Problèmes inverses

Ali Mohammad-Djafari

Cours du DEA Automatique et Traitement du Signal

Laboratoire des Signaux et Systèmes (CNRS-ESE-UPS)
École Supérieure d'Électricité
Plateau de Moulon, 91192 Gif-sur-Yvette Cedex, France.

Version provisoire du: 1er mai 2001 [1]

1. Rapport interne LSS: fichier `poly.tex`

Table des matières

Chapitre 1

Introduction

En sciences expérimentales, rares sont les situations où l'on peut mesurer directement une quantité f à laquelle on s'intéresse, pour étudier sa distribution spatiale $f(\boldsymbol{r})$ à un instant donné ou sa variation temporelle $f(t)$ à une position donnée ou encore sa variation spatio-temporelle $f(\boldsymbol{r},t)$.

Souvent, cette quantité est mesurée par l'intermédiaire d'un simple capteur (ou d'un système de mesure plus complexe et rarement idéal) qui a pour rôle de fournir une grandeur mesurable g liée à f par une relation simple. Prenons comme exemple la mesure d'une température par un thermomètre classique. La grandeur réellement mesurée est une longueur qui, idéalement, est proportionnelle à la température. Mais, lorsque l'on souhaite étudier la variation temporelle de cette température les difficultés apparaissent : le capteur ne reproduit pas fidèlement la variation de la température lorsque celle-ci est un peu trop rapide (limitation de la bande passante du capteur). On essaie ensuite d'étudier le lien entre la sortie du capteur $g(t)$ et son entrée $f(t)$ en fournissant un *modèle*, souvent linéaire et invariant par translation–*modèle de convolution* –et donc caractérisé par sa *réponse impulsionnelle* $h(t)$ ou sa *réponse fréquentielle* $H(\omega)$. Ensuite on *analyse* ce modèle pour prescrire la plage de son utilisation.

L'établissement d'un lien entre la sortie $g(t)$ et l'entrée $f(t)$ est ce qu'on appelle *modélisation directe.* t La détermination de la réponse impulsionnelle ou de la réponse fréquentielle est le problème de l'*identification.* Étant donné le modèle, l'étude de la sortie pour différentes formes de l'entrée est l'*analyse du problème directe.* Et finalement, étant donné le modèle et la sortie, la détermination de l'entrée est ce qu'on appelle l'*inversion* ou encore la *résolution du problème inverse.*

Lorsque le modèle liant la sortie $g(t)$ et l'entrée $f(t)$ est linéaire et invariant par translation on a un modèle de *convolution*:

$$g(t) = \int f(t')h(t-t')\,\mathrm{d}t'.$$

Étant données l'entrée $f(t)$ et la réponse impulsionnelle $h(t)$, en général, on ne rencontre pas de problème particulier pour déterminer la sortie $g(t)$. Mais, malheureusement, les problèmes réels se formulent ainsi :

1. Ayant mesuré $f(t)$ et $g(t)$, déterminer la réponse impulsionnelle $h(t)$. C'est le problème de *l'identification de la réponse impulsionnelle.*
2. Connaissant $h(t)$ et ayant mesuré $g(t)$, déterminer l'entrée $f(t)$. C'est le problème inverse de la *déconvolution simple.*

3. Ayant seulement mesuré $g(t)$, déterminer à la fois la réponse impulsionnelle $h(t)$ et l'entrée $f(t)$. C'est le problème inverse de la *déconvolution aveugle*.

Prenons un deuxième exemple: l'imagerie tomographique, où on souhaite étudier la distribution spatiale de la densité de matière $f(x,y,z)$ à l'intérieur d'un corps. Pour la mesurer, nous avons besoin d'un intermédiaire: des rayons X par exemple. On illumine ce corps par ces rayons et on mesure les résultats de son interaction avec ces ondes: la perte d'énergie des rayons lors de la traversé du corps, c'est-à-dire les radiographies suivant plusieurs directions $p(r,s,\phi)$. On essaie ensuite d'étudier le lien entre ces mesures $p(r,s,\phi)$ et l'inconnue $f(x,y,z)$ en fournissant un *modèle* liant ces quantités. Le problème inverse consiste alors à déterminer la fonction $f(x,y,z)$ à partir de ces *projections* $p(r,s,\phi)$. Lorsqu'on s'intéresse à une section du corps étudié $f(x,y,z_0)$ on parle de la *reconstruction tomographique* ou de la *reconstruction d'image 2D* et dans le cas général on parle de la *reconstruction 3D*. Dans le cas 2D, une modélisation simple entre $f(x,y)$ et $p(r,\phi)$ est la transformée de Radon (TR):

$$p(r,\phi) = \int_L f(x,y)\,\mathrm{d}l = \iint f(x,y)\delta(r - x\cos\phi - y\sin\phi)\,\mathrm{d}x\,\mathrm{d}y$$

L'objet de ce texte est de fournir un cadre général et compréhensible pour la description et la résolution de ces problèmes.

- Dans le chapitre 2, nous montrons à travers des exemples que les problèmes inverses se trouvent au cœur d'un très grand nombre de domaines des sciences expérimentales et plus particulièrement des sciences pour l'ingénieur.
- Dans le chapitre 3, nous essayons de fournir un cadre général pour les problèmes inverses et de montrer que les différents exemples d'applications que nous avons présentés au chapitre précédent sont des cas particulier de ce cadre général.
- Dans le chapitre 4 [1], nous essayons de définir les notions des problèmes bien-posés et mal-posés et de montrer pourquoi la résolution des problèmes inverses est une tâche difficile.
- Dans le chapitre 5, nous introduisons les notions de pseudo-solutions, solutions au sens des moindres carrés (MC) et solution inverse généralisée (IG). Nous verrons ensuite que, malheureusement, bien que ces notions permettent, sous certaines conditions, de définir une solution unique pour un problème inverse mal-posé, elles ne permettent pas de fournir une solution satisfaisante à ces problèmes.
- Le chapitre 6 est consacré à la notion de régularisation et à l'étude des différentes techniques de régularisation pour la résolution des problèmes inverses linéaires.
- Dans le chapitre 7, nous essaierons de classifier les différentes méthodes d'inversion. Ceci nous permettra d'avoir une vue plus synthétique de ces différentes méthodes et des liens qui existent entre elles.
- Dans le chapitre 8, nous présentons la résolution des problèmes inverses comme un problème d'estimation dans un cadre probabiliste. Nous distinguerons ensuite quatre approches:
  1. Approche estimation au sens des moindres carrés où estimation linéaire en moyenne quadratique où seuls les moments d'ordre un et deux sont utilisés,

1. Les contenus des chapitres 4, 5 et 6, restent pour l'instant très sommaires et sont inspirés du polycopié du Guy Demoment intitulé "Déconvolution des signaux". Ce polycopié est disponible à la bibliothèque de Supélec.

ce qui revient implicitement à faire l'hypothèse que toutes les variables sont gaussiennes. Pour illustrer cette approche nous prendrons comme exemple le filtrage optimal ;

2. Approche estimation au sens du maximum de vraisemblance où nous verrons rapidement que cette approche, bien que plus générale, dans le sens où l'on peut prendre en compte explicitement la loi des mesures, est peu satisfaisante pour la résolution des problèmes inverses ;
3. Approche maximum d'entropie où, après un rappel sur le principe du maximum d'entropie et son utilité pour l'attribution d'une loi de probabilité, nous verrons que cette approche permet de fournir une nouvelle interprétation de la notion de régularisation et permet de répondre à un certain nombre de questions concernant le choix des fonctionnelles de régularisation ;
4. Approche estimation bayésienne où nous verrons comment elle peut fournir un cadre fédérateur pour la définition d'une solution régularisée englobant tous les critères précédents. Nous verrons ensuite que cette approche est plus cohérente, plus générale, plus simple et plus efficace pour aborder les problèmes inverses.

– Le chapitre 9 est entièrement réservé à l'approche bayésienne. Dans ce chapitre, nous détaillerons cette approche et présenterons un certain nombre de cas particuliers qui illustreront la généralité et l'universalité de cette approche.
– Dans le chapitre 11, nous donnons un aperçu général de la modélisation markovienne des signaux et des images. Les modèles markoviens et particulièrement les champs de Gibbs sont devenus les outils indispensables pour la traduction d'une information *a priori* plus élaborée que la douceur ou la positivité en une loi de probabilité. En effet, la modélisation markovienne permet de construire des modèles qui peuvent prendre en compte des discontinuités dans un signal ou une image. Cette outil prend particulièrement son importance en traitement d'image où l'on peut construire les modèles composites (bords, contours, intensité) pour les images et les utiliser lors de la segmentation, restauration ou reconstruction d'image.
– Dans le chapitre 12, nous aborderons le problème de l'estimation des hyperparamètres, et en particulier, la détermination du paramètre de régularisation. Dans un premier temps, nous décrirons brièvement les méthodes classiques du type : adéquation aux données, la courbe en L, la validation croisée, et la validation croisée généralisée. Ensuite, nous présenterons le problème général de l'estimation des hyperparamètres dans le cadre bayésien.
– Dans les chapitres 13, 14 et 15 nous présenterons un certain nombre d'*études de cas* comme la déconvolution des signaux, la restauration d'image, la reconstruction d'images, ...

Un certain nombre d'annexes sont présentées à la fin du document qui ont pour but soit de compléter certains détails ou d'apporter des éléments et des outils qui sont nécessaires lorsqu'un ingénieur veut aborder la résolution de problèmes inverses.

Et, finalement, une bibliographie classifiée concluera ce document.

Chapitre 2

Exemples de problèmes inverses

L'objet de ce chapitre est de montrer que les problèmes inverses se trouvent au cœur d'un très grand nombre de domaines de la science expérimentale et plus particulièrement de la science pour l'ingénieur. Bien entendu, la description de chaque application ne peut qu'être très brève.

2.1 Instrumentation

Un capteur ou un appareil de mesure serait idéal si sa sortie $g(t)$ était proportionnelle à son entrée $f(t)$ pour tout t, c'est-à-dire lorsque

$$\forall t \quad g(t) = af(t) + b,$$

a et b étant deux constantes fixées.

Ceci est rare. Le mieux que l'on puisse demander est que sa sortie soit une fonction linéaire de son entrée :

$$g(t) = \int f(\tau)h(t,\tau)\, d\tau.$$

Mieux encore, c'est le cas où sa réponse $h(t,\tau)$ ne varie pas en cours du temps, ce qui veut dire $h(t,\tau) = h(t-\tau)$. On a ainsi

$$g(t) = \int f(\tau)h(t-\tau)\, d\tau.$$

Il s'agit d'une équation de convolution. La fonction $h(t)$ est appelée la réponse impulsionnelle. Lorsque le système est causal, i.e; $h(t) = 0, \forall t < 0$, on a

$$g(t) = \int_0^t f(\tau)h(t-\tau)\, d\tau.$$

Un capteur serait parfait si

$$h(t) \simeq a\delta(t) \quad \longrightarrow \quad f(t) = \frac{1}{a}g(t)$$

Lorsque ceci n'est pas le cas, les ingénieurs ont l'habitude de travailler avec la réponse indicielle du capteur :

$$s(t) = \int_0^t h(t-\tau)\, d\tau$$

sur laquelle ils mesurent, le *temps de montée*, le *dépassement*, le *déphasage*, *la résolution maximale*, le critère de Raleigh, etc. Et lorsque ces grandeurs ne sont pas satisfaisantes ils changent le capteur. Mais cela peut être coûteux et même parfois techniquement impossible.

Que peut-on faire alors? Il y a vingt ans, cette question n'avait pas de réponse pratique. Aujourd'hui, avec les techniques numériques de l'inversion, on peut proposer la *déconvolution*. L'idée de base est de placer derrière le capteur un microprocesseur qui, partant des échantillons de la réponse impulsionnelle $h(t)$ et ceux de la sortie du capteur $g(t)$ fournirais $\widehat{f}(t)$ qui est beaucoup plus proche de $f(t)$ que $g(t)$, dépassant ainsi le critère de Raleigh. En effet, $\widehat{f}(t)$ aura une meilleure résolution que $g(t)$. On peut même par des techniques adaptatives compenser les effets d'une dégradation des caractéristiques du capteur en réestimant un certain nombre de paramètres de sa réponse impulsionnelle au cours du temps.

Ainsi, au lieu d'exiger du capteur d'avoir un temps de montée de telle valeur, un dépassement de telle autre valeur, etc., on lui imposerait d'être *linéaire*, *fidèle*, et *précis*. Le traitement numérique permettra alors de compenser le manque de résolution ou la limitation de la bande passante, etc.

$$\left\{ \begin{array}{c} \text{linéarité} \\ \text{fidélité} \\ \text{précision} \end{array} \right. \quad + \quad \text{Déconvolution} \quad = \quad \left\{ \begin{array}{l} \text{Amélioration de la résolution} \\ \text{Dépassement du critère de Raleigh} \end{array} \right.$$

2.2 Tomographie à rayons X

Si un objet de matière homogène est illuminé par des rayons X et qu'un rayon mono–énergétique de faible largeur traverse cet objet, l'intensité du rayon mesuré à la sortie s'obtient par

$$I = I_0 \exp\left[-\mu L\right]$$

où I_0 est l'intensité du rayon à l'entrée, L est la distance parcourue par le rayon, et μ est la constante d'atténuation linéaire de l'objet qui dépend de la densité de matière et de sa composition nucléaire.

