

## Chapitre 7

# Classification des méthodes d'inversion

L'objet de ce chapitre est de donner un aperçu général des différentes approches pour la résolution des problèmes inverses, ce qui nous permettra de mieux situer ces différentes méthodes. Dans le cas général, on peut distinguer deux approches : dans la première, on travaille directement dans le cadre fonctionnelle et en dimension infinie (fonctions et opérateurs) ; dans la deuxième, en revanche, dès le départ on discrétise les équations et on travaille en dimensions finie.

Dans ce chapitre, j'ai voulu faire une classification un peu plus fine en distinguant quatre approches :

- *Méthodes analytiques* où, soit directement soit après passage dans un espace dual, on peut obtenir une expression analytique explicite de la solution en fonction de la grandeur observable.
- *Méthodes de développement en série* où l'on décompose la grandeur observable et inconnue en une série de fonctions bien choisie pour pouvoir trouver une relation explicite, simple et inversible entre les coefficients de ces deux développements.
- *Méthode de décomposition sur une base* où, l'on projette la grandeur inconnue sur une base de fonctions bien choisie en fonction de l'opérateur la liant aux mesures, de telle sorte que l'on puisse obtenir une relation explicite et inversible entre les mesures et ces coefficients.
- *Méthodes algébriques* où on projette aussi la grandeur inconnue sur une base de fonctions bien choisie mais indépendantes de l'opérateur. De plus, ici on ne cherche pas forcément une relation inversible entre les mesures et ces coefficients, mais on travaille directement dans cet espace de dimension finie tout en étant conscient qu'il faut apporter de l'information *a priori* pour obtenir une solution satisfaisante au problème.

L'objet de ce chapitre est de détailler ces approches pour montrer les limites de chacune.

## 7.1 Méthodes analytiques

Dans ces méthodes, on fait l'hypothèse que les mesures sont des échantillons d'une fonction qui est liée à l'objet par une équation intégrale. On cherche alors une relation analytique explicite pour l'inverser. L'outil ici est l'analyse fonctionnelle. Une fois cette équation analytique trouvée, on la calcule numériquement. Evidemment, cette approche ne peut être utilisée qu'au cas par cas et en fonction de la nature de l'opérateur liant l'objet à la grandeur observable.

Pour illustrer le principe de ces méthodes considérons par exemple, le cas de la reconstruction d'image en tomographie à rayons X. Si l'on modélise le lien entre les mesures (projections) et l'objet par l'équation intégrale de la transformée de RADON  $\mathcal{R}$  :

$$p(r, \phi) = \iint f(x, y) \delta(r - x \cos \phi - y \sin \phi) dx dy$$

alors on a une expression analytique pour son inverse :

$$f(x, y) = \frac{1}{2\pi^2} \int_0^\pi \int_0^\infty \frac{\frac{\partial p(r, \phi)}{\partial r}}{(r - x \cos \phi - y \sin \phi)} dr d\phi$$

qui est classiquement décomposée en trois opérateurs :

$$\begin{aligned} \text{Dérivation } \mathcal{D} : \quad & \bar{p}(r, \phi) = \frac{\partial p(r, \phi)}{\partial r} \\ \text{Transformée de Hilbert } \mathcal{H} : \quad & g(r', \phi) = \frac{1}{\pi} \int_0^\infty \frac{\bar{p}(r, \phi)}{(r - r')} dr \\ \text{Rétroprojection } \mathcal{B} : \quad & f(x, y) = \frac{1}{2\pi} \int_0^\pi g(r' = x \cos \phi + y \sin \phi, \phi) d\phi \end{aligned}$$

ce qui permet d'écrire :

$$f = \mathcal{B} \mathcal{H} \mathcal{D} \mathcal{R} f$$

On peut alors s'imaginer que le problème est résolu et qu'il suffit de calculer cette expression intégrale d'une manière numérique en approximant l'équation intégrale par une somme appropriée. En effet, si l'on connaissait parfaitement  $p(r, \phi)$  pour tout  $r$  et tout  $\phi$  alors cela pourrait fournir une solution satisfaisante. Mais, en pratique on connaît  $p(r, \phi)$  pour des valeurs discrètes de  $r$  et de  $\phi$  et cela avec une précision finie. Alors, deux difficultés surviennent : la première est le calcul de la dérivée  $\frac{\partial p(r, \phi)}{\partial r}$  et la seconde, l'intégration par rapport à  $r$  due à la singularité du noyau de l'intégrale.

C'est pourquoi, dans cet exemple précis de reconstruction d'image en tomographie X, bien que l'expression de l'inversion était connue depuis longtemps, rares sont les méthodes de reconstruction qui effectuent l'inversion directement de cette manière. Il existe cependant un très grand nombre de travaux basés sur cette équation, mais où l'opération de dérivation  $\mathcal{D} = \frac{\partial p(r, \phi)}{\partial r}$  est approximée par un filtrage passe haut et l'opération d'intégration par rapport à la variable  $r$  qui correspond à une transformation de Hilbert  $\mathcal{H}$  est aussi effectuée dans le domaine de FOURIER, et finalement, l'intégration par rapport à la variable  $\phi$  est reconnue comme une opération de *rétroprojection*  $\mathcal{B}$ .

