0}v_{f_j}^{-1}\right\},\\
&p(\boldsymbol{v}_c|\alpha_{c_0},\beta_{c_0})=\mathcal{IG}(\boldsymbol{v}_c|\alpha_{c_0},\beta_{c_0})\propto\prod_{j=1}^{N}v_{c_j}^{-(\alpha_{c_0}+1)}\prod_{j=1}^{N}\exp\left\{-\beta_{c_0}v_{c_j}^{-1}\right\},
\end{aligned}
\tag{6.32}
$$

where

$$
\boldsymbol{m}_f=\frac{\sum_{r'\in N_r}(1-q_{r'})f_{r'}}{\sum_{r'\in N_r}(1-q_{r'})},
\tag{6.33}
$$

and

$$
\boldsymbol{q}=\frac{\left|\ddot{\boldsymbol{f}}\right|-\mathbf{min}\left(\left|\ddot{\boldsymbol{f}}\right|\right)}{\mathbf{max}\left(\left|\ddot{\boldsymbol{f}}\right|\right)-\mathbf{min}\left(\left|\ddot{\boldsymbol{f}}\right|\right)}.
\tag{6.34}
$$

According to the Bayes rule, the posterior distribution is :

$$
\begin{aligned}
&p(\boldsymbol{f},\ddot{\boldsymbol{f}},\boldsymbol{v}_\epsilon,\boldsymbol{v}_\tau,\boldsymbol{v}_f,\boldsymbol{v}_c|\boldsymbol{g},\ddot{\boldsymbol{g}})\\
\propto&p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{v}_\epsilon)p(\ddot{\boldsymbol{g}}|\ddot{\boldsymbol{f}},\boldsymbol{v}_\tau)p(\boldsymbol{f}|\boldsymbol{m}_f,\boldsymbol{v}_f)p(\ddot{\boldsymbol{f}}|\boldsymbol{v}_c)p(\boldsymbol{v}_\epsilon|\alpha_{\epsilon_0},\beta_{\epsilon_0})\\
&p(\boldsymbol{v}_\tau|\alpha_{\tau_0},\beta_{\tau_0})p(\boldsymbol{v}_f|\alpha_{f_0},\beta_{f_0})p(\boldsymbol{v}_c|\alpha_{c_0},\beta_{c_0})\\
=&|\boldsymbol{V}_\epsilon|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\right)^T\boldsymbol{V}_\epsilon^{-1}\left(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\right)\right\}\\
&|\boldsymbol{V}_\tau|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left(\ddot{\boldsymbol{g}}-\boldsymbol{H}\ddot{\boldsymbol{f}}\right)^T\boldsymbol{V}_\tau^{-1}\left(\ddot{\boldsymbol{g}}-\boldsymbol{H}\ddot{\boldsymbol{f}}\right)\right\}\\
&|\boldsymbol{V}_f|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left(\boldsymbol{f}-\boldsymbol{m}_f\right)^T\boldsymbol{V}_f^{-1}\left(\boldsymbol{f}-\boldsymbol{m}_f\right)\right\}|\boldsymbol{V}_c|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\ddot{\boldsymbol{f}}^T\boldsymbol{V}_c^{-1}\ddot{\boldsymbol{f}}\right\}\\
&\prod_{i=1}^{M}v_{\epsilon_i}^{-(\alpha_{\epsilon_0}+1)}\prod_{i=1}^{M}\exp\left\{-\beta_{\epsilon_0}v_{\epsilon_i}^{-1}\right\}\prod_{i=1}^{M}v_{\tau_i}^{-(\alpha_{\tau_0}+1)}\prod_{i=1}^{M}\exp\left\{-\beta_{\tau_0}v_{\tau_i}^{-1}\right\}\\
&\prod_{j=1}^{N}v_{f_j}^{-(\alpha_{f_0}+1)}\prod_{j=1}^{N}\exp\left\{-\beta_{f_0}v_{f_j}^{-1}\right\}\prod_{j=1}^{N}v_{c_j}^{-(\alpha_{c_0}+1)}\prod_{j=1}^{N}\exp\left\{-\beta_{c_0}v_{c_j}^{-1}\right\}.
\end{aligned}
\tag{6.35}
$$

6.5.2 JMAP Estimation

By using the JMAP estimation, the unknown variables are estimated iteratively. The maximization of the posterior distribution is transferred into the minimization of the negative logarithm of the posterior distribution, considering the exponential form of the distribution :

$$\begin{aligned}\left(\widehat{\boldsymbol{f}}, \widehat{\ddot{\boldsymbol{f}}}, \widehat{\boldsymbol{v}}_\epsilon, \widehat{\boldsymbol{v}}_\tau, \widehat{\boldsymbol{v}}_f, \widehat{\boldsymbol{v}}_c\right) &= \arg \max_{\boldsymbol{f}, \ddot{\boldsymbol{f}}, \boldsymbol{v}_\epsilon, \boldsymbol{v}_\tau, \boldsymbol{v}_f, \boldsymbol{v}_c} \left\{ p(\boldsymbol{f}, \ddot{\boldsymbol{f}}, \boldsymbol{v}_\epsilon, \boldsymbol{v}_\tau, \boldsymbol{v}_f, \boldsymbol{v}_c | \boldsymbol{g}, \ddot{\boldsymbol{g}}) \right\} \\ &= \arg \min_{\boldsymbol{f}, \ddot{\boldsymbol{f}}, \boldsymbol{v}_\epsilon, \boldsymbol{v}_\tau, \boldsymbol{v}_f, \boldsymbol{v}_c} \left\{ J(\boldsymbol{f}, \ddot{\boldsymbol{f}}, \boldsymbol{v}_\epsilon, \boldsymbol{v}_\tau, \boldsymbol{v}_f, \boldsymbol{v}_c) \right\},\end{aligned} \tag{6.36}$$

where

$$\begin{aligned}&J(\boldsymbol{f}, \ddot{\boldsymbol{f}}, \boldsymbol{v}_\epsilon, \boldsymbol{v}_\tau, \boldsymbol{v}_f, \boldsymbol{v}_c) \\ =&\frac{1}{2}\sum_{i=1}^{M} \ln v_{\epsilon_i} + \frac{1}{2}(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f})^T \boldsymbol{V}_\epsilon^{-1} (\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}) \\ &+ \frac{1}{2}\sum_{i=1}^{M} \ln v_{\tau_i} + \frac{1}{2}\left(\ddot{\boldsymbol{g}} - \boldsymbol{H}\ddot{\boldsymbol{f}}\right)^T \boldsymbol{V}_\tau^{-1} \left(\ddot{\boldsymbol{g}} - \boldsymbol{H}\ddot{\boldsymbol{f}}\right) \\ &+ \frac{1}{2}\sum_{j=1}^{N} \ln v_{f_j} + \frac{1}{2}(\boldsymbol{f} - \boldsymbol{m}_f)^T \boldsymbol{V}_f^{-1} (\boldsymbol{f} - \boldsymbol{m}_f) \\ &+ \frac{1}{2}\sum_{j=1}^{N} \ln v_{c_j} + \frac{1}{2}\ddot{\boldsymbol{f}}^T \boldsymbol{V}_c^{-1} \ddot{\boldsymbol{f}} + (\alpha_{\epsilon_0}+1)\sum_{i=1}^{M} \ln v_{\epsilon_i} + \beta_{\epsilon_0}\sum_{i=1}^{M} v_{\epsilon_i}^{-1} + (\alpha_{\tau_0}+1)\sum_{i=1}^{M} \ln v_{\tau_i} + \beta_{\tau_0}\sum_{i=1}^{M} v_{\tau_i}^{-1} \\ &+ (\alpha_{f_0}+1)\sum_{j=1}^{N} \ln v_{f_j} + \beta_{f_0}\sum_{j=1}^{N} v_{f_j}^{-1} + (\alpha_{c_0}+1)\sum_{j=1}^{N} \ln v_{c_j} + \beta_{c_0}\sum_{j=1}^{N} v_{c_j}^{-1}.\end{aligned} \tag{6.37}$$

By estimating the unknown variables alternately, we estimate the unknowns by minimizing their corresponding criterions :

$$\begin{aligned}J(\boldsymbol{f}) &= \frac{1}{2}(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f})^T \boldsymbol{V}_\epsilon^{-1} (\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}) + \frac{1}{2}(\boldsymbol{f} - \boldsymbol{m}_f)^T \boldsymbol{V}_f^{-1} (\boldsymbol{f} - \boldsymbol{m}_f), \\ J(\ddot{\boldsymbol{f}}) &= \frac{1}{2}\left(\ddot{\boldsymbol{g}} - \boldsymbol{H}\ddot{\boldsymbol{f}}\right)^T \boldsymbol{V}_\tau^{-1} (\ddot{\boldsymbol{g}} - \boldsymbol{H}\boldsymbol{f}) + \frac{1}{2}\ddot{\boldsymbol{f}}^T \boldsymbol{V}_c^{-1} \ddot{\boldsymbol{f}}, \\ J(v_{\epsilon_i}) &= \frac{1}{2}\ln v_{\epsilon_i} + \frac{1}{2}[\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}]_i^2 v_{\epsilon_i}^{-1} + (\alpha_{\epsilon_0}+1)\ln v_{\epsilon_i} + \beta_{\epsilon_0} v_{\epsilon_i}^{-1}, \\ J(v_{\tau_i}) &= \frac{1}{2}\ln v_{\tau_i} + \frac{1}{2}\left[\ddot{\boldsymbol{g}} - \boldsymbol{H}\ddot{\boldsymbol{f}}\right]_i^2 v_{\tau_i}^{-1} + (\alpha_{\tau_0}+1)\ln v_{\tau_i} + \beta_{\tau_0} v_{\tau_i}^{-1}, \\ J(v_{f_j}) &= \frac{1}{2}\ln v_{f_j} + \frac{1}{2}[\boldsymbol{f} - \boldsymbol{m}_f]_j^2 v_{f_j}^{-1} + (\alpha_{f_0}+1)\ln v_{f_j} + \beta_{f_0} v_{f_j}^{-1}, \\ J(v_{c_j}) &= \frac{1}{2}\ln v_{c_j} + \frac{1}{2}\ddot{f}_j^2 v_{c_j}^{-1} + (\alpha_{c_0}+1)\ln v_{c_j} + \beta_{c_0} v_{c_j}^{-1}.\end{aligned} \tag{6.38}$$

By calculating the zero points of the derivation of each criterion, we get the updating rules

for all the variables :

$$\begin{aligned}
\widehat{\boldsymbol{f}} &= \left(\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{H} + \boldsymbol{V}_f^{-1}\right)^{-1}\left(\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{g} + \boldsymbol{V}_f^{-1}\boldsymbol{m}_f\right),\\
\widehat{\ddot{\boldsymbol{f}}} &= \left(\boldsymbol{H}^T\boldsymbol{V}_\tau^{-1}\boldsymbol{H} + \boldsymbol{V}_c^{-1}\right)^{-1}\boldsymbol{H}^T\boldsymbol{V}_\tau^{-1}\ddot{\boldsymbol{g}},\\
\widehat{v}_{\epsilon_i} &= \frac{\beta_{\epsilon_0} + \frac{1}{2}\left[\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\right]_i^2}{\alpha_{\epsilon_0} + \frac{3}{2}},\\
\widehat{v}_{\tau i} &= \frac{\beta_{\tau_0} + \frac{1}{2}\left[\ddot{\boldsymbol{g}} - \boldsymbol{H}\ddot{\boldsymbol{f}}\right]_i^2}{\alpha_{\tau_0} + \frac{3}{2}},\\
\widehat{v}_{f_j} &= \frac{\beta_{f_0} + \frac{1}{2}\left[\boldsymbol{f} - \boldsymbol{m}_f\right]_j^2}{\alpha_{f_0} + \frac{3}{2}},\\
\widehat{v}_{cj} &= \frac{\beta_{c_0} + \frac{1}{2}\ddot{f}_j^2}{\alpha_{c_0} + \frac{3}{2}}.
\end{aligned} \tag{6.39}$$

The updating of the variables $\widehat{v}_{\epsilon_i}$, $\widehat{v}_{f_j}$ and $\widehat{v}_{cj}$ are explicit, but as we have pointed out in Chapter 5 that, the inversion of the non-diagonal big size matrix $\left(\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{H} + \boldsymbol{V}_f^{-1}\right)$ and $\left(\boldsymbol{H}^T\boldsymbol{V}_\epsilon^{-1}\boldsymbol{H} + \boldsymbol{V}_c^{-1}\right)$ are not realizable, we use the gradient descent algorithm to optimize the unknown variables $\widehat{\boldsymbol{f}}$ and $\widehat{\ddot{\boldsymbol{f}}}$:

$$\begin{aligned}
\widehat{\boldsymbol{f}}^{(k+1)} &= \widehat{\boldsymbol{f}}^{(k)} - \gamma_{\boldsymbol{f}}\nabla J(\widehat{\boldsymbol{f}}^{(k)}),\\
\widehat{\ddot{\boldsymbol{f}}}^{(k+1)} &= \widehat{\ddot{\boldsymbol{f}}}^{(k)} - \gamma_{\ddot{\boldsymbol{f}}}\nabla J(\widehat{\ddot{\boldsymbol{f}}}^{(k)}),
\end{aligned} \tag{6.40}$$

where $\nabla J(\cdot)$ represents the derivation of the corresponding criterion $J(\cdot)$, and $\gamma_\cdot$ represents the descent step length. The descent step length is obtained by minimizing the following criterion :

$$\begin{aligned}
\gamma_{\boldsymbol{f}} &= \arg\min_{\gamma_{\boldsymbol{f}}}\left\{J(\widehat{\boldsymbol{f}} - \gamma_{\boldsymbol{f}}\nabla J(\widehat{\boldsymbol{f}}))\right\},\\
\gamma_{\ddot{\boldsymbol{f}}} &= \arg\min_{\gamma_{\ddot{\boldsymbol{f}}}}\left\{J(\widehat{\ddot{\boldsymbol{f}}} - \gamma_{\ddot{\boldsymbol{f}}}\nabla J(\widehat{\ddot{\boldsymbol{f}}}))\right\},
\end{aligned} \tag{6.41}$$

and the expressions of $\gamma_{\boldsymbol{f}}$ and $\gamma_{\ddot{\boldsymbol{f}}}$ are respectively :

$$\gamma_{\boldsymbol{f}} = \frac{\|\nabla J(\boldsymbol{f})\|^2}{\|\boldsymbol{Y}_\epsilon\boldsymbol{H}\nabla J(\boldsymbol{f})\|^2 + \|\boldsymbol{Y}_f\nabla J(\boldsymbol{f})\|^2}, \text{ where } \boldsymbol{Y}_\epsilon = \boldsymbol{V}_\epsilon^{-\frac{1}{2}} \text{ and } \boldsymbol{Y}_f = \boldsymbol{V}_f^{-\frac{1}{2}}, \tag{6.42}$$

$$\gamma_{\ddot{\boldsymbol{f}}} = \frac{\left\|\nabla J(\ddot{\boldsymbol{f}})\right\|^2}{\left\|\boldsymbol{Y}_\epsilon\boldsymbol{H}\nabla J(\ddot{\boldsymbol{f}})\right\|^2 + \left\|\boldsymbol{Y}_c\nabla J(\ddot{\boldsymbol{f}})\right\|^2}, \text{ where } \boldsymbol{Y}_c = \boldsymbol{V}_c^{-\frac{1}{2}}. \tag{6.43}$$

The optimization of all the variables by using ROCC method follows the following iterative

updating rules :

$$\text{iter}: \ \widehat{\boldsymbol{f}}^{(k+1)} = \widehat{\boldsymbol{f}}^{(k)} - \gamma_{\boldsymbol{f}} \nabla J(\widehat{\boldsymbol{f}}^{(k)}), \tag{6.44}$$

$$\text{iter}: \ \widehat{\ddot{\boldsymbol{f}}}^{(k+1)} = \widehat{\ddot{\boldsymbol{f}}}^{(k)} - \gamma_{\ddot{\boldsymbol{f}}} \nabla J(\widehat{\ddot{\boldsymbol{f}}}^{(k)}), \tag{6.45}$$

$$\widehat{v}_{\epsilon_i} = \frac{\beta_{\epsilon_0} + \frac{1}{2}\left[\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\right]_i^2}{\alpha_{\epsilon_0} + \frac{3}{2}},$$

$$\widehat{v}_{\tau i} = \frac{\beta_{\tau_0} + \frac{1}{2}\left[\ddot{\boldsymbol{g}} - \boldsymbol{H}\ddot{\boldsymbol{f}}\right]_i^2}{\alpha_{\tau_0} + \frac{3}{2}}, \tag{6.46}$$

$$\widehat{v}_{f_j} = \frac{\beta_{f_0} + \frac{1}{2}\left[\boldsymbol{f} - \boldsymbol{m}_f\right]_j^2}{\alpha_{f_0} + \frac{3}{2}}, \tag{6.47}$$

$$\widehat{v}_{cj} = \frac{\beta_{c_0} + \frac{1}{2}\ddot{f}_j^2}{\alpha_{c_0} + \frac{3}{2}}. \tag{6.48}$$

In this proposed method, the estimated Laplacian in each iteration is used in the optimization of the phantom, via the non-homogeneous Markovian model. The normalized Laplacian is considered for the weighted mean in the Gaussian distribution. By doing so, the neighbor pixels are distinguished via the value of the Laplacian, hence the homogeneous areas and the contours areas are separately considered. On the other hand, the optimized object $\widehat{\boldsymbol{f}}$ is used to construct the initialization variable for the optimization of Laplacian $\widehat{\ddot{\boldsymbol{f}}}$ in each iteration.

Figure 6.6 shows the Directed Acyclic Graph (DAG) which presents the principal relationship of the updating of all the variables in the iterative updating rule in Eq.(6.44)-Eq.(6.48).

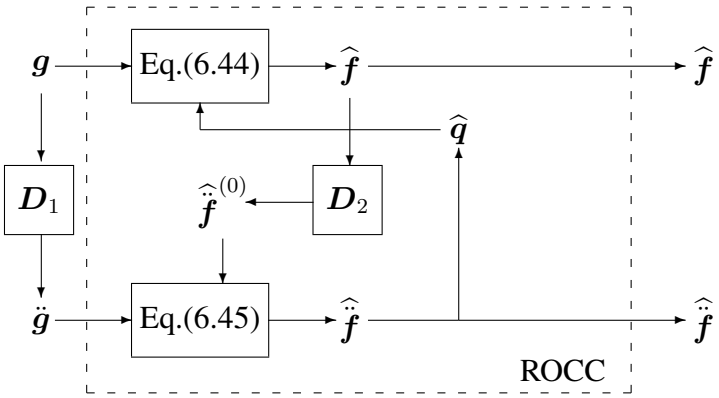


FIGURE 6.6 – DAG of proposed ROCC model.

6.6 Experiment results

Here we show the experimental results of 3D X-ray CT reconstruction of both simulated phantom and real object. Comparison with conventional methods are also derived.

6.6.1 Implementation

In the simulation, the 3D "Shepp Logan" phantom of size 256^3 and the 3D "Head" object of size 256^3 are considered. Parallel projection and back-projection are implemented via the ASTRA toolbox [vAPC+16]. For each projection, 256^2 detectors are used to receive the attenuated radiation, and hence an image of size 256^2 is obtained for each projection. In our notation, N and M represent respectively the size of object $\boldsymbol{f}$ and projection dataset $\boldsymbol{g}$. n_p represents the number of projections. According to the definition, we have $M = 256^2 n_p$.

In the proposed ROCC method, two types of gradient are used : the one of the object $\boldsymbol{f}$, and the one of the data $\boldsymbol{g}$. First of all, we consider the 2D case. The 3D case then can be derived with the same token.