Considérons maintenant une section transversale d'un objet non homogène perpendiculaire à l'axe oz, et supposons que les rayons traversent cette section. En caractérisant l'objet par la distribution de sa constante d'atténuation linéaire μ, fonction continue des deux variables d'espace; $\mu(x,y) = f(x,y)$; on a alors

$$\frac{I}{I_0} = \exp\left[-\int_L f(x,y)\,\mathrm{d}l\right]$$

ou bien

$$-\ln\left(\frac{I}{I_0}\right) = \int_L f(x,y)\,\mathrm{d}l$$

où $\mathrm{d}l$ est l'élément de longueur sur le parcours L. Ainsi, si on déplace en parallèle l'émetteur (source S) et le récepteur (détecteur D) sur une ligne droite faisant un angle ϕ avec l'axe ox on obtient une projection $p(r,\phi)$:

$$p(r,\phi) = \int_L f(x,y)\,\mathrm{d}l$$

En se référant aux figures 2.2 et 2.2, on voit que cette équation peut aussi s'écrire sous la forme

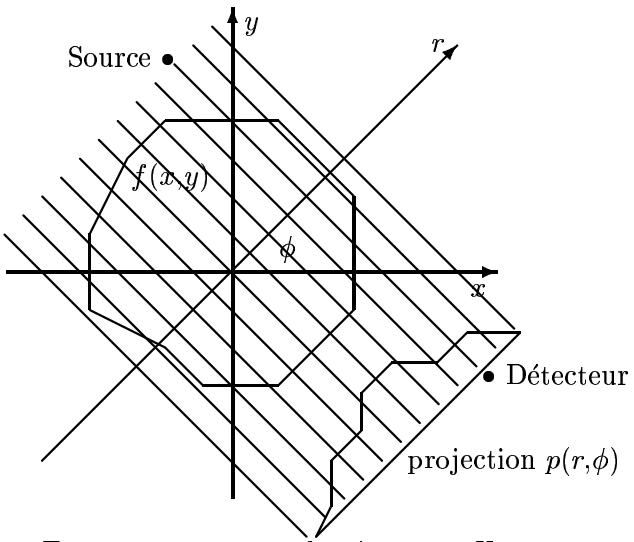


FIG. 2.1 – *Tomographie à rayons X*

$$p(r,\phi) = \iint_D f(x,y)\,\delta(r - x\cos\phi - y\sin\phi)\,\mathrm{d}x\,\mathrm{d}y$$

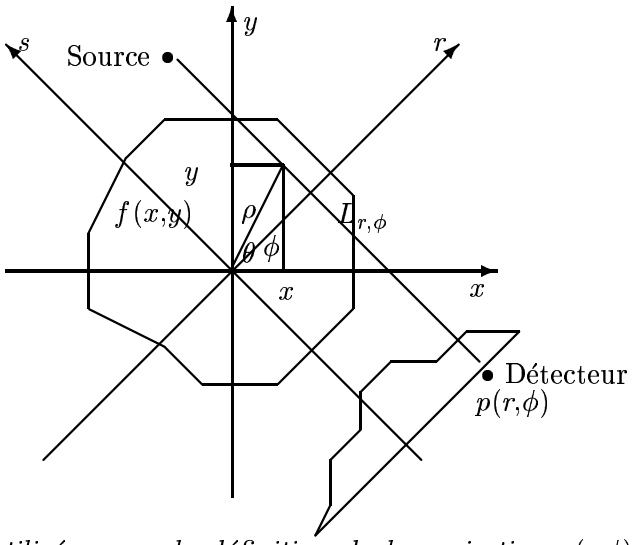


FIG. 2.2 – *Notations utilisées pour la définition de la projection $p(r,\phi)$ d'une fonction $f(x,y)$*

Pour des variables continues r et ϕ, $p(r,\phi)$ est, par définition, la transformée de Radon de la fonction $f(x,y)$. J. RADON a montré l'existence et l'unicité de cette transformée et de son inverse pour des fonctions continues à supports compacts. Ceci peut être résumé ainsi :

- Si $f(x,y)$ est continue et à support compact, alors $p(r,\phi) = \mathcal{R}\{f(x,y)\}$ est déterminée d'une façon unique par le résultat de l'intégration sur toutes les lignes droites $L_{r,\phi}$ dans le plan $(x,y) \in \mathbf{R}^2$, $r \in \mathbb{R}$, $\phi \in [0,\pi]$.
- La connaissance de $p(r,\phi)$ en tout point $r \in \mathbf{R}$ et $\phi \in [0,\pi]$ permet de déterminer d'une façon unique la fonction $f(x,y)$ par :

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

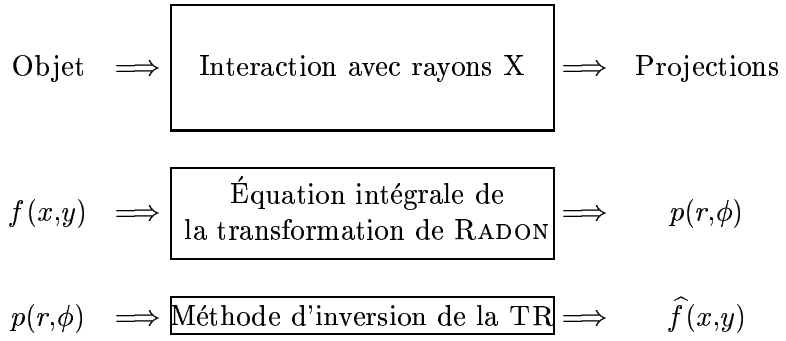


FIG. 2.3 – *Reconstruction d'image en tomographie à rayons X*

Voici ici un exemple de problème inverse pour lequel nous avons directement une expression analytique pour sa solution et on peut s'imaginer qu'il n'y a plus de problème.

En pratique cependant, les mesures sont discrètes. En particulier les projections sont obtenues pour les valeurs discrètes de ϕ. Il est alors plus convenable de dire que chaque projection $p(r,\phi_i)$ nous fournit un échantillonnage de la TR de la fonction $f(x,y)$. La multiplication des projections obtenues dans des conditions différentes nous permet de "remplir" progressivement le domaine de la transformée. Si ce remplissage pouvait être dense et uniforme alors l'inversion pourrait être effectuée par la transformation inverse pour récupérer l'objet. Mais la connaissance de tous les échantillons n'est malheureusement pas possible dans une situation réelle et on ne peut mesurer qu'un sous-ensemble de ces échantillons. Nous verrons que c'est là que se situe la difficulté de la résolution du problème.

Pour mieux comprendre les fondements des méthodes de reconstruction d'image il est utile de noter qu'il existe une relation fondamentale entre la TR et la TF d'une fonction. Elle se résume ainsi :
Si on note par $P(\Omega,\phi)$ la TF de $p(r,\phi)$ par rapport à la variable r :

$$P(\Omega,\phi) = \int p(r,\phi) \exp\left[-j\Omega r\right] \mathrm{d}r$$

et par $F(\omega_x,\omega_y)$ la TF 2D de $f(x,y)$:

$$F(\omega_x,\omega_y) = \iint f(x,y) \exp\left[-j\omega_x x,\omega_y y\right] \mathrm{d}x\, \mathrm{d}y$$

alors, il est facile de montrer la relation suivante plus connue sous le nom du *Théorème de projection* :

$$F(\omega_x,\omega_y) = P(\Omega,\phi) \quad \text{pour} \quad \omega_x = \Omega\cos\phi \quad \text{et} \quad \omega_y = \Omega\sin\phi$$

Cette relation permet de transformer le problème de la reconstruction d'image $f(x,y)$ à partir de ses projections $p(r,\phi)$ en un problème de *synthèse de* FOURIER, qui consiste à déterminer la fonction $f(x,y)$ à partir d'une connaissance partielle de sa transformée de FOURIER $F(\omega_x,\omega_y)$.

Encore une fois, on peut constater que si l'on pouvait connaître parfaitement toutes les projections on pourrait alors remplir le domaine de FOURIER de l'objet (connaître $F(\omega_x,\omega_y)$) et ensuite par une transformation de FOURIER inverse on pourrait déterminer la fonction $f(x,y)$. Mais, en pratique, ceci est rarement possible et le passage dans le domaine de FOURIER ne change rien.

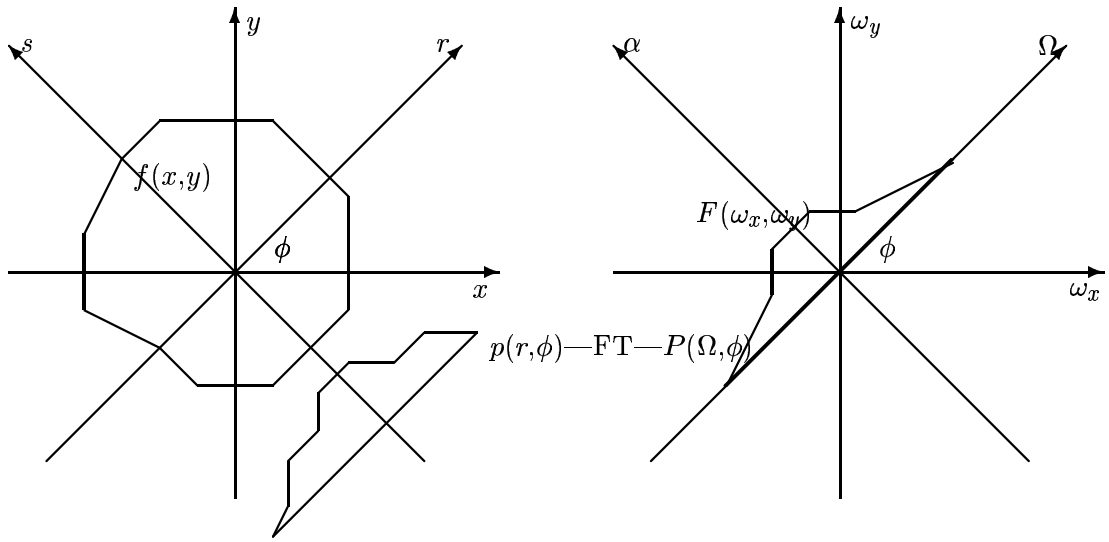


FIG. 2.4 – *Théorème de projection : Relation entre la* TF *1–D de la projection* $p(r,\phi)$ *et la* TF *2–D de* $f(x,y)$

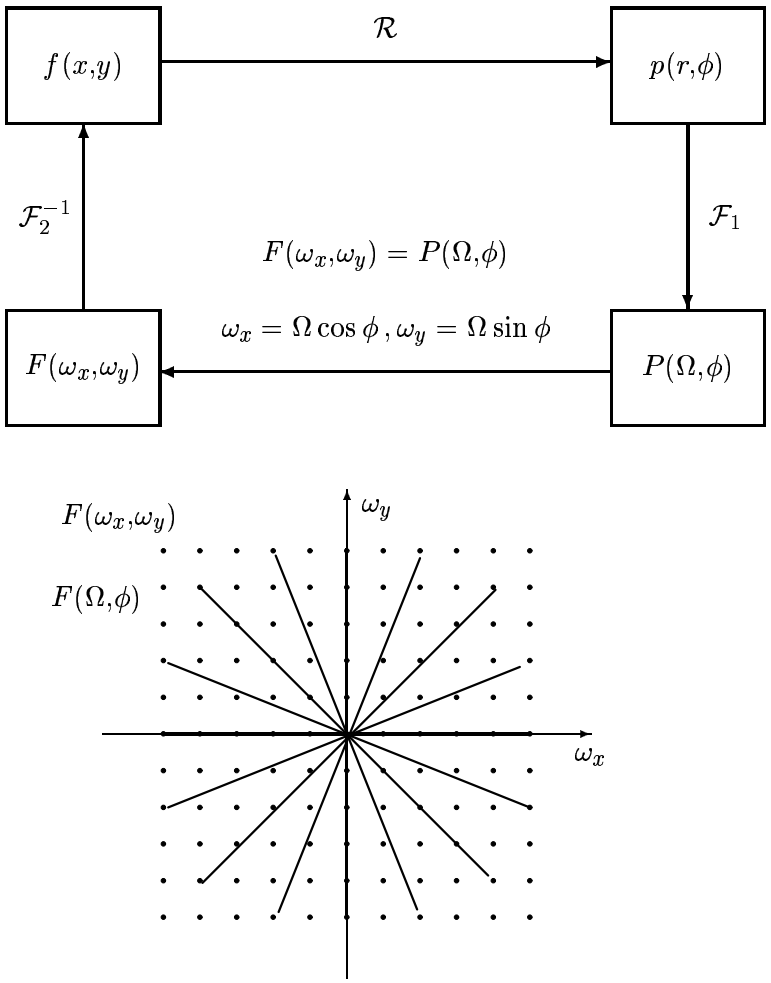


FIG. 2.5 – *Reconstruction d'image en tomographie X en passant dans le domaine de* FOURIER.