En effet, notant que la dérivation par rapport à la variable  $r$  dans le domaine spatial correspond à une multiplication par  $\Omega$  dans l'espace dual et la transformation de Hilbert correspond à un changement de signe dans l'espace dual, les deux opérations peuvent s'effectuer dans cet espace par un filtre dont la fonction de transfert (réponse fréquentielle)

est  $|\Omega|$ . D'autre part, en dimension infinie, tous ces opérateurs sont linéaires et symétriques et peuvent commuter, ce qui permet, par exemple écrire

$$f = \mathcal{B} \mathcal{H} \mathcal{D} \mathcal{R} f = \mathcal{H} \mathcal{D} \mathcal{B} \mathcal{R} f$$

Ainsi, on peut obtenir d'autres implantations algorithmiques pour calculer cette solution. Pour pouvoir résumer les différentes méthodes qui peuvent être utilisées pour le calcul de la solution de la TR inverse, notons les opérateurs suivants :

$$\begin{aligned} \text{Transformée de RADON } \mathcal{R} : & \quad p(r, \phi) = \iint f(x, y) \delta(r - x \cos \phi - y \sin \phi) dx dy \\ \text{Dérivation } \mathcal{D} : & \quad \bar{p}(r, \phi) = \frac{\partial p(r, \phi)}{\partial r} \\ \text{Transformée de Hilbert } \mathcal{H} : & \quad g(r', \phi) = \frac{1}{\pi} \int_0^\infty \frac{\bar{p}(r, \phi)}{(r - r')} dr \\ \text{Rétroprojection } \mathcal{B} : & \quad f(x, y) = \frac{1}{2\pi} \int_0^\pi g(r' = x \cos \phi + y \sin \phi, \phi) d\phi \\ \text{Transformée de FOURIER 1D } \mathcal{F}_1 : & \quad P(\Omega, \phi) = \int p(r, \phi) \exp[-j\Omega r] dr \\ \text{TF inverse 1D } \mathcal{F}_1^{-1} : & \quad p(r, \phi) = \int P(\Omega, \phi) \exp[+j\Omega r] d\Omega \\ \text{TF 2D } \mathcal{F}_2 : & \quad F(\omega_x, \omega_y) = \iint f(x, y) \exp[-j(\omega_x x + \omega_y y)] dx dy \\ \text{TF inverse 2D } \mathcal{F}_2^{-1} : & \quad f(x, y) = \iint F(\omega_x, \omega_y) \exp[-j(\omega_x x + \omega_y y)] d\omega_x d\omega_y \end{aligned}$$

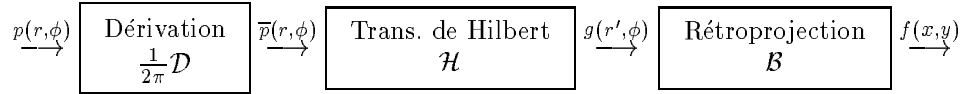
On peut alors résumer les différents algorithmes qui peuvent être utilisés pour le calcul de la solution de la TR inverse :

Malheureusement, bien que ces méthodes aient été utilisées avec succès dans certaines applications, en particulier en imagerie médicale, leur utilisation pratique reste plus limitée dans d'autres applications par exemple en contrôle non destructif. En effet, souvent en imagerie médicale on peut tourner autour de l'objet et on peut disposer d'un très grand nombre de mesures réparties d'une manière uniforme autour de l'objet, mais dans les applications telles qu'en contrôle non destructif, les mesures ne sont pas équiréparties sur l'espace de la transformée, et elles sont très limitées en nombre et entachées de bruit. Le résultat est que : les solutions analytiques ainsi calculées sont rarement satisfaisantes.

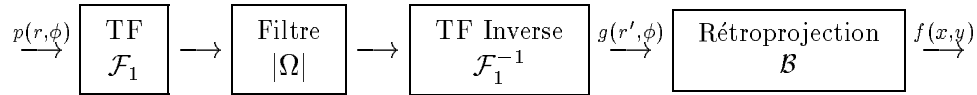
Les figures de la page suivante illustrent ces difficultés.

De plus, modéliser la relation entre les données et les inconnues par une transformation telle que l'on puisse trouver une transformation inverse analytique nécessite souvent des approximations qui deviennent rapidement assez irréalistes. Par exemple, vouloir prendre en compte la largeur du détecteur dans cet exemple met en cause le modèle simple de la transformée de RADON entre les projections et l'objet.

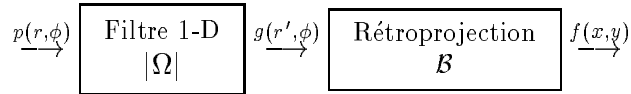
- Inversion directe :



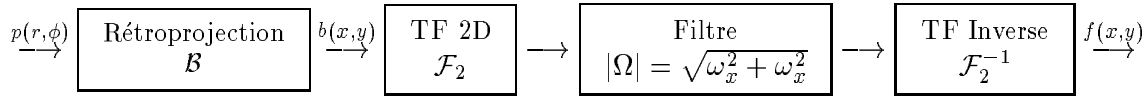
- Rétroprojection des projections filtrées :



- Filtrage par convolution et rétroprojection :



- Rétroprojection suivie de filtrage en 2D :



- Rétroprojection suivie d'un filtrage par convolution en 2D :

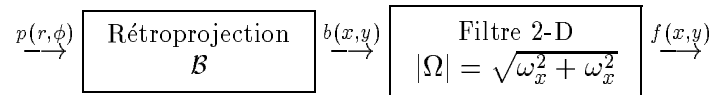


FIG. 7.1 - Méthodes analytiques pour l'inversion de la TR.

