



Reconstruction d'images de tomographie

Isabelle BLOCH

Télécom ParisTech, département TSI, CNRS UMR 5141, 46 rue Barrault, 75634
Paris Cedex 13

Tél : 01 45 81 75 85, Fax : 01 45 81 37 94, E-mail : Isabelle.Bloch@enst.fr

Table des matières

Introduction	3
1 Principe physique de la tomographie à rayons X	3
1.1 Atténuation	3
1.2 Systèmes d'acquisition	4
1.3 Transformation de Radon, rétroprojection	6
1.4 Limites	7
2 Inversion analytique théorique	9
2.1 Théorème de projection et inversion directe	9
2.2 Théorème de rétroprojection et inversion	10
2.3 Introduction à la rétroprojection filtrée	10
3 Discrétisation des méthodes analytiques	12
3.1 Position du problème	12
3.2 TFD et rétroprojection discrète	13
3.3 Inversion directe discrète	13
3.4 Déconvolution 2D discrète	16
3.5 Rétroprojection filtrée discrète	16
4 Méthodes algébriques	18
4.1 Matrice de projection	19
4.2 Méthode ART	21
4.3 Méthode SIRT	22
4.4 Variantes	23
5 Régularisation	24
5.1 Moindres carrés, inverse généralisée	24
5.2 Stabilisation, conditionnement	25
5.3 Application aux méthodes algébriques	27
6 Extensions	27
Références	28

Introduction

Ce cours résume les principales méthodes de reconstruction tomographique. Il vise à illustrer, sur un exemple d'application réelle, un certain nombre de principes et de méthodes étudiés dans les cours de base du traitement du signal et des images (restauration, déconvolution, filtrage, échantillonnage, etc.). Ce texte regroupe les principaux résultats sans en donner les démonstrations.

Nous rappellerons tout d'abord les principes physiques de la tomographie à rayons X, ou tomodynamométrie (partie 1), afin d'aboutir au système d'équations à résoudre, exprimant les liens entre la fonction de densité des tissus en tout point en fonction de projections. La résolution, ou inversion, permettant de calculer la densité à partir des projections, sera d'abord décrite dans le cadre théorique continu idéal (partie 2). Puis le problème de la discrétisation sera abordé selon deux points de vue : une première classe de méthodes, dites méthodes analytiques, consiste à discrétiser directement les formules d'inversion continues (partie 3); une deuxième classe de méthodes, dites méthodes algébriques, consiste à discrétiser les équations initiales de projection et à se ramener à la résolution d'un système linéaire (partie 4). Dans la partie 5, nous verrons que le problème de la reconstruction tomographique est souvent mal posé, et présenterons des méthodes de régularisation. Enfin, en guise de conclusion, quelques extensions de ces méthodes seront mentionnées dans la partie 6.

Notons que si le phénomène physique d'atténuation est spécifique de la tomographie à rayons X, les méthodes de reconstruction sont en revanche plus générales et sont appliquées également dans d'autres systèmes d'imagerie, dans lesquelles des équations analogues expriment une fonction à reconstruire en fonction de projections. C'est le cas par exemple de la tomographie d'émission de simples photons utilisée en médecine nucléaire.

1 Principe physique de la tomographie à rayons X

1.1 Atténuation

Un faisceau monochromatique de rayons X traversant des tissus biologiques subit une variation d'énergie. C'est le phénomène d'atténuation. De manière plus précise, l'atténuation linéaire f dans un volume élémentaire dv est définie par :

$$\log \frac{I + dI}{I} = -f dv, \quad (1)$$

où I est l'intensité du faisceau à l'entrée du volume dv et $I + dI$ l'intensité à la sortie de ce volume.

En intégrant cette équation sur une droite D représentant le trajet rectiligne des rayons, on obtient l'intensité transmise I en fonction de l'intensité incidente I_0 par :

$$I = I_0 \exp\left(-\int_D f(x, y) dv\right), \quad (2)$$

où $f(x, y)$ représente l'atténuation au point (x, y) du volume traversé. C'est la fonction à reconstruire en tout point de l'espace. Pour cela il faut disposer d'un certain nombre de mesures. Les mesures accessibles sont de la forme de l'équation 2, et fournissent une information de type *projection*.

Le principe de l'acquisition de ces mesures est illustré par la figure 1. Les axes x et y sont les axes du repère cartésien de la coupe du patient observée¹, et donc de l'image à reconstruire. Les axes u et v sont liés à une direction du faisceau de rayons X : un faisceau envoyé dans la direction v fournit une suite de points de données (valeurs de projections) sur la droite u d'angle θ . En chaque point de cette droite, la valeur mesurée $p_\theta(u)$ correspond à l'intégrale du coefficient d'atténuation le long d'une droite parallèle à l'axe v (selon l'équation 2). En faisant varier θ , on acquiert des projections tout autour du patient.

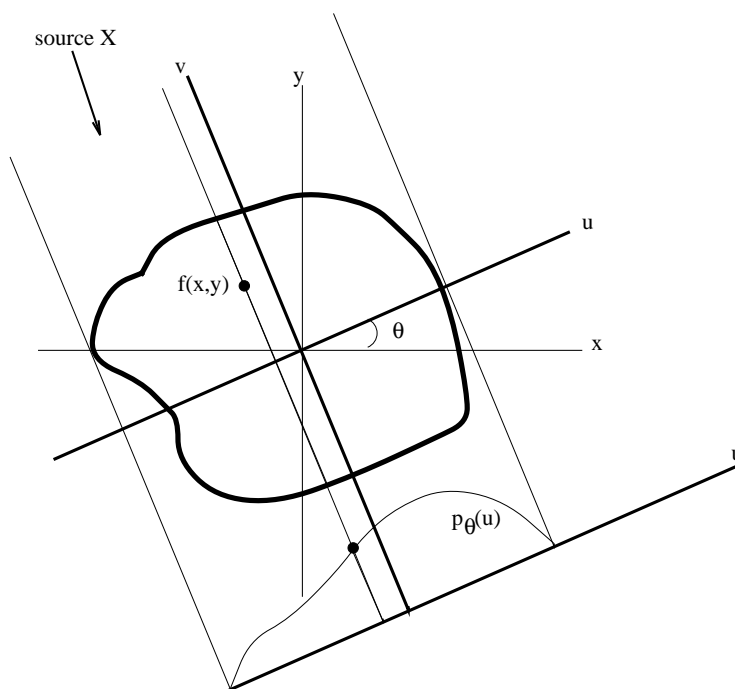


FIG. 1 – Principe de projection : la mesure $p_\theta(u)$ correspond à l'intégrale de la fonction à reconstruire le long d'une droite orthogonale à l'axe u .

1.2 Systèmes d'acquisition

Les principales générations de systèmes d'acquisition (*scanners X*) sont illustrés sur la figure 2. Les systèmes de la première génération comportaient une source et un détecteur. Cet ensemble était déplacé en translation afin d'acquérir l'ensemble des mesures correspondant à un angle fixé. Puis l'ensemble source et détecteur subissait

¹Les images sont acquises coupe par coupe, chacune étant orthogonale à l'axe du patient (coupes axiales), d'où le nom de tomographie. L'ensemble des coupes acquises à des niveaux anatomiques différents constitue un volume 3D, en général anisotrope.

une rotation afin d'acquérir la ligne de mesures suivante, pour un nouvel angle. Les scanners de la deuxième génération comportaient une source émettant un faisceau étroit de rayons X, reçus par quelques détecteurs. Puis l'ensemble subissait translation puis rotation comme pour la première génération. A partir de la troisième génération, la translation est supprimée. La source émet un faisceau large, recouvrant toute la section du patient. Dans les systèmes de troisième génération, les rayons sont reçus par un ensemble de détecteurs. L'ensemble source et détecteurs subit des rotations successives pour acquérir des projections selon tous les angles. Enfin, dans la quatrième génération, les détecteurs sont disposés en couronne tout autour du patient et sont fixes. Seule la source subit la rotation.