2.3 Imagerie à ondes diffractées : cadre général

D'une manière générale lorsqu'on cherche à *voir* l'intérieur d'un objet (sans le couper), on peut envisager deux procédés :

- soit l'illuminer par une onde (rayons X, ultrasons ou micro ondes) de l'extérieur et mesurer son interaction avec celle-ci; c'est l'imagerie active;
- soit profiter éventuellement de son propre rayonnement lorsque cela est possible (infra rouge, radioactivité) et mesurer les effets de ces rayons à l'extérieur de l'objet; c'est l'imagerie passive.

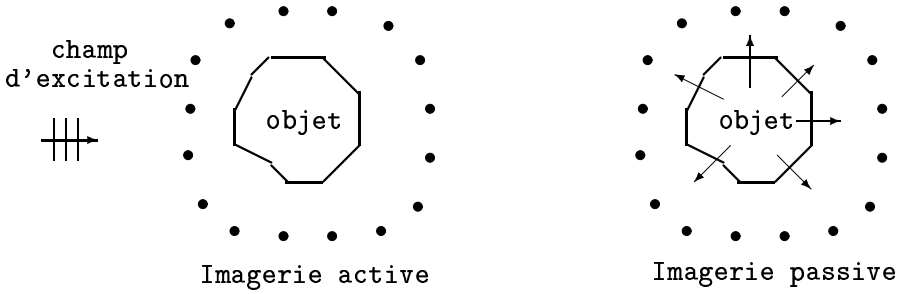


FIG. 2.6 – *Imagerie active et passive.*

Dans le cas d'imagerie active que nous considérons ici, on peut envisager trois cas :

- Imagerie par transmission où la ligne de mesure se trouve du côté opposé du champ incident ;
- Imagerie par réflexion où la ligne de mesure se trouve du même côté que celle du champ incident ;
- Imagerie mixte où les points de mesure se trouvent tout autour de l'objet.

Dans tous les cas, il faut d'abord établir une relation mathématique (modèle) entre ces mesures et une propriété interne de l'objet qui le caractérise, et ensuite inverser cette relation.

D'une manière plus générale, dans toute technique d'imagerie, il y a trois problèmes fondamentaux à résoudre :

- un problème direct, qui consiste à établir un modèle mathématique suffisamment simple décrivant le mieux possible la relation entre les grandeurs mesurées et les grandeurs inconnues auxquelles nous nous intéressons ;
- un problème d'instrumentation, qui consiste à construire un système d'imagerie : générer les ondes, concevoir les capteurs et mesurer les résultats des interactions de l'onde avec l'objet ;
- un problème inverse, qui consiste à développer des méthodes d'inversion qui, partant des données mesurées, reconstruisent ces images qui caractérisent l'objet étudié. Bien entendu, nous voulons que ces résultats soient, à la fois, robustes vis-à-vis de l'inévitable bruit de mesure et des erreurs d'approximation, de la discrétisation et de la quantification qui sont nécessaires pour le calcul numérique, et qu'ils aient en même temps la plus grande résolution spatiale possible.

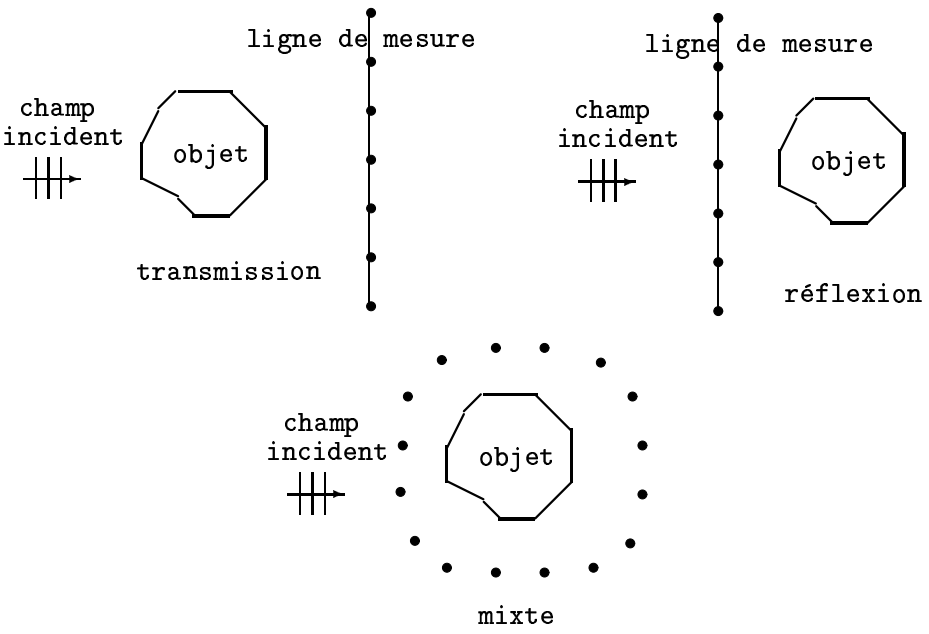


FIG. 2.7 – *Différents modes d'imagerie : par réflexion, par transmission et mode mixte.*

Dans la section précédente, nous avons vu comment les rayons X peuvent servir pour obtenir une carte de distribution de la densité de matière à l'intérieur d'un objet. Ici, nous allons voir qu'il est possible de remplacer les rayons X par d'autres types d'ondes, par exemple les ultrasons ou les micro ondes. Mais, en général, contrairement aux rayons X et due au fait que les longueurs d'ondes utilisées sont de l'ordre de dimensions de l'objet, l'établissement d'une relation entre les mesures (champs diffractés) et l'objet devient plus complexe.

Ici, sans entrer trop dans le détail, nous allons décrire le cadre général qui nous permettra d'identifier les problèmes inverses associés.

Lorsqu'un objet $f(\boldsymbol{r})$ est illuminé par une onde incidente $\phi_0(\boldsymbol{r})$ le champ total $\phi(\boldsymbol{r})$ peut être modélisé par la somme du champ incident $\phi_0(\boldsymbol{r})$ et du champ diffracté $\phi_s(\boldsymbol{r})$ qui dépend à la fois de l'objet et du champ incident :

$$\phi(\boldsymbol{r}) = \phi_0(\boldsymbol{r}) + \int_D G_o(\boldsymbol{r},\boldsymbol{r}')\phi(\boldsymbol{r}')\, f(\boldsymbol{r}')\, \mathrm{d}\boldsymbol{r}'$$

et le champ diffracté à l'extérieur de l'objet $g(\boldsymbol{u})$ que l'on mesure est une fonction de l'objet et du champ total à l'intérieur de l'objet:

$$g(\boldsymbol{r}') = \int_D G_m(\boldsymbol{r},\boldsymbol{r}')\, f(\boldsymbol{r})\, \phi(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}.$$

Dans ces deux équations G_o et G_m sont des fonctions de GREEN. Notons que la première équation est une équation implicite en ϕ et que la deuxième équation est bilinéaire en f et en ϕ.

Le problème de l'imagerie dans ce contexte est de fournir une estimation de $f(\boldsymbol{r})$ à partir d'un nombre fini de mesures $g(\boldsymbol{r}_i'), i = 1, \cdots, M$. Comme on peut constater, la

relation entre g et f n'est pas linéaire, ce qui nous donne un exemple type d'un problème inverse non linéaire.

Le cas de l'approximation de BORN étant la situation de faibles perturbations où l'on peut remplacer $\phi(\boldsymbol{r})$ par $\phi_0(\boldsymbol{r})$ dans cette dernière équation. Cette approximation permet d'établir une relation linéaire entre f et g et ainsi transformer le problème inverse non linéaire à un problème inverse linéaire :

$$g(\boldsymbol{r}') = \int_D G_m(\boldsymbol{r},\boldsymbol{r}')\, f(\boldsymbol{r})\, \phi_0(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}.$$

Par la suite nous allons décrire les techniques d'imagerie basées sur cette approximation.

2.4 Imagerie micro ondes : cas 2D

Pour comprendre les principales idées de l'imagerie à ondes diffractées, nous allons considérer le cas particulier de l'imagerie micro ondes. Nous nous placerons dans le cas d'un problème 2D. Pour cela considérer le cas où un objet cylindrique est illuminé par une onde plane polarisée rectiligne selon l'axe oz.

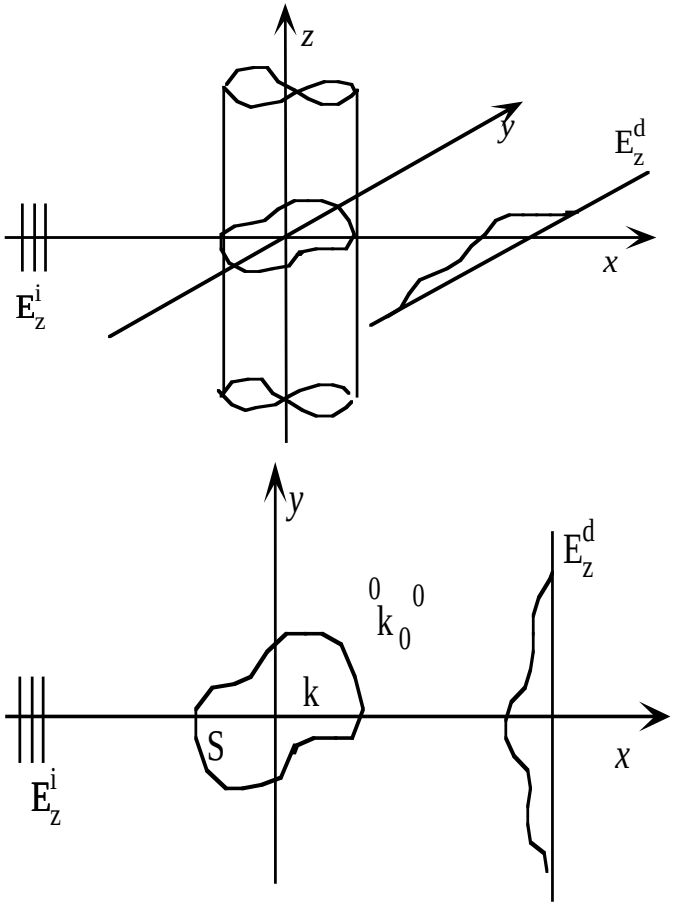


FIG. 2.8 – *Géométrie de l'imagerie micro ondes.*

Si on fait l'hypothèse que l'objet est homogène suivant l'axe oz ($\frac{\partial}{\partial z} = 0$) alors les champs incident et diffracté n'auront qu'une seule composante et on a les relations suivantes :

$$E_z(\boldsymbol{r}) = E_z^i(\boldsymbol{r}) + E_z^d(\boldsymbol{r})$$

$$E_z^d(\boldsymbol{r}) = \int G(\boldsymbol{r} - \boldsymbol{r}')J(\boldsymbol{r}')\,\mathrm{d}\boldsymbol{r}'$$

où

$$J(\boldsymbol{r}') = [k^2(\boldsymbol{r}') - k_0^2]E_z(\boldsymbol{r}') \quad \text{et} \quad k(\boldsymbol{r}') = \sqrt{\omega^2\mu_0\epsilon(\boldsymbol{r}') - j\omega\mu_0\sigma(\boldsymbol{r}')}$$

Deux cas peuvent alors être considérés :

- On s'intéresse à la quantité $J_n(\boldsymbol{r}') = J(\boldsymbol{r}')/E_z^i(\boldsymbol{r}')$ qui représente la densité des courants induits dans l'objet (normalisé par rapport au champ incident) et qui dépend à la fois de l'objet et du champ incident, mais qui a la propriété

$$f(\boldsymbol{r}') = [k^2(\boldsymbol{r}') - k_0^2] = 0 \longrightarrow J(\boldsymbol{r}') = 0.$$

On a alors

$$E_z^d(\boldsymbol{r}) = \int G(\boldsymbol{r} - \boldsymbol{r}') J_n(\boldsymbol{r}') E_z^i(\boldsymbol{r}') \, \mathrm{d}\boldsymbol{r}'$$

- On s'intéresse à la quantité $o(\boldsymbol{r}') = [k^2(\boldsymbol{r}') - k_0^2]$ qui représente directement l'objet et on a alors

$$E_z^d(\boldsymbol{r}) = \int G(\boldsymbol{r} - \boldsymbol{r}') o(\boldsymbol{r}') E_z(\boldsymbol{r}') \, \mathrm{d}\boldsymbol{r}'$$

Mais en se mettant dans le cadre de l'approximation de BORN, c'est-à-dire $E_z(\boldsymbol{r}') \simeq E_z^i(\boldsymbol{r}')$, alors on a

$$E_z^d(\boldsymbol{r}) = \int G(\boldsymbol{r} - \boldsymbol{r}') o(\boldsymbol{r}') E_z^i(\boldsymbol{r}') \, \mathrm{d}\boldsymbol{r}'$$