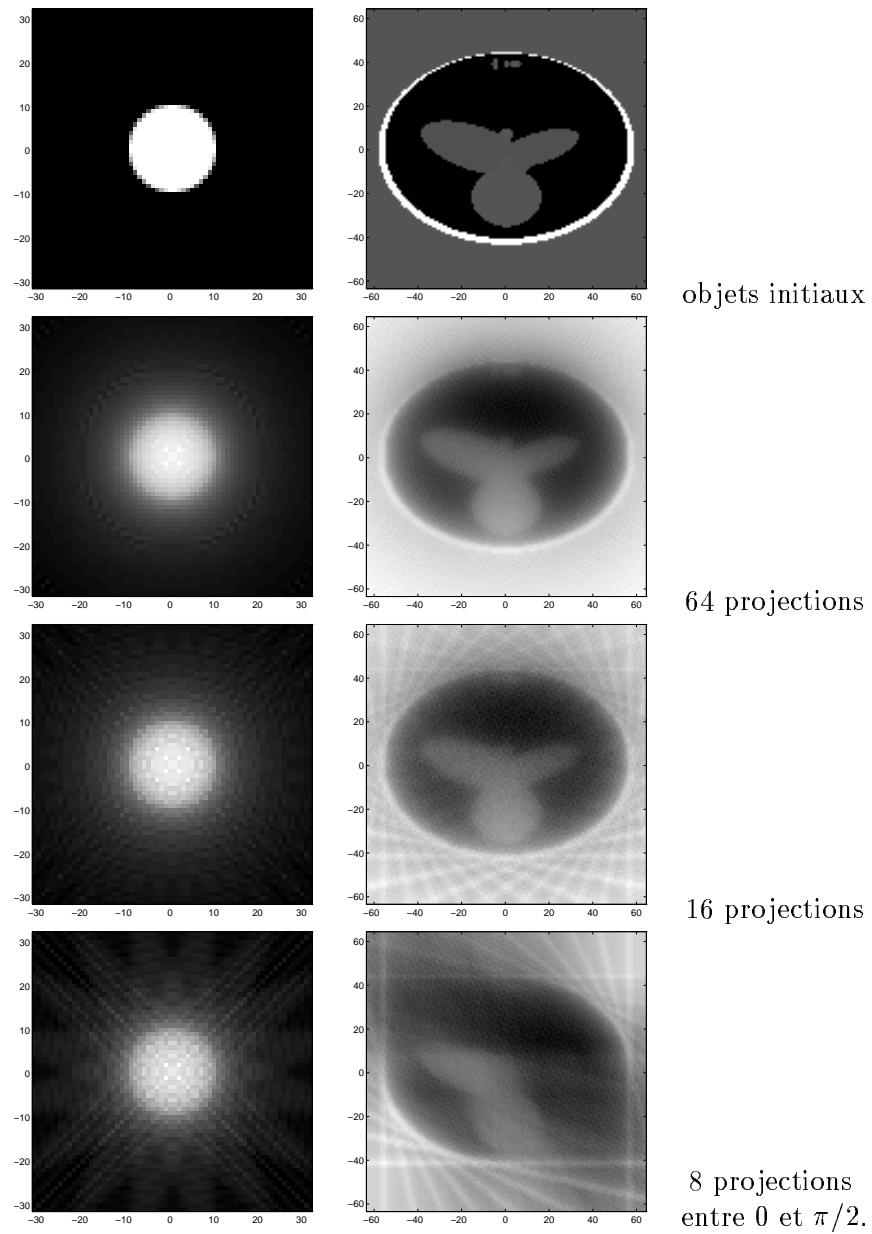


FIG. 7.2 - *Reconstruction par rétroprojection filtrée : Trois cas sont considérés : 64 projections, 16 projections réparties entre 0 et  $\pi$  et 8 projections réparties entre 0 et  $\pi/2$ .*

## 7.2 Méthodes de développement en série

Dans ces méthodes, on développe en série les fonctions  $f(\mathbf{r})$  et  $g(\mathbf{u})$

$$f(\mathbf{r}) = \sum_{k=1}^{\infty} f_k b_k(\mathbf{r}), \quad g(\mathbf{u}) = \sum_{k=1}^{\infty} g_k b_k(\mathbf{u}),$$

en choisissant les fonctions  $b_k$  telles que l'on puisse trouver une relation analytique simple et inversible entre les coefficients de ces développements. Ensuite, à partir des mesures  $g(\mathbf{u}_i)$ ,  $i = 1, \dots, K$  on détermine  $K$  coefficients  $g_k$ ,  $k = 1, \dots, K$  et on en déduit les  $K$  coefficients  $f_k$ ,  $k = 1, \dots, K$ . Finalement, on reconstruit  $\hat{f}(\mathbf{r})$  par

$$\hat{f}(\mathbf{r}) = \sum_{k=1}^K f_k b_k(\mathbf{r}).$$

Le choix des  $b_k(\mathbf{r})$  est crucial pour la mise en œuvre de cette méthode. Il dépend du contexte.

Pour illustrer le principe de ces méthodes nous allons considérer deux exemples : le premier est le problème de la déconvolution et le deuxième encore celui de la reconstruction d'image en tomographie à rayons X.

### Déconvolution :

Considérons le problème de la déconvolution des signaux :

$$g(r) = \int f(r') h(r, r') dr'$$

Sachant que les fonctions exponentielles sont les fonctions propres d'un tel opérateur, on décide de développer les deux fonctions  $f(r)$  et  $g(r)$  en une série de ces fonctions :

$$\begin{aligned} f(r) &= \sum_k f_k \exp[-jk\omega_0 r] \\ g(r) &= \sum_k g_k \exp[-jk\omega_0 r] \end{aligned}$$

Il est alors facile de montrer qu'il y a une relation très facile entre les coefficients  $f_k$  et  $g_k$  qui est :

$$g_k = H_k(\omega_0) f_k$$

où

$$H_k(\omega_0) = \int h(r) \exp[+jk\omega_0 r] dr$$

On peut alors envisager une méthode d'inversion en trois étapes :

1. À partir des mesures  $g(r)$  on estime les  $K$  coefficients  $g_k$  par

$$g_k = \frac{1}{2\pi} \int g(r) \exp[-jk\omega_0 r] dr, \quad k = 1 \dots, K$$

2. On calcule les  $K$  coefficients  $f_k$  en utilisant

$$f_k = \frac{g_k}{H_k(\omega_0)}$$

3. On calcule la solution par

$$f(r) = \sum_k^K f_k \exp[-jk\omega_0 r]$$

On peut facilement voir que lorsque  $\omega_0 = \frac{2\pi}{T}$  avec  $T$  la période d'échantillonnage des signaux  $f(r)$ ,  $g(r)$  et  $h(r)$ , l'on retrouve la notion de filtrage inverse. En effet,  $f_k$ ,  $g_k$  sont, respectivement, les coefficients de la décomposition en série de FOURIER des fonctions  $f(r)$  et  $g(r)$  et  $H_k(\omega_0)$  est la réponse fréquentielle du système. La principale difficulté, comme on le sait, se trouve dans l'étape 2 lorsque certains coefficients  $H_k(\omega_0)$  deviennent proche de zéro (amplification du bruit).