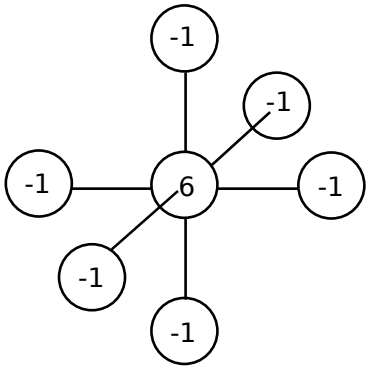


FIGURE 6.7 – Laplacian kernel in 3D

In 3D cases, the Laplacian of $\boldsymbol{f}$ is the convolution of the object with a 3D Laplacian operator, shown in Figure 6.7. In this operator, there is a value 6 in the middle, and a -1 on the left, right, top, bottom, front and behind. It is represented by the operator :

$$\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \tag{6.49}$$

The derivation of $\boldsymbol{g}$ in 3D cases is the 2D derivation of each projection direction.

In each iteration, after the optimization of $\boldsymbol{f}$ and $\ddot{\boldsymbol{f}}$, the corresponding $\boldsymbol{m}_f$, Eq.(6.33), and $\boldsymbol{q}$, Eq.(6.34), are calculated. They are used in the optimization of the other variables.

The implementation algorithm of the ROCC method is illustrated in Algorithm 6.

6.6.2 Simulation results of reconstructing 3D phantom with ROCC method

The simulated Shepp Logan phantom and the real Head object are used in the simulations. Both of these two objects have 256^3 voxels. The datasets are obtained by projecting the object using parallel beam radiations. For each projection direction, 256^2 detectors are used to receive the attenuated photons and hence an image of size 256×256 is obtained for each projection direction. The projections are distributed evenly between $0°$ and $180°$.

Algorithm 6 Summary of the JMAP algorithm for ROCC method

Input : $\boldsymbol{H},\ \boldsymbol{D},\ \boldsymbol{g}$
Output : $\widehat{\boldsymbol{f}},\ \widehat{\ddot{\boldsymbol{f}}}$
Initialization :
$\ddot{\boldsymbol{g}} = \mathrm{Laplacian}(\boldsymbol{g})$
$\widehat{\boldsymbol{f}}^{(0)} \leftarrow$ normalized FDK
repeat
    Compute $\widehat{\boldsymbol{m}}_f$ according to Eq.(6.33)
    Compute $\widehat{\boldsymbol{q}}$ according to Eq.(6.34)
    repeat
        Compute $\widehat{\gamma}_{\boldsymbol{f}}$ according to Eq.(6.42)
        Compute $\nabla J(\widehat{\boldsymbol{f}})$ from Eq.(6.38)
        $\widehat{\boldsymbol{f}} = \widehat{\boldsymbol{f}} - \widehat{\gamma}_{\boldsymbol{f}} \nabla J(\widehat{\boldsymbol{f}})$
    until convergence or maximum iteration
    $\widehat{\ddot{\boldsymbol{f}}} = \mathrm{Laplacian}(\widehat{\boldsymbol{f}})$
    repeat
        Compute $\widehat{\gamma}_{\ddot{\boldsymbol{f}}}$ according to Eq.(6.43)
        Compute $\nabla J(\widehat{\ddot{\boldsymbol{f}}})$ from Eq.(6.38)
        $\widehat{\ddot{\boldsymbol{f}}} = \widehat{\ddot{\boldsymbol{f}}} - \widehat{\gamma}_{\ddot{\boldsymbol{f}}} \nabla J(\widehat{\ddot{\boldsymbol{f}}})$
    until convergence or maximum iteration
    Compute $v_{\epsilon_i}, \forall i \in [1:M]$ according to Eq.(6.39)
    Compute $v_{\tau_i}, \forall i \in [1:M]$ according to Eq.(6.39)
    Compute $v_{f_j}, \forall j \in [1:N]$ according to Eq.(6.39)
    Compute $v_{c_j}, \forall j \in [1:N]$ according to Eq.(6.39)
until convergence or maximum iteration

6.6.3 Initializations

The initialization of the variables and hyperparameters of the ROCC method is discussed in this section. The reconstruction result of the object by using the FBP technique is used to initialize the variable $\widehat{\boldsymbol{f}}$. Another hidden variable, $\widehat{\ddot{\boldsymbol{f}}}$, is initialized as the Laplacian of the object $\widehat{\boldsymbol{f}}$:

$$\widehat{\ddot{\boldsymbol{f}}} = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \widehat{\boldsymbol{f}}. \tag{6.50}$$

The initialization of hyperparameters α_{ϵ_0}, β_{ϵ_0}, α_{τ_0}, β_{τ_0}, α_{f_0}, β_{f_0}, α_{c_0} and β_{c_0} depends on the chosen prior model and the prior property of the parameters.

Initialization of α_{ϵ_0} and β_{ϵ_0}

α_{ϵ_0} and β_{ϵ_0} are the hyperparameters of the prior model of the noise $\boldsymbol{\epsilon}$. In the simulation, a Gaussian noise is added to datasets and consequently $\boldsymbol{\epsilon}$ belong to a Gaussian distribution. However, in the real applications, the model of the noise $\boldsymbol{\epsilon}$ is not known. In positron tomography the noise is normally modeled by a Poisson distribution, but in X-ray CT the Gaussian distribution is used to model noise. So that α and β are chosen such that $\mathcal{S}t_g$ is approximate to a Gaussian distribution.

As presented in Section 4.4.2, the $\mathcal{S}t_g$ distribution approaches to a Gaussian distribution when the value of α_{ϵ_0} and β_{ϵ_0} are both big.

Also, a relationship between α_{ϵ_0} and β_{ϵ_0} is obtained via the variance of the noise v_ϵ. In the simulations, the value of v_ϵ is obtained via the value of the SNR and the variance of the projection dataset v_g. As presented in Eq.(5.84), we have :

$$\beta_{\epsilon_0} = \frac{\|\boldsymbol{g}\|_2^2}{M} \times \frac{1}{1 + 10^{SNR/10}} \times (\alpha_{\epsilon_0} - 1)\,. \tag{6.51}$$

Therefore it is sufficient to initialize only the value of α_{ϵ_0}.

Figure 6.8 shows the influence of the value of hyperparameter α_{ϵ_0} on the performance of reconstruction in terms of RMSE of the reconstructed object. We can see from the figures that, for different cases of number of projections or different noise in the dataset, we always have better results when the value of α_{ϵ_0} has a bigger value. Furthermore, when the value of α_{ϵ_0} is bigger than a threshold value, its influence on the RMSE is relatively insensitive. This relative insensitiveness facilitates the initialization of the hyperparameter α_{ϵ_0}.

Initialization of α_{τ_0} and β_{τ_0}

The noise $\boldsymbol{\tau}$ is also modeled by a Gaussian distribution. For the same reason as for $\boldsymbol{\epsilon}$, the hyperparameters α_{τ_0} and β_{τ_0} are related by :

$$\beta_{\tau_0} = \frac{\|\ddot{\boldsymbol{g}}\|_2^2}{M} \times \frac{1}{1 + 10^{SNR/10}} \times (\alpha_{\tau_0} - 1)\,, \tag{6.52}$$

where $\ddot{\boldsymbol{g}}$ is the Laplacian of dataset $\boldsymbol{g}$.

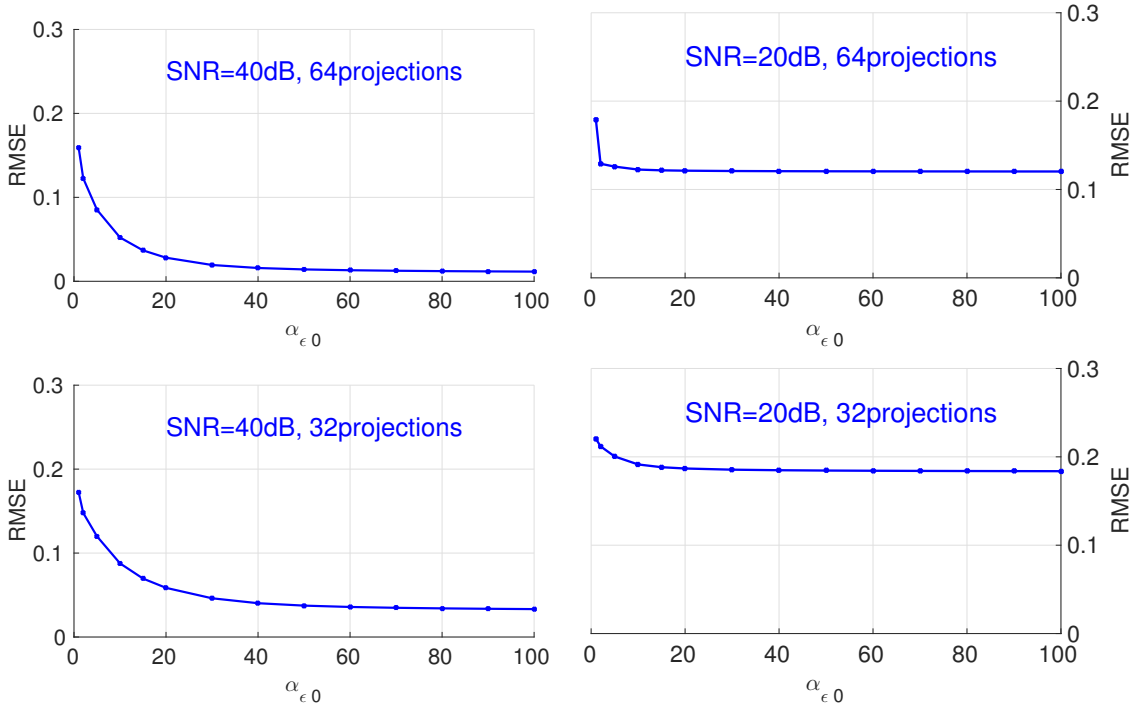


FIGURE 6.8 – Influence of hyperparameter α_{ϵ_0} in ROCC method on RMSE of final reconstruction results for different number of projections and different noise.

Figure 6.9 shows the influence of the value of α_{τ_0} on the performance of reconstruction in terms of RMSE of reconstructed object. In these figures, we can see that with different value of α_{τ_0}, the RMSE of the reconstruction is the same. In other words, the value α_{τ_0} has almost no influence on the final results of reconstruction.

Initialization of α_{f_0} and β_{f_0}

The hyperparameters α_{f_0} and β_{f_0} appear in the Inverse Gamma distribution of the parameter $\boldsymbol{v}_f$. Consequently, the variable $\boldsymbol{f}$ belongs to a $\mathcal{S}t_g$ model. A rough value of the variance of the variable $\boldsymbol{f}$ can be obtained via the initialization of the variable $\boldsymbol{f}$: $\widehat{\boldsymbol{v}}_f = \mathrm{Var}[\widehat{\boldsymbol{f}}]$. According to the property of the $\mathcal{S}t_g$ distribution, we have $\widehat{\boldsymbol{v}}_f|\alpha_{f_0}, \beta_{f_0} = \beta_{f_0}/(\alpha_{f_0} - 1)$, therefore we obtain :

$$\beta_{f_0} = \widehat{\boldsymbol{v}}_f \times (\alpha_{f_0} - 1)\,. \tag{6.53}$$

As there exist a relationship between these two hyperparameters, it is sufficient to initialize α_{f_0}. In our work, the object $\boldsymbol{f}$ under consideration, which is modeled by the $\mathcal{S}t_g$ distribution, is not sparse. Theoretically the value of α_{f_0} and β_{f_0} should be initialized with a bigger value in order to have better results.

Figure 6.10 shows the RMSE of the reconstruction results with different values of α_{f_0} with different number of projections and different noise level in dataset. The simulation results proved our proposition. When the dataset has a high SNR=40dB, the initilization of α_{f_0} is not sensitive. While there is a low SNR=20dB, the reconstruction has a better result when α_{f_0} is bigger.

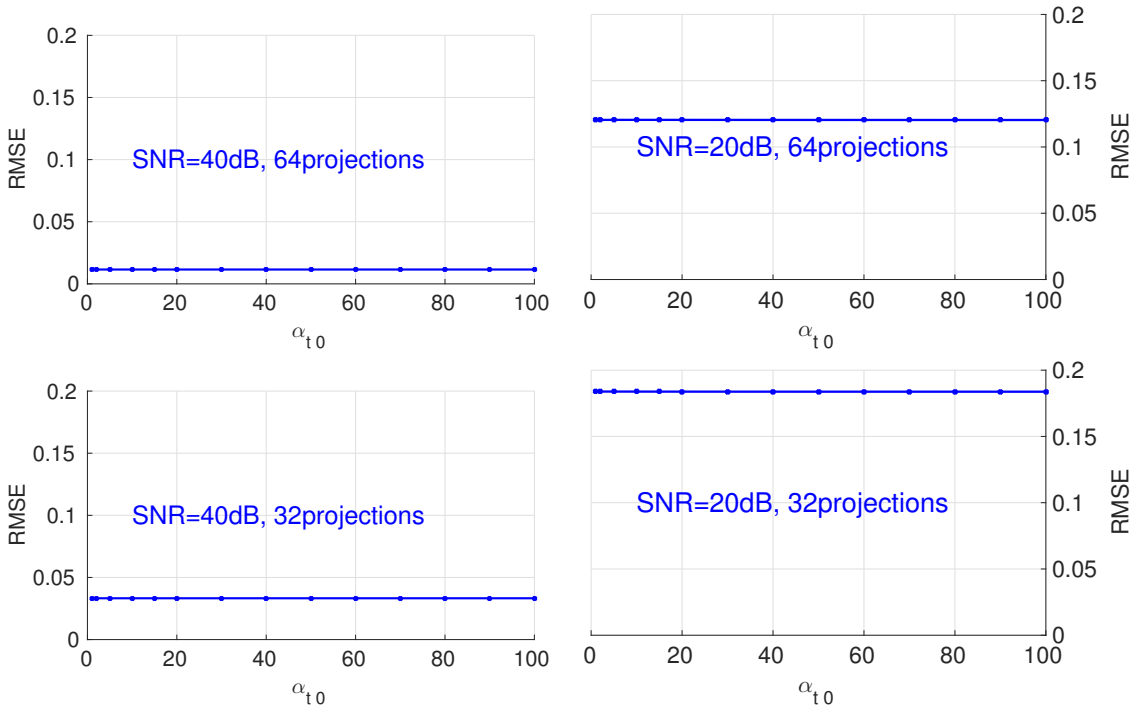


FIGURE 6.9 – Influence of hyperparameter α_{τ_0} in ROCC method on RMSE of final reconstruction results for different number of projections and different noise.

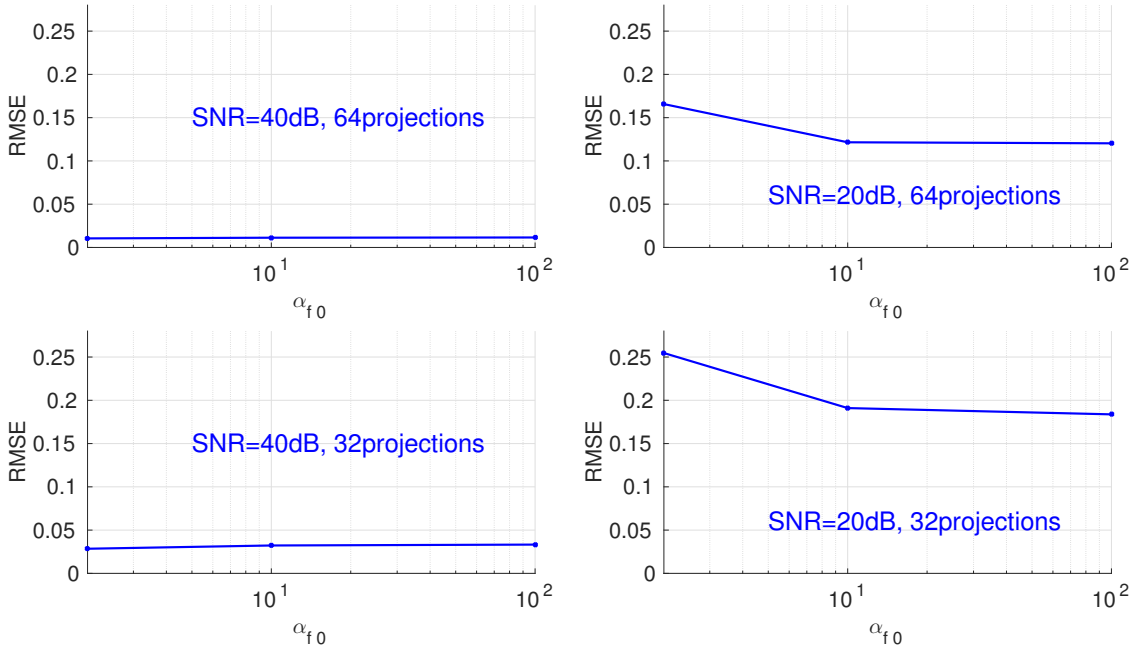


FIGURE 6.10 – Influence of hyperparameter α_{f_0} in ROCC method on RMSE of final reconstruction results for different number of projections and different noise.

Initialization of α_{c_0} and β_{c_0}

In ROCC method, the Laplacian of the object, $\ddot{\boldsymbol{f}}$, is modeled by a $\mathcal{S}t_g$ distribution depending on two hyper-parameters α_{c_0} and β_{c_0}. Here we analysis the initialization of these two hyper-parameters simultaneously. The variable $\ddot{\boldsymbol{f}}$ is supposed to have a sparse structure, and could be realized with small values for both α_{c_0} and β_{c_0}. Consequently, the ROCC method is expected to have a better result when α_{c_0} and β_{c_0} are smaller.

Figure 6.11 and Figure 6.12 show the influence of the initialization of α_{c_0} and β_{c_0}. In Figure 6.11, different colors correspond to different values of β_{c_0}, and in Figure 6.12 different colors correspond to different values of α_{c_0}. We can see from the results that, for different number of projections and different noise levels in dataset, the value of α_{c_0} and β_{c_0} have almost no influence on the RMSE of reconstruction.

This property is very interesting. As we have mentioned above, when the value of α_{c_0} and β_{c_0} are smaller, the variable $\ddot{\boldsymbol{f}}$ will have a sparser structure. In the ROCC method, the variable $\ddot{\boldsymbol{f}}$ is modeled by this $\mathcal{S}t_g$ distribution. However, $\ddot{\boldsymbol{f}}$ also depends on the observed data $\ddot{\boldsymbol{g}}$. The simulation results show us that the values of α_{c_0} and β_{c_0} don't have much influence of the quality of reconstruction because the relationship with $\ddot{\boldsymbol{g}}$ plays a more important role in this method.

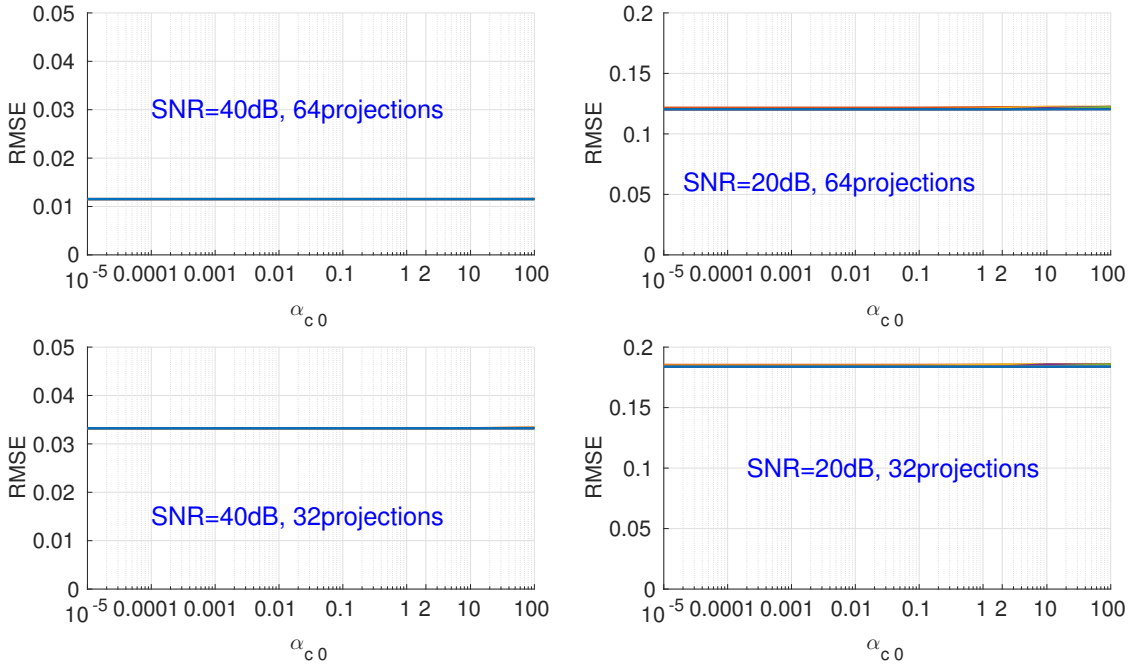


FIGURE 6.11 – Influence of hyperparameter α_{c_0} in ROCC method on RMSE of final reconstruction results for different number of projections and different noise.