Dans les deux cas, si on note par $\phi(\boldsymbol{r}) = E_z^d(\boldsymbol{r})$, $\phi_0(\boldsymbol{r}') = E_z^i(\boldsymbol{r}')$ et par $f(\boldsymbol{r}')$ soit $J_n(\boldsymbol{r}')$ soit $o(\boldsymbol{r}')$, alors les deux problèmes se confondent à

$$\phi(\boldsymbol{r}) = \int G(\boldsymbol{r} - \boldsymbol{r}') f(\boldsymbol{r}') \phi_0(\boldsymbol{r}') \, \mathrm{d}\boldsymbol{r}'$$

Maintenant, remplaçant l'expression de la fonction de GREEN dans le cas 2D :

$$G(\boldsymbol{r} - \boldsymbol{r}') = \frac{j}{4} k_0^2 H_0(k_0 |\boldsymbol{r} - \boldsymbol{r}'|)$$

Regardant de plus près cette équation on constate qu'il s'agit d'une opération de convolution entre $G(\boldsymbol{r})$ et le produit $f(\boldsymbol{r}')\phi_0(\boldsymbol{r}')$. Passant dans le domaine de FOURIER on a

$$\mathrm{TF}\{\phi\} = \mathrm{TF}\{G\} \quad [\mathrm{TF}\{f\} * \mathrm{TF}\{\phi_0\}]$$

ou encore

$$\widehat{\phi}(\boldsymbol{\omega}) = \widehat{G}(\boldsymbol{\omega}) \left[\widehat{f}(\boldsymbol{\omega}) * \widehat{\phi}_0(\boldsymbol{\omega}) \right]$$

et où

$$\widehat{G}(\boldsymbol{\omega}) = \frac{k_0^2}{k_0^2 - \omega^2} \quad \text{avec} \quad \omega^2 = |\boldsymbol{\omega}|^2.$$

Supposons maintenant que le champ incident est une onde plane se propageant en direction de l'axe ox : $\phi_0 = \exp[j\boldsymbol{k}_0 \boldsymbol{r}]$ avec $\boldsymbol{k}_0 = [k_0,0,0]$, on a alors :

$$\widehat{\phi_0}(\boldsymbol{\omega}) = \delta(\boldsymbol{\omega} - \boldsymbol{k}_0)$$

et

$$\widehat{\phi}(\boldsymbol{\omega}) = \frac{k_0^2}{k_0^2 - \omega^2} \, \widehat{f}(\boldsymbol{\omega} - \boldsymbol{k}_0)$$

Ainsi, il y a une relation entre la TF spatiale de l'objet et la TF spatiale du champ diffracté. Mais pour être plus précis, mettons nous dans le cas 2D où l'on mesure le champ diffracté sur une ligne droite perpendiculaire à la direction de propagation du champ incident. On peut alors montrer la relation suivante :

$$\widehat{f}(\omega_x = -k_0 + \sqrt{k_0^2 - \omega_y^2}, \omega_y) = \widehat{\phi}(\omega_y)$$

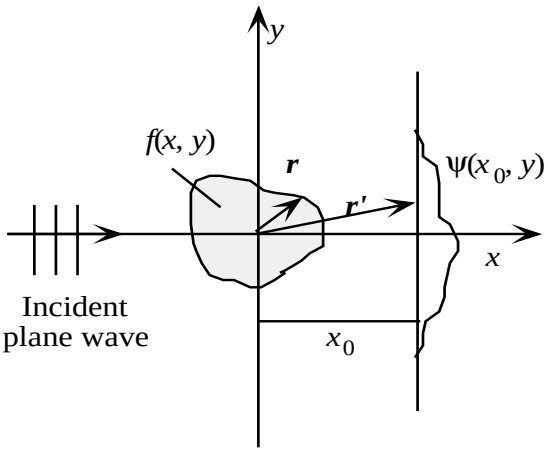


FIG. 2.9 – *Géométrie de la tomographie de diffraction 2D.*

En d'autres termes, les valeurs de la TF 1D du champ diffracté mesuré sur une ligne droite perpendiculaire à la direction de propagation d'une onde plane, sont égales aux valeurs de la TF 2D de l'objet sur un demi-cercle de rayon k_0 centré au point $\boldsymbol{k} = (k_0,0)$ (voir figure). videmment, lorsque la direction de propagation fait un angle θ avec l'axe ox, on a la même relation, sauf que le demi-cercle aussi sera tourné d'un angle θ.

La mise en évidence de cette relation permet de transformer le problème de la reconstruction d'objet $f(x,y)$ à partir de mesures des champs diffractés $\phi(r,\theta_i)$ en un problème de synthèse de FOURIER qui consiste en la détermination de $f(x,y)$ à partir d'une connaissance partielle de sa TF. De plus, on peut faire l'analogie avec le problème de la reconstruction d'image en tomographie X. En effet en tomographie X, les projections nous fournissaient les valeurs de la TF de l'objet sur des lignes droites, alors qu'ici, les champs diffractés nous fournissent les valeurs de la TF de l'objet sur des demi-cercles dont le rayon est inversement proportionnel à la longueur d'onde utilisée.

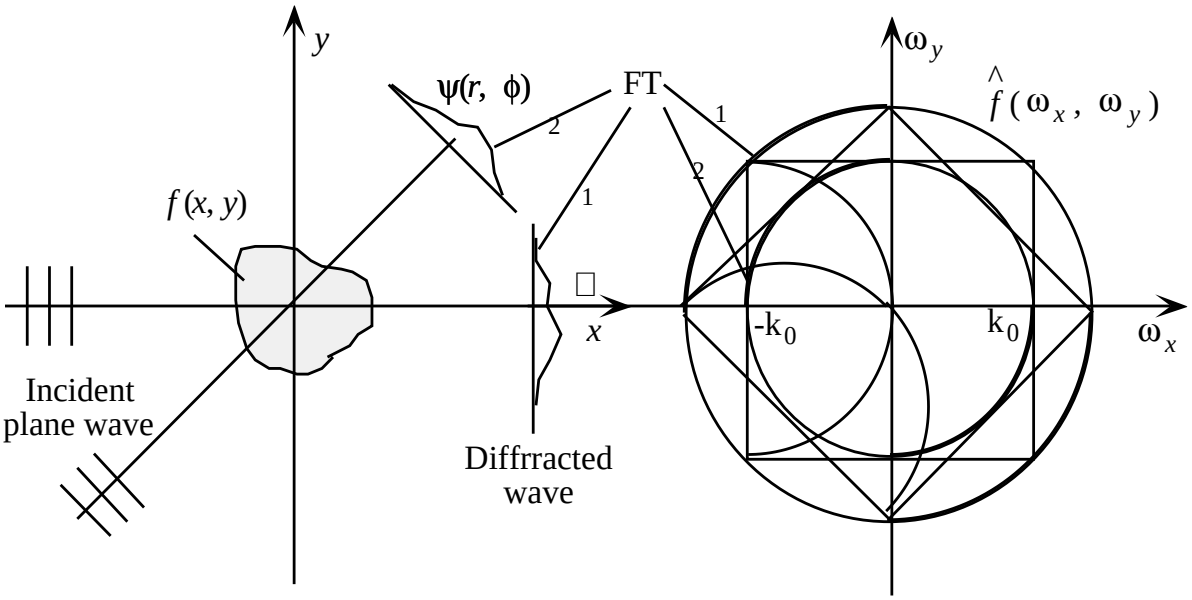


FIG. 2.10 – *Synthèse de* FOURIER *en tomographie de diffraction 2D.*

2.5 Imagerie micro ondes : cas 3D

Les mêmes raisonnements et les mêmes équations peuvent être obtenues dans le cas de l'imagerie 3D, si l'onde incidente est plane et si le plan sur lequel les mesures du champ diffracté sont effectuées, on aura la même relation dans le cas 3D.

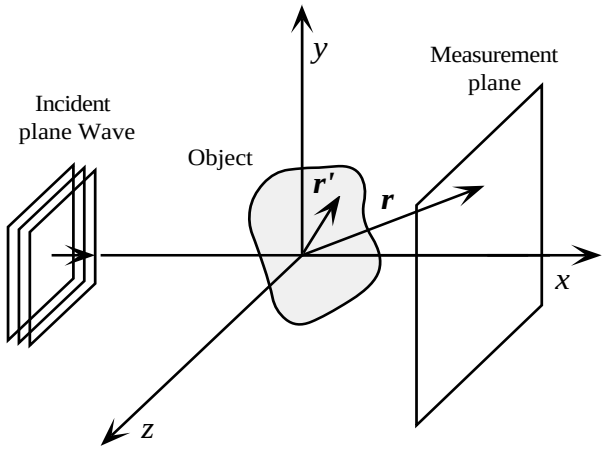


FIG. 2.11 – *Géométrie de la tomographie de diffraction 3D.*

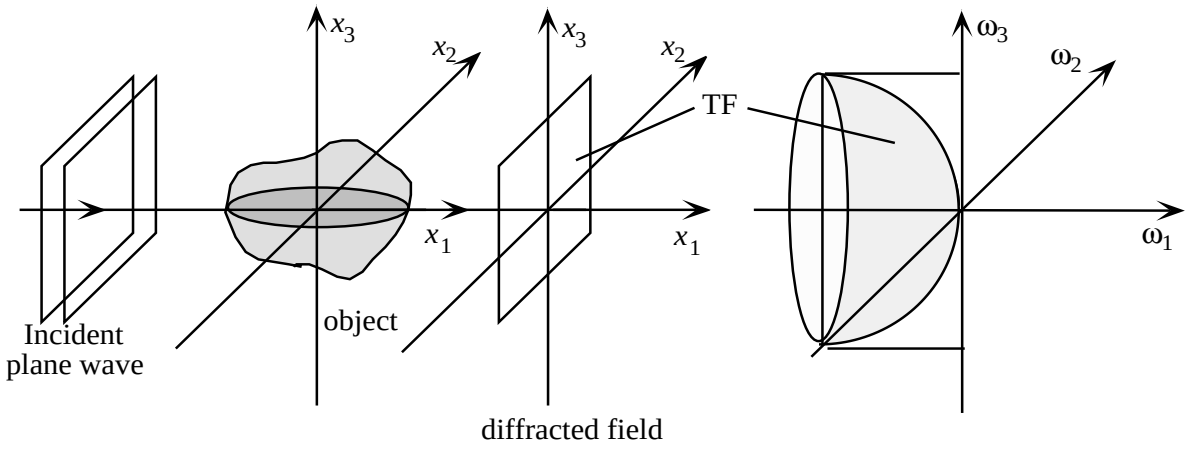


FIG. 2.12 – *Synthèse de* FOURIER *en tomographie de diffraction 3D.*

2.6 Imagerie par courants de Foucault

Les techniques de contrôle non destructif utilisent couramment les courants de FOUCAULT pour détecter une anomalie dans un milieu conducteur. L'imagerie par courants de FOUCAULT est plus récente. Son objectif est, une fois une anomalie détectée, de fournir une image qui représente la variation de la conductivité à l'intérieur d'un objet conducteur. Cette image peut alors être utilisée pour fournir un diagnostic plus précis sur le défaut.

Bien que les fréquences utilisées pour exciter l'objet par cette technique sont de l'ordre du kHz à MHz, on peut encore utiliser le cadre général développé en imagerie micro ondes pour obtenir des équations liant l'objet aux mesures. Trois différences importantes cependant existent :

- En tomographie de diffraction, en général, l'objet étudié est dans un milieu homogène où se trouvent les capteurs (émetteurs et récepteurs). En imagerie par courants de FOUCAULT le défaut se trouve à l'intérieur d'un milieu homogène (bloc métallique), qui est différent du milieu où se trouvent les capteurs (air).
- En tomographie de diffraction, en général, on peut négliger l'atténuation des ondes, alors qu' en imagerie par courants de FOUCAULT cela n'est pas le cas. En effet, les ondes se propagent dans un milieu conducteur où les deux phénomènes de diffusion et de propagation sont d'importance égale (la constante de propagation est un nombre complexe dont la partie réelle et la partie imaginaire sont de même ordre de grandeur).
- En tomographie de diffraction, en général, on peut tourner autour de l'objet, ce qui permet de remplir le domaine de FOURIER de l'objet. En imagerie par courants de FOUCAULT, en général, il n'est pas possible de tourner autour de l'objet. C'est pourquoi, ici, on modifie la fréquence temporelle du champ d'excitation pour obtenir des données complémentaires, ce qui permet dans un sens de remplir partiellement le domaine de FOURIER de l'objet.

Pour donner une description simple de cette technique considérons le cas 2D de la figure (2.6).