### Reconstruction d'image en tomographie à rayons X :

Nous avons vu que la relation entre un objet  $f(x, y)$  et ses projections  $p(r, \phi)$  en tomographie à rayons X est donnée par la TR. Pour illustrer l'utilisation de la méthode de développement en série dans ce cas, écrivons la TR en coordonnées polaires  $(\rho, \theta)$  :

$$p(r, \phi) = \int_0^\infty \int_0^{2\pi} f(\rho, \theta) \delta(r - \rho \cos(\phi - \theta)) \rho \, d\rho \, d\theta$$

Notant que  $p(r, \phi)$  est périodique en  $\phi$  et  $f(\rho, \theta)$  est périodique en  $\theta$ , on peut les développer en série de FOURIER :

$$p(r, \phi) = \sum_k p_k(r) \exp[-jk\phi], \quad f(\rho, \theta) = \sum_k f_k(\rho) \exp[-jk\theta],$$

et on peut établir les relations suivantes entre les coefficients  $p_k(r)$  et  $f_k(\rho)$  :

$$\begin{aligned} p_k(r) &= 2 \int_r^\infty f_k(\rho) T_k\left(\frac{r}{\rho}\right) \left(1 - \frac{r^2}{\rho^2}\right)^{-1/2} d\rho \\ g_k(\rho) &= \frac{-1}{\pi\rho} \int_\rho^\infty \frac{\partial p_k(r)}{\partial r} T_k\left(\frac{r}{\rho}\right) \left(\frac{r^2}{\rho^2} - 1\right)^{-1/2} dr \end{aligned}$$

avec  $T_k(t) = \cos[k \cos^{-1} t]$ .

On peut alors envisager une méthode d'inversion en trois étapes :

1. À partir des mesures  $p(r, \phi)$  on estime les  $K$  coefficients  $p_k(r)$  par

$$p_k(r) = \frac{1}{2\pi} \int_{-\pi}^{\pi} p(r, \phi) \exp[-jk\phi] \, d\phi, \quad k = 1 \dots, K$$

2. On calcule les  $K$  coefficients  $f_k(\rho)$  en utilisant

$$g_k(\rho) = \frac{-1}{\pi\rho} \int_\rho^\infty \frac{\partial p_k(r)}{\partial r} T_k\left(\frac{r}{\rho}\right) \left(\frac{r^2}{\rho^2} - 1\right)^{-1/2} dr, \quad k = 1 \dots, K$$

3. On calcule la solution par

$$f(\rho, \theta) = \sum_k^K f_k(\rho) \exp[-jk\theta]$$

Notez que l'étape la plus délicate dans ce cas est la deuxième, car elle nécessite le calcul d'une intégrale avec un noyau ayant des singularités.

Dans le cas général, malheureusement, ces méthodes ont aussi une utilisation pratique limitée car :

1. Dans la plupart des applications réelles on ne dispose pas d'expression analytique pour le noyau de l'équation intégrale. De plus on n'a pas, en général, de symétrie de telle sorte que l'on puisse choisir les fonctions  $b_k(\mathbf{r})$  satisfaisant à une relation analytique, simple et inversible avec  $g_k$  ;
2. La solution ainsi obtenue est très sensible d'une part au choix du seuil de troncature  $K$  de la série et d'autre part à la manière d'approximer le calcul de l'intégrale qui nous donne ces coefficients ;
3. La solution ainsi obtenue peut être très sensible aux erreurs de modélisation comme par exemple l'erreur géométrique dans les applications d'imagerie.



### 7.3 Méthodes de décomposition sur une base

Dans ces méthodes, on fait l'hypothèse que la solution du problème, c'est à dire la fonction objet, peut être décomposée en un nombre fini de fonctions de base  $b_j(\mathbf{r})$ :

$$f(\mathbf{r}) = \sum_{n=1}^N f_n b_n(\mathbf{r})$$

Ces fonctions sont choisies en rapport avec les propriétés du noyau  $h(\mathbf{r}, \mathbf{r}')$  de l'équation intégrale, de telle sorte qu'une fois la fonction objet décomposée sur cette base et remplacée dans l'équation d'observation :

$$\begin{aligned} g(\mathbf{r}_m) &= \int f(\mathbf{r}') h(\mathbf{r}_m, \mathbf{r}') d\mathbf{r}' = \int \sum_{n=1}^N f_n b_n(\mathbf{r}') h(\mathbf{r}_m, \mathbf{r}') d\mathbf{r}' \\ &= \sum_{n=1}^N f_n \int b_n(\mathbf{r}') h(\mathbf{r}_m, \mathbf{r}') d\mathbf{r}' = \sum_{n=1}^N f_n H_{ij}, \quad i = 1, \dots, M, \end{aligned}$$

on puisse calculer analytiquement les coefficients  $H_{ij}$ .

Notons que cette relation peut s'écrire sous la forme :

$$\mathbf{g} = \mathbf{H} \mathbf{f}$$

où  $\mathbf{H}$  est une matrice avec les éléments  $H_{ij}$ ,  $\mathbf{g} = [g_1, \dots, g_M]$  et  $\mathbf{f} = [f_1, \dots, f_N]$ .

Ensuite, deux approches sont courantes :

1. Choisir  $N = M$  et espérer inverser directement la matrice  $\mathbf{H}$  et résoudre ainsi ce système d'équations pour obtenir les coefficients de la base  $f_j$  ;
2. Choisir  $N < M$  et essayer de résoudre ce système au sens des moindres carrés ou au sens de l'inversion généralisée.