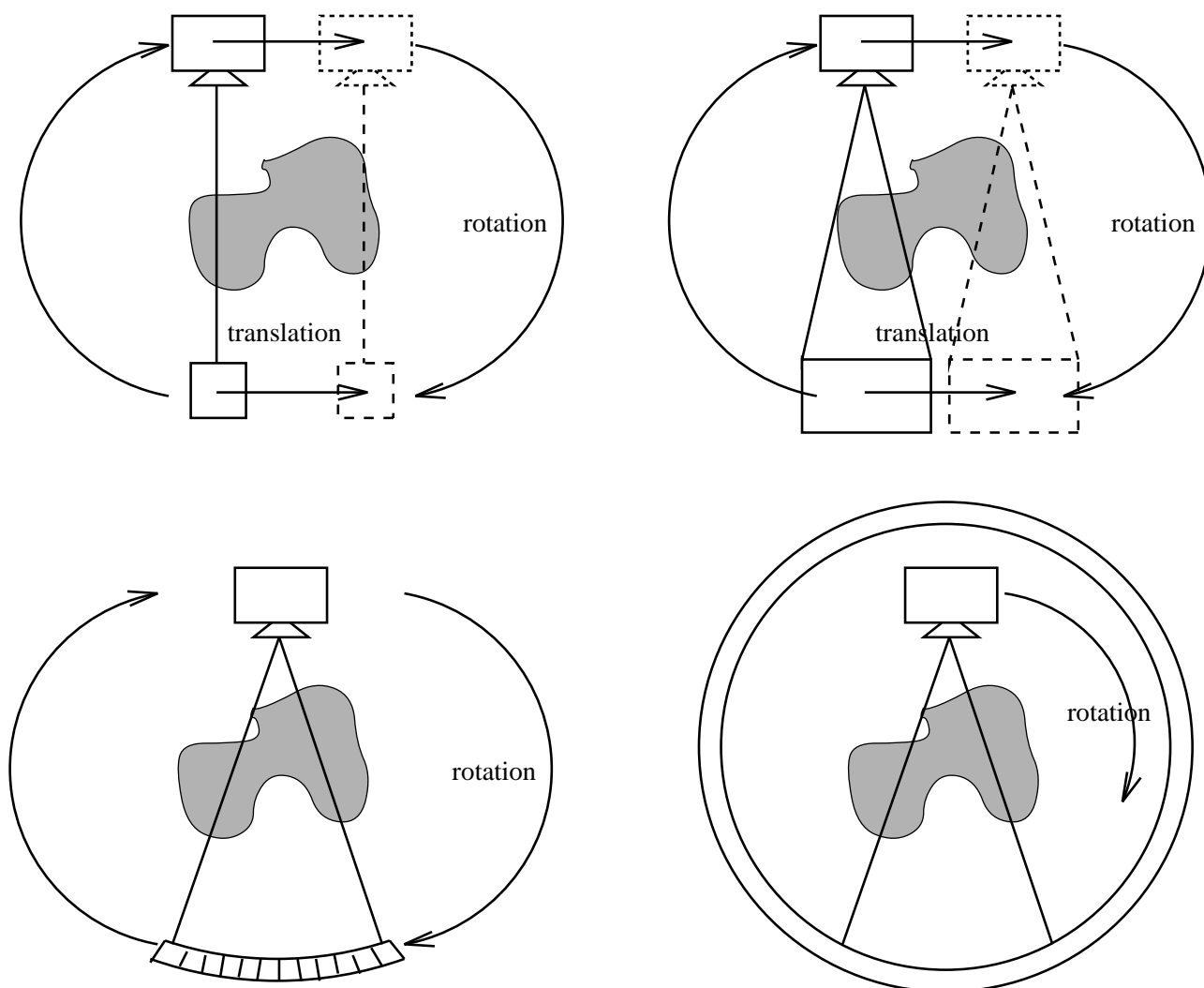


FIG. 2 – Les 4 générations de scanners X.

Dans toute la suite, nous nous placerons dans le cas d'une géométrie parallèle, c'est-à-dire dans laquelle les rayons associés à un angle de projection sont tous parallèles.

1.3 Transformation de Radon, rétroprojection

L'opérateur R qui exprime les projections de f est appelé transformation de Radon. De manière plus précise, il s'exprime au point (u, θ) par :

$$\begin{aligned} R[f](u, \theta) &= p_\theta(u) \\ &= \int_{D_\theta} f(u \cos \theta - v \sin \theta, u \sin \theta + v \cos \theta) dv. \end{aligned} \quad (3)$$

Notons que cette transformation est connue depuis 1917, donc bien avant l'invention du scanner !

Si l'on représente les valeurs de $p_\theta(u)$ dans un plan d'axes θ et u , par exemple par des niveaux de gris, on obtient un *sinogramme*, tel que celui de la figure 3. Ce sinogramme, ensemble de sinusoides, n'est pas l'image du patient, et en particulier il n'est pas du tout interprétable directement par un médecin. Les images du patient sont des images calculées à partir des projections. C'est donc bien un problème de reconstruction qu'il faut résoudre.

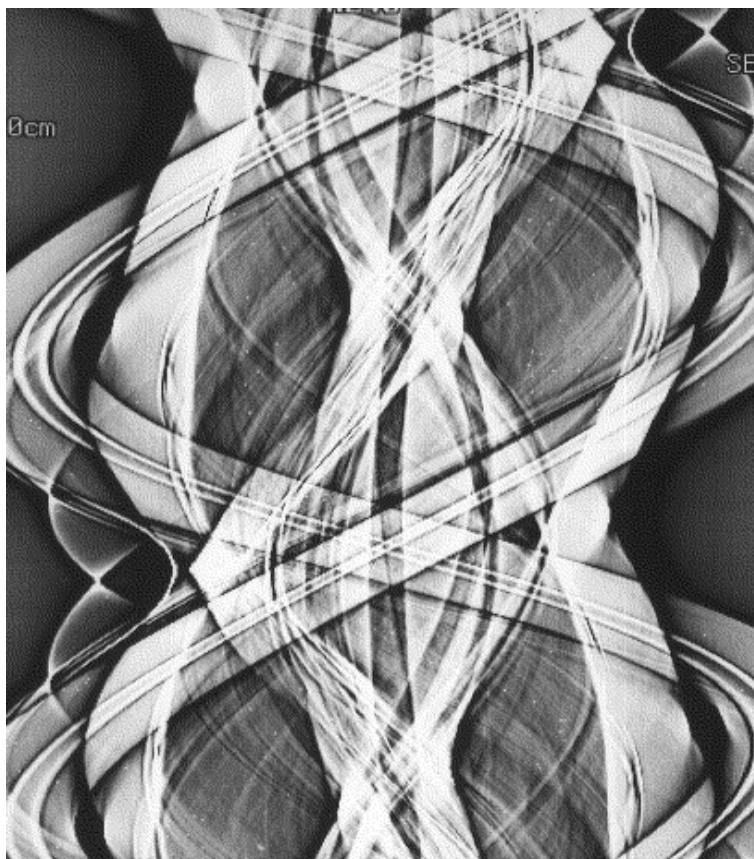


FIG. 3 – Exemple de sinogramme.

En remarquant que $p_\theta(u) = p_{\theta+\pi}(-u)$, le problème de la reconstruction s'exprime

alors de la manière suivante : étant donné un ensemble de mesures de projection

$$\{p_\theta(u), \theta \in [0, \pi[, u \in \mathbb{R}\},$$

retrouver f en tout point de l'espace, c'est-à-dire calculer

$$\{f(x, y), (x, y) \in \mathbb{R}^2\}.$$

Une manière simple de retrouver en tout point du plan une valeur d'atténuation pourrait consister à attribuer la valeur $p_\theta(u)$ à tout point placé sur le rayon de projection ayant donné cette valeur, puis à sommer toutes les contributions issues de toutes les projections. Ce principe, illustré sur la figure 4, est appelé rétroprojection. Il s'exprime précisément de la manière suivante. La rétroprojection en (x, y) d'une projection est la valeur de la projection d'angle θ au point sur lequel se projette (x, y) , et vaut :

$$h_\theta(x, y) = p_\theta(x \cos \theta + y \sin \theta). \quad (4)$$

La rétroprojection de toutes les projections définit l'opérateur B , dit opérateur de rétroprojection, obtenu en sommant sur tous les angles les expressions données par l'équation 4 :

$$\begin{aligned} B[p](x, y) &= \int_0^\pi h_\theta(x, y) d\theta \\ &= \int_0^\pi p_\theta(x \cos \theta + y \sin \theta) d\theta. \end{aligned} \quad (5)$$

L'image obtenue n'est pas l'image cherchée f , mais une version floue de f . Nous verrons plus loin l'expression mathématique de ce flou. L'opérateur de rétroprojection n'est donc pas l'inverse de l'opérateur de Radon².

La reconstruction ainsi obtenue, qui n'est donc pas exacte, a un intérêt historique puisque c'était la première méthode qui a été utilisée. Elle avait l'avantage de pouvoir être mise en œuvre de manière analogique.

1.4 Limites

La modélisation qui précède suppose implicitement qu'un certain nombre d'hypothèses soient vérifiées, qui ne sont en réalité pas toujours réalisées.

Tout d'abord, le phénomène physique est imparfaitement modélisé et un certain nombre de limites physiques perturbent donc la reconstruction. L'équation 2 suppose que le faisceau de rayons X soit monochromatique et infiniment fin, ce qui n'est pas le cas en réalité. Cette équation n'est donc qu'une approximation. De plus, ce modèle ne tient pas compte du durcissement du faisceau ni du phénomène de diffusion. Le durcissement du faisceau provient du fait que tous les rayons qui le constituent n'ont

²L'opérateur de rétroprojection peut en fait s'interpréter comme l'adjoint de l'opérateur de Radon.