Supposons que le milieu conducteur est suffisamment épais et que le défaut se trouve près de la surface pour ne considérer que deux milieux D_1 (air) et D_2 (métal) avec une interface que nous supposerons plane. Supposons aussi que la région D_2 est amagnétique. Partant des équations de Helmoltz, on peut obtenir une équation intégrale liant la composante normale du champ électrique E_z ou les composantes tangentielles du champ magnétique (H_x et H_y) aux sources de courants induites par le défaut. Pour cela, nous noterons :

- $\vec{E}_0$, $\vec{H}_0$, les champs électrique et magnétique en l'absence du défaut ;
- $\vec{E}_f$, $\vec{H}_f$, les champs en présence du défaut ;
- $\vec{E}_s = \vec{E}_0 - E_f$, $\vec{H}_s = \vec{H}_0 - \vec{H}_f$, les différences des champs en l'absence et en présence du défaut ;
- $k_1^2 = \omega^2\mu_0\epsilon_0$, la constante de propagation dans la région R_1 ;
- $k_2^2 = \omega^2\mu_0\epsilon_0 + j\omega\mu_0\sigma_0 \simeq j\omega\mu_0\sigma_0$, la constante de propagation dans la région R_2 ;
- $k_\Omega^2(x,y) = \omega^2\mu_0\epsilon_0 + j\omega\mu_0\sigma(x,y)$, la constante de propagation dans la région du défaut ;
- $f(x,y) = \frac{k_2^2 - k^2(x,y)}{k_2^2}$, la fonction du contraste qui caractérise le défaut ;

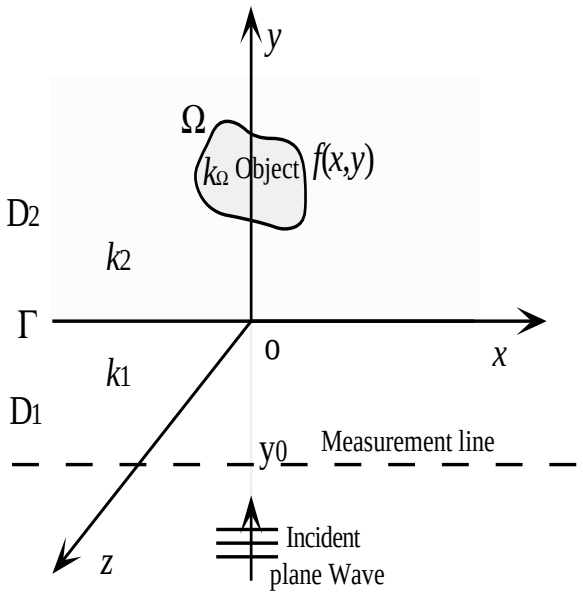


FIG. 2.13 – *Géométrie de l'imagerie par courants de* FOUCAULT.

- $J_\Omega(x,y) = k_2^2 f(x,y) E(x,y)$, les sources de courants équivalentes induites dans le défaut ;
- $J_{\Omega_n}(x,y) = \frac{J_\Omega(x,y)}{E_0(x,y)}$, une quantité normalisée par rapport au champ incident E_0.

On peut alors montrer que le champ d'anomalie mesuré sur des capteurs placés sur une ligne droite $y = y_0$ parallèle à la surface de l'interface Γ s'écrit :

$$E_s(x,y_0) = \iint J_\Omega(x',y')\, G(x - x', y_0 - y')\, \mathrm{d}x'\, \mathrm{d}y'$$

Supposant aussi que le champ incident est une onde plane se propageant dans la direction perpendiculaire de l'interface Γ, on peut écrire :

$$E_0(x,y) = E_0(y) = \begin{cases} \exp[+jk_1 y] + R\exp[-jk_1 y] & y > 0 \\ T\exp[+jk_2 y] & y < 0 \end{cases}$$

où R et T sont des coefficients de réflexion et de transmission.

Par la suite, pour être bref, remplaçant l'expression de $E_0(x,y)$ et l'expression de la fonction de GREEN dans l'équation du champ d'anomalie, on peut montrer la relation suivante :

$$\widehat{E}_s(u,y_0) = \frac{2T}{\delta^2} \frac{\exp[-j\beta_1 y_0]}{\beta_1 + \beta_2} \iint_\Omega f(x',y') \exp\left[-j[ux' - (k_2 + \beta_2)y']\right] \mathrm{d}x'\, \mathrm{d}y',$$

ou d'une manière équivalente

$$\frac{\delta^2\,(\beta_1 + \beta_2)}{2T\,\exp[-j\beta_1 y_0]}\, \widehat{E}_s(u,y_0) = \iint_\Omega f(x',y') \exp\left[-j[ux' - (k_2 + \beta_2)y']\right] \mathrm{d}x'\, \mathrm{d}y'$$

où :

$$\widehat{E}(u,y_0) = \mathrm{TF1D}\left\{E_s(x,y_0)\right\} = \int E_s(x,y_0) \exp[-jux]\, \mathrm{d}x$$

et

$$\begin{array}{ll} T = \frac{2k_1}{k_1+k_2} \simeq \frac{2k_1}{k_2}, & \delta = \sqrt{\frac{2}{\omega\mu\sigma}} \\ k_1^2 = \omega^2\mu\epsilon, & k_2^2 = \omega^2\mu\epsilon + j\omega\mu\sigma \simeq j\omega\mu\sigma, \quad k_2 = \frac{1}{\delta}(1+j) \\ \beta_1 = \sqrt{k_1^2 - u^2}, & \beta_2 = \sqrt{k_2^2 - u^2} \\ f(x,y) = \frac{k_2^2 - k^2(x,y)}{k_2^2} \simeq -\frac{\sigma(x,y)-\sigma_0}{\sigma_0} & \end{array}$$

Pour mieux interpréter ces équations, notons par

$$s = k_2 + \beta_2 = k_2 + \sqrt{k_2^2 - u^2}$$

et par

$$\widehat{f}(u,s) = \mathrm{TFL}\{f(x,y)\} = \iint_D f(x,y) \exp\left[-jux - sy\right] \mathrm{d}x\,\mathrm{d}y.$$

la transformée de FOURIER-LAPLACE (TFL) de $f(x,y)$. On a alors :

$$\widehat{E}_s(u,y_0) = \frac{2T}{\delta^2} \frac{\exp\left[-j\beta_1 y_0\right]}{\beta_1 + \beta_2} \widehat{f}(u,s)$$

Ainsi, les différentes mesures du champs $E_s(x,y_0)$ à des fréquences temporelles ω_i différentes, nous fournit des valeurs de la TFL de l'objet $\widehat{f}(u,s)$ pour les différentes valeurs de u et de s. Le problème devient ainsi celui de la *synthèse de* FOURIER-LAPLACE qui consiste en la détermination de la fonction $f(x,y)$ à partir d'une connaissance partielle des valeurs de sa TFL.

Le travail avec la TFL n'étant pas facile, on peut envisager des approximations :

1. Négliger entièrement l'atténuation :
   Il s'agit de négliger la partie imaginaire de k_2 et la partie réelle de s, c'est à dire :

$$\begin{aligned} k_2 &= \frac{1}{\delta}(1+j) \simeq \frac{1}{\delta} \\ s &= k_2 + \beta_2 = k_2 + \sqrt{k_2^2 - u^2} \simeq \frac{1}{\delta} + \sqrt{\frac{1}{\delta^2}} \simeq \frac{1}{\delta}\left(1 + \sqrt{1 - (\delta u)^2}\right) \end{aligned}$$

   ce qui nous permet d'écrire :

$$\widehat{E}_s(u,y_0) = \frac{2T}{\delta^2} \frac{\exp\left[-j\beta_1 y_0\right]}{\beta_1 + \beta_2}\ \widehat{f}\left(u, v = -\frac{1}{\delta}\left(1 + \sqrt{1 - (\delta u)^2}\right)\right)$$

   où $\widehat{f}(u,v)$ est la TF2D de $f(x,y)$.
2. Négliger partiellement l'atténuation :
   Il s'agit de négliger la partie imaginaire de $\beta_2 = \sqrt{k_2^2 - u^2}$, c'est à dire :

$$\begin{aligned} s \simeq k_2 + \mathcal{I}\mathrm{m}\,(\beta_2) &= k_2 + \mathcal{I}\mathrm{m}\left(\sqrt{k_2^2 - u^2}\right) \\ &= \frac{1}{\delta}(1+j) + \mathcal{I}\mathrm{m}\left(\sqrt{j\frac{2}{\delta^2} - u^2}\right) \\ &= \frac{1}{\delta}\left[1 + \frac{1}{\sqrt{2}}\sqrt{\sqrt{4 + (\delta u)^4} - (\delta u)^2}\right] + j\frac{1}{\delta} \end{aligned}$$

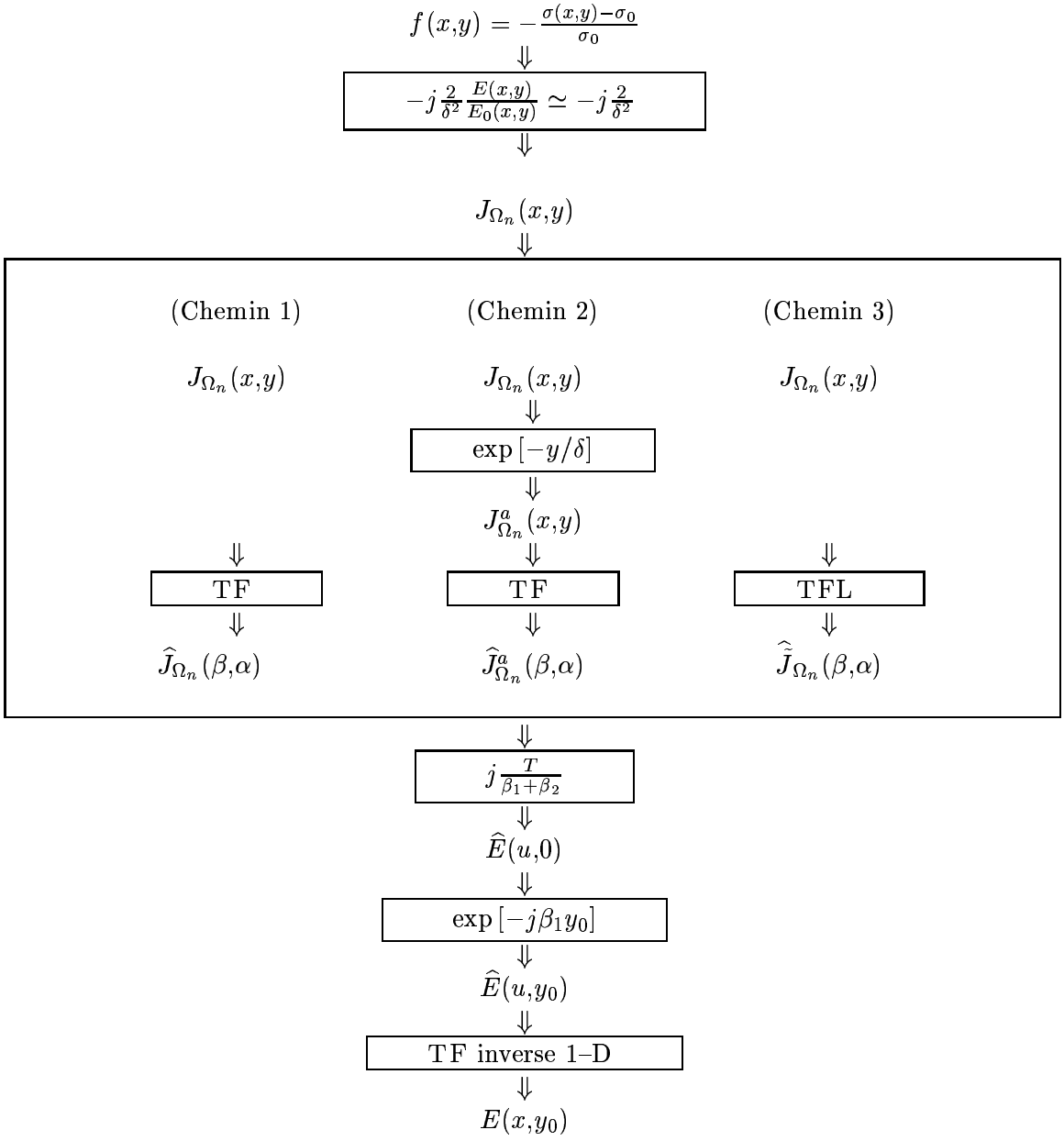
ce qui nous permet d'écrire :

$$\widehat{E}_s(u,y_0) = \frac{2T}{\delta^2}\frac{\exp\left[-j\beta_1 y_0\right]}{\beta_1+\beta_2}\ \widehat{f}\left(u,v=-\frac{1}{\delta}\left(1+\frac{1}{\sqrt{2}}\sqrt{\sqrt{4+(\delta u)^4}-(\delta u)^2}\right)\right)$$

où $\widehat{f}(u,v)$ est la TF2D de $f(x,y)$.