Pour illustrer cette approche, considérons un exemple de problème inverse que l'on rencontre en imagerie par microscopie électronique pour imager la distribution  $\rho(r)$  de la densité de la charge électrique du noyau d'un atome en mesurant une quantité  $F_c(q)$  appelée le facteur de forme. Le lien entre ces deux quantités est décrit par la relation intégrale suivante :

$$F_c(q) = 4\pi \int_0^\infty r^2 J_0(qr) \rho(r) dr \quad (7.1)$$

où  $J_0$  est la fonction de Bessel d'ordre zéro et  $q$  est la valeur absolue du moment de transfert. Lors de l'expérimentation on peut mesurer  $F_c(q)$  pour un nombre très réduit de  $\mathbf{q} = \{q_1, \dots, q_M\}$  et l'objectif de l'imagerie est de déterminer la fonction  $\rho(r)$ .

Dans cet exemple, supposant que  $\rho(r)$  s'annule au delà d'un rayon  $R_c$ , un choix judicieux de fonctions de base est :

$$\rho(r) = \begin{cases} \sum_{n=1}^N f_n J_0(q_n r) & r \leq R_c \\ 0 & r > R_c \end{cases} \quad (7.2)$$

Avec ce choix, et en utilisant une des propriétés d'orthogonalité des fonctions de Bessel :

$$\int_0^{R_c} r^2 J_l(q_n r) J_l(q_m r) dr = \frac{R_c^3}{2} J_{l+1}^2(q_n R_c) \delta_{n,m} \quad (7.3)$$

il n'est pas difficile de montrer que :

$$\mathbf{F}_c = \mathbf{A} \mathbf{f} \quad (7.4)$$

avec  $\mathbf{q} = \{q_1, \dots, q_M\}$ ,  $\mathbf{F}_c = \{F_c(q_1), \dots, F_c(q_M)\}$ ,  $\mathbf{f} = \{f_1, \dots, f_N\}$  et

$$A_{mn} = \frac{4\pi R_c^2}{q_m} \frac{(-1)^n}{(q_m R_c)^2 - (n\pi)^2} \sin(q_m R_c). \quad (7.5)$$

En effet, remplaçant (7.2) dans (7.1) on a

$$\begin{aligned} F_c(q_m) &= 4\pi \int_0^{R_c} r^2 J_0(q_m r) \sum_{n=1}^N f_n J_0(q_n r) dr \\ &= 4\pi \sum_{n=1}^N f_n \int_0^{R_c} r^2 J_0(q_m r) J_0(q_n r) dr \\ &= \sum_{n=1}^N f_n A_{mn} \end{aligned}$$

avec  $A_{mn}$  définie par (7.5).

Notons que, si  $M = N$  et si les mesures pouvaient être faites exactement aux points  $q_n = \frac{n\pi}{R_c}$ , alors la matrice  $\mathbf{A}$  serait une matrice diagonale avec les éléments diagonaux  $A_{nn} = \frac{R_c^3}{2} J_1^2(q_n R_c)$ . On pourrait alors obtenir directement les coefficients  $f_n$  par

$$f_n = \frac{F_c(q_n)}{2\pi R_c^3 J_1^2(q_n R_c)}. \quad (7.6)$$

Ceci est due au fait que si  $q_n = \frac{n\pi}{R_c}$  la base choisie est une base orthogonale.

En général cependant, les mesures sont faites aux points qui ne sont pas équirépartis et la matrice  $\mathbf{A}$  n'est plus diagonale. De plus, il n'y a aucune raison de choisir  $M = N$ , car on peut toujours essayer d'estimer  $\mathbf{f}$  au moins par MC :

$$\begin{aligned} \hat{\mathbf{f}} &= (\mathbf{A}^t \mathbf{A})^{-1} \mathbf{A}^t \mathbf{F}_c \quad \text{si } M > N \\ &= \mathbf{A}^t (\mathbf{A} \mathbf{A}^t)^{-1} \mathbf{F}_c \quad \text{si } M < N \end{aligned}$$

ou par MC de norme minimale (MCNM) :

$$\begin{aligned} \hat{\mathbf{f}} &= (\mathbf{A}^t \mathbf{A} + \lambda \mathbf{I})^{-1} \mathbf{A}^t \mathbf{F}_c \quad \text{si } M > N \\ &= \mathbf{A}^t (\mathbf{A} \mathbf{A}^t + \lambda \mathbf{I})^{-1} \mathbf{F}_c \quad \text{si } M < N \end{aligned}$$

ou encore par d'autres méthodes.

Les figures qui suivent montrent quelques exemples de simulation pour ce problème.

Ces méthodes ont une portée un peu plus générale que les méthodes analytiques précédentes; elles peuvent être utilisées lorsqu'on a une expression analytique du noyau de l'équation intégrale qui lie les mesures à l'objet et lorsqu'il y a une base convenable

telle que la fonction désirée puisse se décomposer sur cette base avec un nombre faible de coefficients.

Cependant, ces méthodes aussi ont une utilisation pratique limitée car :

1. dans la plupart des applications réelles on ne dispose pas d'expression analytique pour le noyau de l'équation intégrale (prenez par exemple le cas de la déconvolution des signaux ou celui de la restauration d'image);
2. Il est assez difficile de trouver une base orthogonale et complète;
3. La solution ainsi obtenue est, en général, très sensible au choix du seuil de troncature  $N$  de la série.
4. Lorsque  $N$  est grand, la matrice  $\mathbf{A}$  peut être très mal conditionnée.
5. Lorsqu'on a une information *a priori* sur la solution (par exemple la positivité) il est difficile de transmettre cette information sur les coefficients.