6.6.4 Simulation results

The initialization of the variables and hyperparameters are defined as following :

- $\widehat{\boldsymbol{f}}^{(0)}$ is the reconstructed object from the dataset $\boldsymbol{g}$ by using the FBP method.
- $\widehat{\ddot{\boldsymbol{f}}}^{(0)}$ is the Laplacian of the initialized $\widehat{\boldsymbol{f}}^{(0)}$.
- $\alpha_{\epsilon_0} = 100$, and $\beta_{\epsilon_0} = v_\epsilon \times (\alpha_{\epsilon_0} - 1) = \frac{\|\boldsymbol{g}\|_2^2}{M} \times \frac{1}{1+10^{SNR/10}} \times (\alpha_{\epsilon_0} - 1)$.

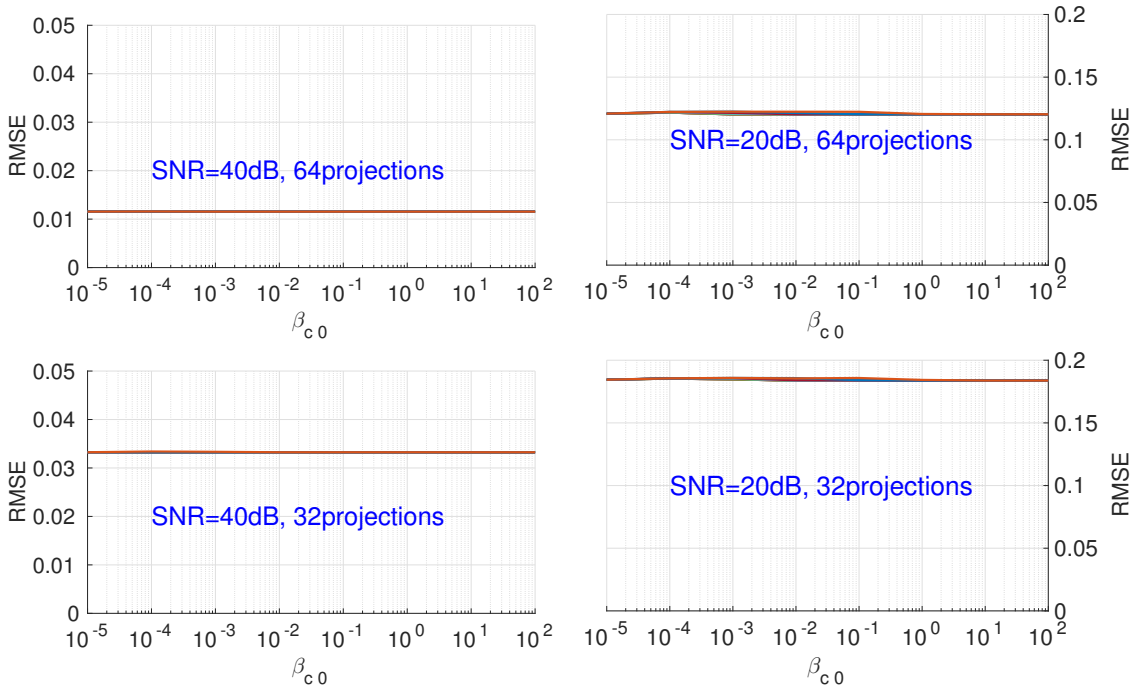


FIGURE 6.12 – Influence of hyperparameter β_{c_0} in ROCC method on RMSE of final reconstruction results for different number of projections and different noise.

- $\alpha_{\tau_0} = 100$, and $\beta_{\tau_0} = v_\tau \times (\alpha_{\tau_0} - 1) = \frac{\|\ddot{\boldsymbol{g}}\|_2^2}{M} \times \frac{1}{1+10^{SNR/10}} \times (\alpha_{\tau_0} - 1)$.
- $\alpha_{f_0} = 100$, and $\beta_{f_0} = v_f \times (\alpha_{f_0} - 1)$.
- $\alpha_{c_0} = 0.01$ and $\beta_{c_0} = 0.01$.

Table 6.1 shows the comparison of different metrics of evaluation of reconstruction by using different methods. Reconstruction with 180, 90, 60, 45, 36 and 18 projections are tested, and they correspond to respectively 1 projection in every 1°, 2°, 3°, 4°, 5° and 10°.

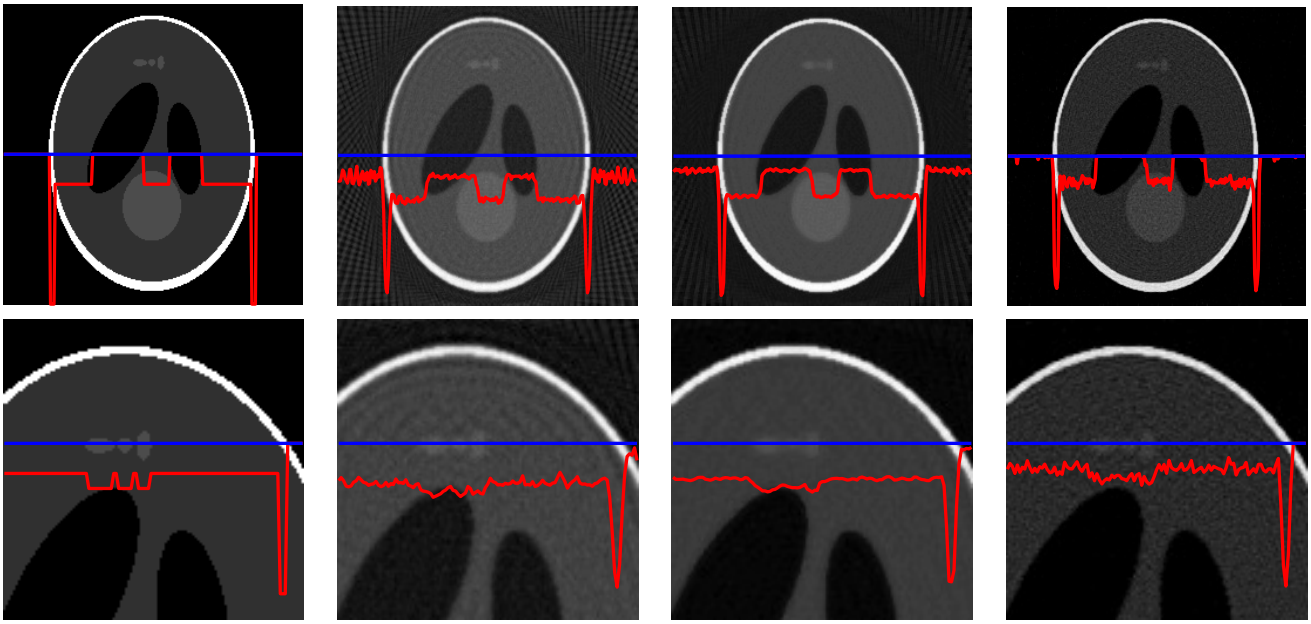
The ROCC method is compared with the QR and TV methods. To evaluate the simulation results, the Relative Means Square Error (RMSE), the Improvement of SNR (ISNR), the Peak SNR (PSNR) and the Structure Similarity of Image (SSIM) of the reconstruction are measured and compared. The results show that TV performs better when a low SNR dataset is used in the reconstruction. However, in the high SNR cases, and when the number of projections is insufficient, the ROCC method outperforms the TV method. The ROCC method is more robust than the two state-of-the-art methods when number of projections is limited.

Figure 6.13 represents the middle slice of the reconstructed Shepp Logan phantom by using different methods, using projection sinogram of 60 projections and SNR=40dB. From the figures we can see that in the reconstructed phantom by using QR method, the edge of the phantom and the contours of different blocks in the phantom are blurred, and in the background the artifacts are obvious. The reconstructed phantom using TV and ROCC methods have clearer contours than the QR method. When comparing with TV method, we see that TV method leads to results with more blurry contours and smoother homogeneous areas. The ROCC method gives clearer contours than the two state-of-the-art methods.

In Figure 6.14, the middle slices of the reconstructed Shepp Logan phantom using different methods are presented. In this simulation, the projection sinogram with 36 projections

TABLE 6.1 – Comparison of RMSE, ISNR, PSNR and SSIM of reconstructed phantom with 50 globle iterations (10 iterations for gradient descent in each of globle iterations) by using QR, TV or ROCC methods. The values of regularization parameters are respectively $\lambda_{QR} = 10$ and $\lambda_{TV} = 50$ for SNR=40dB, $\lambda_{QR} = 600$ and $\lambda_{TV} = 100$ for SNR=20dB.

	256*256*256											
	180 projections						90 projections					
	40dB			20dB			40dB			20dB		
	QR	TV	ROCC	QR	TV	ROCC	QR	TV	ROCC	QR	TV	ROCC
RMSE	0.0236	0.0114	**0.0114**	0.1309	**0.0209**	0.0688	0.0401	0.0212	**0.0177**	0.1558	**0.0491**	0.0914
ISNR	5.5584	8.7217	**8.6386**	7.2024	**15.1775**	9.9937	6.6136	9.3832	**10.1605**	8.4583	**13.4765**	10.7756
PSNR	30.0675	33.2308	**33.2443**	22.6318	**30.6069**	25.4231	27.7743	30.5439	**31.3211**	21.8754	**26.8937**	24.1928
SSIM	0.9999	0.9999	**0.9999**	0.9992	**0.9999**	0.9996	0.9997	0.9999	**0.9999**	0.9990	**0.9997**	0.9994
	60 projections						45 projections					
	40dB			20dB			40dB			20dB		
	QR	TV	ROCC	QR	TV	ROCC	QR	TV	ROCC	QR	TV	ROCC
RMSE	0.0636	0.0321	**0.0218**	0.1656	**0.0753**	0.1102	0.0904	0.0474	**0.0259**	0.1854	**0.0901**	0.1300
ISNR	9.3826	12.3480	**14.0209**	9.1492	**12.5701**	10.9191	10.3301	13.1308	**15.7625**	10.0137	**13.1476**	11.5572
PSNR	25.7693	28.7347	**30.4076**	21.6116	**25.0325**	23.3814	24.2404	27.0412	**29.6728**	21.1195	**24.2535**	22.6631
SSIM	0.9996	0.9995	**0.9999**	0.9990	**0.9995**	0.9993	0.9994	0.9997	**0.9998**	0.9988	**0.9994**	0.9992
	36 projections						18 projections					
	40dB			20dB			40dB			20dB		
	QR	TV	ROCC	QR	TV	ROCC	QR	TV	ROCC	QR	TV	ROCC
RMSE	0.1177	0.0680	**0.0301**	0.1957	**0.1116**	0.1451	0.2581	0.2104	**0.0566**	0.2907	0.2313	**0.2252**
ISNR	10.6591	13.0424	**16.5786**	10.8633	**13.3032**	12.1617	10.7122	11.5992	**16.8020**	10.8088	11.8022	**11.8279**
PSNR	23.0949	25.4783	**29.0145**	20.8865	**23.3264**	22.1849	19.6263	20.5133	**26.2758**	19.1085	20.1020	**20.2758**
SSIM	0.9993	0.9996	**0.9998**	0.9988	**0.9993**	0.9991	0.9983	0.9987	**0.9996**	0.9981	0.9985	**0.9986**



Original QR TV ROCC

FIGURE 6.13 – The reconstructed Shepp Logan phantom of size 256^3 with 60 projections and SNR=40dB, by using different reconstruction methods. Red curves represent the profile at the position of the blue line.

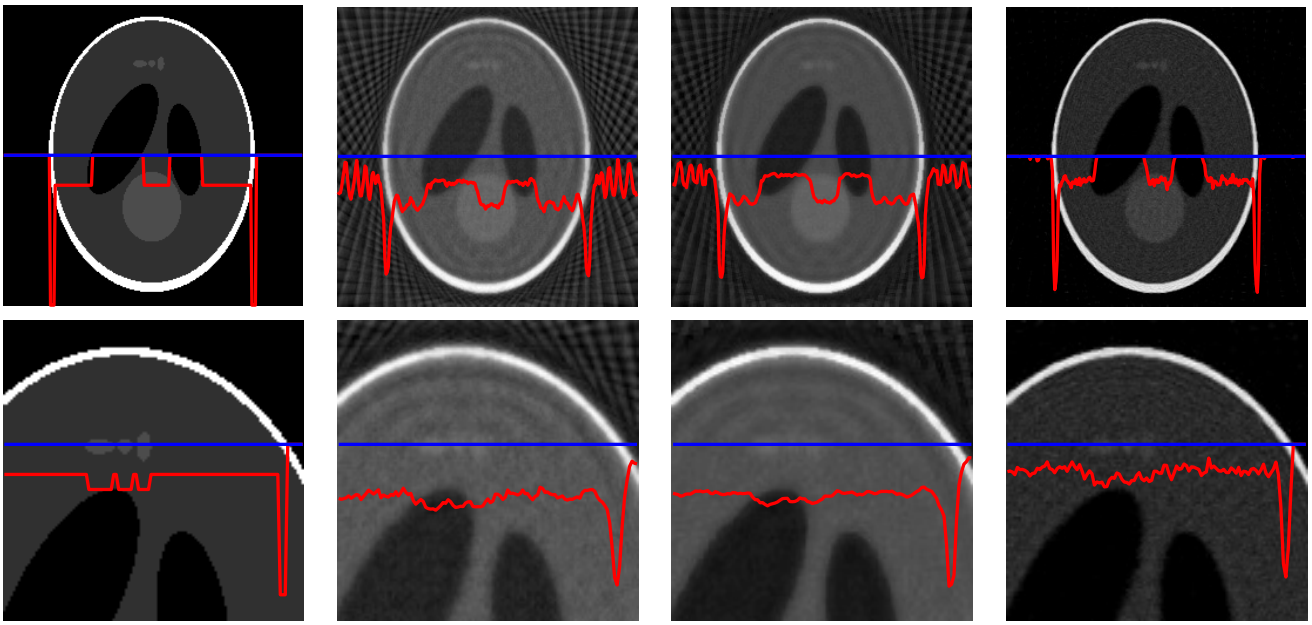


FIGURE 6.14 – The reconstructed Shepp Logan phantom of size 256^3 with 36 projections and SNR=40dB, by using different reconstruction methods. Red curves represent the profile at the position of the blue line.

and SNR=40dB is used. As shown in the results, by using the QR and TV, the contours of the reconstructed phantom are blurry, and in the object there exist some circle artifacts. In the reconstructed phantom by using ROCC method, the contours are clearer than the other state-of-the-art methods.

Figure 6.15 and Figure 6.16 present a profile of the middle slice of the reconstructed Shepp Logan phantom. By comparing the ROCC method with the state-of-the-art methods, we can see that the shape of the reconstructed phantom is closer to the original one at the contour areas by using the ROCC method. When the number of projections is insufficient, ROCC method stays more robust than the other two state-of-the-art methods.

In Figure 6.17 and Figure 6.18, we compared the ROCC method with QR and TV method in terms of the RMSE of reconstructed object. The evaluation are compared with projection sinograms with different number of projections. As we can see in the figures, the RMSE is smaller when the number of projections is larger. When there are sufficient number of projections, both these three methods perform very well. When there are insufficient number of projections, the ROCC method stays more robust than QR and TV methods. In the cases where there is a high SNR, the ROCC method stays robust when there are more than 36 projections.

6.7 Comparison of computation time

The computation time is a crucial problem in the real applications, especially for the iterative methods for the three dimensional objects. In this section, we show the comparison of the computation time by using the TV, HHBM and ROCC methods.

In Table 6.2, the reconstruction time by using different methods are compared for a Shepp

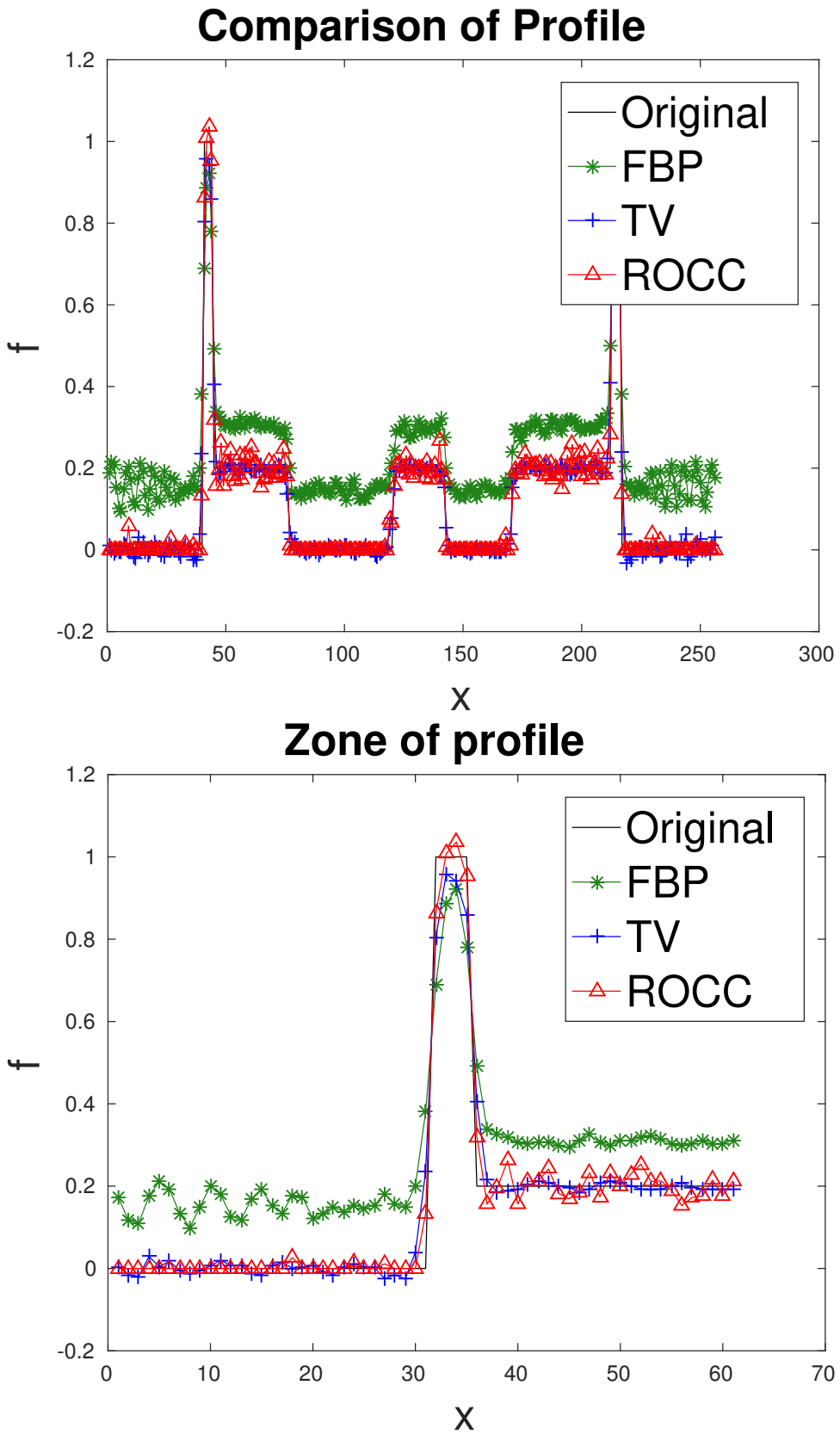


FIGURE 6.15 – The profiles of the middle slice of reconstructed Shepp Logan phantom of size 256^3 with dataset of 60 projections and SNR=40dB, by using different reconstruction methods. The bottom figure is a zone of contour of the top figure.