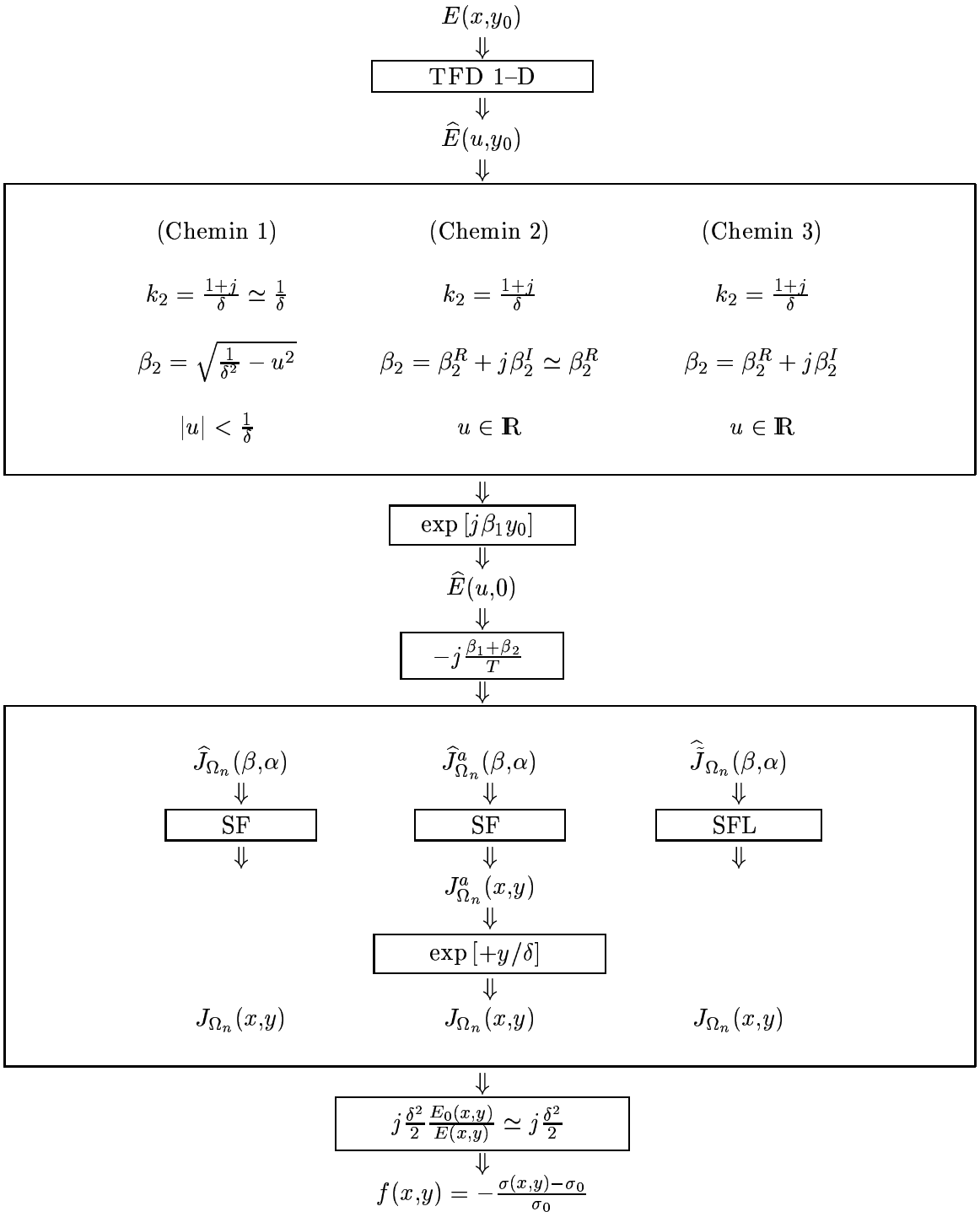
On peut résumer ces différentes situations dans le schéma bloc suivant:

Le schéma bloc de différentes approximations du problème direct en imagerie par courants de FOUCAULT.



Lors de l'inversion on doit effectuer le trajet inverse. Ceci est résumé dans le schéma bloc suivant :

Le schéma bloc des différentes étapes de l'inversion
en imagerie par courant de FOUCAULT.



Ainsi les étapes qui nous permettent de simuler le problème direct de l'imagerie par courants de FOUCAULT sont :

- Calculer la source de courant normalisée

$$J_{\Omega_n}(x,y) = -j\frac{2}{\delta^2}\frac{E(x,y)}{E_0(x,y)}f(x,y)$$

- Suivant les trois chemins :
  - calculer la TF de $J_{\Omega_n}(x,y)$ sur des contours algébriques définis par :

$$\begin{cases} u &= \alpha \\ v &= -\frac{1}{\delta}\left[1+\sqrt{1-(\delta\alpha)^2}\right] \end{cases},$$

  - calculer la TF de $J_{\Omega_n}(x,y)\exp\left[-\frac{y}{\delta}\right]$ sur des contours algébriques définis par :

$$\begin{cases} u &= \alpha \\ v &= -\frac{1}{\delta}\left[1+\frac{1}{\sqrt{2}}\sqrt{\sqrt{(\delta u)^4+4}-(\delta u)^2}\right] \end{cases},$$

  - calculer la TFL de $J_{\Omega_n}(x,y)$ sur sur des contours algébriques définis par :

$$\begin{cases} u &= \alpha \\ v &= -j\left[k_2+\sqrt{k_2^2-\alpha^2}\right] \end{cases}$$

- Calculer $\widehat{E}(u,0)$ à partir de l'une des quantités calculées dans l'étape précédente :
- Calculer $E(x,0)$ et finalement $E(x,y_0)$.

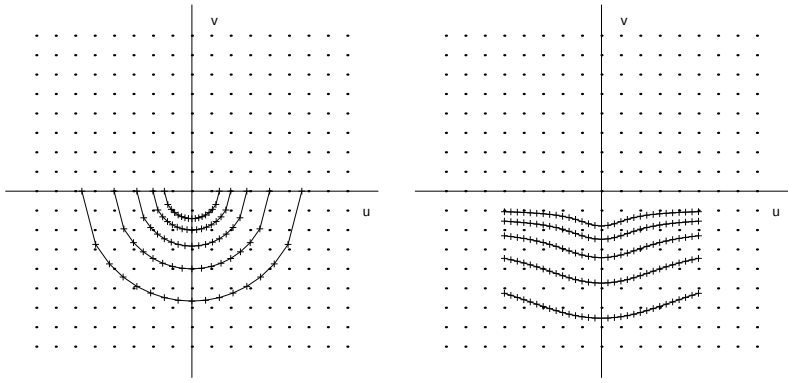


FIG. 2.14 – *Synthèse de* FOURIER *en imagerie par courants de* FOUCAULT.

2.7 Imagerie à émission de positons

En médecine nucléaire l'objectif de la tomographie d'émission est de déterminer la distribution d'un isotope radioactif d'une substance existante ou injectée dans le corps en mesurant le nombre de photons émis par ces substances.

Il existe deux types de tomographie à émission dépendant du type d'isotope utilisé:

- Tomographie à émission de positons (TEP): le radioélément est un émetteur de positons, qui après un parcours moyen très restreint va interagir avec la matière, de manière à émettre deux rayonnements gamma à 180 degrés l'un de l'autre. Le système de détection devra mesurer le nombre de photons qui coïncident, c'est à dire estimer la présence d'un positon par la détection simultanée des deux rayonnement gamma sur deux détecteurs opposés.
- Tomographie à émission d'un seul photon (SPECT): le radioélément est un émetteur gamma ordinaire, dit mono photonique, qui émet son rayonnement dans tous les sens. À la détection, l'ensemble des collimateurs mesurent dans une direction donnée le rayonnement émis. Ce type d'imagerie est appelé SPECT, de l'anglais (Single Photon Emitter Computed Tomographie).

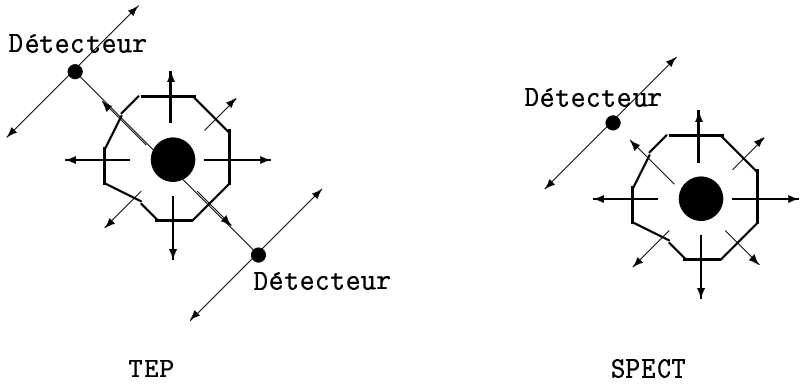


FIG. 2.15 – *Principe de la tomographie d'émission :* TEP *et* SPECT.

Dans la tomographie à émission de positons, le positon éjecté perd une très grande partie de son énergie dans une distance de quelques millimètres et forme un rayon gamma en rencontrant un électron se trouvant sur son chemin. Cette rencontre se produisant presque au repos, chaque rayon gamma a une énergie de 0.511 MEV et les photons émis repartent dans des directions opposées sur une ligne droite. Si un ensemble de détecteurs en anneau est placé autour de l'objet, deux détecteurs se trouvant face à face sur cette ligne, détecteront simultanément une énergie de 0.511 MEV. Ainsi, la direction des émissions est connue et les nombres de photons par unité de temps mesurés par chaque détecteur est une mesure de la concentration totale des substances radioactives le long de la ligne reliant les deux détecteurs.

Ainsi un modèle simple reliant la distribution de la densité de la matière radioactive $f(x,y)$ aux mesures est :

$$p(r,\phi) = \int f(x,y)\,\mathrm{d}l$$

Bien entendu, dans des situations réelles, on ne peut pas négliger l'atténuation et un

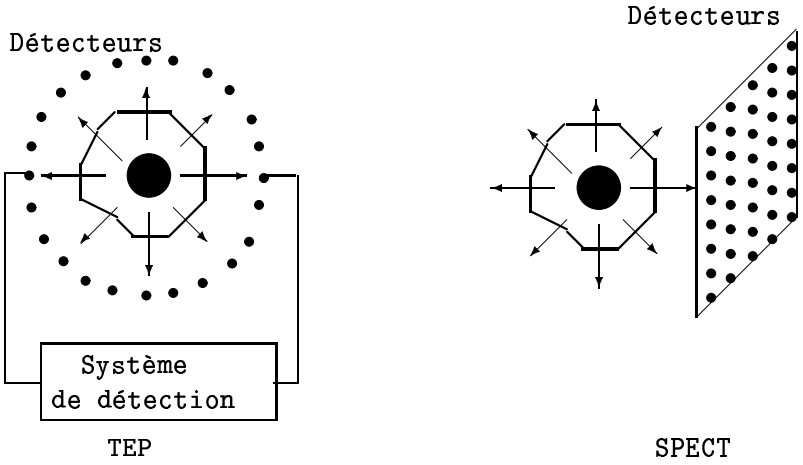


FIG. 2.16 – *Système de mesures en tomographie d'émission :* TEP *et* SPECT.

modèle plus approprié est :

$$p(r,\phi) = \int f(x,y)a(x,y)\,\mathrm{d}l$$

où $a(x,y)$ est une fonction dépendant de la densité de la matière $f(x,y)$. Le problème est donc non linéaire. Une manière de traiter ce problème est, de supposer dans un premier temps $a(x,y)$ constant et estimer alors $f(x,y)$, puis déterminer $a(x,y)$ par un filtrage passe–bas de $f(x,y)$ et recommencer (linéarisations successives).

Une autre différence entre la tomographie à rayons X et la TEP est que dans la première les détecteurs utilisés sont des photomultiplicateurs qui, sous l'action du flux de photons importants que représente le rayonnement X de la source, fournissent sur un temps court une réponse intégrée, c'est à dire, un courant. Dans ces conditions, le rôle des fluctuations statistiques est faible et le rapport signal à bruit est élevé. Alors qu'en tomographie d'émission les détecteurs sont des compteurs à scintillation qui fournissent des impulsions individualisées liées à des flux de photons faibles, dont le taux est soumis à des fluctuations statistiques importantes.

2.8 Imagerie à résonance magnétique nucléaire RMN

L'énergie magnétique du moment nucléaire $\vec{\mu}$ dans un champ magnétique extérieur $\vec{H}$ est de la forme :

$$E = -\vec{\mu}.\vec{H}$$

A l'équilibre thermique, l'aimantation macroscopique $\vec{\mu}$ d'un ensemble de N spins nucléaires identiques est parallèle au champ extérieur. Si $\vec{\mu}$est écarté de son orientation d'équilibre (en le soumettant à une impulsion de radiofréquence) d'un angle θ, son évolution ultérieure consistera en une précession autour du $\vec{H}$de fréquence $\omega_0 = \gamma|\vec{H}|$ où γ est le rapport gyromagnétique et ω_0 est la fréquence de Larmor. Du fait de la relaxation transversale l'amplitude du signal de précession libre décroît exponentiellement au cours du temps.

Le principe de base de l'imagerie est d'observer le signal de précession libre en présence d'un gradient statique du champ magnétique, superposé au champ statique homogène principal. Comme la fréquence de Larmor en chaque point est proportionnelle au champ en ce point, l'utilisation d'un gradient réalise un codage de la position par la fréquence.