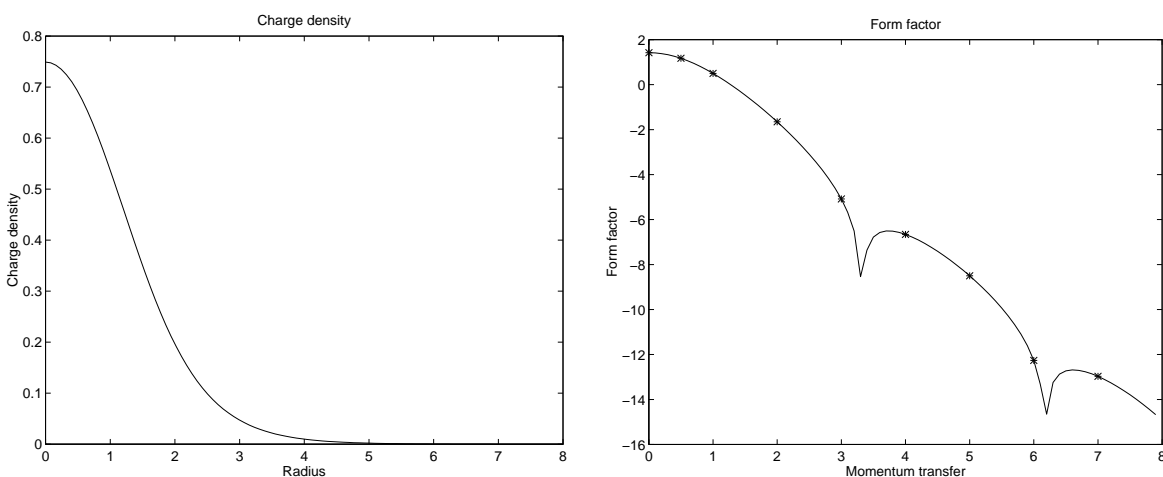


Figure a: La densité de charge théorique  $\rho(r)$  [gauche], le facteur de forme théorique  $\log \|F_c(q)\|$  et les données utilisées pour les simulations [droite].

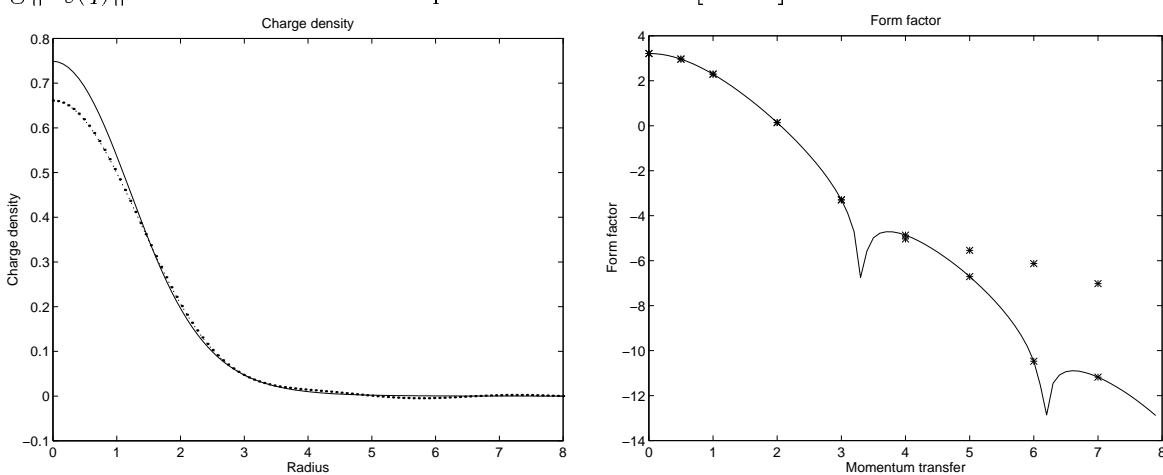


Figure b. Résultat de reconstruction de  $\hat{\rho}(r)$  obtenus par MC (points) et par MCNM (tirets) pour  $N = 5$  [gauche] et les facteurs de formes correspondants  $\log \|\hat{F}_c(q)\|$  [droite]. Dans ce cas les deux solutions sont pratiquement identiques car la matrice  $\mathbf{A}^t \mathbf{A}$  est bien conditionnée.

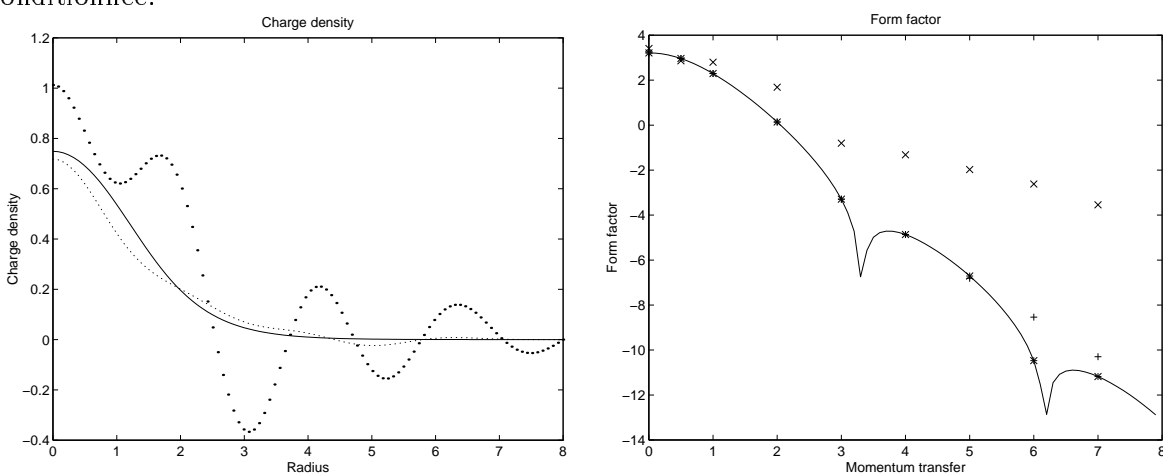


Figure c. Résultat de reconstruction de  $\hat{\rho}(r)$  obtenus par MC (points) et par MCNM (tirets) pour  $N = 10$  [gauche] et les facteurs de formes correspondants  $\log \|\hat{F}_c(q)\|$  [droite]. Dans ce cas la matrice  $\mathbf{A} \mathbf{A}^t$  est mal conditionnée, alors que la matrice  $\mathbf{A} \mathbf{A}^t + \lambda \mathbf{I}$  est bien conditionnée. C'est pourquoi, la solution MC n'est plus satisfaisante, alors que la solution MCNM est encore satisfaisante.