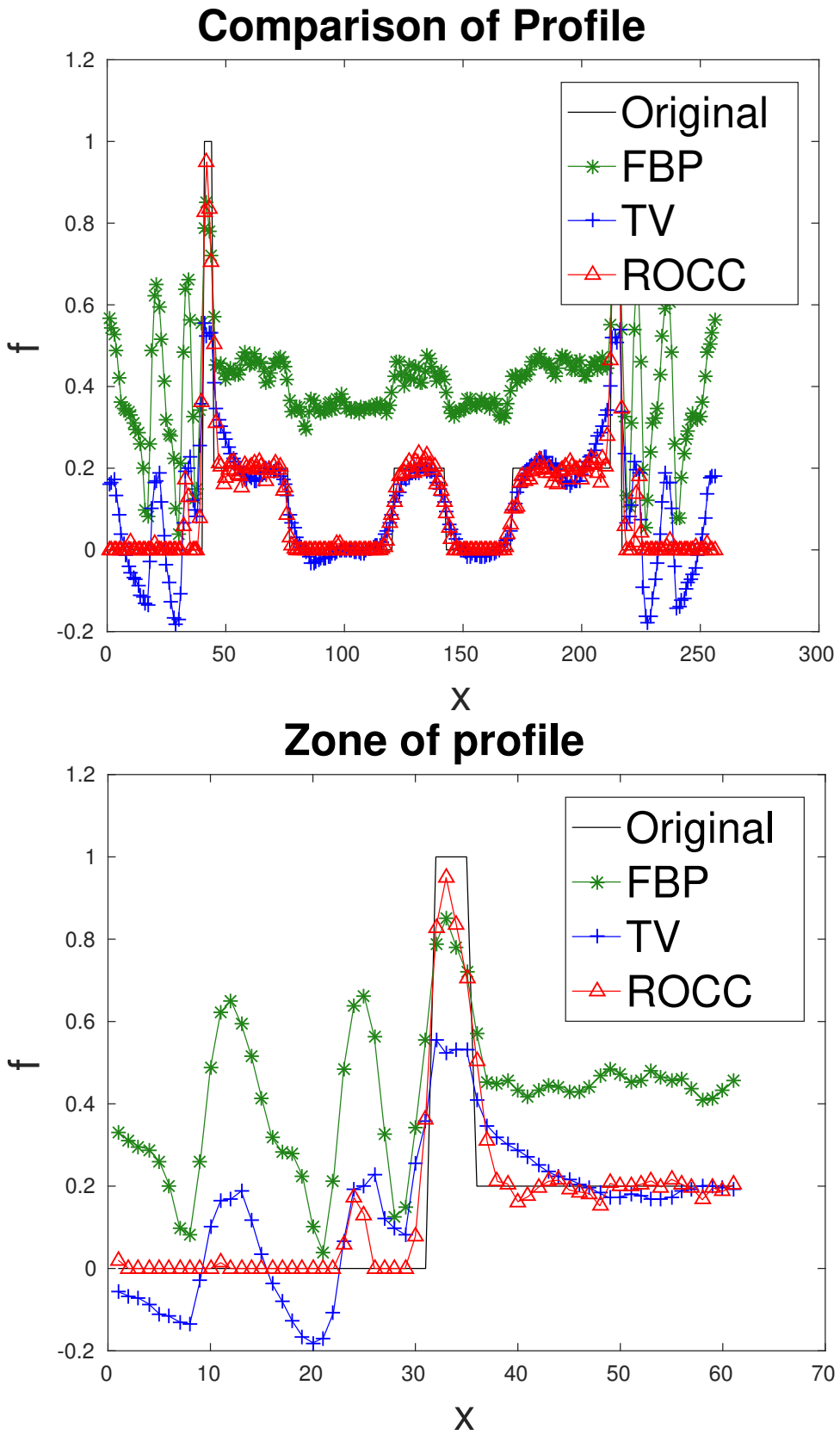


FIGURE 6.16 – The profiles of the middle slice of reconstructed Shepp Logan phantom of size 256^3 with dataset of 18 projections and SNR=40dB, by using different reconstruction methods. The bottom figure is a zone of contour of the top figure.

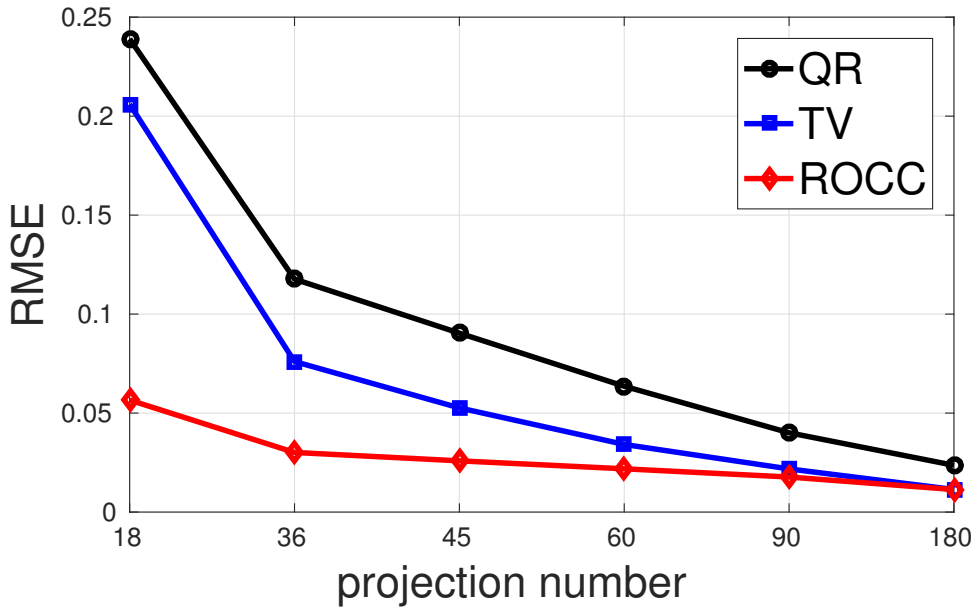


FIGURE 6.17 – The Relative Mean Square Error (RMSE) of the reconstructed Shepp Logan phantom with the dataset of different numbers of projections and SNR=40dB, by using different reconstruction methods.

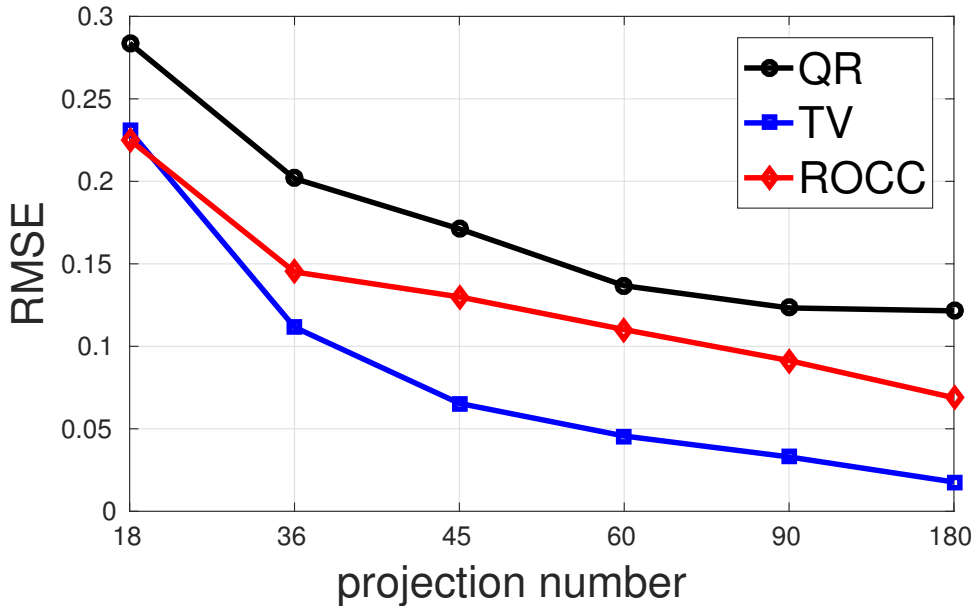


FIGURE 6.18 – The Relative Mean Square Error (RMSE) of the reconstructed Shepp Logan phantom with the dataset of different numbers of projections and SNR=20dB, by using different reconstruction methods.

TABLE 6.2 – Comparison of computation time by using different methods. * The threshold of RMSE implies the critical condition of the end of iteration. When RMSE of the reconstruction is smaller than the threshold, the iteration stops. ** For the TV method, the regularization parameter is an optimized parameter, which is fixed by comparing the RMSE of reconstruction with different parameter values. In the table, the number in the parentheses represents the number of iterations till convergence.

Projection number	128	64	32
Threshold of RMSE*	0.02	0.04	0.1
TV**	722.8s (40)†	754.65s (43)	682.08s (42)
HHBM	1532.3s (13)	822.56s (8)	496.3s (5)
ROCC	**673.2s** (20)	**440.7s** (15)	**82.65s** (3)

Logan phantom of size 256^3, by applying respectively 128, 64 and 32 projections. The computation time represents the total time needed such that the RMSE of estimation $\widehat{\boldsymbol{f}}$ is smaller than the threshold of RMSE. For TV method, the optimal regularization parameter is fixed by comparing the RMSE of reconstruction results with different parameter values. The threshold of RMSE of the critical condition is chosen according to the RMSE of the final reconstruction results of all the reconstruction methods.

In this table, we can conclude that the ROCC method has a much faster convergence rate than the other two methods. For the cases of 128 projections and 64 projections, the HHBM method is slower than the TV method. However, in order to fix the regularization parameter for TV method, a test has to be done for each different parameter value, and normally more than ten times tests are needed. Consequently, the computation time for TV method is much more than the presented time.

6.8 Conclusion

In this chapter we present a Bayesian method which uses a Markovian prior model for the piecewise continuous variable. From the property of the Radon transform, the Laplacian of the object $\ddot{\boldsymbol{f}}$ is related with the Laplacian of the projection sinogram $\ddot{\boldsymbol{g}}$ by a direct model. By using simultaneously the two direct functions $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$ and $\ddot{\boldsymbol{g}} = \boldsymbol{H}\ddot{\boldsymbol{f}} + \boldsymbol{\tau}$, these two variables are estimated. Meanwhile, the updated normalized Laplacian value is used as a weight coefficient in the prior model of $\boldsymbol{f}$ in each iteration. At last, all the variables and parameters are estimated simultaneously via the posterior distribution by using the JMAP estimation algorithm.

The initialization for the hyperparameters is proposed, and the insensitiveness of them to the reconstruction is proved by the simulations.

From the simulations, we conclude that the proposed ROCC method performs better for preserving the contours comparing with the conventional QR and TV methods. When the dataset is biased with a noise of SNR=40dB, the relative mean square error of the reconstructed object is smaller than QR and TV methods, and the advantage is more obvious when there is less number of projections. The RMSE value of the ROCC method by using 36 projections is the same level as the TV method by using 90 projections.

This method is suitable for the parallel beam projection CT reconstruction, where the Radon

transform is used. The future work focus on the extension from the parallel beam to the cone beam projections.

7

Conclusions and Perspectives

The 3D X-ray CT reconstruction has been a hot topic in the computational imaging and medical imaging during several decades, and has been developed rapidly thanks to the development of computation science. In both medical and industrial applications, researchers' attention are changing to the aspect of the data limitation problems. The limited angle or limited number of detectors involves to an ill-posed problem. Motivated by these challenges, in this thesis, we developed several Bayesian methods for X-ray CT reconstruction.

7.1 Conclusions

In this thesis, we proposed mainly two Bayesian methods for 3D CT reconstruction. These methods are based on the context of the Non Destructive Testing (NDT) in industrial applications. The large data size in 3D applications and the expensive computational costs are considered in the algorithms.

In Chapter 5, we proposed a Bayesian method based on the forward system model $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$. In this method the novelty is that we considered a hierarchical structure in the system model, with $\boldsymbol{f} = \boldsymbol{D}\boldsymbol{z} + \boldsymbol{\xi}$. With the additional level of the system, the variable $\boldsymbol{f}$ is represented by a hidden variable $\boldsymbol{z}$. For the choice of the prior model, the considered property then changes from the modelization of the piecewise continuous variable to the modelization of the sparse hidden variable. A generalized Student-t distribution is introduced in this Chapter. Thanks to the fact that the $\mathcal{S}t_g$ distribution is a marginal of a Normal-Inverse Gamma distribution, the estimation algorithms are easy to apply.

In the iterative algorithms, including the Bayesian inference, the initialization for the parameters is crucial for the final results of the reconstruction. In Section 5.4.2 and Section 6.6.3, we presented the strategy of initializing the hyper-parameters for the optimization algorithm in the Bayesian inference with the proposed model. This initialization strategy is experimentally demonstrated. In the analysis of the hyper-parameters, we observed a relative weak influence of

the hyper-parameters on the behavior of the corresponding iterative algorithm. The interest of this relative weak dependency is that it offers a practical way to insure the initialization of the algorithm which is typically not trivial.

In Section 5.5, we presented another orthonormal transformation, the Dual Tree Complex Wavelet Transformation (DT-CWT). This transformation is compared with the discrete Haar transformation. In the comparison, we find that the DT-CWT transformation will not cause the bloc artifacts in the reconstructed figures and is more robust to the contamination in the coefficients. When using in the proposed Bayesian method, the DT-CWT has a better performance of reconstruction than HT. The disadvantage is that the computation of the DT-CWT is relatively more costly than using the Haar transformation.

In Section 5.8, we proposed to use another forward model in this Bayesian approach. In this model, the additive noise is separated into two parts, one is modeled by a Gaussian distribution and another modeled by a generalized Student-t distribution. The forward model considered for the projection system is

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\rho} + \boldsymbol{\epsilon}. \tag{7.1}$$

In this forward system, a more complicated noise model is considered. The simulation results show that this model with the consideration of a more complex noise model derives better reconstruction results.

In Chapter 6, another strategy for the reconstruction of X-ray CT objects is presented. In this method, the relationship between the object $\boldsymbol{f}$, the projection sinogram of the object $\boldsymbol{g}$, the Laplacian of the object $\ddot{\boldsymbol{f}}$ and the Laplacian of the projection sinogram $\ddot{\boldsymbol{g}}$ are considered. According to the Radon transform system, the relations of the direct projection system that we take into consideration are :

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}, \tag{7.2}$$

$$\ddot{\boldsymbol{g}} = \boldsymbol{H}\ddot{\boldsymbol{f}} + \boldsymbol{\tau}, \tag{7.3}$$

with some relations between $\boldsymbol{g}$ and $\ddot{\boldsymbol{g}}$ and between $\boldsymbol{f}$ and $\ddot{\boldsymbol{f}}$.

By considering these relationships, the Laplacian of the object $\ddot{\boldsymbol{f}}$ is considered as a hidden variable in the system model. A non-homogeneous Markovian model is used for the piecewise continuous object $\boldsymbol{f}$. In this Markovian model, the hidden variable $\ddot{\boldsymbol{f}}$ is used as a parameter. By using this Markovian model considering the value of $\ddot{\boldsymbol{f}}$, the contours and homogeneous areas are modeled with the same model with different parameters, and therefore both the continuity of the homogeneous areas and the discontinuity of the contours are both considered. This strategy preserves the edges when reconstructing the object.

7.2 Perspectives

For the future work several aspects are considered. First of all, the computational time should take into consideration. The GPU processor is only used for the projection operator and back-projection operator in this thesis. In order to accelerate the computation, more computations in the program can be implemented by using GPU processor. For example, the DT-CWT transformation presented in Chapter 5 has very good properties comparing with the Haar transformation, but has a more complicated computation. Once we implement the transformation by

using the GPU processor, the utilization of the DT-CWT or many other transformations will be less costly.

In this thesis, we used the Maximum A Posterior estimator for the optimization when the posterior distribution is obtained. There exists some other estimation algorithms, and some gives more information for the variables. For example, the Posterior Mean estimation can be used to estimate the variables. By using the Variational Bayesian Approach (VBA) method, the posterior distribution is approached by another separable distribution :

$$p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}) = q(\boldsymbol{f}, \boldsymbol{\theta}) = q_1(\boldsymbol{f})q_2(\boldsymbol{\theta}), \tag{7.4}$$

where $\boldsymbol{\theta}$ represents all the hidden variables and parameters of the model. The mean and variance of the approximating distribution are estimated.

In the VBA method, the approaching distributions are computed by minimizing the Kullback Leibler divergence. During the computation of VBA method, a calculation of the diagonal elements of the large size matrix $\boldsymbol{H}^T\boldsymbol{H}$ is needed, shown in Appendix A. Some simulation results of the VBA method are presented in [WMDGD16] for the reconstruction of 2D Shepp Logan phantom by considering a stationary noise. Because of the huge size of the matrix $\boldsymbol{H}$ in the 3D problem, it is very costly if we compute directly by projection and back-projection to obtain each diagonal element. Now our group is studying on this subject and searching for algorithms which calculate approximately the diagonal elements by using the projection and back-projection properties.

In this thesis, the generalized Student-t distribution is used to define the sparse variables. As we have mentioned, the $\mathcal{S}t_g$ distribution has a heavy tailed structure. At the same time, it can have a very small variance thanks to the fact that there are two parameters for the control of the shape of this distribution. Another very useful property in the Bayesian context is that, it can be expressed as a marginal of the bivariate Normal-Inverse Gamma distribution. As the Normal distribution and the Inverse Gamma distribution are conjugate, the estimation of variables in the Bayesian methods are more simple.

In future work, in place of using the Normal-Inverse Gamma mixture distributions, we look for other similar distributions, for example the Normal variance mixture distributions. We will compare different mixture distributions and compare their properties, concerning the sparsity enforcing property of the marginal distribution and also the influence of the hyper-parameters of the marginal distributions.

One of the most important aspects in our future work is to study the influence of the initialization of hyper-parameters. In this thesis, we have observed a relative weak influence of the hyper-parameters of the Bayesian model under the prior models. In the future work we will study on the reason of this weak influence of the hyper-parameters of this model, and see if the same property appears when using the other Normal variance mixture prior models.

In a long term future work, we will search for new prior models for the X-ray CT reconstruction. We will concentrate on the ill-posed cases where the number of projection is extremely limited. For that, we should optimize the prior model and avoid the artifacts appearing in the reconstruction results. We are also going to investigate the relation of the regularization methods and the Bayesian methods. There are many regularization methods which works very well for the ill-posed reconstruction problem if a suitable regularization parameter is chosen. We can study, from the Bayesian point of view, the principle of the regularization. At the same time, we can benefit the advantages of the Bayesian method that the parameters are estimated during

iterations and the hyper-parameters are relatively insensitive to the reconstruction results.