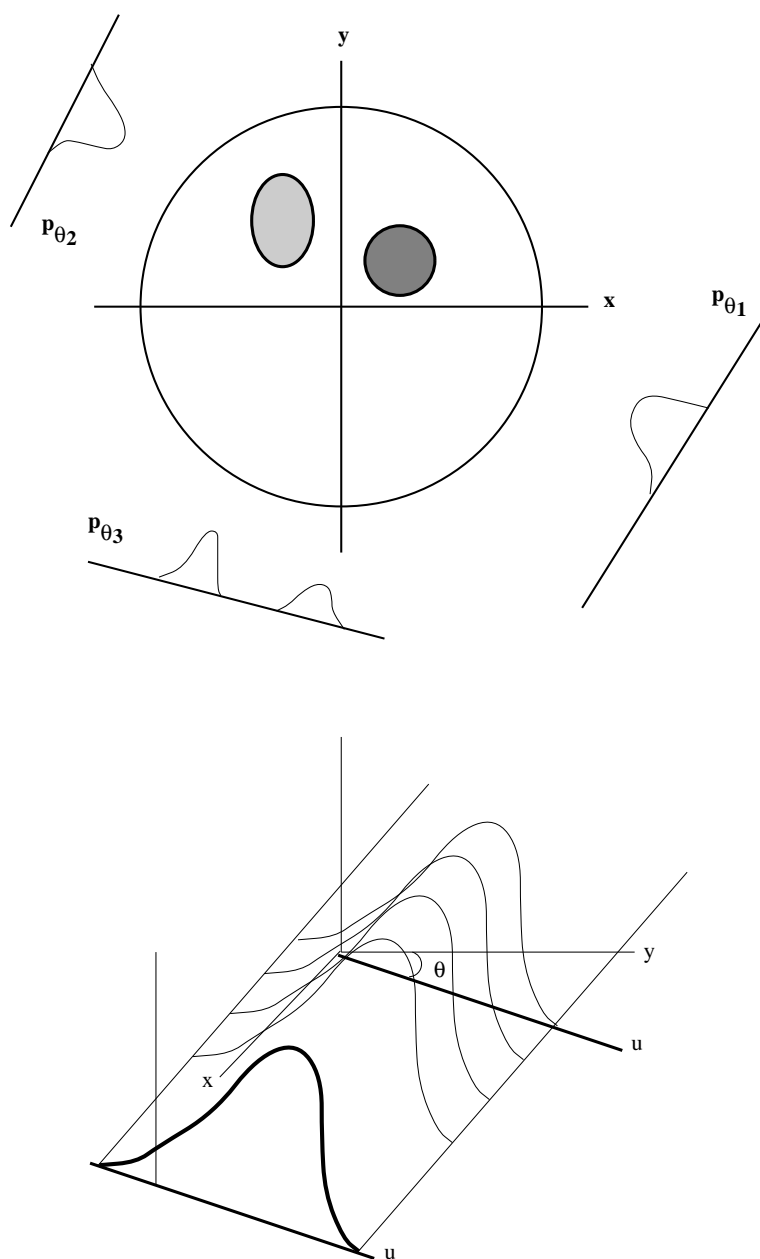


FIG. 4 – Acquisition de plusieurs projections et principe de rétroprojection d'une projection.

pas la même énergie. Les rayons de faible énergie disparaissent au cours de la traversée des tissus, qui ne laissent passer que ceux de plus forte énergie. Le phénomène de diffusion peut être limité grâce à l'utilisation de grilles qui sélectionnent les rayons ayant l'orientation désirée, éliminant ainsi les rayons diffusés d'orientation quelconque.

Le processus de reconstruction suppose que le patient soit immobile dans le repère de référence. Les mouvements éventuels du patient entraînent des erreurs dans la localisation des rayons de projection par rapport aux mesures effectuées et produisent donc des artefacts sur les images reconstruites.

Une autre limite est liée aux acquisitions où les données sont incomplètes. En particulier pour l'imagerie du cœur, un faible nombre de projections est généralement acquis (fenêtres d'exploration limitées, durée de l'examen, acquisition temporelle, etc.).

Enfin, les données sont généralement bruitées (bruit de mesures, imperfections du modèle, etc.).

Toutes ces limites conduisent à un problème de reconstruction tomographique mal posé, sur lequel nous reviendrons dans la partie 5. Pour l'instant, nous nous plaçons dans le cas idéal sans tenir compte de ces limites.

2 Inversion analytique théorique

Ce chapitre est consacré à la description succincte des trois méthodes principales de reconstruction dans le cas idéal, continu. Chacune de ces méthodes repose sur des relations analytiques particulières entre f et ses projections.

2.1 Théorème de projection et inversion directe

La première méthode repose sur le théorème de projection. Celui-ci stipule que la transformation de Fourier (1D) d'une projection d'angle θ au point U correspond à la transformée de Fourier 2D de f au point de fréquences spatiales $(U \cos \theta, U \sin \theta)$, ce qui s'écrit :

$$TF[p_\theta](U) = TF[f](U \cos \theta, U \sin \theta). \quad (6)$$

La transformée de Fourier d'une projection d'angle θ fournit donc une ligne d'angle θ dans le plan de Fourier de f .

Ce théorème permet d'en déduire le schéma de reconstruction suivant :

$$\begin{array}{c} \{p_\theta(u)\} \\ \Downarrow \\ \{TF[p_\theta](U)\} \\ \Downarrow \\ TF[f](X, Y) \\ \Downarrow \\ f \text{ par TF inverse} \end{array}$$

En effet, θ variant entre 0 et π , tout le plan de Fourier de f est bien couvert par les transformées de Fourier des p_θ et la transformée de Fourier de f peut être connue en tout point (X, Y) dans un repère cartésien. Cette propriété ne sera plus vérifiée dans le cas discret et nous verrons dans la partie 3 les problèmes de rééchantillonnage afin de passer d'un échantillonnage polaire à un échantillonnage cartésien.

Cette méthode, appelée inversion directe, nécessite d'effectuer autant de transformées de Fourier 1D que d'angles de projection, et une transformée de Fourier inverse 2D.

2.2 Théorème de rétroprojection et inversion

Nous avons vu dans la partie 1 que la rétroprojection de l'ensemble des projections fournissait une version floue de f . Le théorème de rétroprojection établit de manière précise la relation entre f et la rétroprojection $B[p]$:

$$B[p](x, y) = (f * h)(x, y), \quad (7)$$

avec :

$$h(x, y) = \frac{1}{\sqrt{x^2 + y^2}}. \quad (8)$$

La rétroprojection est donc une convolution de f par un filtre dont la réponse impulsionnelle est illustrée sur la figure 5. Cela explique donc le flou de la rétroprojection.

Ce théorème permet d'établir une deuxième méthode de reconstruction, par déconvolution :

$$f = TFI [TF(B[p]) \cdot \rho], \quad (9)$$

avec :

$$\rho(X, Y) = \sqrt{X^2 + Y^2}. \quad (10)$$

Cette méthode nécessite des opérations bidimensionnelles de filtrage et de transformée de Fourier et est donc assez lourde à mettre en œuvre. C'est la raison pour laquelle elle est assez peu utilisée.

2.3 Introduction à la rétroprojection filtrée

La troisième méthode de reconstruction, dite de rétroprojection filtrée, repose sur le résultat suivant :

$$f = B[\tilde{p}], \quad (11)$$

avec :

$$\tilde{p}_\theta = TFI [TF[p_\theta](U) \cdot |U|]. \quad (12)$$

Ce théorème exprime f comme la rétroprojection des projections filtrées par le filtre rampe $|U|$. Ce résultat s'obtient en écrivant que f est la transformée de Fourier inverse de sa transformée de Fourier et en utilisant le théorème de projection (équation 6).

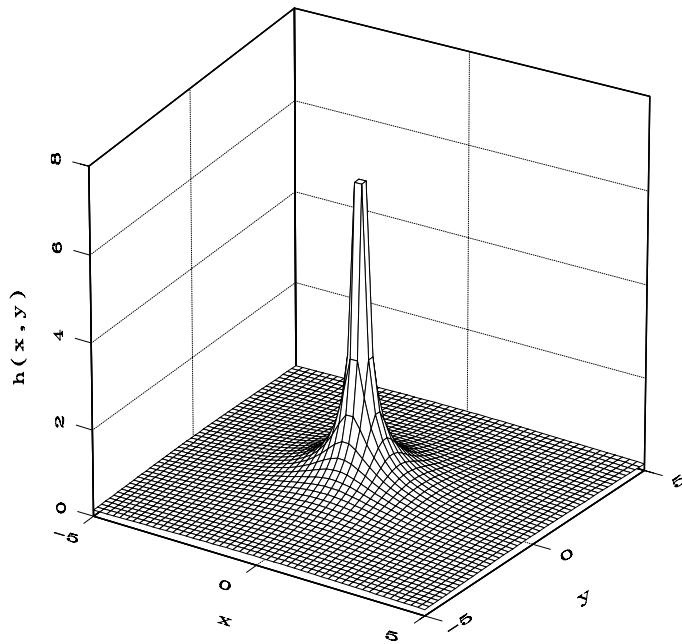


FIG. 5 – Réponse impulsionnelle du filtre $h(x, y)$ expliquant le flou de la rétroprojection : $h(x, y) = (x^2 + y^2)^{-1/2}$.

Notons que le filtre qui intervient ici est l'équivalent 1D du filtre $\rho(X, Y) = \sqrt{X^2 + Y^2}$ utilisé dans la méthode de déconvolution.

Ce théorème conduit au schéma de reconstruction suivant :

filtrage des projections (1D) par l'équation 12

↓

rétroprojection des projections filtrées (équation 11)

Cette fois, le filtrage ne nécessite que des opérations 1D, et la rétroprojection est une simple sommation des contributions de chaque projection. C'est la méthode la moins lourde en calculs et donc la plus utilisée.

En pratique, le filtrage est souvent effectué par $H(U) = |U| \cdot W(U)$ où $W(U)$ est un filtre passe-bas, ce qui permet de limiter le bruit, mais détériore la résolution spatiale. Il y a donc un compromis à trouver entre résolution spatiale et bruit. Nous reviendrons sur les filtres utilisés dans la partie 3.

3 Discrétisation des méthodes analytiques

3.1 Position du problème

Jusqu'à maintenant, nous nous sommes placés dans le cas idéal continu : nous avons travaillé dans un domaine continu infini (\mathbb{R}^2), et cherché à reconstruire une fonction f continue à partir de projections p_θ continues, connues pour tout θ de l'intervalle $[0, \pi[$. Ces conditions ne sont évidemment pas réalisables en pratique.