FIG. 7.3 - Illustration de la méthode de décomposition sur une base.

## 7.4 Méthodes algébriques déterministes

Dans ces méthodes on commence par discrétiser la fonction objet  $f(\mathbf{r})$  et par conséquent l'équation intégrale avec la résolution maximale souhaitée. Une manière de faire cette discrétisation est, comme dans le cas précédent, de décomposer la fonction  $f(\mathbf{r})$  sur une base :

$$f(\mathbf{r}) = \sum_{n=1}^N f_n b_n(\mathbf{r}),$$

sauf qu'ici les fonctions de base sont choisies parmi

$$b_n(\mathbf{r}) = \delta(\mathbf{r} - n\Delta) \quad (7.7)$$

ou

$$b_n(\mathbf{r}) = \begin{cases} 1 & \text{si } \mathbf{r} \in [(n-1)\Delta, n\Delta] \\ 0 & \text{ailleurs} \end{cases} \quad (7.8)$$

ou encore (pour un échantillonnage non régulier)

$$b_n(\mathbf{r}) = \begin{cases} 1 & \text{si } \mathbf{r} \in D_n \\ 0 & \text{ailleurs} \end{cases} \quad (7.9)$$

avec  $D_n$  un ensemble de régions dans le domaine de la fonction telle que  $D_i \cap D_j = \emptyset, \forall i \neq j$  et  $\cup D_j = S$ .

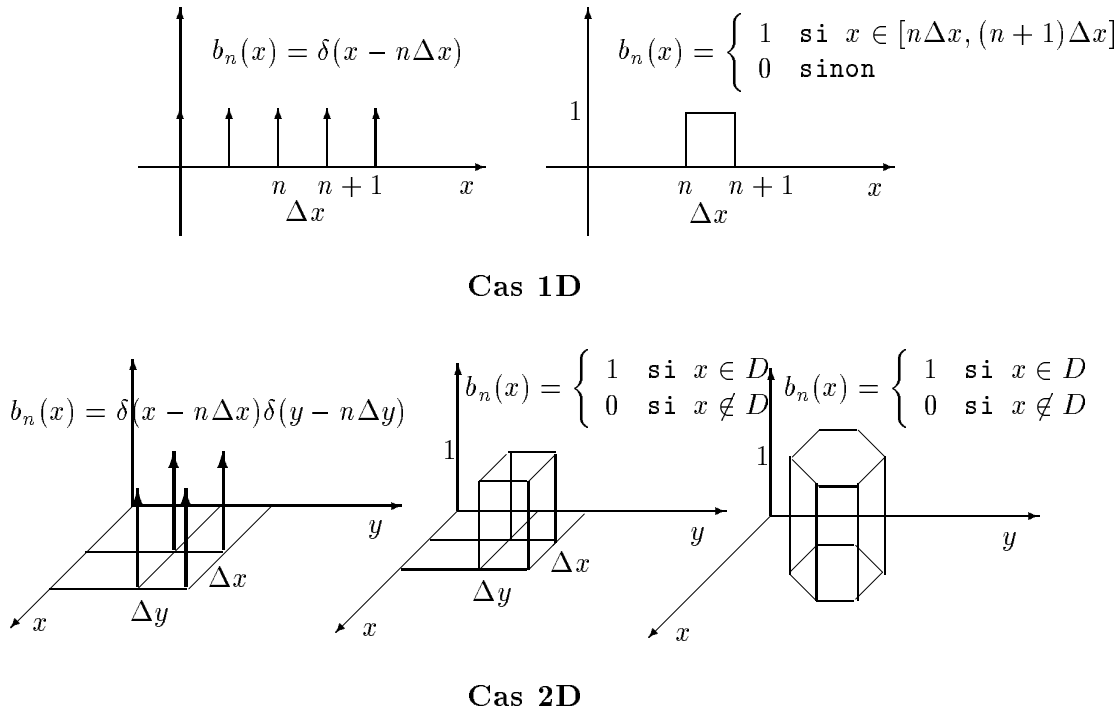


FIG. 7.4 - Échantillonnage et fonctions bases.

Les pas d'échantillonnage  $\Delta = \{\Delta_1, \Delta_2, \dots\}$  sont choisis en faisant un compromis entre la résolution maximale souhaitée et le coût de calcul.

On obtient ainsi, comme dans le cas précédent, un système d'équations linéaires de la forme :

$$\mathbf{g} = \mathbf{H} \mathbf{f},$$

où  $\mathbf{f}$  est un vecteur de dimension  $N$  contenant les coefficients  $f_n$ . La différence essentielle est dans la signification de ces coefficients. En choisissant les fonctions de base (7.7), ces coefficients représentent les échantillons de la fonction  $f$ , alors qu'en choisissant les fonctions de base (7.8) ou (7.9), ces coefficients représentent les valeurs moyenne de la fonction  $f$  dans des régions  $[(n-1)\Delta, n\Delta]$  ou  $D_j$ .

Ainsi, contrairement aux méthodes décrites précédemment, dans ces méthodes cette discrétisation se fait sans trop se soucier du choix des fonctions de base, du nombre de ces fonctions (on suppose en général que l'objet est à support borné et on choisit un pas de discrétisation de l'espace le plus petit correspondant à la résolution souhaitée), du mauvais conditionnement de la matrice  $\mathbf{H}$  et des erreurs de discrétisation. En effet, la résolution de ces problèmes est laissée à la charge de la méthode de régularisation qui suit.

Il serait en effet illusoire de vouloir se ramener à une équation de la forme  $\mathbf{g} = \mathbf{H} \mathbf{f}$  avec  $\mathbf{H}$  une matrice carrée et inversible et vouloir proposer la solution  $\hat{\mathbf{f}} = \mathbf{H}^{-1} \mathbf{g}$ .