A

Posterior Mean estimation via Variational Bayesian Approach (VBA) algorithm for HHBM method

The Posterior Mean estimation for the hierarchical Haar transformation based Bayesian method is demonstrated in this Chapter. First of all, the posterior distribution is calculated via the likelihood and the prior distributions :

$$
\begin{aligned}
& p\left(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_\epsilon, \boldsymbol{v}_\xi, \boldsymbol{v}_z | \boldsymbol{g}\right) \\
& \propto p\left(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{v}_\epsilon\right) p\left(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{v}_\xi\right) p\left(\boldsymbol{z}|\boldsymbol{v}_z\right) p\left(\boldsymbol{v}_z|\alpha_{z_0}, \beta_{z_0}\right) p\left(\boldsymbol{v}_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}\right) p\left(\boldsymbol{v}_\xi|\alpha_{\xi_0}, \beta_{\xi_0}\right) \\
& \propto \mathcal{N}\left(\boldsymbol{g}|\boldsymbol{H}\boldsymbol{f}, \boldsymbol{V}_\epsilon\right) \mathcal{N}\left(\boldsymbol{f}|\boldsymbol{D}\boldsymbol{z}, \boldsymbol{V}_\xi\right) \mathcal{N}\left(\boldsymbol{z}|0, \boldsymbol{V}_z\right) \mathcal{IG}\left(\boldsymbol{z}|\alpha_{z_0}, \beta_{z_0}\right) \mathcal{IG}\left(\boldsymbol{v}_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}\right) \mathcal{IG}\left(\boldsymbol{v}_\xi|\alpha_{\xi_0}, \beta_{\xi_0}\right) \\
& = \left|\boldsymbol{V}_\epsilon\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\right)^T \boldsymbol{V}_\epsilon^{-1}\left(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\right)\right\} \left|\boldsymbol{V}_\xi\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\right)^T \boldsymbol{V}_\xi^{-1}\left(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\right)\right\} \\
& \quad \left|\boldsymbol{V}_z\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{z}^T \boldsymbol{V}_z^{-1}\boldsymbol{z}\right\} \prod_{j=1}^{N}\left(v_{z_j}^{-(\alpha_{z_0}+1)} \exp\left\{-\beta_{z_0} v_{z_j}^{-1}\right\}\right) \\
& \quad \prod_{i=1}^{M}\left(v_{\epsilon_i}^{-(\alpha_{\epsilon_0}+1)} \exp\left\{-\beta_{\epsilon_0} v_{\epsilon_i}^{-1}\right\}\right) \prod_{j=1}^{N}\left(v_{\xi_j}^{-(\alpha_{\xi_0}+1)} \exp\left\{-\beta_{\xi_0} v_{\xi_j}^{-1}\right\}\right) \\
& = \prod_{i=1}^{M} v_{\epsilon_i}^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\right)^T \boldsymbol{V}_\epsilon^{-1}\left(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\right)\right\} \prod_{j=1}^{N} v_{\xi_j}^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\right)^T \boldsymbol{V}_\xi^{-1}\left(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\right)\right\} \\
& \quad \prod_{j=1}^{N} v_{z_j}^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{z}^T \boldsymbol{V}_z^{-1}\boldsymbol{z}\right\} \prod_{j=1}^{N} v_{z_j}^{-(\alpha_{z_0}+1)} \prod_{j=1}^{N} \exp\left\{-\beta_{z_0} v_{z_j}^{-1}\right\} \\
& \quad \prod_{i=1}^{M} v_{\epsilon_i}^{-(\alpha_{\epsilon_0}+1)} \prod_{i=1}^{M} \exp\left\{-\beta_{\epsilon_0} v_{\epsilon_i}^{-1}\right\} \prod_{j=1}^{N} v_{\xi_j}^{-(\alpha_{\xi_0}+1)} \prod_{j=1}^{N} \exp\left\{-\beta_{\xi_0} v_{\xi_j}^{-1}\right\}.
\end{aligned}
\tag{A.1}
$$

First of all we consider a simplified model, in which the variance of noise $\boldsymbol{\epsilon}$ and $\boldsymbol{\xi}$ are scalars, so we have :

$$\begin{aligned}
&p(\boldsymbol{f}, \boldsymbol{z}, v_\epsilon, v_\xi, \boldsymbol{v}_z|\boldsymbol{g}) \\
&\propto v_\epsilon^{-\frac{M}{2}} \exp\left\{-\frac{1}{2} v_\epsilon^{-1} \left\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\right\|_2^2\right\} v_\xi^{-\frac{N}{2}} \exp\left\{-\frac{1}{2} v_\xi^{-1} \left\|\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\right\|_2^2\right\} \\
&\prod_{j=1}^{N} v_{z_j}^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{z}^T \boldsymbol{V}_z^{-1} \boldsymbol{z}\right\} \prod_{j=1}^{N} v_{z_j}^{-(\alpha_{z_0}+1)} \prod_{j=1}^{N} \exp\left\{-\beta_{z_0} v_{z_j}^{-1}\right\} \\
&\prod_{i=1}^{M} v_{\epsilon_i}^{-(\alpha_{\epsilon_0}+1)} \prod_{i=1}^{M} \exp\left\{-\beta_{\epsilon_0} v_{\epsilon_i}^{-1}\right\} \prod_{j=1}^{N} v_{\xi_j}^{-(\alpha_{\xi_0}+1)} \prod_{j=1}^{N} \exp\left\{-\beta_{\xi_0} v_{\xi_j}^{-1}\right\}
\end{aligned} \tag{A.2}$$

With the VBA method, the posterior law is approximated by a separable law $q\left(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_z, \boldsymbol{v}_\epsilon, \boldsymbol{v}_\xi\right)$:

$$p\left(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_z, v_\epsilon, v_\xi|\boldsymbol{g}\right) \simeq q\left(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_z, \boldsymbol{v}_\epsilon, \boldsymbol{v}_\xi\right) = \prod_{j=1}^{N} q_{1j}\left(f_j\right) \prod_{j=1}^{N} q_{2j}\left(z_j\right) q_{3j}\left(v_{z_j}\right) q_4\left(v_\epsilon\right) q_5\left(v_\xi\right)$$

The approximate distribution $q\left(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_z, v_\epsilon, v_\xi\right)$ is obtained via the minimization of the Kullback Leibler divergence.

$$KL\left(q:p\right) = \int\int\cdots\int q\left(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_z, v_\epsilon, v_\xi\right) \ln \frac{q\left(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_z, v_\epsilon, v_\xi\right)}{p\left(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_z, v_\epsilon, v_\xi \mid \boldsymbol{g}\right)} \mathrm{d}\boldsymbol{f}\ \mathrm{d}\boldsymbol{z}\ \mathrm{d}\boldsymbol{v}_z\ \mathrm{d}v_\epsilon\ \mathrm{d}v_\xi.$$

Now we want to minimize the term $KL\left(q:p\right)$, for a hierarchical problem, we will optimize the parameters alternatively.

$$q_{1j}\left(f_j\right) \propto \exp\left\{\left\langle \ln p\left(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_z, v_\epsilon, v_\xi|\boldsymbol{g}\right)\right\rangle_{q_{\substack{1,m\\m\neq j}}\left(f_m\right) q_2(\boldsymbol{z}) q_3(\boldsymbol{v}_z) q_4(v_\epsilon) q_5\left(v_\xi\right)}\right\} \tag{A.3}$$

$$q_{2j}\left(z_j\right) \propto \exp\left\{\left\langle \ln p\left(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_z, v_\epsilon, v_\xi|\boldsymbol{g}\right)\right\rangle_{q_1(\boldsymbol{f}) q_{\substack{2,m\\m\neq j}}\left(z_m\right) q_3(\boldsymbol{v}_z) q_4(v_\epsilon) q_5\left(v_\xi\right)}\right\} \tag{A.4}$$

$$q_{3j}\left(v_{z_j}\right) \propto \exp\left\{\left\langle \ln p\left(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_z, v_\epsilon, v_\xi|\boldsymbol{g}\right)\right\rangle_{q_1(\boldsymbol{f}) q_2(\boldsymbol{z}) q_{\substack{3,m\\m\neq j}}\left(v_{zm}\right) q_4(v_\epsilon) q_5\left(v_\xi\right)}\right\} \tag{A.5}$$

$$q_4\left(v_\epsilon\right) \propto \exp\left\{\left\langle \ln p\left(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_z, v_\epsilon, v_\xi|\boldsymbol{g}\right)\right\rangle_{q_1(\boldsymbol{f}) q_2(\boldsymbol{z}) q_3(\boldsymbol{v}_z) q_5\left(v_\xi\right)}\right\} \tag{A.6}$$

$$q_5\left(v_\xi\right) \propto \exp\left\{\left\langle \ln p\left(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_z, v_\epsilon, v_\xi|\boldsymbol{g}\right)\right\rangle_{q_1(\boldsymbol{f}) q_2(\boldsymbol{z}) q_3(\boldsymbol{v}_z) q_4(v_\epsilon)}\right\}, \tag{A.7}$$

where the defenition of the notation $\left\langle u\left(x\right)\right\rangle_{v(x)}$ is that $\left\langle u\left(x\right)\right\rangle_{v(x)} = \int u\left(x\right) v\left(x\right) dx$.

In the following expressions, the notation $\mathcal{C}$ represents a constant.

A.1 $q_{1j}\left(f_j\right)$

First step is to estimate the distribution on f_j :

$$
\begin{aligned}
q_{1j}\left(f_j\right) &\propto \exp\left\{\left\langle -\frac{1}{2}v_\epsilon^{-1}\left\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\right\|_2^2-\frac{1}{2}v_\xi^{-1}\left\|\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\right\|_2^2\right\rangle_{q_{\substack{1,m\\m\neq j}}(f_m)q_2(\boldsymbol{z})q_3(\boldsymbol{v}_z)q_4(v_\epsilon)q_5\left(v_\xi\right)}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left\langle v_\epsilon^{-1}\right\rangle_{q_4(v_\epsilon)}\left\langle\left\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\right\|_2^2\right\rangle_{q_{\substack{1,m\\m\neq j}}(f_m)}-\frac{1}{2}\left\langle v_\xi^{-1}\right\rangle_{q_5\left(v_\xi\right)}\left\langle\left\|\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\right\|\right\rangle_{q_{\substack{1,m\\m\neq j}}(f_m)q_2(\boldsymbol{z})}\right\}.
\end{aligned}
\tag{A.8}
$$

A.1.1 Computation of $\left\langle\left\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\right\|_2^2\right\rangle_{q_{\substack{1,m\\m\neq j}}(f_m)}$

We have :

$$
\begin{aligned}
&\left\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\right\| \\
=&\left|g_1-H_{11}f_1-H_{12}f_2-\cdots-H_{1N}f_N\right|^2+\left|g_2-H_{21}f_1-H_{22}f_2-\cdots-H_{2N}f_N\right|^2+ \\
&\cdots+\left|g_M-H_{M1}f_1-H_{M2}f_2-\cdots-H_{MN}f_N\right|^2 \\
=&H_{1j}^2f_j^2+2H_{1j}f_j(H_{11}f_1+H_{12}f_2+\cdots+H_{1N}f_N-H_{1j}f_j)-2H_{1j}f_jg_1 \\
&+H_{2j}^2f_j^2+2H_{2j}f_j(H_{21}f_1+H_{22}f_2+\cdots+H_{2N}f_N-H_{2j}f_j)-2H_{2j}f_jg_2 \\
&+H_{Mj}^2f_j^2+2H_{Mj}f_j(H_{M1}f_1+H_{M2}f_2+\cdots+H_{MN}f_N-H_{Mj}f_j)-2H_{Mj}f_jg_M+\mathcal{C} \\
=&\left(H_{1j}^2+H_{2j}^2+\cdots+H_{Mj}^2\right)f_j^2-2\left(H_{1j}g_1+H_{2j}g_2+\cdots+H_{Mj}g_M\right)f_j \\
&+2f_j\left\{\left(H_{1j}H_{12}+H_{2j}H_{22}+\cdots+H_{Mj}H_{M2}\right)f_2\right. \\
&+\cdots \\
&+\left(H_{1j}H_{1j-1}+H_{2j}H_{2j-1}+\cdots+H_{Mj}H_{Mj-1}\right)f_{j-1} \\
&+\left(H_{1j}H_{1j+1}+H_{2j}H_{2j+1}+\cdots+H_{Mj}H_{Mj+1}\right)f_{j+1} \\
&+\cdots \\
&\left.+\left(H_{1j}H_{1N}+H_{2j}H_{2N}+\cdots+H_{Mj}H_{MN}\right)f_N\right\}+\mathcal{C}
\end{aligned}
\tag{A.9}
$$

Where $\mathcal{C}$ is a polynome which has no relation with f_j, so that we have :

$$\begin{aligned}\left\langle \|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|^2\right\rangle_{\substack{q_{\,1,m\\ m\neq j}}(f_m)} =& \left(H_{1j}^2+H_{2j}^2+\cdots+H_{Mj}^2\right)f_j^2+2f_j\Big\{\\ &-\left(H_{1j}g_1+H_{2j}g_2+\cdots+H_{Mj}g_M\right)\\ &+\left(H_{1j}H_{11}+H_{2j}H_{21}+\cdots+H_{Mj}H_{M1}\right)\left\langle f_1\right\rangle_{q_{11}(f_1)}\\ &+\left(H_{1j}H_{12}+H_{2j}H_{22}+\cdots+H_{Mj}H_{M2}\right)\left\langle f_2\right\rangle_{q_{12}(f_2)}\\ &+\cdots\\ &+\left(H_{1j}H_{1j-1}+H_{2j}H_{2j-1}+\cdots+H_{Mj}H_{Mj-1}\right)\left\langle f_{j-1}\right\rangle_{q_{1j-1}(f_{j-1})}\\ &+\left(H_{1j}H_{1j+1}+H_{2j}H_{2j+1}+\cdots+H_{Mj}H_{Mj+1}\right)\left\langle f_{j+1}\right\rangle_{q_{1j+1}(f_{j+1})}\\ &+\cdots\\ &+\left(H_{1j}H_{1N}+H_{2j}H_{2N}+\cdots+H_{Mj}H_{MN}\right)\left\langle f_N\right\rangle_{q_{1N}(f_N)}\Big\}+\mathcal{C}\\ =&\mathcal{A}_f f_j^2+2\mathcal{B}_f f_j+\mathcal{C},\end{aligned} \tag{A.10}$$

where

$$\begin{aligned}\mathcal{A}_f &= H_{1j}^2+H_{2j}^2+\cdots+H_{Mj}^2=\left[\boldsymbol{H}^T\boldsymbol{H}\right]_{jj},\\ \mathcal{B}_f &= -\left(\boldsymbol{H}^T\boldsymbol{g}\right)_j+\left(\boldsymbol{H}^t\boldsymbol{H}\right)_{j1}\left\langle f_1\right\rangle_{q_{11}(f_1)}+\left(\boldsymbol{H}^t\boldsymbol{H}\right)_{j2}\left\langle f_2\right\rangle_{q_{12}(f_2)}+\cdots\\ &\quad+\left(\boldsymbol{H}^t\boldsymbol{H}\right)_{jN}\left\langle f_N\right\rangle_{q_{1N}(f_N)}-\left(\boldsymbol{H}^t\boldsymbol{H}\right)_{jj}\left\langle f_j\right\rangle_{q_{1j}(f_j)}\\ &= -\left(\boldsymbol{H}^T\boldsymbol{g}\right)_j+\left(\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{\mu}_f\right)_j-\left[\boldsymbol{H}^T\boldsymbol{H}\right]_{jj}\mu_{f_j} \qquad (A.11)\\ &= -\left[\boldsymbol{H}^T\left(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{\mu}_f\right)\right]_j-\left[\boldsymbol{H}^T\boldsymbol{H}\right]_{jj}\mu_{f_j}. \qquad (A.12)\end{aligned}$$

A.1.2 Computation of $\left\langle \|\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\|_2^2\right\rangle_{\substack{q_{\,1,m\\ m\neq j}}(f_m)q_2(\boldsymbol{z})}$

We have :

$$\begin{aligned}&\|\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\|_2^2\\ =&\left|f_1-\left(D_{11}z_1+D_{12}z_2+\cdots\right)\right|^2+\left|f_2-\left(D_{21}z_1+D_{22}z_2+\cdots\right)\right|^2+\cdots+\\ &\left|f_j-\left(D_{j1}z_1+D_{j2}z_2+\cdots\right)\right|^2+\cdots+\left|f_N-\left(D_{N1}z_1+D_{N2}z_2+\cdots\right)\right|^2\\ =&f_j^2-2\left(\boldsymbol{D}\boldsymbol{z}\right)_j f_j+\mathcal{C}\end{aligned} \tag{A.13}$$

A.1.3 Computation of $q_{1j}(f_j)$

From the previous computation we obtain :

$$\begin{aligned}q_{1j}(f_j) &\propto \exp\left\{-\frac{1}{2}\left\langle v_\epsilon^{-1}\right\rangle_{q_4(v_\epsilon)}\left(\mathcal{A}_f f_j^2+2\mathcal{B}_f f_j\right)-\frac{1}{2}\left\langle v_\xi^{-1}\right\rangle_{q_5(v_\xi)}\left(f_j^2-2\left(\boldsymbol{D}\boldsymbol{\mu}_z\right)_j f_j\right)\right\}\\ &\propto \exp\left\{-\frac{1}{2}J_f\right\},\end{aligned} \tag{A.14}$$

where

$$\begin{aligned}
J_f &= \left\langle v_\epsilon^{-1}\right\rangle_{q_4(v_\epsilon)}\left(\mathcal{A}_f f_j^2 + 2\mathcal{B}_f f_j\right) + \left\langle v_\xi^{-1}\right\rangle_{q_5\left(v_\xi\right)}\left(f_j^2 - 2\left(\boldsymbol{D}\boldsymbol{\mu}_z\right)_j f_j\right) \\
&= \left(\left\langle v_\epsilon^{-1}\right\rangle_{q_4(v_\epsilon)}\mathcal{A}_f + \left\langle v_\xi^{-1}\right\rangle_{q_5\left(v_\xi\right)}\right) f_j^2 + 2\left(\mathcal{B}_f\left\langle v_\epsilon^{-1}\right\rangle_{q_4(v_\epsilon)} - \left\langle v_\xi^{-1}\right\rangle_{q_5\left(v_\xi\right)}\left(\boldsymbol{D}\boldsymbol{\mu}_z\right)_j\right) f_j \\
&= \left(\left\langle v_\epsilon^{-1}\right\rangle_{q_4(v_\epsilon)}\mathcal{A}_f + \left\langle v_\xi^{-1}\right\rangle_{q_5\left(v_\xi\right)}\right)\left\{f_j^2 + 2\frac{\left\langle v_\epsilon^{-1}\right\rangle_{q_4(v_\epsilon)}\mathcal{B}_f - \left\langle v_\xi^{-1}\right\rangle_{q_5\left(v_\xi\right)}\left[\boldsymbol{D}\boldsymbol{\mu}_z\right]_j}{\left\langle v_\epsilon^{-1}\right\rangle_{q_4(v_\epsilon)}\mathcal{A}_f + \left\langle v_\xi^{-1}\right\rangle_{q_5\left(v_\xi\right)}} f_j\right\}.
\end{aligned}$$

So we find that f_j belongs to a Gaussian distribution, with mean and variance equal to :

$$\widehat{\mu}_{f_j} = -\frac{\left\langle \widehat{v_\epsilon}^{-1}\right\rangle_{q_4(v_\epsilon)}\widehat{\mathcal{B}}_f - \left\langle \widehat{v_\xi}^{-1}\right\rangle_{q_5\left(v_\xi\right)}\left[\boldsymbol{D}\boldsymbol{\mu}_z\right]_j}{\left\langle \widehat{v_\epsilon}^{-1}\right\rangle_{q_4(v_\epsilon)}\mathcal{A}_f + \left\langle \widehat{v_\xi}^{-1}\right\rangle_{q_5\left(v_\xi\right)}} \tag{A.15}$$

$$\widehat{\sigma}_{f_j} = \frac{1}{\left\langle \widehat{v_\epsilon}^{-1}\right\rangle_{q_4(v_\epsilon)}\mathcal{A}_f + \left\langle \widehat{v_\xi}^{-1}\right\rangle_{q_5\left(v_\xi\right)}} \tag{A.16}$$