Les systèmes d'acquisition permettent d'obtenir des projections p_θ pour un nombre fini d'angles, notés θ_k . Le nombre limité de détecteurs entraîne que ces projections sont échantillonnées et connues simplement en des points discrets u_k . Il serait illusoire avec de telles données de vouloir reconstruire une fonction f continue dans \mathbb{R}^2 . La fonction f ne sera reconstruite que sur une grille discrète, en un nombre fini de points. C'est aussi la limite qu'imposent les algorithmes numériques de reconstruction.

À la différence de ce qui a été vu dans la partie 1, le problème de reconstruction se pose alors de la manière suivante : étant donné un ensemble de mesures de projection

$$\{p_{\theta_k}(u_l), 0 \leq l < NP, 0 \leq k < M\},$$

retrouver f en tout point d'une grille discrète finie, c'est-à-dire calculer

$$\{f(x_i, y_j), 0 \leq i < N, 0 \leq j < N\},$$

avec :

$$\theta_k = k\Delta\theta,$$

$$\Delta\theta = \frac{\pi}{M},$$

$$u_l = ld,$$

$$x_i = i\Delta x,$$

$$y_j = j\Delta y,$$

où $\Delta\theta$ est le pas d'échantillonnage des angles de rotation, M le nombre d'angles, d le pas d'échantillonnage sur chaque droite de projection, et Δx (respectivement Δy) le pas d'échantillonnage en x (respectivement en y) dans le plan de reconstruction.

Les méthodes de reconstruction de f selon ce schéma peuvent être réparties en deux classes.

1. La première classe de méthodes consiste à définir les équivalents discrets des opérateurs (transformée de Radon, rétroprojection, transformée de Fourier, etc.) puis à discrétiser les formules d'inversion découlant des trois théorèmes de la partie 2. Ce sont les méthodes dites analytiques.
2. La deuxième classe de méthodes repose sur une approche complètement différente : cette fois, c'est directement l'équation de projection qui est discrétisée, fournissant ainsi un système d'équations linéaires. La résolution de ce système s'effectue par des méthodes itératives, la résolution directe n'étant pas réalisable. Ces méthodes sont dites algébriques.

Ce chapitre est consacré aux méthodes analytiques. Les méthodes algébriques seront vues dans le chapitre 4.

3.2 TFD et rétroprojection discrète

Rappelons que la transformée de Fourier discrète (TFD), bien étudiée en traitement du signal, s'exprime, pour une fonction f connue en N points $(f_0, f_1, \dots, f_{N-1})$, comme la suite des valeurs $(F_0, F_1, \dots, F_{N-1})$ données par :

$$F_k = \sum_{l=0}^{N-1} f_l \exp\left(\frac{-2\pi}{N}lk\right). \quad (13)$$

Notons que la relation entre la TFD et la transformée de Fourier continue doit être considérée avec précautions, à cause des problèmes de recouvrement de spectre éventuels. Pour éviter ces problèmes, la théorie classique de l'échantillonnage impose que les fonctions doivent vérifier les conditions de Shannon, ce qui suppose qu'elles soient à spectre limité. Nous nous placerons sous cette hypothèse dans la suite.

L'opérateur de rétroprojection discret est défini par :

$$B[p](x_i, y_j) = \frac{\pi}{M} \sum_{k=0}^{M-1} p_{\theta_k}(x_i \cos \theta_k + y_j \sin \theta_k). \quad (14)$$

Il s'agit d'une formule d'approximation de l'intégrale de la formule 5. Cette approximation est optimale si les angles des projections sont équidistants (ce que nous avons supposé ici).

Cette équation fait intervenir les valeurs de projections aux points $x_i \cos \theta_k + y_j \sin \theta_k$ qui ne coïncident pas forcément avec un des points d'échantillonnage u_l des projections. Il va donc falloir calculer ces valeurs par interpolation. Une première solution consiste à calculer chaque $p_{\theta_k}(x_i \cos \theta_k + y_j \sin \theta_k)$ nécessaire par interpolation linéaire ou d'ordre supérieur. Une deuxième solution consiste, lors d'un pré-traitement, à interpoler des échantillons des projections avec une résolution suffisamment fine pour qu'ensuite on puisse estimer $p_{\theta_k}(x_i \cos \theta_k + y_j \sin \theta_k)$ simplement par la valeur du plus proche voisin.

3.3 Inversion directe discrète

La méthode d'inversion directe repose sur un équivalent discret du théorème de projection :

$$TFD[p_{\theta_k}](U_l) = TFD[f](U_l \cos \theta_k, U_l \sin \theta_k). \quad (15)$$

Cette formule conduit directement au schéma de reconstruction suivant :

$$\{p_{\theta_k}(u_l)\}$$

↓

$$\{TFD[p_\theta](U_i)\}$$

↓

estimation de $TF[f]$ sur une grille polaire

↓

interpolation polaire / cartésienne

↓

f par TFD inverse

La différence essentielle avec le schéma continu réside dans la connaissance du plan de Fourier de f . Alors que dans le cas continu, la transformée de Fourier de f est connue en tout point, ici on n'en connaît qu'un échantillonnage en coordonnées polaires. Les méthodes de transformée de Fourier inverse en coordonnées polaires sont difficiles à mettre en œuvre, et il n'existe pas d'algorithme rapide similaire à ceux de la TFD en coordonnées cartésiennes. Les méthodes utilisées en pratique consistent donc à rééchantillonner le plan de Fourier de f en coordonnées cartésiennes et à utiliser une TFD inverse classique.

La difficulté provient de la densité variable de l'échantillonnage polaire en fonction de la distance au centre du repère (figure 6).

Toujours dans l'hypothèse d'un spectre limité, l'échantillonnage doit être de bonne qualité dans un disque de rayon B , avec la condition de Shannon :

$$B = \frac{1}{2d},$$

d étant le pas d'échantillonnage des projections. La fonction f reconstruite sera donc une approximation à spectre limité.

Puisque NP points sont acquis sur une droite de projection, le pas d'échantillonnage radial dans le plan de Fourier est donné par :

$$\rho = \frac{2B}{NP} = \frac{1}{dNP},$$

(NP points sur un diamètre de longueur $2B$).

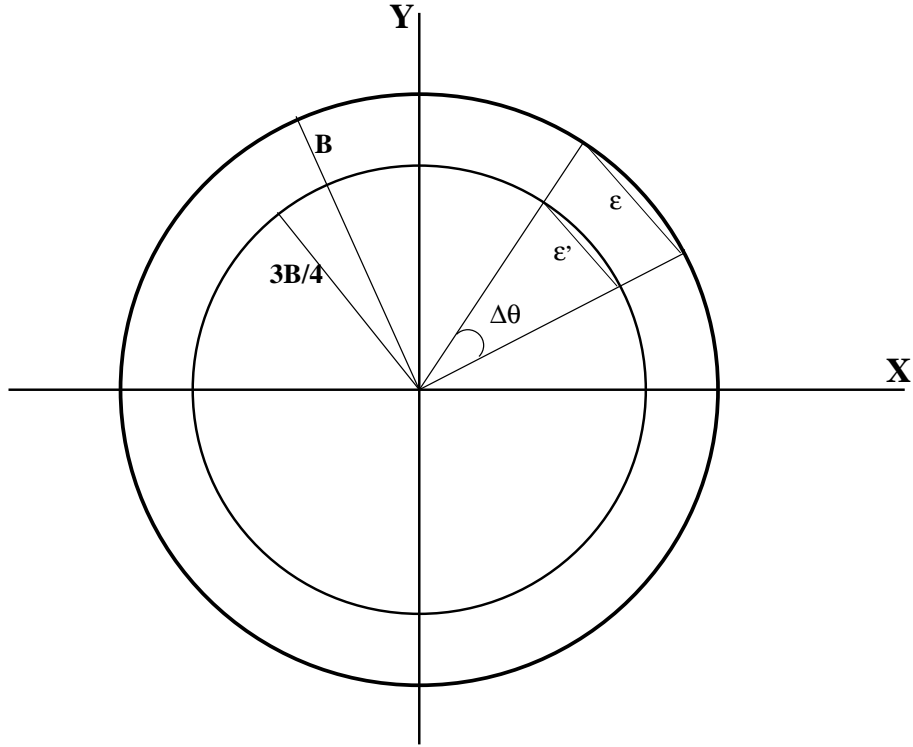


FIG. 6 – Echantillonnage dans le plan de Fourier de f .

On impose alors que l'échantillonnage azimutal le plus mauvais (c'est-à-dire à une distance B de l'origine) soit égal au pas d'échantillonnage radial :

$$\varepsilon = \rho.$$

On en déduit alors le pas d'échantillonnage des angles de rotation :

$$\Delta\theta = \frac{2}{NP}.$$

Une contrainte plus souple consiste à imposer simplement que le pas d'échantillonnage azimutal aux trois quarts de la bande soit égal au pas d'échantillonnage radial :

$$\varepsilon' = \rho_{3B/4} = \frac{3}{4}B\Delta\theta = \frac{2B}{NP}.$$

Le pas d'échantillonnage des angles de rotation vaut alors :

$$\Delta\theta = \frac{8}{3NP}.$$

On en déduit alors la valeur M du nombre de projections à réaliser, qui est de l'ordre de NP .

En pratique, les ordres de grandeurs sont les suivants : on réalise 512 projections échantillonnées sur 512 points et on reconstruit une image de taille 512×512 .