Il serait de même illusoire de se contenter des solutions du type pseudo-solutions

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \{\Delta(\mathbf{g}, \mathbf{H} \mathbf{f})\}$$

où  $\Delta(., .)$  est une mesure de distance dans l'espace des mesures, ou les solutions au sens des moindres carrés :

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \{\|\mathbf{g} - \mathbf{H} \mathbf{f}\|^2\}$$

qui satisfont l'équation normale

$$\mathbf{H}^t \mathbf{H} \hat{\mathbf{f}} = \mathbf{H}^t \mathbf{g}$$

ou encore, la solution inverse généralisée (IG) qui est la solution de norme minimale de cette équation normale. Ces solutions restent très sensibles aux erreurs de discrétisation et de mesure  $\mathbf{b}$ . Il faut introduire une information *a priori* de manière explicite sur la solution et sur les erreurs de mesure; il faut régulariser.

Parmi les méthodes de régularisation déterministes, on peut citer les méthodes basées sur la décomposition tronquée des valeurs singulières, les méthodes itératives de minimisation d'un critère du type  $J(\mathbf{f}) = \|\mathbf{g} - \mathbf{H} \mathbf{f}\|^2$ , avec un nombre limité d'itérations ou avec des contraintes supplémentaires du type positivité par exemple, ou encore des méthodes qui minimisent un critère mixte du type

$$J(\mathbf{f}) = \|\mathbf{g} - \mathbf{H} \mathbf{f}\|^2 + \lambda H(\mathbf{f}),$$

où  $H(\mathbf{f})$  est une fonctionnelle régularisante et  $\lambda$  est le coefficient de régularisation.

De manière encore plus générale on peut définir la solution régularisée comme celle qui minimise un critère de la forme

$$J(\mathbf{f}) = \Delta_1(\mathbf{g}, \mathbf{H} \mathbf{f}) + \lambda \Delta_2(\mathbf{f}, \mathbf{f}_0),$$

où  $\Delta_1$  et  $\Delta_2$  sont deux distances et  $\mathbf{f}_0$  est une solution par défaut (*a priori*).

Parmi ces méthodes, la dernière a été utilisée avec succès dans un grand nombre d'applications. Cependant elle souffre des limitations suivantes :

1. le choix de la fonctionnelle  $H(\mathbf{f})$  est assez arbitraire et on manque d'arguments pour le choix du paramètre de régularisation  $\lambda$  ;

2. le bruit  $\mathbf{b}$  est pris en compte mais de manière implicite ; la prise en compte explicite des caractéristiques du bruit n'est pas aisée ;
3. il n'est pas aisé non plus de prendre en compte explicitement des informations statistiques sur la solution.

Pour remédier à ces difficultés il n'y a pas d'autres solutions que de considérer la résolution du problème inverse comme un problème d'estimation probabiliste.

## 7.5 Méthodes probabilistes

L'idée de base dans ces méthodes est de dire que lorsqu'on cherche à résoudre d'une manière satisfaisante un problème inverse il faut

- prendre en compte explicitement les erreurs : non seulement celles de la discrétisation mais aussi celles due aux mesures et celles due à modélisation ;
- prendre en compte les informations *a priori* ; et
- caractériser l'incertitude qui reste dans la solution proposée.

L'outil mathématique *théorie des probabilités* est alors l'outil idéal. On utilise cet outil pour représenter l'incertitude sur les mesures et/ou le manque de connaissance parfaite sur une grandeur par une loi de probabilité, et ensuite on utilise la notion d'estimation pour déterminer les grandeurs inconnues. On peut distinguer alors plusieurs approches ou classes de méthodes :

1. L'approche ou les méthodes qui n'utilisent que les moments – jusqu'à l'ordre deux, ce qui implicitement revient à faire l'hypothèse gaussienne – d'ordre plus élevé. On cherche à lier les moments de la grandeur observée aux moments de la grandeur inconnue afin de les déterminer en inversant ce lien. On peut citer en exemple le filtrage optimal en restauration des signaux. Le principal avantage de ces méthodes est la facilité de mise en œuvre et le fait que la formulation du problème peut se faire en continu ou en discret.
2. L'approche estimation *statistique classique* où l'on prend en compte explicitement le caractère incertain des mesures en la caractérisant par une loi de probabilité conditionnelle  $p(\mathbf{g}|\mathbf{f})$ . On considère ensuite  $p(\mathbf{g}|\mathbf{f})$  comme une fonction de  $\mathbf{f}$  que l'on nomme la fonction *vraisemblance*  $V(\mathbf{f})$ . On définit alors un critère basé sur cette fonction, par exemple le maximum de vraisemblance (MV), qui permet alors d'estimer  $\mathbf{f}$ .

Mais malheureusement, ces méthodes ne peuvent pas fournir de solutions satisfaisantes à des problèmes inverses, où souvent le nombre de paramètres à estimer est aussi grand, sinon plus que le nombre de points de mesures.

3. L'approche estimation *statistique bayésienne* où non seulement on prend en compte l'incertitude sur les mesures, mais aussi on attribue une loi *a priori* aux inconnues  $\mathbf{f}$  qui traduit notre information *a priori* sur ces inconnues. Ainsi, on peut combiner l'information contenue dans les données et cette information *a priori* si nécessaire dans des problèmes inverses.

On retrouve en effet la principale idée de la régularisation qui consiste à combiner l'information contenue dans les données et celle contenue dans l'*a priori*.

4. Enfin, on peut aussi noter l'approche basée sur le principe du maximum d'entropie. L'idée de base peut se résumer dans les étapes suivantes :
- on définit une loi *a priori*  $p_0(\mathbf{f})$  ou plus généralement une mesure de référence pour traduire notre information *a priori* sur ces inconnues.
  - on considère les mesures comme des contraintes sur la loi *a posteriori*  $p(\mathbf{f})$ , mais ces mesures n'étant pas suffisantes pour déterminer d'une manière unique  $p(\mathbf{f})$ , on choisirait la loi  $p(\mathbf{f})$  qui est la plus proche de  $p_0(\mathbf{f})$  au sens de la mesure de Kulback, et finalement,
  - on définit comme solution du problème inverse la moyenne de cette loi.

Nous allons détailler ces approches dans le chapitre suivant.