A.2 Computation of $q_{2j}\left(z_j\right)$

From the criterion we have :

$$\begin{aligned}
q_{2j}\left(z_j\right) &\propto \exp\left\{\left\langle -\frac{1}{2}v_\xi^{-1}\left\|\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}\right\|_2^2 - \frac{1}{2}\boldsymbol{z}^T\boldsymbol{V}_z^{-1}\boldsymbol{z}\right\rangle_{q_1(\boldsymbol{f})q_{\substack{2,m\\m\neq j}}(z_m)q_3(\boldsymbol{v}_z)q_4(v_\epsilon)q_5\left(v_\xi\right)}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left\langle v_\xi^{-1}\right\rangle_{q_5\left(v_\xi\right)}\left\langle\left\|\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}\right\|_2^2\right\rangle - \frac{1}{2}\left\langle v_{z_j}^{-1}\right\rangle_{q_{3j}\left(v_{z_j}\right)} z_j^2\right\}.
\end{aligned} \tag{A.17}$$

With the same calculation as we've used for $\left\langle\left\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\right\|_2^2\right\rangle_{q_{\substack{1,m\\m\neq j}}(f_m)}$, we obtain :

$$\left\langle\left\|\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}\right\|_2^2\right\rangle_{q_1(\boldsymbol{f})q_{\substack{2,m\\m\neq j}}(z_m)} = \mathcal{A}_z z_j^2 + 2\mathcal{B}_z z_j + \mathcal{C}, \tag{A.18}$$

where

$$\mathcal{A}_z = \left[\boldsymbol{D}^T\boldsymbol{D}\right]_{jj} \tag{A.19}$$

$$\mathcal{B}_z = -\left[\boldsymbol{D}^T\boldsymbol{\mu}_f\right]_j + \left[\boldsymbol{D}^T\boldsymbol{D}\boldsymbol{\mu}_z\right]_j - \left[\boldsymbol{D}^T\boldsymbol{D}\right]_{jj}\mu_{z_j}. \tag{A.20}$$

So that we have :

$$\begin{aligned}
q_{2j}\left(z_j\right) &\propto \exp\left\{-\frac{1}{2}\left\langle v_\xi^{-1}\right\rangle_{q_5\left(v_\xi\right)}\left(\mathcal{A}_z z_j^2 + 2\mathcal{B}_z z_j\right) - \frac{1}{2}\left\langle v_{z_j}^{-1}\right\rangle_{q_{3j}\left(v_{z_j}\right)} z_j^2\right\} \\
&\propto \exp\left\{-\frac{1}{2}J_z\right\}
\end{aligned} \tag{A.21}$$

where

$$\begin{aligned} J_z &= \left\langle v_\xi^{-1}\right\rangle_{q_5\left(v_\xi\right)}\left(\mathcal{A}_z z_j^2 + 2\mathcal{B}_z z_j\right) + \left\langle v_{z_j}^{-1}\right\rangle_{q_{3j}\left(v_{z_j}\right)} z_j^2 \\ &= \left(\left\langle v_\xi^{-1}\right\rangle_{q_5\left(v_\xi\right)}\mathcal{A}_z + \left\langle v_{z_j}^{-1}\right\rangle_{q_{3j}\left(v_{z_j}\right)}\right) z_j^2 + 2\left\langle v_\xi^{-1}\right\rangle_{q_5\left(v_\xi\right)}\mathcal{B}_z z_j \\ &= \left(\left\langle v_\xi^{-1}\right\rangle_{q_5\left(v_\xi\right)}\mathcal{A}_z + \left\langle v_{z_j}^{-1}\right\rangle_{q_{3j}\left(v_{z_j}\right)}\right)\left(z_j^2 + 2\frac{\left\langle v_\xi^{-1}\right\rangle_{q_5\left(v_\xi\right)}\mathcal{B}_z}{\left\langle v_\xi^{-1}\right\rangle_{q_5\left(v_\xi\right)}\mathcal{A}_z + \left\langle v_{z_j}^{-1}\right\rangle_{q_{3j}\left(v_{z_j}\right)}} z_j\right). \end{aligned} \tag{A.22}$$

From the formulation we can see that z_j belong to a Gaussian distribution with mean and variance equal to :

$$\widehat{\mu}_{z_j} = -\frac{\left\langle \widehat{v_\xi}^{-1}\right\rangle_{q_5\left(v_\xi\right)}\widehat{\mathcal{B}}_z}{\left\langle \widehat{v_\xi}^{-1}\right\rangle_{q_5\left(v_\xi\right)}\mathcal{A}_z + \left\langle \widehat{v}_{z_j}^{-1}\right\rangle_{q_{3j}\left(v_{z_j}\right)}} \tag{A.23}$$

$$\widehat{\sigma}_{z_j} = \frac{1}{\left\langle \widehat{v_\xi}^{-1}\right\rangle_{q_5\left(v_\xi\right)}\mathcal{A}_z + \left\langle \widehat{v}_{z_j}^{-1}\right\rangle_{q_{3j}\left(v_{z_j}\right)}} \tag{A.24}$$

A.3 Computation of $q_{3j}\left(v_{z_j}\right)$

From Eq.(A.5), we have :

$$\begin{aligned} q_{3j}\left(v_{z_j}\right) &\propto \exp\left\{\left\langle -\frac{1}{2}\ln v_{z_j} - \frac{1}{2}v_{z_j}^{-1}z_j^2 - \left(\alpha_{z_0}+1\right)\ln v_{z_j} - \beta_{z_0}v_{z_j}^{-1}\right\rangle_{q_1(\boldsymbol{f})q_2(\boldsymbol{z})q_{\substack{3,m\\m\neq j}}(v_{z_m})q_4(v_\epsilon)q_5\left(v_\xi\right)}\right\} \\ &\propto \exp\left\{-\left(\alpha_{z_0}+\frac{1}{2}+1\right)\ln v_{z_j} - \left(\beta_{z_0}+\frac{1}{2}\left\langle z_j^2\right\rangle_{q_{2j}(z_j)}\right)v_{z_j}^{-1}\right\} \end{aligned} \tag{A.25}$$

From the definition of variance we have $\left\langle z_j^2\right\rangle_{q_{2j}(z_j)} = \mathrm{Var}(z_j) + \left\langle z_j\right\rangle_{q_{2j}(z_j)}^2 = \mu_{z_j}^2 + \sigma_{z_j}$, so that :

$$\widehat{\alpha}_{z_j} = \alpha_{z_0} + \frac{1}{2} \tag{A.26}$$

$$\widehat{\beta}_{z_j} = \beta_{z_0} + \frac{1}{2}\left(\widehat{\mu}_{z_j}^2 + \widehat{\sigma}_{z_j}\right) \tag{A.27}$$

A.4 Computation of $q_4\left(v_\epsilon\right)$

From Eq.(A.6) we have :

$$\begin{aligned} q_4\left(v_\epsilon\right) &\propto \exp\left\{\left\langle -\left(\alpha_{\epsilon_0}+\frac{M}{2}+1\right)\ln v_\epsilon - \beta_{\epsilon_0}v_\epsilon^{-1} - \frac{1}{2}v_\epsilon^{-1}\left\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\right\|_2^2\right\rangle_{q_1(\boldsymbol{f})q_2(\boldsymbol{z})q_3(\boldsymbol{v}_z)q_5\left(v_\xi\right)}\right\} \\ &\propto \exp\left\{-\left(\alpha_{\epsilon_0}+\frac{M}{2}+1\right)\ln v_\epsilon - \left(\beta_{\epsilon_0}+\frac{1}{2}\left\langle\left\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\right\|_2^2\right\rangle_{q_1(\boldsymbol{f})}\right)v_\epsilon^{-1}\right\} \end{aligned} \tag{A.28}$$

A.4.1 Computation of $\left\langle\left\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\right\|_2^2\right\rangle_{q_1(\boldsymbol{f})}$

The computation begin with :

$$\begin{aligned} &\left\langle\left\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\right\|^2\right\rangle_{q_1(f)} = \left\langle\boldsymbol{g}^t\boldsymbol{g}-2\boldsymbol{g}^t\boldsymbol{H}\boldsymbol{f}+\boldsymbol{f}^t\boldsymbol{H}^t\boldsymbol{H}\boldsymbol{f}\right\rangle_{q_1(f)} \\ =&\left\langle\boldsymbol{g}^t\boldsymbol{g}-2\boldsymbol{g}^t\boldsymbol{H}\boldsymbol{f}+\boldsymbol{f}^t\boldsymbol{H}^t\boldsymbol{H}\boldsymbol{\mu}+\boldsymbol{\mu}^t\boldsymbol{H}^t\boldsymbol{H}\boldsymbol{f}-\boldsymbol{\mu}^t\boldsymbol{H}^t\boldsymbol{H}\boldsymbol{\mu}+\left(\boldsymbol{f}^t-\boldsymbol{\mu}^t\right)\boldsymbol{H}^t\boldsymbol{H}\left(\boldsymbol{f}-\boldsymbol{\mu}\right)\right\rangle_{q_1(f)} \\ =&\boldsymbol{g}^t\boldsymbol{g}-2\boldsymbol{g}^t\boldsymbol{H}\boldsymbol{\mu}+2\boldsymbol{\mu}^t\boldsymbol{H}^t\boldsymbol{H}\boldsymbol{\mu}-\boldsymbol{\mu}^t\boldsymbol{H}^t\boldsymbol{H}\boldsymbol{\mu}+\left\langle\left(\boldsymbol{f}-\boldsymbol{\mu}\right)^t\boldsymbol{H}^t\boldsymbol{H}\left(\boldsymbol{f}-\boldsymbol{\mu}\right)\right\rangle_{q_1(f)}. \end{aligned} \tag{A.29}$$

Because the term $\left(\boldsymbol{f}-\boldsymbol{\mu}\right)^t\boldsymbol{H}^t\boldsymbol{H}\left(\boldsymbol{f}-\boldsymbol{\mu}\right)$ is a scalar, the $Trace$ of it is the same.

$$\left\langle\left\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\right\|^2\right\rangle_{q_1(f)} =\boldsymbol{g}^t\boldsymbol{g}-2\boldsymbol{g}^t\boldsymbol{H}\boldsymbol{\mu}+\boldsymbol{\mu}^t\boldsymbol{H}^t\boldsymbol{H}\boldsymbol{\mu}+\left\langle Tr\left(\left(\boldsymbol{f}-\boldsymbol{\mu}\right)^t\boldsymbol{H}^t\boldsymbol{H}\left(\boldsymbol{f}-\boldsymbol{\mu}\right)\right)\right\rangle_{q_1(f)}. \tag{A.30}$$

By using the property that : $Tr\left(\boldsymbol{A}\boldsymbol{B}\right)=Tr\left(\boldsymbol{B}\boldsymbol{A}\right)$, we obtain :

$$\begin{aligned} \left\langle\left\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\right\|^2\right\rangle_{q_1(f)} =&\boldsymbol{g}^t\boldsymbol{g}-2\boldsymbol{g}^t\boldsymbol{H}\boldsymbol{\mu}+\boldsymbol{\mu}^t\boldsymbol{H}^t\boldsymbol{H}\boldsymbol{\mu}+\left\langle Tr\left(\boldsymbol{H}^t\boldsymbol{H}\left(\boldsymbol{f}-\boldsymbol{\mu}\right)\left(\boldsymbol{f}-\boldsymbol{\mu}\right)^t\right)\right\rangle_{q_1(f)} \\ =&\left\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{\mu}\right\|^2+Tr\left(\boldsymbol{H}^t\boldsymbol{H}\left\langle\left(\boldsymbol{f}-\boldsymbol{\mu}\right)\left(\boldsymbol{f}-\boldsymbol{\mu}\right)^t\right\rangle_{q_1(f)}\right) \\ =&\left\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{\mu}\right\|^2+Tr\left(\boldsymbol{H}^t\boldsymbol{H}\boldsymbol{\Sigma}_{\boldsymbol{f}}\right). \end{aligned}$$

Finally we find that v_ϵ belong to an Inverse Gamma distribution with parameters equal to :

$$\widehat{\alpha}_\epsilon = \alpha_{\epsilon_0}+\frac{M}{2} \tag{A.31}$$

$$\widehat{\beta}_\epsilon = \beta_{\epsilon_0}+\frac{1}{2}\left(\left\|\boldsymbol{g}-\boldsymbol{H}\widehat{\boldsymbol{\mu}}_f\right\|_2^2+Tr\left(\boldsymbol{H}^T\boldsymbol{H}\widehat{\Sigma}_f\right)\right) \tag{A.32}$$

where $\widehat{\Sigma}_f=\text{diag}\left[\left[\cdots,\widehat{\sigma}_{f_j},\cdots\right]\right]$.

A.5 Computation of $q_5\left(v_\xi\right)$

From Eq.(A.7), we obtain :

$$\begin{aligned} q_5\left(v_\xi\right) &\propto \exp\left\{\left\langle -\left(\alpha_{\xi_0}+\frac{N}{2}+1\right)\ln v_\xi - \beta_{\xi_0}v_\xi^{-1} - \frac{1}{2}v_\xi^{-1}\left\|\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\right\|_2^2 \right\rangle_{q_1(\boldsymbol{f})q_2(\boldsymbol{z})q_3(\boldsymbol{v}_z)q_4(v_\epsilon)}\right\} \\ &\propto \exp\left\{-\left(\alpha_{\xi_0}+\frac{N}{2}+1\right)\ln v_\xi - \left(\beta_{\xi_0}+\frac{1}{2}\left\langle\left\|\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\right\|_2^2\right\rangle_{q_1(\boldsymbol{f})q_2(\boldsymbol{z})}\right)v_\xi^{-1}\right\} \end{aligned} \tag{A.33}$$

By using the same method as for the calculation of $\left\langle \|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|_2^2 \right\rangle_{q_1(\boldsymbol{f})}$, we have :

$$\left\langle \left\|\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\right\|_2^2 \right\rangle_{q_1(\boldsymbol{f})q_2(\boldsymbol{z})} = \left\|\boldsymbol{\mu}_f-\boldsymbol{D}\boldsymbol{\mu}_z\right\|_2^2 + Tr\left(\boldsymbol{D}^T\boldsymbol{D}\Sigma_z\right)$$

And finally we have that v_ξ belongs to an Inverse Gamma distribution with the parameters :

$$\widehat{\alpha}_\xi = \alpha_{\xi_0}+\frac{N}{2} \tag{A.34}$$

$$\widehat{\beta}_\xi = \beta_{\xi_0}+\frac{1}{2}\left(\left\|\widehat{\boldsymbol{\mu}}_f-\boldsymbol{D}\widehat{\boldsymbol{\mu}}_z\right\|_2^2 + Tr\left(\boldsymbol{D}^T\boldsymbol{D}\widehat{\Sigma}_z\right)\right). \tag{A.35}$$

where $\widehat{\Sigma}_z = \text{diag}\left[\left[\cdots,\widehat{\sigma}_{z_j},\cdots\right]\right]$.

A.6 The iterative optimization by using the VBA algorithm

From the previous calculation, we obtain the updating formulations for all the parameters of the approximated distribution of each variable :

$$q_{1j}\left(f_j\right) = \mathcal{N}(f_j|\widehat{\mu}_{f_j},\widehat{\sigma}_{f_j}) \text{ where } \begin{cases} \widehat{\mu}_{f_j} = -\dfrac{\left\langle \widehat{v_\epsilon}^{-1}\right\rangle_{q_4(v_\epsilon)}\widehat{\mathcal{B}}_f - \left\langle \widehat{v_\xi}^{-1}\right\rangle_{q_5\left(v_\xi\right)}\left[\boldsymbol{D}\widehat{\boldsymbol{\mu}}_z\right]_j}{\left\langle \widehat{v_\epsilon}^{-1}\right\rangle_{q_4(v_\epsilon)}\mathcal{A}_f + \left\langle \widehat{v_\xi}^{-1}\right\rangle_{q_5\left(v_\xi\right)}} \\ \widehat{\sigma}_{f_j} = \dfrac{1}{\left\langle \widehat{v_\epsilon}^{-1}\right\rangle_{q_4(v_\epsilon)}\mathcal{A}_f + \left\langle \widehat{v_\xi}^{-1}\right\rangle_{q_5\left(v_\xi\right)}} \end{cases}$$

$$q_{2j}\left(z_j\right) = \mathcal{N}(z_j|\widehat{\mu}_{z_j},\widehat{\sigma}_{z_j}) \text{ where } \begin{cases} \widehat{\mu}_{z_j} = -\dfrac{\left\langle \widehat{v_\xi}^{-1}\right\rangle_{q_5\left(v_\xi\right)}\widehat{\mathcal{B}}_z}{\left\langle \widehat{v_\xi}^{-1}\right\rangle_{q_5\left(v_\xi\right)}\mathcal{A}_z + \left\langle \widehat{v}_{z_j}^{-1}\right\rangle_{q_{3j}\left(v_{z_j}\right)}} \\ \widehat{\sigma}_{z_j} = \dfrac{1}{\left\langle \widehat{v_\xi}^{-1}\right\rangle_{q_5\left(v_\xi\right)}\mathcal{A}_z + \left\langle \widehat{v}_{z_j}^{-1}\right\rangle_{q_{3j}\left(v_{z_j}\right)}} \end{cases}$$

$$q_{3j}\left(v_{z_j}\right) = \mathcal{IG}(v_{z_j}|\widehat{\alpha}_{z_j},\widehat{\beta}_{z_j}) \text{ where } \begin{cases} \widehat{\alpha}_{z_j} = \alpha_{z_0}+\frac{1}{2} \\ \widehat{\beta}_{z_j} = \beta_{z_0}+\frac{1}{2}\left(\widehat{\mu}_{z_j}^2+\widehat{\sigma}_{z_j}\right) \end{cases}$$

$$q_4\left(v_\epsilon\right) = \mathcal{IG}(v_\epsilon|\widehat{\alpha}_\epsilon,\widehat{\beta}_\epsilon) \text{ where } \begin{cases} \widehat{\alpha}_\epsilon = \alpha_{\epsilon_0}+\frac{M}{2} \\ \widehat{\beta}_\epsilon = \beta_{\epsilon_0}+\frac{1}{2}\left(\left\|\boldsymbol{g}-\boldsymbol{H}\widehat{\boldsymbol{\mu}}_f\right\|_2^2 + Tr\left(\boldsymbol{H}^T\boldsymbol{H}\widehat{\Sigma}_f\right)\right) \end{cases}$$

$$q_5\left(v_\xi\right) = \mathcal{IG}(v_\xi|\widehat{\alpha}_\xi,\widehat{\beta}_\xi) \text{ where } \begin{cases} \widehat{\alpha}_\xi = \alpha_{\xi_0}+\frac{N}{2} \\ \widehat{\beta}_\xi = \beta_{\xi_0}+\frac{1}{2}\left(\left\|\widehat{\boldsymbol{\mu}}_f-\boldsymbol{D}\widehat{\boldsymbol{\mu}}_z\right\|_2^2 + Tr\left(\boldsymbol{D}^T\boldsymbol{D}\widehat{\Sigma}_z\right)\right) \end{cases}$$

where

$$\begin{aligned}
&\mathcal{A}_f = \left[\boldsymbol{H}^T\boldsymbol{H}\right]_{jj}, \\
&\widehat{\mathcal{B}}_f = -[\boldsymbol{H}^T(\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{\mu}}_f)]_j - \left[\boldsymbol{H}^T\boldsymbol{H}\right]_{jj}\widehat{\mu}_{f_j}, \\
&\mathcal{A}_z = \left[\boldsymbol{D}^T\boldsymbol{D}\right]_{jj}, \\
&\widehat{\mathcal{B}}_z = -\left[\boldsymbol{D}^T\widehat{\boldsymbol{\mu}}_f\right]_j + \left[\boldsymbol{D}^T\boldsymbol{D}\widehat{\boldsymbol{\mu}}_z\right]_j - \left[\boldsymbol{D}^T\boldsymbol{D}\right]_{jj}\widehat{\mu}_{z_j}.
\end{aligned}$$

The computation of $\left\langle \widehat{v}_{z_j}^{-1} \right\rangle_{q_{3j}\left(v_{z_j}\right)}$, $\left\langle \widehat{v}_{\epsilon}^{-1} \right\rangle_{q_4(v_\epsilon)}$ and $\left\langle \widehat{v}_{\xi}^{-1} \right\rangle_{q_5\left(v_\xi\right)}$ depends on the estimated values of the hyper-parameters $\left(\widehat{\alpha}_{z_j}, \widehat{\beta}_{z_j}\right)$, $\left(\widehat{\alpha}_{\epsilon}, \widehat{\beta}_{\epsilon}\right)$ and $\left(\widehat{\alpha}_{\xi}, \widehat{\beta}_{\xi}\right)$. It is shown in the results that the three variances v_{z_j}, v_ϵ and v_ξ are all belonging to an Inverse Gamma distribution. Now let x belonging to an Inverse Gamma distribution, so :

$$p(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left\{-\frac{\beta}{x}\right\} \mathrm{d}x. \tag{A.36}$$

So we have :

$$\begin{aligned}
\left\langle x^{-1} \right\rangle &= \int \frac{1}{x}\frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left\{-\frac{\beta}{x}\right\} \mathrm{d}x \\
&= \int \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-2} \exp\left\{-\frac{\beta}{x}\right\} \mathrm{d}x \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int x^{-(\alpha+1)-1} \exp\left\{-\frac{\beta}{x}\right\} \mathrm{d}x \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}} \int \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)} x^{-(\alpha+1)-1} \exp\left\{-\frac{\beta}{x}\right\} \mathrm{d}x \\
&= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \frac{1}{\beta} \int p(x|\alpha+1,\beta)\, \mathrm{d}x \\
&= \frac{\alpha}{\beta}
\end{aligned} \tag{A.37}$$

So we have :

$$\begin{aligned}
&\left\langle \widehat{v}_{z_j}^{-1} \right\rangle_{q_{3j}\left(v_{z_j}\right)} = \widehat{\alpha}_{z_j}/\widehat{\beta}_{z_j}, \\
&\left\langle \widehat{v}_{\epsilon}^{-1} \right\rangle_{q_4(v_\epsilon)} = \widehat{\alpha}_{\epsilon}/\widehat{\beta}_{\epsilon}, \\
&\left\langle \widehat{v}_{\xi}^{-1} \right\rangle_{q_5\left(v_\xi\right)} = \widehat{\alpha}_{\xi}/\widehat{\beta}_{\xi}.
\end{aligned} \tag{A.38}$$

B

My contributions

This Appendix shows my main contributions during the thesis. The publications and the details are listed in Table B.1.