3.4 Déconvolution 2D discrète

La méthode discrète reposant sur l'équivalent du théorème de rétroprojection consiste à :

1. effectuer la rétroprojection discrète de toutes les projections, selon l'équation 14,
2. effectuer la déconvolution par TFD : celle-ci est généralement calculée sur une image plus grande (afin d'éviter les problèmes de repliement de spectre), et avec un filtre de déconvolution combiné avec un filtre passe-bas pour limiter le bruit.

Cette méthode étant lourde en temps de calcul, elle n'est pas couramment utilisée et nous ne la développerons donc pas plus.

3.5 Rétroprojection filtrée discrète

La méthode discrète de reconstruction par rétroprojection filtrée repose sur la discrétisation des formules 11 et 12 : chaque projection est filtrée puis rétroprojetée. Dans le cas discret, nous avons vu qu'il fallait se placer dans l'hypothèse d'un spectre limité à un rayon B :

$$B = \frac{1}{2d}.$$

Le filtre s'exprime donc par :

$$\begin{aligned} TF(k)(U) &= |U| \text{ si } |U| < B, \\ &= 0 \text{ sinon.} \end{aligned} \quad (16)$$

Différents filtres vérifiant cette condition sont possibles.

Le plus simple, proposé par Ramachandran et Lakshminarayanan est le filtre \hat{k} vérifiant :

$$TF(\hat{k})(U) = |U|Rect_B(U),$$

où $Rect_B(U)$ est la fonction valant 1 sur $[-B, B]$ et 0 ailleurs. Ce filtre peut être calculé explicitement en remarquant que

$$|U|Rect_B(U) = B[Rect_B(U) - Trgl_B(U)],$$

où $Trgl_B(U)$ est la fonction triangle sur $[-B, B]$. On obtient alors :

$$\hat{k}(u) = 2B^2 \left(\frac{\sin(2\pi Bu)}{2\pi Bu} \right) - B^2 \left(\frac{\sin(\pi Bu)}{\pi Bu} \right)^2. \quad (17)$$

Le filtre numérique correspondant, appelé filtre RAM-LAK, vaut :

$$k\left(\frac{m}{2B}\right) = \begin{cases} B^2 & \text{si } m = 0, \\ 0 & \text{si } m \text{ pair et } m \neq 0, \\ \frac{-4B^2}{m^2\pi^2} & \text{si } m \text{ impair.} \end{cases} \quad (18)$$

Un autre filtre très utilisé, proposé par Shepp et Logan, consiste à atténuer les hautes fréquences :

$$TF(\hat{k})(U) = |U| Rect_B(U) \frac{\sin(\frac{\pi U}{2B})}{\frac{\pi U}{2B}}. \quad (19)$$

La version numérique de ce filtre s'exprime par :

$$k(\frac{m}{2B}) = \frac{-4B^2}{\pi^2(4m^2 - 1)}. \quad (20)$$

Les transformées de Fourier des principaux filtres utilisés sont illustrés sur la figure 7. Les filtres numériques RAM-LAK et de Shepp et Logan sont illustrés sur la figure 8. On pourra constater que les coefficients du filtre s'atténuent très vite.

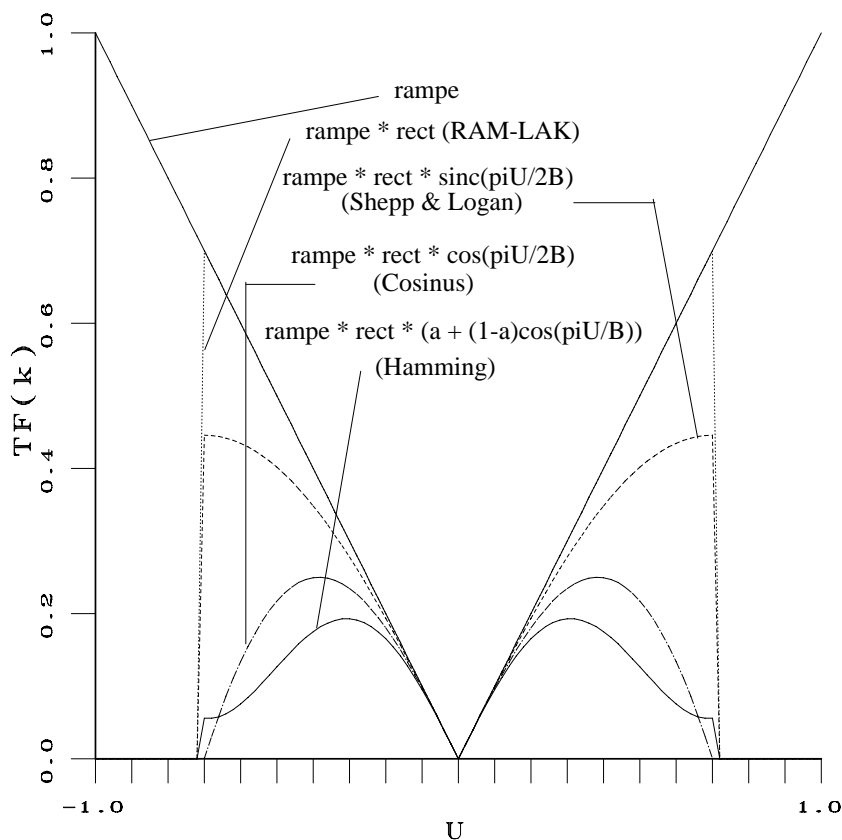


FIG. 7 – Quelques filtres de reconstruction.

D'autres filtres sont également utilisés, tels que les filtres cosinus, de Hamming, de Wiener, etc. Ces fenêtres plus douces permettent de diminuer le bruit (les hautes fréquences, peu fiables, sont atténuées), en évitant les oscillations dans les hautes fréquences engendrées par la fonction rectangle. Il faut cependant trouver un bon compromis afin de ne pas obtenir une image trop lissée.

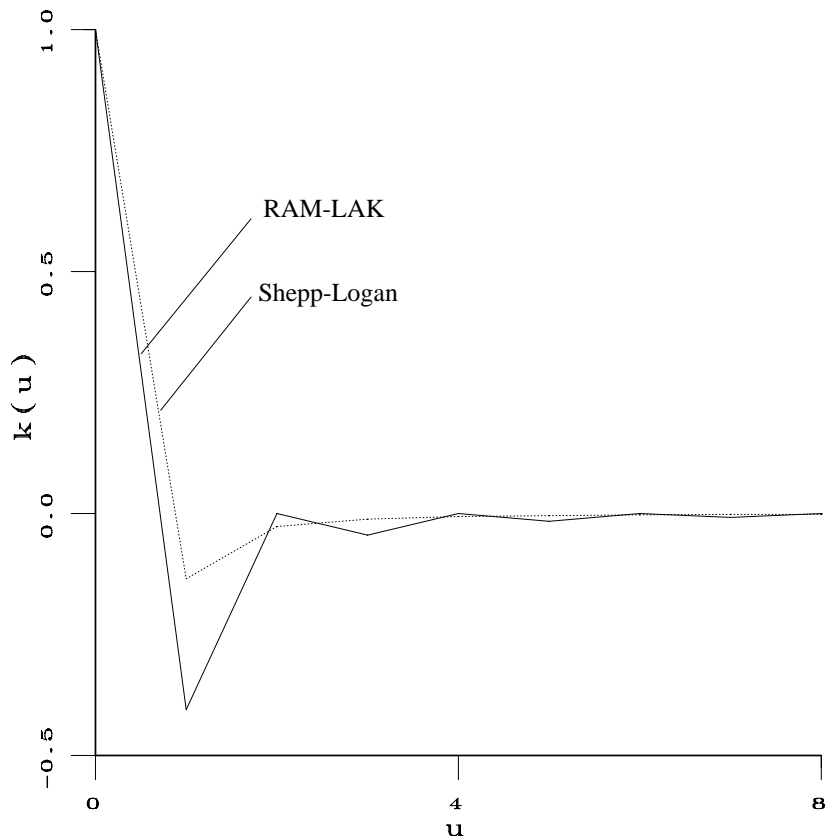


FIG. 8 – Filtres numériques (RAM-LAK et Shepp-Logan).

La mise en œuvre peut être effectuée soit par convolution discrète directe, soit dans le domaine de Fourier par FFT.

Cette méthode présente de nombreux avantages, qui en font la méthode la plus répandue. Essentiellement, les traitements sont 1D uniquement. De plus, il est possible d'effectuer le traitement de chaque projection dès qu'elle est acquise, c'est-à-dire de la filtrer et de la rétroprojeter en additionnant sa contribution à celle des projections déjà traitées. La reconstruction se fait donc simultanément à l'acquisition.

4 Méthodes algébriques

Dans cette partie, nous présentons le principe des méthodes algébriques. La notion de matrice de projection est introduite, puis les deux méthodes principales d'inversion itérative sont décrites.

4.1 Matrice de projection

Le principe des méthodes algébriques consiste à rechercher f directement dans un espace de dimension finie, sous la forme d'une combinaison linéaire de fonctions de base φ_i :

$$f(x, y) = \sum_{i=1}^n f_i \varphi_i(x, y). \quad (21)$$

La base la plus utilisée est celle des fonctions caractéristiques des pixels, définies par :

$$\varphi_i(x, y) = \begin{cases} 1 & \text{si } (x, y) = \text{pixel } i, \\ 0 & \text{sinon,} \end{cases} \quad (22)$$

les pixels de l'image étant numérotés dans un ordre arbitraire (par exemple dans l'ordre de balayage vidéo).

D'autres fonctions de base peuvent être utilisées, par exemple des fonctions polynomiales ou des fonctions adaptées à la géométrie de l'acquisition (typiquement à une géométrie conique). Dans la suite, nous resterons dans le cas simple de la base des pixels.