Considérons un échantillon macroscopique entouré d'une bobine de résonance. Après une impulsion de $\theta = \pi/2$ modifiant la direction du champ magnétique d'un angle de $\pi/2$, le signal de précession libre que l'on observe provient de tous les spins de cette espèce dans l'échantillon. L'objectif de l'imagerie RMN est d'obtenir, séparément, la contribution à ce signal des divers points de l'échantillon.

A la suite d'une impulsion de $\theta = \pi/2$ à la fréquence ω, les aimantations transversales de chaque point de l'échantillon sont toutes parallèles à une direction ox du référentiel tournant à la fréquence ω. Appliquant, ensuite, un gradient $\boldsymbol{G}(t)$ dans le référentiel tournant à la fréquence ω, on mesure un signal de précession libre $s(t)$ qui nous fournit des renseignements sur la distribution spatiale de la densité de spin magnétique dans l'objet.

Le cas idéal le plus simple est celui d'un objet linéaire (la distribution de la densité de spin dans l'objet variant linéairement le long de l'axe) : le gradient étant connu, il suffit de calculer la transformée de FOURIER du signal de précession libre, dont l'amplitude à chaque fréquence donne la contribution de chaque point au signal.

Dans le cas plus général d'un objet à deux ou trois dimensions, on peut montrer qu'il y a un lien entre le signal temporel de précession libre $s(t)$ et la distribution de la densité de spin magnétique dans l'objet $f(\boldsymbol{r})$ qui est donnée par la relation suivante :

$$s(t) = \iiint f(\boldsymbol{r}) \exp\left[j(\int \boldsymbol{G}(t').\boldsymbol{r}\, \mathrm{d}t')\right] \exp\left[\frac{-t}{T}\right] \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z$$

Le terme $\exp\left[\frac{-t}{T}\right]$ est due à l'effet de la relaxation transversale.

Cette relation peut être résumée par le schéma suivant :

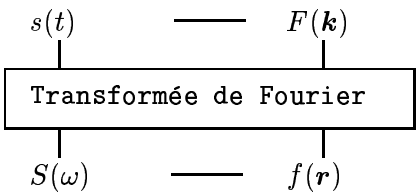


FIG. 2.17 – *Principe de l'imagerie RMN.*

où $S(\omega)$ est la transformée de FOURIER temporelle de $s(t)$ (une distribution de fréquences temporelles) et $F(\boldsymbol{k})$ est la transformée de FOURIER spatiale de $f(\boldsymbol{r})$.

En coordonnées cartésiennes on a :

$$s(t) = \iiint f(x,y,z) \exp\left[j(k_x(t)x + k_y(t)y + k_z(t)z)\right] \exp\left[\frac{-t}{T}\right] \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$$

où

$$k_x(t) = \gamma \int_0^t G_x(t')\,\mathrm{d}t', \quad k_y(t) = \gamma \int_0^t G_y(t')\,\mathrm{d}t', \quad k_z(t) = \gamma \int_0^t G_z(t')\,\mathrm{d}t'$$

et où $G_x(t)$, $G_y(t)$ et $G_z(t)$ sont les composantes du gradient du champ magnétique au point (x,y,z).

Ainsi, à un instant donné $s(t)$ est proportionnelle à la valeur de la TF spatiale 3D de $f(x,y,z)$ au point correspondant $\boldsymbol{k} = (k_x,k_y,k_z)$. Ainsi, en changeant les trois composantes $G_x(t)$, $G_y(t)$ et $G_z(t)$ du gradient du champ magnétique on peut remplir le domaine de FOURIER de l'objet.

Simplifions ces relations pour le cas d'un problème en 2D et négligeons pour l'instant le terme exponentiel. On a ainsi

$$s(t) = \iint f(x,y) \exp\left[j(k_x(t)x + k_y(t)y)\right] \mathrm{d}x\,\mathrm{d}y$$

$$F(u,v) = \iint f(x,y) \exp\left[j(ux + vy)\right] \mathrm{d}x\,\mathrm{d}y$$

$$s(t) = F(u,v), \forall u = k_x(t), \quad v = k_y(t)$$

L'organisation des séquences d'observation doit donc être agencée de façon à fournir une carte complète de l'espace réciproque de l'objet étudié. Deux méthodes sont actuellement d'usage.

Dans la première, on applique un gradient de champ constant et on répète l'expérience en changeant chaque fois son orientation. Le signal complexe de précession libre $s(t)$ à chaque instant d'échantillonnage t nous fournit donc deux points de l'espace réciproques de l'objet dans la direction du gradient. La carte de l'espace de FOURIER de l'objet ainsi obtenue est de la même forme que dans la tomographie à rayons X, mais ici la grandeur mesurée nous fournit directement ces renseignements.

On constate que ces points divergent le long des rayons. L'obtention d'une bonne densité des points aux grandes valeurs de $|\boldsymbol{k}|$ s'accompagnent d'une densité bien plus grande aux faibles valeurs de $|\boldsymbol{k}|$, autrement dit, l'obtention d'une bonne résolution à courte distance s'accompagne d'une sur-information à grande distance et vis versa.

Cette méthode est actuellement supplantée par une méthode dite de codage de phase dont le principe est le suivant : Après l'impulsion de $\pi/2$, on applique pendant un temps t un gradient G_y, sans effectuer d'observation; on annule, ensuite, le gradient G_y et on applique un gradient G_x. On observe alors le signal en fonction du temps t, durant un intervalle du temps T, suivant cette commutation du gradient. Le signal observé est:

$$s(t) = \iint f(x,y) \exp\left[jG_x x + G_y y)\right] \mathrm{d}x\,\mathrm{d}y = F(-G_x t, -G_y t).$$

En modifiant la valeur du gradient G_y on remplit l'espace de FOURIER directement sur des coordonnées cartésiennes. Cependant à cause de la limitation de la valeur du gradient G_y, et l'amplitude décroissante de $s(t)$, due à l'effet de la relaxation, le remplissage reste incomplet.

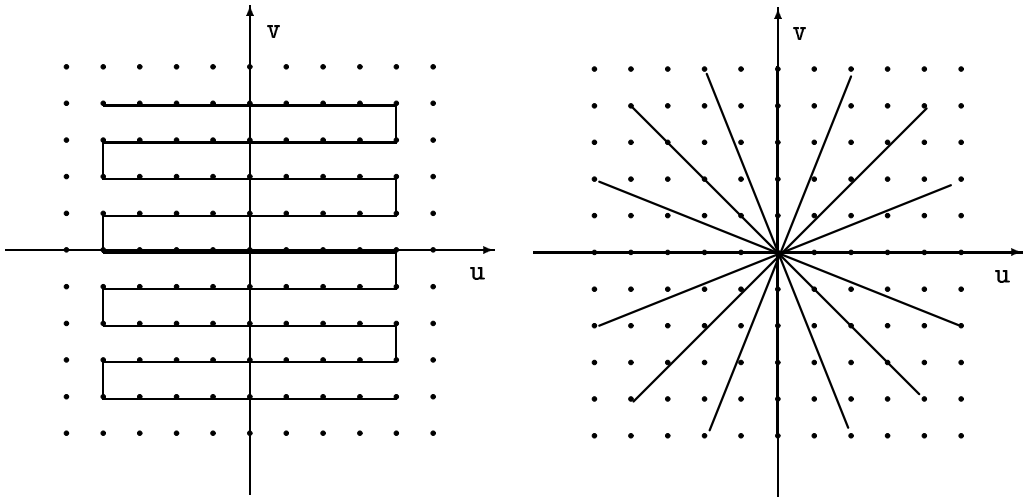


FIG. 2.18 – *Modes de remplissage du domaine de* FOURIER *en imagerie RMN.*

2.9 Imagerie radar à ouverture synthétique : SAR

L'imagerie radar à focalisation synthétique est basée sur le principe de l'effet DOPPLER. On peut montrer que le signal DOPPLER après détection synchrone et filtrage nous renseigne sur l'intersection avec une ligne droite de la TF spatiale en 2-D de l'objet illuminé par des ondes radar. Par exemple si l'on se place dans la géométrie de la figure (2.9), en supposant que h est petit, et si on modélise la réflectivité de l'objet par une fonction complexe $f(x,y) = |f(x,y)| \exp[j\phi(x,y)]$ et si le signal émis $s(t)$ est une impulsion modulée linéairement en fréquence de la forme

$$s(t) = \begin{cases} \exp[j(\omega_0 t + at^2] & |t| \leq T/2, \\ 0 & \text{ailleurs} \end{cases}$$

alors le signal reçu est de la forme

$$r_\theta(t) = A \int_{-L}^{L} p_\theta(x') s\left(t - \frac{2(R+x')}{c}\right) \mathrm{d}x'$$

où

$$p_\theta(x') = \int f(x' \cos\theta - y' \sin\theta, x' \sin\theta + y' \cos\theta)\, \mathrm{d}y'.$$

En remplaçant cette dernière équation dans l'équation précédente et en notant $\tau_0 = \frac{2R}{c}$, le signal reçu, après la détection synchrone et filtrage passe-bas, est de la forme :

$$c_\theta(t) = \frac{A}{2} \int_{-L}^{L} p_\theta(x') \exp\left[-j\frac{2}{c}(\omega_0 + 2a(t-\tau_0))x'\right] \mathrm{d}x'$$

Cette équation nous montre que le signal mesuré $c_\theta(t)$ est relié à la TF de la projection $p_\theta(x')$ par

$$c_\theta(t) = \frac{A}{2} P_\theta\left(\frac{2}{c}(\omega_0 + 2a(t-\tau_0)\right).$$

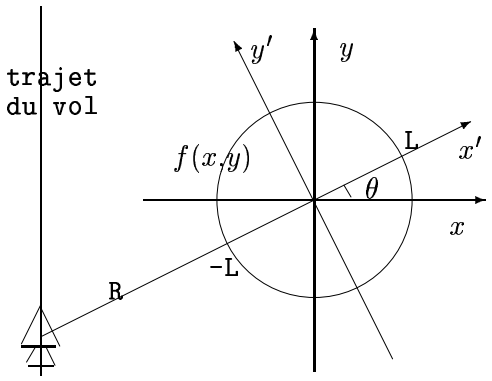


FIG. 2.19 – *La géométrie de l'imagerie radar.*

Ce signal est disponible pour $-\frac{T}{2} + \frac{2(R-L)}{c} \leq t \leq \frac{T}{2} + \frac{2(R+L)}{c}$, ce qui veut dire que l'on dispose pour chaque projection de données sur la TF spatiale de l'objet dans un intervalle $[k_{x_1}, k_{x_2}]$ donné par

$$k_{x_1} = \frac{2}{c}(\omega_0 - aT + \frac{4aL}{c}) \leq k_x \leq k_{x_1} = \frac{2}{c}(\omega_0 + aT - \frac{4aL}{c})$$

De plus l'angle θ est en général limité à $|\theta| \leq \theta_M$. Ainsi l'ensemble des signaux mesurés nous fournit des renseignements sur $F(k_x,k_y)$ sur un anneau comme le montre la figure 2.20 Dans des applications pratiques on a en général $\omega_0 aT >> \frac{4aL}{c}$, ce qui veut dire que k_{x_1} et k_{x_2} sont proportionnels respectivement à la fréquence la plus petite et à la fréquence la plus grande dans le signal émis, et on a

$$k_{x_1} = \frac{2}{c}(\omega_0 - aT), \quad \text{et} \quad k_{x_2} = \frac{2}{c}(\omega_0 + aT)$$

Comme on peut le constater sur la figure 2, le remplissage du domaine de FOURIER de l'objet est très incomplet. De plus, les données se trouvent ici aussi sur des contours qui ne correspondent pas à un maillage cartésien.

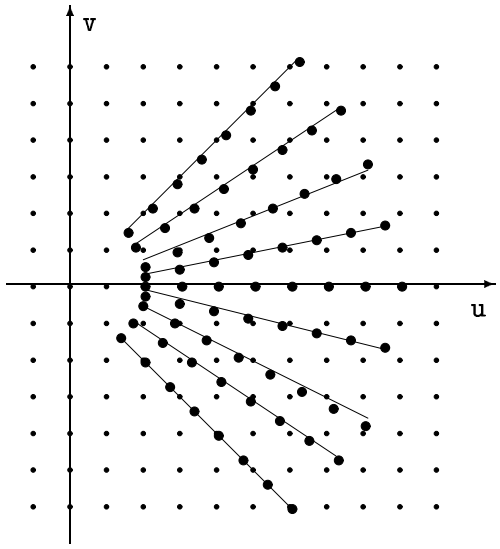


FIG. 2.20 – *Remplissage du domaine de* FOURIER *en imagerie radar.*

2.10 Imagerie en géophysique

L'équation différentielle de la propagation d'une onde dans un milieu non-homogène peut être transformée en une équation intégrale (équation de Lippmann-Schwinger) de la forme:

$$P_s(\boldsymbol{r},\omega) = k^2 \int_V f(\boldsymbol{r}')P(\boldsymbol{r}',\omega)G_0(\boldsymbol{r},\boldsymbol{r}',\omega)\,\mathrm{d}\boldsymbol{r}'$$

dans laquelle P_s et P sont respectivement le champ total et le champ diffracté, G_0 est la fonction de GREEN, $f(\boldsymbol{r}) = n^2(\boldsymbol{r}) - 1$ avec $n(\boldsymbol{r}) = \frac{c}{v(\boldsymbol{r})}$ et $c = \frac{\omega}{k}$. Du fait que $P(\boldsymbol{r}',\omega)$ est fonction des inconnues $v(\boldsymbol{r})$ ou $n(\boldsymbol{r})$, on ne peut pas résoudre cette équation directement. Mais si on se place dans le cadre de l'approximation de BORN, c'est-à-dire un objet peu diffractant, on peut remplacer $P(\boldsymbol{r}',\omega)$ par l'onde incidente $P_0(\boldsymbol{r}',\omega)$.