My work begins from the use of the basic Bayesian methods in CT reconstruction. ① [WGMD15] used the Bayesian method in 2D case with a stationary noise. Several conventional prior models are used and compared. In this paper, the work mainly concentrate on the use of the Bayesian method on the X-ray CT reconstruction problem. 2D simulations are realized by using Bayesian method.

In ② [WMDGD16], a hierarchical prior model is proposed to account for the sparse structure of the Haar Transformation coefficient of the image. We call this method the Hierarchical structured Haar based Bayesian Method (HHBM). In this method, the transformation coefficient is considered as a hidden variable, and is estimated simultaneously during reconstruction. The noise is modeled by a stationary Gaussian model with unknown variance, which is modeled by an Inverse Gamma distribution. The JMAP and the VBA methods are used for the optimization of the variables. The VBA method is realized in the 2D simulation with stationary noise model, however it is still computationally expensive.

In ③ [WMDG17b], the HHBM method is developed, by adding an additive uncertainty variable in the Haar Transformation prior model : $\boldsymbol{f} = \boldsymbol{H}\boldsymbol{z} + \boldsymbol{\xi}$. The simulation for 2D CT reconstruction is done. In this paper, a non-stationary noise model is considered. The variance of the noise of the forward projection model and the variance of the uncertainties of the forward transformation model are considered as unfixed parameters, and are modeled by Inverse Gamma distributions. The details and extensions of this conference paper is presented in the published journal article ④ [WMDG17d]. Then, in ⑤ [WMDGD17], the HHBM method is developed for 3D CT reconstruction, in which the ASTRA toolbox is used in order to accelerate the computation. A French version of this article is presented in ⑥ [WGMD17].

In the above published papers, the discrete Haar transformation coefficient is considered because it is computationally cheap and is suitable for the strict piecewise-constant or piecewise-continuous objects. However, sometimes the object is not strictly piecewise-constant, and some

other transformations may optimize the property of the proposed Bayesian method. In ⑦ [WMDG17a], the DT-CWT transformation coefficient is used instead of Haar transformation. A comparison of the performance of the two transformations is presented.

In ⑧ [WMDG17c], the ROCC method is proposed. In this method, We proposed a non-homogeneous Markovian prior model to define the piecewise homogeneous object. In this model, the prior distribution of the variable $\boldsymbol{f}$ depends on the gradient of the neighbours. Meanwhile, the gradient of the object is considered as a hidden variable in this model. In this article, a stationary noise model is considered and the 2D CT reconstruction is simulated. The 3D reconstruction with a non-stationary noise model is extended in the thesis.

In the submitted ⑨ IEEE journal paper, the HHBM method is considered in a 3D framework, for both simulated and real datasets. The semi-supervised extension for the initialization is demonstrated. The influence of the initialization for hyper-parameters and the theoretical basis for initializing the hyper-parameters are presented.

The following table shows the guideline.

Reference	Method	Sparse coefficient	Noise model ϵ	Optimization	Additional uncertainty ξ	Simulation realization
① MaxEnt 2015 [WGMD15]	Bayesian method	×	Stationary	JMAP	Non	2D
② ICASSP 2016 [WMDGD16]	HHBM	Haar	Stationary	JMAP/VBA	Non	2D
③ MaxEnt 2016 [WMDG17b] ④ FI journal [WMDG17d]	HHBM	Haar	Non-stationary	JMAP	Yes	2D
⑤ Fully3D 2017 [WMDGD17] ⑥ Gretsi 2017 [WGMD17]	HHBM	Haar	Non-stationary	JMAP	Yes	3D
⑦ Eusipco 2017 [WMDG17a]	HHBM	DT-CWT	Non-stationary	JMAP	Yes	2D
⑧ ICASSP 2017 [WMDG17c]	ROCC	gradient	Stationary	JMAP	×	2D
⑨ IEEE journal (submitted)	HHBM	Haar	Non-stationary Hyperparameter analyse	JMAP	Yes	3D

TABLE B.1 – Main contributions

C

Author's publications during 3 years of PhD

The list of my publications can be found online : https ://cv.archives-ouvertes.fr/li-wang-l2s.

Articles published in journals :

- [WMDG17d] **L.Wang**, A.Mohammad-Djafari and N.Gac, X-ray Computed Tomography using a sparsity enforcing prior model based on Haar transformation in a Bayesian framework, Fundamenta Informaticae, 2017, 155 (4), pp.449-480.
- **L.Wang**, A.Mohammad-Djafari, N.Gac and M.Dumitru, Sparse transformation coefficient based hierarchical Bayesian method for 3D X-ray Computed Tomography, IEEE Transaction on Image Processing. (Submitted)

Article in preparation to be submitted in journals :

- M.Dumitru, **L.Wang** and A.Mohammad-Djafari, A generalization of Student-t based on Normal-Inverse Gamma distribution and its application in controlling the sparsity rate.

Papers published in international conferences :

- [WGMD15] **L.Wang**, N.Gac and A.Mohammad-Djafari, Bayesian 3D X-ray computed tomography image reconstruction with a scaled Gaussian mixture prior model, in AIP Conference Proceedings (Vol. 1641, No. 1, pp. 556-563), AIP.
- [WMDGD16] **L.Wang**, A.Mohammad-Djafari, N.Gac and M.Dumitru, Computed tomography reconstruction based on a hierarchical model and variational Bayesian method, in Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International

Conference on (pp. 883-887), IEEE.

- [WMDG17b] **L.Wang**, A.Mohammad-Djafari and N.Gac, Bayesian X-ray computed tomography using a three-level hierarchical prior model, in AIP Conference Proceedings (Vol. 1853, No. 1, p. 060003), AIP Publishing.
- [WMDG17c] **L.Wang**, A.Mohammad-Djafari and N.Gac, X-ray Computed Tomography simultaneous image reconstruction and contour detection using a hierarchical Markovian model, in Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on (pp. 6070-6074), IEEE.
- [WMDGD17] **L.Wang**, A.Mohammad-Djafari, N.Gac and M.Dumitru, 3D X-ray Computed Tomography reconstruction using sparsity enforcing hierarchical model based on Haar transformation, in The 2017 International Conference on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine.
- [WMDG17a] **L.Wang**, A.Mohammad-Djafari and N.Gac, Bayesian method with sparsity enforcing prior of dual-tree complex wavelet transform coefficients for X-ray CT image reconstruction, in The 25th European Signal Processing Conference (EUSIPCO).
- [DWMDG17] M.Dumitru, **L.Wang**, A.Mohammad-Djafari and N.Gac, Model selection in the sparsity context for inverse problems in Bayesian framework, in 37th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering.
- [DLGMD17] M.Dumitru, **L.Wang**, N.Gac and A.Mohammad-Djafari, Performance comparison of Bayesian iterative algorithms for three classes of sparsity enforcing priors with application in computed tomography, in Image Processing (ICIP), 2017 IEEE International Conference on, IEEE.
- [DGWMD17] M.Dumitru, N.Gac, **L.Wang** and A.Mohammad-Djafari, Unsupervised sparsity enforcing iterative algorithms for 3D image reconstruction in X-ray computed tomography, in The 2017 International Conference on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine.

Papers published in national conferences :

- [WGMD17] **L.Wang**, N.Gac and A.Mohammad-Djafari, Reconstruction 3D en tomographie à rayons X à l'aide d'un modèle a priori hiérarchique utilisant la transformation de Haar, in Colloque GRETSI 2017.
- [DWGMD17] M.Dumitru, **L.Wang**, N.Gac and A.Mohammad-Djafari, Comparaison des performances d'algorithmes itératifs bayésiens basés sur trois classes de modeles a priori parcimonieux appliqués à la reconstruction tomographique, in Colloque GRETSI 2017.

Bibliographie

[ABDF10] Manya V Afonso, José M Bioucas-Dias, and Mário AT Figueiredo. Fast image recovery using variable splitting and constrained optimization. *IEEE Transactions on Image Processing*, 19(9) :2345–2356, 2010.

[ABDF11] Manya V Afonso, José M Bioucas-Dias, and Mário AT Figueiredo. An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Transactions on Image Processing*, 20(3) :681–695, 2011.

[ABMD92] Marc Antonini, Michel Barlaud, Pierre Mathieu, and Ingrid Daubechies. Image coding using wavelet transform. *IEEE Transactions on Image Processing*, 1(2) :205–220, 1992.

[AGU76] J Ambrose, MR Gooding, and D Uttley. EMI scan in the management of head injuries. *The Lancet*, 307(7964) :847–848, 1976.

[AK84] Anders H Andersen and Avinash C Kak. Simultaneous algebraic reconstruction technique (SART) : a superior implementation of the ART algorithm. *Ultrasonic Imaging*, 6(1) :81–94, 1984.

[AMD10] Hacheme Ayasso and Ali Mohammad-Djafari. Joint NDT image restoration and segmentation using Gauss–Markov–Potts prior models and variational Bayesian computation. *IEEE Transactions on Image Processing*, 19(9) :2265–2277, 2010.

[Arr99] Simon R Arridge. Optical tomography in medical imaging. *Inverse Problems*, 15(2) :R41, 1999.

[AU96] Akram Aldroubi and Michael Unser. *Wavelets in medicine and biology*. CRC press, 1996.

[AZ05] Mohammad Reza Ay and Habib Zaidi. Development and validation of MCNP4C-based Monte Carlo simulator for fan-and cone-beam x-ray CT. *Physics in Medicine and Biology*, 50(20) :4863, 2005.

[Bar07] Richard G Baraniuk. Compressive sensing [lecture notes]. *IEEE signal processing magazine*, 24(4) :118–121, 2007.

[Bat92] Roberto Battiti. First-and second-order methods for learning : between steepest descent and Newton's method. *Neural computation*, 4(2) :141–166, 1992.

[BKK12] Marcel Beister, Daniel Kolditz, and Willi A Kalender. Iterative reconstruction methods in X-ray CT. *Physica medica*, 28(2) :94–108, 2012.

[BKW96] Mickey Bhatia, William Clement Karl, and Alan S Willsky. A wavelet-based method for multiscale tomographic reconstruction. *IEEE Transactions on Medical Imaging*, 15(1) :92–101, 1996.

[BLO95] X Boespflug, BFN Long, and S Occhietti. CAT-scan in marine stratigraphy : a quantitative approach. *Marine Geology*, 122(4) :281–301, 1995.

[BMD08] Nadia Bali and Ali Mohammad-Djafari. Bayesian approach with hidden Markov modeling and mean field approximation for hyperspectral data analysis. *IEEE Transactions on Image Processing*, 17(2) :217–225, 2008.

[Bot10] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*, pages 177–186. Springer, 2010.

[BS05] Natalia Bochkina and Theofanis Sapatinas. On the posterior median estimators of possibly sparse sequences. *Annals of the Institute of Statistical Mathematics*, 57(2) :315–351, 2005.

[BS11] Kees Joost Batenburg and Jan Sijbers. DART : a practical reconstruction algorithm for discrete tomography. *IEEE Transactions on Image Processing*, 20(9) :2542–2553, 2011.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[BYL+00] Su Bangliang, Zhang Yiheng, Peng Lihui, Yao Danya, and Zhang Baofen. The use of simultaneous iterative reconstruction technique for electrical capacitance tomography. *Chemical Engineering Journal*, 77(1) :37–41, 2000.

[CB01] Adrian Corduneanu and Christopher M Bishop. Variational Bayesian model selection for mixture distributions. In *Artificial intelligence and Statistics*, volume 2001, pages 27–34. Waltham, MA : Morgan Kaufmann, 2001.

[CDDY06] Emmanuel Candes, Laurent Demanet, David Donoho, and Lexing Ying. Fast discrete curvelet transforms. *Multiscale Modeling & Simulation*, 5(3) :861–899, 2006.

[CDSB03] Roger L Claypoole, Geoffrey M Davis, Wim Sweldens, and Richard G Baraniuk. Nonlinear wavelet transforms for image coding via lifting. *IEEE Transactions on Image Processing*, 12(12) :1449–1459, 2003.

[CGM99] Tony F Chan, Gene H Golub, and Pep Mulet. A nonlinear primal-dual method for total variation-based image restoration. *SIAM journal on scientific computing*, 20(6) :1964–1977, 1999.

[CKL01] Yong Choi, Ja-Yong Koo, and Nam-Yong Lee. Image reconstruction using the wavelet transform for positron emission tomography. *IEEE Transactions on Medical Imaging*, 20(11) :1188–1193, 2001.

[CL97] Antonin Chambolle and Pierre-Louis Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(2) :167–188, 1997.

[CLMT02] Srdjan Coric, Miriam Leeser, Eric Miller, and Marc Trepanier. Parallel-beam backprojection : an FPGA implementation optimized for medical imaging. In *Proceedings of the 2002 ACM/SIGDA tenth international symposium on Field-programmable gate arrays*, pages 217–226. ACM, 2002.

[CM15] Nabil Chetih and Zoubeida Messali. Tomographic image reconstruction using filtered back projection (FBP) and algebraic reconstruction technique (ART). In *Control, Engineering & Information Technology (CEIT), 2015 3rd International Conference on*, pages 1–6. IEEE, 2015.

[CMRS00] D Calvetti, S Morigi, L Reichel, and F Sgallari. Tikhonov regularization and the L-curve for large discrete ill-posed problems. *Journal of Computational and Applied Mathematics*, 123(1) :423–446, 2000.

[DB95] Alexander H Delaney and Yoram Bresler. Multiresolution tomographic reconstruction using wavelets. *IEEE Transactions on Image Processing*, 4(6) :799–813, 1995.

[Dea07] Stanley R Deans. *The Radon transform and some of its applications*. Courier Corporation, 2007.

[DGWMD17] Mircea Dumitru, Nicolas Gac, Li Wang, and Ali Mohammad-Djafari. Unsupervised sparsity enforcing iterative algorithms for 3D image reconstruction in X-ray computed tomography. In *The 2017 International Conference on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2017.

[DLGMD17] Mircea Dumitru, Wang Li, Nicolas Gac, and Ali Mohammad-Djafari. Performance comparison of Bayesian iterative algorithms for three classes of sparsity enforcing priors with application in computed tomography. In *2017 IEEE International Conference on Image Processing*, 2017.

[Don06] David L Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4) :1289–1306, 2006.

[DS07] Lucia Dettori and Lindsay Semler. A comparison of wavelet, ridgelet, and curvelet-based texture classification algorithms in computed tomography. *Computers in biology and medicine*, 37(4) :486–498, 2007.

[Dum16] Mircea Dumitru. *A Bayesian approach for periodic components estimation for chronobiological signals*. PhD thesis, Paris Saclay, 2016.

[DV05] Minh N Do and Martin Vetterli. The contourlet transform : an efficient directional multiresolution image representation. *IEEE Transactions on Image Processing*, 14(12) :2091–2106, 2005.

[DWGMD17] Mircea Dumitru, Li Wang, Nicolas Gac, and Ali Mohammad-Djafari. Comparaison des performances d'algorithmes itératifs bayésiens basés sur trois classes de modeles a priori parcimonieux appliqués à la reconstruction tomographique. In *GRETSI*, 2017.

[DWMDG17] Mircea Dumitru, Li Wang, Ali Mohammad-Djafari, and Nicolas Gac. Model selection in the sparsity context for inverse problems in Bayesian framework. In *37th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, 2017.

[EMR07] Michael Elad, Peyman Milanfar, and Ron Rubinstein. Analysis versus synthesis in signal priors. *Inverse Problems*, 23(3) :947, 2007.

[Eri69] William A Ericson. A note on the posterior mean of a population mean. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 332–334, 1969.

[FB99] Jeffrey A Fessler and Scott D Booth. Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction. *IEEE Transactions on Image Processing*, 8(5) :688–699, 1999.

[FBS02] Thomas Frese, Charles A Bouman, and Ken Sauer. Adaptive wavelet graph model for Bayesian tomographic reconstruction. *IEEE Transactions on Image Processing*, 11(7) :756–770, 2002.

[FDK84] LA Feldkamp, LC Davis, and JW Kress. Practical cone-beam algorithm. *JOSA A*, 1(6) :612–619, 1984.

[Fis92] Andreas Fischer. A special Newton-type optimization method. *Optimization*, 24(3-4) :269–284, 1992.

[FR12] Charles W Fox and Stephen J Roberts. A tutorial on variational Bayesian inference. *Artificial intelligence review*, 38(2) :85–95, 2012.

[FSH15] Matthias Feurer, Jost Tobias Springenberg, and Frank Hutter. Initializing Bayesian Hyperparameter Optimization via Meta-learning. 2015.

[GBDBM74] J Gawler, JWD Bull, GH Du Boulay, and John Marshall. Computer-assisted tomography (EMI scanner) : Its place in investigation of suspected intracranial tumours. *The Lancet*, 304(7878) :419–423, 1974.

[GBH70] Richard Gordon, Robert Bender, and Gabor T Herman. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *Journal of theoretical Biology*, 29(3) :471IN1477–476IN2481, 1970.

[GCSR14] Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. *Bayesian data analysis*, volume 2. Chapman & Hall/CRC Boca Raton, FL, USA, 2014.

[GHW79] Gene H Golub, Michael Heath, and Grace Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2) :215–223, 1979.