La valeur d'une projection en un point vaut alors :

$$p_j = \sum_{i=1}^n R_{ji} f_i,$$

avec

$$R_{ji} = \int \varphi_i(u_l \cos \theta_k - v \sin \theta_k, u_l \sin \theta_k + v \cos \theta_k) dv,$$

si p_j désigne la valeur de la projection d'angle θ_k au point u_l .

Le problème s'exprime donc sous la forme matricielle suivante :

$$P = Rf, \quad (23)$$

où R est appelée matrice de projection.

La signification des termes de l'équation 23 est la suivante :

1. P est le vecteur de mesures, chaque composante étant une valeur de projection (un des $p_\theta(u_l)$, rangés dans un ordre arbitraire); elle est de taille m égale au produit du nombre de projections par le nombre de points par projection, soit $M \times NP$;
2. f est le vecteur des valeurs recherchées en chaque pixel, chaque composante étant la valeur d'atténuation en un pixel de l'image; elle est de taille n égale au nombre de pixels, soit $N \times N$;
3. R est la matrice de projection, de taille $m \times n$. Cette matrice ne dépend que de l'acquisition et pas des données. Ses coefficients peuvent donc être calculés une

fois pour toute. Le coefficient R_{ji} est caractéristique de l'intersection du rayon j avec le pixel i . Dans le cas simple considéré ici, on a :

$$R_{ji} = \begin{cases} 1 & \text{si le rayon } j \text{ rencontre le pixel } i, \\ 0 & \text{sinon .} \end{cases} \quad (24)$$

Une version plus sophistiquée (voir figure 9) consiste à choisir des coefficients proportionnels à la surface d'intersection entre la surface élémentaire représentant le pixel et le rayon (ayant une certaine largeur) :

$$R_{ji} \propto \text{surface de recouvrement rayon } j / \text{pixel } i. \quad (25)$$

La première solution ne nécessite qu'un bit de stockage par coefficient, la seconde étant plus précise mais plus gourmande en mémoire.

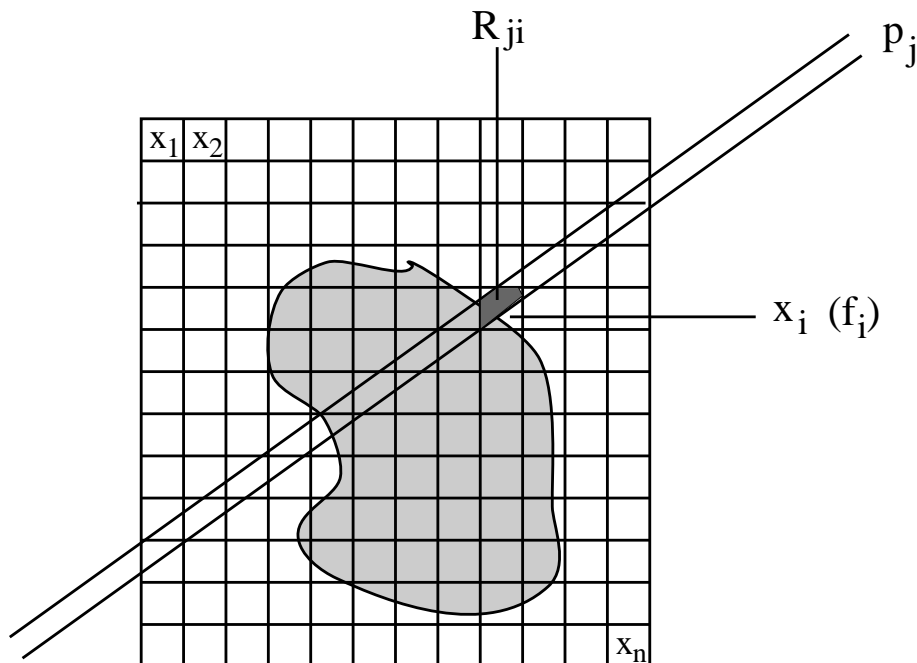


FIG. 9 – Géométrie des méthodes algébriques.

L'inversion directe de la matrice permettrait de résoudre le problème de reconstruction une fois pour toutes, pour une géométrie d'acquisition donnée (puisque R ne dépend que de cette géométrie). Cela semble d'autant plus intéressant que la matrice est très creuse. La reconstruction d'un jeu de données ne nécessiterait plus qu'un simple produit matriciel.

Cependant, cette méthode n'est pas réalisable en pratique sur les systèmes de tomographie à rayons X^3 , en raison de la taille de la matrice (250000×250000 au moins), du fait qu'elle n'est pas forcément carrée, et du bruit qui rend le problème mal conditionné, voire singulier.

³L'inversion directe est en revanche utilisable pour des systèmes expérimentaux de tomographie d'impédance, permettant d'acquérir pour l'instant très peu de données.

On préfère donc des méthodes itératives, dont les deux principales sont décrites dans les parties suivantes. Leur principe général consiste à partir d'une initialisation $f^{(0)}$ de f et à corriger itérativement l'estimation, en estimant $f^{(k)}$ à l'itération k à partir des différences entre les projections de $f^{(k-1)}$ et les données P (projections réelles de f).

4.2 Méthode ART

La méthode ART (algebraic reconstruction method) consiste à corriger les coefficient f_i de f en utilisant une projection à chaque fois. De manière géométrique, cela s'interprète comme la recherche de l'intersection entre des hyperplans dont les équations sont celles du système 23. Ces hyperplans s'interprètent comme des contraintes que doit respecter la solution. La recherche s'effectue par projections successives de la solution à une itération donnée sur une des contraintes. Ainsi, au moins une des équations du système est satisfaite à chaque itération. Cette interprétation géométrique est illustrée sur la figure 10 dans le cas où $m = n = 2$. La convergence de la méthode peut également être illustrée par des considérations géométriques sur le triangle dont les sommets sont les solutions à trois itérations successives.

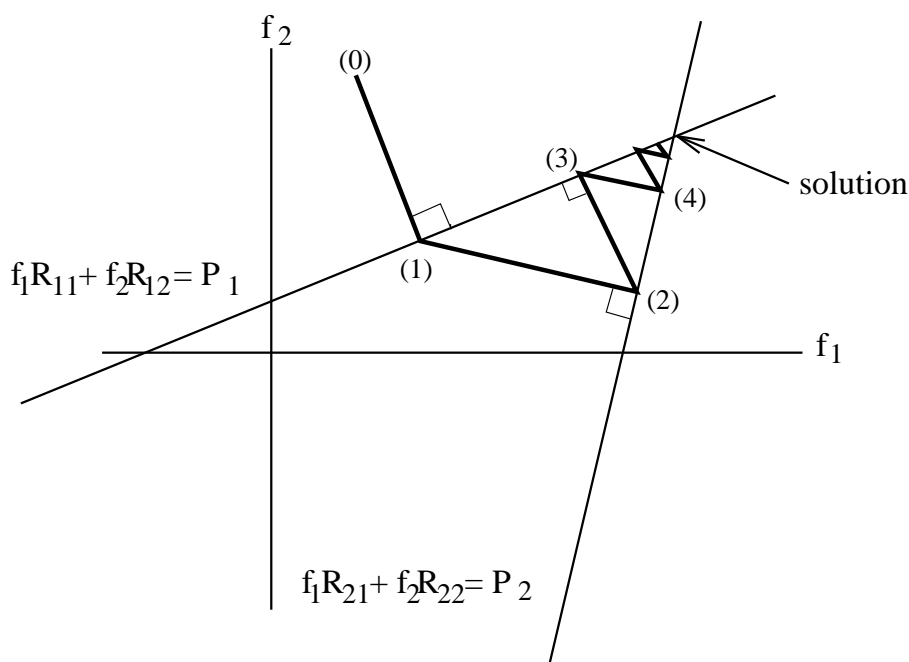


FIG. 10 – Principe de la méthode ART.

L'expression mathématique de la correction selon la méthode ART s'obtient aisément à partir de l'équation de la projection d'un point sur un hyperplan :

$$f_i^{(k)} = f_i^{(k-1)} + R_{ji} \frac{p_j - R_j f^{(k-1)}}{\|R_j\|^2}. \quad (26)$$

Cette équation s'interprète de la manière suivante : chaque composante i du vecteur $f^{(k)}$ à l'itération k est corrigée en ajoutant à la valeur $f_i^{(k-1)}$ obtenue à l'itération précédente un coefficient qui vaut 0 si le rayon j utilisé pour la correction ne traverse pas le point x_i (dans ce cas R_{ji} vaut 0) et qui est sinon proportionnel à la différence entre la donnée p_j (la vraie projection) et la projection recalculée à partir de $f^{(k-1)}$, égale à $R_j f^{(k-1)}$. Le coefficient de normalisation $\|R_j\|^2$ est la norme de la ligne de la matrice R correspondant à la donnée j , c'est-à-dire dans le cas simple considéré ici il est égal au nombre de pixels traversés par le rayon j . En résumé, le rayon j permet de corriger tous les pixels qu'il traverse. À chaque itération, un rayon différent est choisi, par exemple selon la règle suivante :

$$j = k[m].$$

Lorsque le système 23 admet au moins une solution, l'algorithme converge vers la solution de variance minimale lorsqu'il est initialisé à 0.

En revanche, lorsque les données sont bruitées, le système peut ne pas avoir de solution, et un phénomène d'oscillations se produit, comme l'illustre la figure 11.

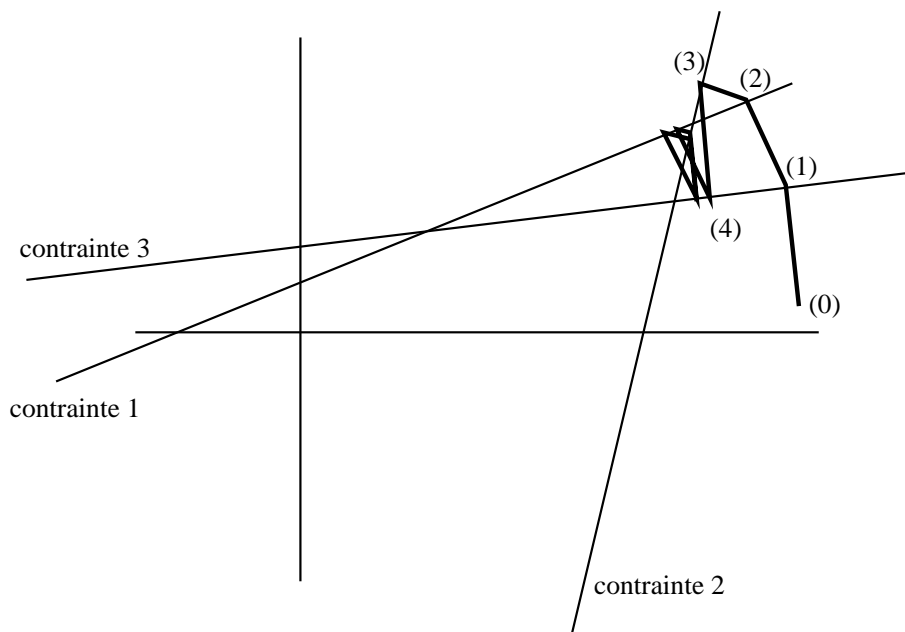


FIG. 11 – Problème d'oscillations lorsque les données sont bruitées.

4.3 Méthode SIRT

La méthode SIRT (simultaneous iterative reconstruction technique) consiste à effectuer la correction de chaque f_i en utilisant tous les rayons passant par x_i , comme l'illustre la figure 12.

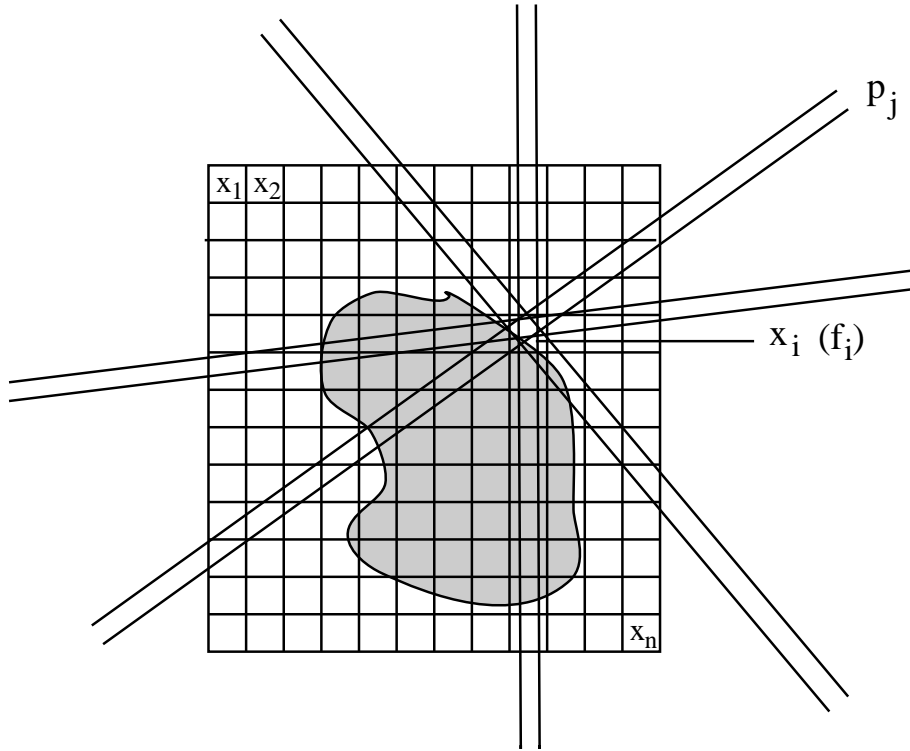


FIG. 12 – Principe de la méthode SIRT.

L'équation permettant d'évaluer $f_i^{(k)}$ par correction de $f_i^{(k-1)}$ est la suivante :

$$f_i^{(k)} = f_i^{(k-1)} + \frac{\sum_j p_j}{\sum_j \sum_i R_{ji}} - \frac{\sum_j R_{ji} f_i^{(k-1)}}{\sum_j \|R_{ji}\|^2}, \quad (27)$$

la sommation portant sur l'ensemble des indices j tels que le rayon j traverse le pixel x_i .

Cette méthode est plus lente que l'ART et nécessite plus de mémoire.

4.4 Variantes

Il existe de nombreuses variantes des algorithmes ART et SIRT. Mentionnons simplement :

- les versions multiplicatives, où la correction s'effectue en fonction du rapport de la projection donnée sur la projection recalculée,
- une variante de l'ART dans laquelle des contraintes de positivité doivent être satisfaites à chaque itération (la solution recherchée étant positive),
- une variante de l'ART permettant de tenir compte du bruit en remplaçant le système d'équations 23 par un système d'inégalités :

$$P - \varepsilon < Rf < P + \varepsilon,$$

ce qui est plus conforme à la réalité.

5 Régularisation

Pour qu'un problème soit *bien posé* il faut que 3 conditions soient remplies :

1. il y a au moins une solution pour tout ensemble de données,
2. la solution est unique,
3. la solution dépend continûment des données.

Les deux premières conditions ne sont pas toujours satisfaites en tomographie, en particulier lorsque les données sont incomplètes ou bruitées. La troisième condition signifie qu'un faible écart sur les données ne doit pas donner une solution radicalement différente. C'est un problème de stabilité, fortement lié au conditionnement et à la régularité de la matrice de projection, que nous analyserons dans la suite.

Ces difficultés ne sont pas spécifiques de la tomographie et concernent de manière générale tous les problèmes inverses. L'étude de tels problèmes mal posés a conduit les chercheurs à développer des méthodes de *régularisation* dont les applications en traitement du signal et des images sont très nombreuses (restauration, déconvolution, etc.).

5.1 Moindres carrés, inverse généralisée

Deux problèmes peuvent se poser lors de l'inversion du système $Rf = P$:

1. la matrice R^{-1} n'existe pas (en particulier si R est rectangulaire),
2. la matrice R^{-1} n'est pas continue.

Dans cette partie, nous nous attachons à résoudre le premier problème. Le second sera résolu dans la partie suivante.

Si R n'est pas inversible, le système n'a pas de solution au sens strict. On cherche alors une solution approchée qui minimise une fonctionnelle $C(Rf, P)$ représentant un critère d'écart entre Rf et P . La méthode la plus classique consiste à chercher la solution des *moindres carrés*, en résolvant le système :

$$R^t R f = R^t P. \quad (28)$$

Cette fois la matrice $R^t R$ est de taille $n \times n$. Deux cas doivent être distingués.

- Si elle est de rang n , la solution du système d'équations linéaires 28 est unique et est donnée par :

$$f = (R^t R)^{-1} R^t P, \quad (29)$$

et la matrice $(R^t R)^{-1} R^t$ est *l'inverse généralisée* de R .

- Sinon, le système 28 a une infinité de solutions, parmi lesquelles on choisit en général celle de norme minimale.

5.2 Stabilisation, conditionnement

Le problème de l'existence d'une inverse étant résolu, l'inverse généralisée peut être instable. Pour bien comprendre les phénomènes provoquant ces instabilités, il est commode d'effectuer l'analyse spectrale des opérateurs impliqués.

Soient σ_k^2 les valeurs propres de $R^t R$ et de RR^t , classées par ordre décroissant :

$$\sigma_1 > \sigma_2 > \dots \geq 0.$$

Soient p_k les vecteurs propres de RR^t et f_k ceux de RR^t :

$$RR^t p_k = \sigma_k^2 p_k, \quad R^t R f_k = \sigma_k^2 f_k.$$

On a, pour $\sigma_k \neq 0$:

$$p_k = \sigma_k^{-1} R f_k, \quad f_k = \sigma_k^{-1} R^t p_k.$$

En décomposant P sur la base des p_k , on obtient :

$$\begin{aligned} f &= (R^t R)^{-1} R^t P \\ &= (R^t R)^{-1} R^t \left(\sum_k \langle P \cdot p_k \rangle p_k \right) \\ &= (R^t R)^{-1} \left(\sum_k \langle P \cdot p_k \rangle \sigma_k f_k \right) \\ &= \sum_k \langle P \cdot p_k \rangle \sigma_k^{-1} f_k, \end{aligned}$$

ce qui exprime f comme combinaison linéaire des vecteurs propres de RR^t .

Dans le cas de données bruitées, où les mesures prennent la forme :

$$P + b,$$

où P représente les projections exactes et b le bruit, un calcul analogue conduit à la décomposition suivante de f :

$$f = \sum_k \langle P \cdot p_k \rangle \sigma_k^{-1} f_k + \sum_k \langle b \cdot p_k \rangle \sigma_k^{-1} f_k.$$

Cette expression permet maintenant de bien comprendre le problème d'instabilité : le bruit comporte généralement des hautes fréquences, qui correspondent aux derniers termes du développement. La décomposition aura donc des forts coefficients sur les petites valeurs propres, donc pour les grandes valeurs de σ_k^{-1} , ce qui aura tendance à amplifier les composantes hautes fréquences du bruit, d'où les phénomènes d'instabilité.