[GLPS07] David Gerónimo, Antonio López, Daniel Ponsa, and Angel D Sappa. Haar wavelets and edge orientation histograms for on–board pedestrian detection. In *Iberian Conference on Pattern Recognition and Image Analysis*, pages 418–425. Springer, 2007.

[GO09] Tom Goldstein and Stanley Osher. The split Bregman method for L1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2) :323–343, 2009.

[Gor74] Richard Gordon. A tutorial on ART (algebraic reconstruction techniques). *IEEE Transactions on Nuclear Science*, 21(3) :78–93, 1974.

[GRL+93] Gene Gindi, Anand Rangarajan, M Lee, PJ Hong, and I George Zubal. Bayesian reconstruction for emission tomography via deterministic annealing. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 322–338. Springer, 1993.

[HFU08] Randolf Hanke, Theobald Fuchs, and Norman Uhlmann. X-ray based methods for non-destructive testing and material characterization. *Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment*, 591(1) :14–18, 2008.

[HL94] H Malcolm Hudson and Richard S Larkin. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Transactions on Medical Imaging*, 13(4) :601–609, 1994.

[HME+81] Forrest F Hopkins, Ira L Morgan, Hunter D Ellinger, Rudy V Klinksiek, Glenn A Meyer, and J Neils Thompson. Industrial tomography applications. *IEEE Transactions on Nuclear Science*, 28(2) :1717–1720, 1981.

[HO93] Per Christian Hansen and Dianne Prost O'Leary. The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM Journal on Scientific Computing*, 14(6) :1487–1503, 1993.

[Kel99] Carl T Kelley. *Iterative methods for optimization*. SIAM, 1999.

[Kin98] Nick G Kingsbury. The dual-tree complex wavelet transform : a new technique for shift invariance and directional filters. In *IEEE Digital Signal Processing Workshop*, volume 86, pages 120–131. Citeseer, 1998.

[Kle03] Lev Borisovič Klebanov. *Heavy tailed distributions*. Matfyzpress, 2003.

[KPSV09] Stefan Klein, Josien PW Pluim, Marius Staring, and Max A Viergever. Adaptive stochastic gradient descent optimisation for image registration. *International Journal of Computer Vision*, 81(3) :227–239, 2009.

[KVS+06] Ville Kolehmainen, Antti Vanne, Samuli Siltanen, Seppo Jarvenpaa, Jari P Kaipio, Matti Lassas, and Martti Kalke. Parallelized Bayesian inversion for three-dimensional dental X-ray imaging. *IEEE Transactions on Medical Imaging*, 25(2) :218–228, 2006.

[LNDD07] Ignace Loris, Guust Nolet, Ingrid Daubechies, and FA Dahlen. Tomographic inversion using l1-norm regularization of wavelet coefficients. *Geophysical Journal International*, 170(1) :359–370, 2007.

[LWGR74] RS Ledley, JB Wilson, T Golab, and LS Rotolo. The ACTA-scanner : the whole body computerized transaxial tomograph. *Computers in biology and medicine*, 4(2) :145IN7153–152IN10155, 1974.

[MD96] Ali Mohammad-Djafari. Joint estimation of parameters and hyperparameters in a Bayesian approach of solving inverse problems. In *Image Processing, 1996. Proceedings., International Conference on*, volume 2, pages 473–476. IEEE, 1996.

[MDA15] Dougal Maclaurin, David Duvenaud, and Ryan Adams. Gradient-based hyperparameter optimization through reversible learning. In *International Conference on Machine Learning*, pages 2113–2122, 2015.

[Moo96] Todd K Moon. The expectation-maximization algorithm. *IEEE Signal Processing Magazine*, 13(6) :47–60, 1996.

[NVL+13] Masih Nilchian, Cédric Vonesch, Stamatios Lefkimmiatis, Peter Modregger, Marco Stampanoni, and Michael Unser. Constrained regularized reconstruction of X-ray-DPCI tomograms with weighted-norm. *Optics Express*, 21(26) :32340–32348, 2013.

[NWS14] Deanna Needell, Rachel Ward, and Nati Srebro. Stochastic gradient descent, weighted sampling, and the randomized kaczmarz algorithm. In *Advances in Neural Information Processing Systems*, pages 1017–1025, 2014.

[OMA+76] AK Ommaya, G Murray, J Ambrose, A Richardson, and G Hounsfield. Computerized axial tomography : estimation of spatial and density resolution capability. *The British journal of radiology*, 49(583) :604–611, 1976.

[PA74] R Paxton and J Ambrose. The EMI scanner. A brief review of the first 650 patients. *The British journal of radiology*, 47(561) :530–565, 1974.

[PP10] Maria Petrou and Costas Petrou. *Image processing : the fundamentals*. John Wiley & Sons, 2010.

[PSWS03] Javier Portilla, Vasily Strela, Martin J Wainwright, and Eero P Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Transactions on Image Processing*, 12(11) :1338–1351, 2003.

[PTS+15] Françoise Peyrin, Alina Toma, Bruno Sixou, Loïc Denis, Andrew Burghardt, and Jean-Baptiste Pialat. Semi-blind joint super-resolution/segmentation of 3D trabecular bone images by a TV box approach. In *Signal Processing Conference (EUSIPCO), 2015 23rd European*, pages 2811–2815. IEEE, 2015.

[QL00] Jinyi Qi and Richard M Leahy. Resolution and noise properties of MAP reconstruction for fully 3-D PET. *IEEE Transactions on Medical Imaging*, 19(5) :493–506, 2000.

[Rad17] Johann Radon. Uber die bestimmug von funktionen durch ihre integralwerte laengs geweisser mannigfaltigkeiten. *Berichte Saechsishe Acad. Wissenschaft. Math. Phys., Klass*, 69 :262, 1917.

[RF12] Sathish Ramani and Jeffrey A Fessler. Statistical X-ray CT reconstruction using a splitting-based iterative algorithm with orthonormal wavelets. In *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1008–1011. IEEE, 2012.

[RLR+12] Sathish Ramani, Zhihao Liu, Jeffrey Rosen, Jon-Fredrik Nielsen, and Jeffrey A Fessler. Regularization parameter selection for nonlinear iterative image restoration and MRI reconstruction using GCV and sure-based methods. *IEEE Transactions on Image Processing*, 21(8) :3659–3672, 2012.

[RSD09] Simon Rit, David Sarrut, and Laurent Desbat. Comparison of analytic and algebraic methods for motion-compensated cone-beam CT reconstruction of the thorax. *IEEE Transactions on Medical Imaging*, 28(10) :1513–1525, 2009.

[RVJ+06] Maaria Rantala, Simopekka Vanska, Seppo Jarvenpaa, Martti Kalke, Matti Lassas, Jan Moberg, and Samuli Siltanen. Wavelet-based reconstruction for limited-angle X-ray tomography. *IEEE Transactions on Medical Imaging*, 25(2) :210–217, 2006.

[RW84] Richard A Redner and Homer F Walker. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, 26(2) :195–239, 1984.

[RY14] K Ramamohan Rao and Ping Yip. *Discrete cosine transform : algorithms, advantages, applications*. Academic press, 2014.

[SB92] Ken Sauer and Charles Bouman. Bayesian estimation of transmission tomograms using segmentation based optimization. *IEEE Transactions on Nuclear Science*, 39(4) :1144–1152, 1992.

[SB93] Ken Sauer and Charles Bouman. A local update strategy for iterative reconstruction from projections. *IEEE Transactions on Signal Processing*, 41(2) :534–548, 1993.

[SBK05] Ivan W Selesnick, Richard G Baraniuk, and Nick C Kingsbury. The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*, 22(6) :123–151, 2005.

[SBS98] Suhail S Saquib, Charles A Bouman, and Ken Sauer. ML parameter estimation for Markov random fields with applications to Bayesian tomography. *IEEE Transactions on Image Processing*, 7(7) :1029–1044, 1998.

[SC91] H Jacques Suermondt and Gregory F Cooper. Initialization for the method of conditioning in Bayesian belief networks. *Artificial Intelligence*, 50(1) :83–94, 1991.

[SC00] Predrag Sukovic and Neal H Clinthorne. Penalized weighted least-squares image reconstruction for dual energy X-ray transmission tomography. *IEEE Transactions on Medical Imaging*, 19(11) :1075–1081, 2000.

[Sca87] John A Scales. Tomographic inversion via the conjugate gradient method. *Geophysics*, 52(2) :179–185, 1987.

[SCD02] Jean-Luc Starck, Emmanuel J Candès, and David L Donoho. The curvelet transform for image denoising. *IEEE Transactions on Image Processing*, 11(6) :670–684, 2002.

[SF03] Radomir S Stanković and Bogdan J Falkowski. The Haar wavelet transform : its status and achievements. *Computers & Electrical Engineering*, 29(1) :25–44, 2003.

[SK13] D Sridhar and IV Murali Krishna. Brain tumor classification using discrete cosine transform and probabilistic neural network. In *Signal Processing Image Processing & Pattern Recognition (ICSIPR), 2013 International Conference on*, pages 92–96. IEEE, 2013.

[SP08] Emil Y Sidky and Xiaochuan Pan. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine and Biology*, 53(17) :4777, 2008.

[TET+04] DA Tregouet, S Escolano, L Tiret, A Mallet, and JL Golmard. A new algorithm for haplotype-based association analysis : the stochastic-EM algorithm. *Annals of human genetics*, 68(2) :165–177, 2004.

[Tib96] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[TL90] Jeannot Trampert and Jean-Jacques Leveque. Simultaneous iterative reconstruction technique : physical interpretation based on the generalized least squares solution. *J. geophys. Res*, 95(12) :553–9, 1990.

[TLG08] Dimitris G Tzikas, Aristidis C Likas, and Nikolaos P Galatsanos. The variational approximation for Bayesian inference. *IEEE Signal Processing Magazine*, 25(6) :131–146, 2008.

[Tre01] Volker Tresp. Mixtures of Gaussian processes. In *Advances in neural information processing systems*, pages 654–660, 2001.

[TSP+08] Christian Tenllado, Javier Setoain, Manuel Prieto, Luis Piñuel, and Francisco Tirado. Parallel implementation of the 2d discrete wavelet transform on graphics processing units : Filter bank versus lifting. *IEEE Transactions on Parallel and Distributed Systems*, 19(3) :299–310, 2008.

[TW09] Xue-Cheng Tai and Chunlin Wu. Augmented Lagrangian method, dual methods and split Bregman iteration for ROF model. *Scale space and variational methods in computer vision*, pages 502–513, 2009.

[vAPC+16] Wim van Aarle, Willem Jan Palenstijn, Jeroen Cant, Eline Janssens, Folkert Bleichrodt, Andrei Dabravolski, Jan De Beenhouwer, K Joost Batenburg, and Jan Sijbers. Fast and flexible X-ray tomography using the ASTRA toolbox. *Optics express*, 24(22) :25129–25147, 2016.

[VCR97] MA Vorontsov, GW Carhart, and JC Ricklin. Adaptive phase-distortion correction based on parallel gradient-descent optimization. *Optics letters*, 22(12) :907–909, 1997.

[VSK85] Yehuda Vardi, LA Shepp, and Linda Kaufman. A statistical model for positron emission tomography. *Journal of the American statistical Association*, 80(389) :8–20, 1985.

[WBAW12] Bo Wahlberg, Stephen Boyd, Mariette Annergren, and Yang Wang. An ADMM algorithm for a class of total variation regularized estimation problems. *IFAC Proceedings Volumes*, 45(16) :83–88, 2012.

[WGMD15] Li Wang, Nicolas Gac, and Ali Mohammad-Djafari. Bayesian 3D X-ray computed tomography image reconstruction with a scaled Gaussian mixture prior model. In *AIP Conference Proceedings*, volume 1641, pages 556–563. AIP, 2015.

[WGMD17] Li Wang, Nicolas Gac, and Ali Mohammad-Djafari. Reconstruction 3D en tomographie à rayons X à l'aide d'un modèle a priori hiérarchique utilisant la transformation de Haar. In *Colloque GRETSI 2017*, 2017.

[WLL14] D Wolf, A Lubk, and H Lichte. Weighted simultaneous iterative reconstruction technique for single-axis tomography. *Ultramicroscopy*, 136 :15–25, 2014.

[WMDG17a] Li Wang, Ali Mohammad-Djafari, and Nicolas Gac. Bayesian method with sparsity enforcing prior of dual-tree complex wavelet transform coefficients for X-ray CT image reconstruction. In *The 25th European Signal Processing Conference (EUSIPCO)*, 2017.

[WMDG17b] Li Wang, Ali Mohammad-Djafari, and Nicolas Gac. Bayesian X-ray computed tomography using a three-level hierarchical prior model. In *AIP Conference Proceedings*, volume 1853, page 060003. AIP Publishing, 2017.

[WMDG17c] Li Wang, Ali Mohammad-Djafari, and Nicolas Gac. X-ray Computed Tomography simultaneous image reconstruction and contour detection using a hierarchical Markovian model. In *The 42nd IEEE International Conference on Acoustics, Speech and Signal Processing*, 2017.

[WMDG17d] Li Wang, Ali Mohammad-Djafari, and Nicolas Gac. X-ray Computed Tomography using a sparsity enforcing prior model based on haar transformation in a Bayesian framework. *Fundamenta Informaticae*, 155(4) :449–480, 2017.

[WMDGD16] Li Wang, Ali Mohammad-Djafari, Nicolas Gac, and Mircea Dumitru. Computed tomography reconstruction based on a hierarchical model and variational Bayesian method. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 883–887. IEEE, 2016.

[WMDGD17] Li Wang, Ali Mohammad-Djafari, Nicolas Gac, and Mircea Dumitru. 3D X-ray Computed Tomography reconstruction using sparsity enforcing hierarchical model based on Haar transformation. In *The 2017 International Conference on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2017.

[WSQ09] Guobao Wang, Larry Schultz, and Jinyi Qi. Statistical image reconstruction for muon tomography using a Gaussian scale mixture model. *IEEE Transactions on Nuclear Science*, 56(4) :2480–2486, 2009.

[Wu01] Z Wu. Frequency and noise dependence of the image reconstruction of ground surfaces using the conjugate gradient based algorithm. *IEE Proceedings-Radar, Sonar and Navigation*, 148(4) :211–218, 2001.

[WV+87] Scott L Wellington, Harold J Vinegar, et al. X-ray computerized tomography. *Journal of Petroleum Technology*, 39(08) :885–898, 1987.

[WVVH10] David W Winters, Barry D Van Veen, and Susan C Hagness. A sparsity regularization approach to the electromagnetic inverse scattering problem. *IEEE Transactions on Antennas and Propagation*, 58(1) :145–154, 2010.

[WYYZ08] Yilun Wang, Junfeng Yang, Wotao Yin, and Yin Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3) :248–272, 2008.

[XYM+12] Qiong Xu, Hengyong Yu, Xuanqin Mou, Lei Zhang, Jiang Hsieh, and Ge Wang. Low-dose X-ray CT reconstruction via dictionary learning. *IEEE Transactions on Medical Imaging*, 31(9) :1682–1697, 2012.

[YAT+97] Tetsuya Yuasa, Masahiro Akiba, Tohoru Takeda, Masahiro Kazama, Atsunori Hoshino, Yuki Watanabe, Kazuyuki Hyodo, F Avraham Dilmanian, Takao Akatsuka, and Yuji Itai. Reconstruction method for fluorescent X-ray computed tomography by least-squares method using singular value decomposition. *IEEE Transactions on Nuclear Science*, 44(1) :54–62, 1997.

[ZFR15] Yuling Zheng, Aurélia Fraysse, and Thomas Rodet. Efficient variational Bayesian approximation method based on subspace optimization. *IEEE Transactions on Image Processing*, 24(2) :681–693, 2015.

[ZHL+16] Hao Zhang, Hao Han, Zhengrong Liang, Yifan Hu, Yan Liu, William Moore, Jianhua Ma, and Hongbing Lu. Extracting information from previous full-dose CT scan for knowledge-based Bayesian reconstruction of current low-dose CT images. *IEEE Transactions on Medical Imaging*, 35(3) :860–870, 2016.

[ZLQ97] Zhenyu Zhou, RN Leahy, and Jinyi Qi. Approximate maximum likelihood hyperparameter estimation for Gibbs priors. *IEEE transactions on image processing*, 6(6) :844–861, 1997.

université PARIS-SACLAY
ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre : Reconstruction d'image en tomographie 3D pour des applications en Contrôle Non Destructif (CND)

Mots clés : Tomographie, Problème inverse, Méthode Bayesian, Structure hiérarchique

Résumé : La tomographie 3D est largement utilisée dans l'imagerie médicale ainsi que dans le Contrôle Non Destructif (CND) pour l'industrie. Pour ces deux applications, il est nécessaire de réduire le nombre de projections. Dans certains cas, la reconstruction doit être faite avec un nombre d'angle de projections limité. Nous sommes donc presque toujours dans la situation de problèmes inversés mal posés. En plus, les données mesurées sont toujours avec des erreurs. Le rôle des méthodes probabilistes et de la modélisation a priori devient crucial.

Pour la modélisation a priori, en particulier dans les applications NDT, l'objet à l'examen est composé de plusieurs matériaux homogènes, avec plusieurs blocs continus séparés par des discontinuités et des contours. Ce type d'objet est dit continu par morceaux.

L'objet de cette thèse est sur la reconstruction des objets continus ou constants par morceaux, ou plus généralement homogènes par morceaux. En résumé, deux méthodes principales sont proposées dans le contexte de l'inférence bayésienne.

La première méthode consiste à reconstruire l'objet en imposant que sa transformée de Haar soit parcimonieuse. Un modèle bayésien hiérarchique est proposé. Dans cette méthode, les variables et les paramètres sont estimés et les hyper-paramètres sont initialisés selon la définition des modèles antérieurs.

La deuxième méthode reconstruit les objets en estimant simultanément les contours. L'objet continu par morceaux est modélisé par un modèle markovien non-homogène, qui dépend du gradient de l'objet, et le gradient dépend aussi de l'estimation de l'objet. Cette méthode est également semi-supervisé, avec les paramètres estimés automatiquement.

Ces méthodes sont adaptées aux reconstructions de grande taille de données 3D, dans lesquelles le processeur GPU est utilisé pour accélérer les calculs. Les méthodes sont validées avec des données simulées et des données réelles, et sont comparées avec plusieurs méthodes classiques.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France

université PARIS-SACLAY
ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Title : Fast and accurate 3D X ray image reconstruction for 3D X ray image reconstruction for Non Destructive Testing industrial applications

Keywords : Tomography, Inverse problem, Bayesian method, Hierarchical structure

Abstract : 3D X-ray Computed Tomography (CT) is widely used in medical imaging as well as in Non Destructive Testing (NDT) for industrial applications. In both domains, there is a need to reduce the number of projections. In some cases we may also be limited in angles. Consequently, we are always in the situation of ill-posed inverse problems. What's more, the measured data is always contaminated by noise. Facing these problems, the role of the probabilistic methods and the prior modelling becomes crucial.

For prior modelling, in particular in NDT applications, objects under examination are piecewise continuous, composing of several homogeneous materials with several continuous blocs separated by some discontinuities and contours. The focus of this thesis on the reconstruction of the piecewise continuous or constant, or more generally piecewise homogeneous objects. In summary two main methods are proposed in the context of the Bayesian inference.

The first method consists in reconstructing the object while enforcing the sparsity of the discrete Haar transformation coefficients of the object. A hierarchical Bayesian model is proposed. In this method, the unknown variables and parameters are estimated and the hyper-parameters are initialized according to the definition of prior models.

The second method reconstructs objects while the contours are estimated simultaneously. The piecewise continuous object is modeled by a non-homogeneous Markovian model, which depends on the Laplacian of the object, while the Laplacian also depends on the estimation of the object. In this method, a semi-supervised system model is also achieved, with parameters estimated automatically.

Both methods are adapted to the 3D big data size reconstructions, in which the GPU processor is used to accelerate the computation. The methods are validated with both simulated and real data, and are compared with several conventional state-of-the-art methods.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France