Le fait que les hautes fréquences correspondent aux faibles valeurs propres est un résultat classique du cours de restauration : sous l'hypothèse d'un défaut infini spatialement invariant, il y a équivalence entre la restauration par transformée de Fourier et

la restauration algébrique, les valeurs propres de la matrice du défaut correspondant aux coefficients de sa transformée de Fourier.

Dans le cas où ces phénomènes d'instabilité se produisent, dont on connaît maintenant l'origine, la solution doit être stabilisée. C'est la régularisation.

Une première méthode consiste à effectuer une troncature du développement, afin d'éliminer les termes correspondant aux faibles valeurs propres. Comme nous l'avons vu précédemment pour les méthodes analytiques, il est nécessaire de trouver un bon compromis pour le nombre de termes afin d'avoir une précision suffisante dans la reconstruction sans avoir trop de bruit. On retrouve ici encore une analogie avec les méthodes de restauration, en particulier avec les méthodes de décomposition en valeurs singulières (SVD). Dans ces méthodes, où le signal est restauré d'abord avec les plus fortes valeurs propres, puis en ajoutant progressivement des termes correspondant à des valeurs propres plus faibles, on constate que la qualité de la restauration augmente dans les premières étapes (on gagne en précision). Puis, lorsque des valeurs propres très faibles interviennent (partie mal conditionnée de la matrice), le bruit est amplifié. Enfin, si des valeurs complètement nulles sont introduites (partie singulière de la matrice), tout le signal est perdu.

Une deuxième méthode consiste à amortir les termes correspondant aux faibles valeurs propres, au lieu de les supprimer : on choisit des poids w_k donnant moins d'importance aux derniers termes du développement. La solution prend alors la forme :

$$f = \sum_k w_k \sigma_k^{-1} \langle P.p_k \rangle f_k. \quad (30)$$

Ces deux solutions ont un inconvénient majeur : elles nécessitent de calculer les valeurs propres et les vecteurs propres des matrices $R^t R$ et RR^t . Elles sont donc très lourdes en temps de calcul pour des matrices de taille importante comme c'est le cas en tomographie à rayons X.

Une troisième solution, qui ne présente pas cet inconvénient, consiste à rechercher une solution stable en introduisant des contraintes de douceur. On cherche ainsi à minimiser une fonctionnelle du type :

$$\min J(f) = \|Rf - P\|^2 + \gamma \Gamma(f), \quad (31)$$

dans laquelle $\Gamma(f)$ représente une contrainte de douceur sur f et γ est un multiplicateur de Lagrange. Une fonction Γ simple et classiquement utilisée est :

$$\Gamma(f) = \|f\|^2,$$

correspondant à la recherche d'une solution de norme minimale. La solution est alors donnée par :

$$f = (R^t R + \gamma I)^{-1} R^t P. \quad (32)$$

En reprenant les résultats de l'analyse spectrale, f peut s'écrire sous la forme :

$$f = \sum_k \frac{\sigma_k}{\sigma_k^2 + \gamma} \langle P.p_k \rangle f_k. \quad (33)$$

Cette dernière équation, bien qu'elle ne soit pas utilisée en pratique puisqu'elle nécessiterait à nouveau le calcul des valeurs propres et des vecteurs propres, montre que cette solution est la même que la solution amortie de l'équation 30, avec :

$$w_k = \frac{1}{1 + \frac{\gamma}{\sigma_k^2}}.$$

Pour des fortes valeurs propres, le terme $\frac{\gamma}{\sigma_k^2}$ est faible et le coefficient associé est proche de 1. Pour des faibles valeurs propres, le coefficient tend vers 0. Cette solution est également proche de celle obtenue par le filtre de Wiener, mais celui-ci nécessite de connaître une estimation du bruit et du signal.

Cette troisième solution nécessite bien sûr également de trouver un compromis entre précision et stabilité. Elle présente l'avantage de permettre d'introduire des contraintes variées, prenant en compte en particulier des informations a priori.

5.3 Application aux méthodes algébriques

L'application directe des méthodes algébriques itératives sur des données bruitées peut conduire à des phénomènes assez curieux au cours des itérations : il arrive par exemple que l'algorithme commence par converger, puis se mette à diverger. Les méthodes de régularisation sont alors indispensables.

Inversement, la minimisation de la fonctionnelle $J(f)$ (équation 31) peut être difficile de manière analytique et des méthodes itératives sont alors nécessaires. Les méthodes ART ou SIRT vues dans la partie 4 peuvent être adaptées à ce propos.

Par exemple, l'utilisation de la méthode ART pour résoudre le système :

$$\begin{pmatrix} R & I \\ 0 & R \end{pmatrix} \begin{pmatrix} f \\ r \end{pmatrix} = \begin{pmatrix} P \\ 0 \end{pmatrix}, \quad (34)$$

où r représente le résidu $Rf - p$, qui doit être minimisé, conduit à la solution des moindres carrés de norme minimale.

6 Extensions

Les méthodes décrites ci-dessus supposaient que la géométrie de l'acquisition était une géométrie parallèle, comme celle des systèmes de première génération. Dans le cas d'une géométrie conique (ou divergente), trois méthodes sont envisageables.

1. La première consiste à négliger la divergence. L'erreur induite par cette approximation est considérée négligeable si l'angle du faisceau est faible (typiquement, inférieur à 15 degrés). Cette méthode est donc applicable aux systèmes de deuxième génération, mais plus aux suivants où le faisceau doit recouvrir toute la section du patient afin de supprimer la translation.

2. La deuxième méthode consiste à réorganiser les données en projections parallèles, essentiellement par interpolation.
3. La dernière méthode consiste à reformuler complètement le problème. On constate alors que le théorème de projection n'est pas généralisable, ce qui ne permet pas de disposer de méthode d'inversion directe. Le théorème de rétroprojection peut être adapté aux géométries divergentes, et l'algorithme est alors le même que pour une géométrie parallèle, avec la même lourdeur de calculs. Enfin, les formules de rétroprojection filtrée peuvent être corrigées et donnent lieu à des algorithmes un peu différents. Les méthodes algébriques posent moins de problèmes, grâce à la flexibilité dans le choix des fonctions de base et donc des coefficients de la matrice R , permettant aux méthodes d'être adaptées à différentes géométries.

La plupart des méthodes analytiques et algébriques peuvent également être généralisées au 3D, afin de reconstruire non plus des coupes mais directement un volume 3D.

Nous n'avons pas du tout décrit les méthodes de type statistique, plus récentes, que nous mentionnons ici pour mémoire. Elles consistent à décrire f et P de manière stochastique, et on cherche par exemple à trouver la fonction f de variance minimale, ou encore sous forme bayésienne en estimant les probabilités a priori et conditionnelles selon le principe du maximum d'entropie.

Enfin, il existe des méthodes structurelles, qui font intervenir explicitement des connaissances a priori sur la forme des objets à reconstruire. Ces méthodes trouvent un domaine d'application privilégié en imagerie vasculaire.

Références

- [1] R. N. Bracewell. *The Fourier Transform and its Applications*. Mac Graw Hill, New-York, 1978.
- [2] Y. Censor. Finite Series Expansion Reconstruction Methods. *Proc. of the IEEE*, 71 :409–419, 1983.
- [3] M. E. Davison. The Ill-Conditioned Nature of the Limited Angle Tomography Problem. *SIAM Journal of Applied Mathematics*, 42(3) :428–448, 1983.
- [4] S. Geman and D. Geman. Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6(6) :721–741, 1984.
- [5] R. Gordon. A Tutorial on ART. *IEEE Trans. on Nuclear Sciences*, NS-21 :78–93, 1974.
- [6] R. Gordon and G. T. Herman. Three-dimensional Reconstruction from Projections : a Review of Algorithms. *International Review of Cytology*, 38, 1973.
- [7] C. Hamaker, K. T. Smith, D. C. Solmon, and S. L. Wagner. The Divergent Beam X-Ray Transform. *Rocky Mountain Journal of Mathematics Winter*, 10(1) :253–283, 1980.

- [8] K. M. Hanson and G. W. Wecksung. Bayesian Approach to Limited-Angle Reconstruction in Computed Tomography. *J. Opt. Soc. Am.*, 73(11) :1501–1509, 1983.
- [9] G. T. Herman. *Image Reconstruction from Projections : the Fundamentals of Computerized Tomography*. Academic Press, New-York, 1980.
- [10] G. M. Hounsfield. A Method and Apparatus for Examination of a Body by Radiation such as X or Gamma. *Pat. Spec. No-1283915*, 1972.
- [11] A. C. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*. IEEE Press, New-York, 1988.
- [12] A. K. Louis. Medical Imaging : State of the Art and Future Development. *Inverse Problems*, 8 :709–738, 1992.
- [13] A. K. Louis and F. Natterer. Mathematical Problems of Computerized Tomography. *Proc. of the IEEE*, 71(3) :390–408, 1993.
- [14] F. Natterer. *The Mathematics of Computerized Tomography*. John Wiley and Sons, New-York, 1986.
- [15] B. D. Smith. Cone Beam Tomography : Recent Advances and a Tutorial Review. *Optical Engineering*, 29(5) :524–534, 1990.