Considérons maintenant le problème suivant : un objet (supposé en 2-D par simplicité) est illuminé par une onde plane et on mesure le champ diffracté pour toutes les fréquences sur une ligne contournant l'objet (figure 1). Dans ce cas, si on suppose que l'onde plane reste plane en traversant l'objet, et si on néglige les diffractions multiples, on a les relations suivantes :

$$P_s(\boldsymbol{r}',\omega) = k^2 S(\omega) \iint_S P_0(\boldsymbol{r}',\omega)\, G(\boldsymbol{r},\boldsymbol{r}',\omega)\,\mathrm{d}\boldsymbol{r}'$$

$$A(\boldsymbol{k},\omega) = -\int_R [P_s(\boldsymbol{r},\omega)\nabla G(\boldsymbol{r})\nabla P_s(\boldsymbol{r},\omega)G(\boldsymbol{r})]\,\boldsymbol{n}\,\mathrm{d}l$$

où $S(\omega)$ est la TF temporelle de la source, $P_0(\boldsymbol{r}',\omega) = \exp[j\boldsymbol{k}_0.\boldsymbol{r}']$ est l'onde plane incidente, $A(\boldsymbol{k},\omega)$ est la composante onde plane du champ diffracté. Et on montre facilement que l'on a

$$A(\boldsymbol{k},\omega) = -k^2 S(\omega) \iint_S f(\boldsymbol{r}') \exp[-j(\boldsymbol{k} - \boldsymbol{k}_0)r']\ \mathrm{d}\boldsymbol{r}'$$

Pour ω et $\boldsymbol{k}$ fixes, $A(\boldsymbol{k},\omega)$ nous donne ainsi la valeur en un point de la TF spatiale de l'objet $F(\boldsymbol{k} - \boldsymbol{k}_0)$ et en faisant varier ω et $\boldsymbol{k}$ on remplit le domaine de FOURIER. Ce remplissage est cependant incomplet car la source n'a en général que des composantes en fréquences entre 0 et $\omega_{\max}$, et k varie pour $\alpha_1 \leq \theta \leq \alpha_2$ (figure 2).

D'autre part, le calcul de $A(\boldsymbol{k},\omega)$ nécessite la connaissance de $P_s(\boldsymbol{r},\omega)$ et de son gradient $\nabla P_s(\boldsymbol{r},\omega)$. Mais si on considère le cas où les récepteurs se trouvent sur une ligne droite (figure 3) on a

$$A(\boldsymbol{k},\omega) = 2\boldsymbol{n}.\boldsymbol{k} \int_R jkP_s(x,\omega) \exp[-jk(x_0 + x)]\ \mathrm{d}x$$

et on n'a plus besoin de $\nabla P_s(\boldsymbol{r},\omega)$ pour le calcul de $A(\boldsymbol{k},\omega)$. Ainsi en mesurant le champ diffracté $P_s(\boldsymbol{r},\omega)$ pour $x_1 \leq x \leq x_2$ on calcule $A(\boldsymbol{k},\omega)$ pour $0 \leq \omega \leq \omega_{\max}$ et k tel que $\alpha_1 \leq \theta \leq \alpha_2$ et on remplit ainsi le domaine de FOURIER sur des lignes droites.

Figure 1: Imagerie en géophysique.

Figure 2: Remplissage du domaine de FOURIER en géophysique.

A COMPLETER

2.11 Imagerie en radio astronomie

La synthèse d'ouverture en radio-astronomie consiste à reconstruire la distribution de brillance émise, à une certaine longueur d'onde, par un champ continu de sources célestes. Pour cela, on dispose d'un ensemble de capteurs, ou antennes (des radio-télescopes ou des télescopes optiques), permettant d'échantillonner spatialement ce champ et de recueillir des données. Ce problème est similaire à ceux rencontrès en traitement spatial (radar ou sonar).

L'ensemble des intercorrélations inter-capteurs, appelées aussi *visibilité complexes*, correspondent à des versions dégradées et bruitées de certains coefficients de FOURIER de la distribution de brillance recherchée (théorème de Van Cittert-Zernicke). Les visibilité sont souvent affectées d'une sévère distorsion de phase, liée aux conditions météorologiques et aux turbulences atmosphériques. De plus, la structure lacunaire du réseau ne permet pas d'accéder à toutes les données de FOURIER. Le problème comporte ainsi deux aspects : un problème de synthèse de FOURIER, et, simultanément, un problème de calibration de phase. Il s'agit là d'un problème inverse dont la résolution nécessite des méthodes de traitement spécifiques.

A COMPLETER

Chapitre 3

Cadre général et unification

Dans ce chapitre nous essayons de fournir un cadre général aux problèmes inverses et d'établir les principales notions que nous utiliserons par la suite.

3.1 Présentation du cadre général

Nous avons vu que, dans un grand nombre de problèmes réels en science expérimentale, nous sommes souvent amenés à déterminer une grandeur non observable $f(\boldsymbol{u})$ à partir d'un ensemble fini de *mesures* d'une grandeur observable $g(\boldsymbol{u})$ reliées par un *modèle* :

$$\mathcal{H}\left(g(\boldsymbol{u}),f(\boldsymbol{r})\right) = 0,$$

où $\mathcal{H}$ est un opérateur– linéaire ou non– décrivant la relation théorique entre $g(\boldsymbol{u})$ et $f(\boldsymbol{u})$. En général $\boldsymbol{r} \in D_1 \subset \mathbf{R}^k$ et $\boldsymbol{u} \in D_2 \subset \mathbf{R}^{k'}$, et k n'est pas forcément égale à k'.

Parfois la description du modèle peut contenir d'autres grandeurs intermédiaires, par exemple

$$\left\{ \begin{array}{ll} g(\boldsymbol{u}) & = \mathcal{H}_1\left(f(\boldsymbol{r}),z(\boldsymbol{r})\right) \\ 0 & = \mathcal{H}_2\left(f(\boldsymbol{r}),z(\boldsymbol{r})\right) \end{array} \right. ,$$

où $\mathcal{H}_1$ et $\mathcal{H}_2$ sont deux opérateurs (souvent $\mathcal{H}_1$ un opérateur bilinéaire en f et en z et $\mathcal{H}_2$ un opérateur non linéaire), et où $z(\boldsymbol{r})$ est une grandeur inconnue qui intervient dans le modèle mais qui n'est pas notre centre d'intérêt.

Parfois ce modèle peut être *explicite*:

$$g(\boldsymbol{u}) = \mathcal{H}\left(f(\boldsymbol{r}),z(\boldsymbol{r})\right),$$

et parfois, il est même possible d'éliminer les variables intermédiaires et d'obtenir une relation explicite encore plus simple

$$g(\boldsymbol{u}) = \mathcal{H}\left(f(\boldsymbol{r})\right).$$

La description des équations qui lient ces différentes grandeurs est appelée *modélisation*. Le modèle $\mathcal{H}$ est, en général, caractérisé par un certain nombre de paramètres $\boldsymbol{\theta}$. Connaissant $g(\boldsymbol{r})$ et $f(\boldsymbol{r})$, la détermination de $\boldsymbol{\theta}$ est ce que l'on appelle le problème de *l'identification* ou *l'estimation des paramètres* du modèle.

Connaissant le modèle $\mathcal{H}$ et l'objet $f(\boldsymbol{r})$, le calcul de $g(\boldsymbol{u})$ est ce que l'on appelle le *problème direct*. Inversement, le calcul de $f(\boldsymbol{r})$ connaissant $\mathcal{H}$ et $g(\boldsymbol{u})$ est ce que l'on appelle le *problème inverse*.

Autant, la résolution des problèmes directs, en général, ne pose pas de difficulté particulière, la résolution des autres problèmes est souvent plus délicate, car ceux-ci sont souvent des *problèmes mal posés*. En effet, la résolution de ces problèmes présente, entre autre, la caractéristique structurelle désagréable d'être très sensible aux erreurs sur les données. Or aucun dispositif expérimental n'est complètement affranchi d'une incertitude, dont l'origine la plus simple est la précision finie des mesures. De même, aucun modèle mathématique ne peut exactement refléter la réalité. Il est donc plus réaliste de considérer que l'objet recherché et les mesures sont reliés par une équation de la forme

$$\mathcal{H}\left(g(\boldsymbol{u}),f(\boldsymbol{r}),b(\boldsymbol{u})\right) = 0,$$

où $b(\boldsymbol{u})$ représente les erreurs, appelé communément *bruit*.

Lorsqu'on a un modèle explicite et simple, et qu'on peut faire l'hypothèse que les erreurs n'interviennent qu'à la sortie, on obtient :

$$g(\boldsymbol{u}) = \mathcal{H}(f(\boldsymbol{r})) \diamond b(\boldsymbol{u}).$$

Si de plus on peut supposer que les erreurs sont additives on a :

$$g(\boldsymbol{u}) = \mathcal{H}(f(\boldsymbol{r})) + b(\boldsymbol{u}),$$

Lorsque $\mathcal{H}$ est un opérateur linéaire cette relation sous sa forme générale s'écrit

$$g(\boldsymbol{u}) = \int f(\boldsymbol{r})\, h(\boldsymbol{r},\boldsymbol{u})\, \mathrm{d}\boldsymbol{r} + b(\boldsymbol{u}),$$

où $h(\boldsymbol{r},\boldsymbol{u})$, le noyau de cette équation intégrale de première espèce, est appelé la fonction d'appareil. Alors, dans différents domaines d'applications on peut se trouver devant les situations suivantes :

- Lorsque les variables $\boldsymbol{r}$ et $\boldsymbol{u}$ sont homogènes et

  $$h(\boldsymbol{r},\boldsymbol{u}) = h(\boldsymbol{r} - \boldsymbol{u}),$$

  *i.e.*lorsque le modèle est *invariant par translation* on a une équation de *convolution*:

  $$g(\boldsymbol{u}) = \int f(\boldsymbol{r})\, h(\boldsymbol{r} - \boldsymbol{u})\, \mathrm{d}\boldsymbol{r} + b(\boldsymbol{u}).$$

  La fonction $h(\boldsymbol{r})$ est alors appelée la *réponse impulsionnelle* et le problème inverse correspondant *déconvolution* ou encore *restauration d'image*.
- $h(\boldsymbol{r},\boldsymbol{u})$ peut être séparable :

  $$h(\boldsymbol{r},\boldsymbol{u}) = \prod_k h_k(r_k,u_k)$$

  C'est, par exemple, le cas de la *synthèse de* FOURIER

  $$g(\boldsymbol{u}) = \int f(\boldsymbol{r}) \exp\left[-j < \boldsymbol{r},\boldsymbol{u} >\right] \mathrm{d}\boldsymbol{r},$$

  où

  $$h(\boldsymbol{r},\boldsymbol{u}) = \exp\left[-j < \boldsymbol{r},\boldsymbol{u} >\right] = \exp\left[-j \sum_k r_k u_k\right].$$
- $h(\boldsymbol{r},\boldsymbol{u})$ peut être à la fois séparable et invariant par translation :

  $$h(\boldsymbol{r},\boldsymbol{u}) = \prod_k h_k(r_k - u_k)$$

  C'est, par exemple, le cas de la restauration d'image lorsque la réponse impulsionnelle est séparable.
- $h(\boldsymbol{r},\boldsymbol{u})$ peut être de la forme :

  $$h(\boldsymbol{r},\boldsymbol{u}) = \delta(c(\boldsymbol{r},\boldsymbol{u}))$$

  où $\delta(.)$ est la fonction DIRAC et $c(\boldsymbol{r},\boldsymbol{u}) = 0$ décrit un sous-espace dans le domaine de la fonction $f(\boldsymbol{r})$. C'est par exemple le cas de la *reconstruction d'image* en tomographie X où on a $\boldsymbol{u} = (r,\phi)$, $\boldsymbol{r} = (x,y)$ et

  $$g(r,\phi) = \iint f(x,y)\delta(r - x\cos\phi - y\sin\phi)\, \mathrm{d}x\, \mathrm{d}y$$

  et où $r - x\cos\phi - y\sin\phi = 0$ décrit une ligne droite dans l'espace (x,y) du domaine de la fonction $f(x,y)$.

– $h(\boldsymbol{r},\boldsymbol{u})$ peut être de la forme :

$$h(\boldsymbol{r},\boldsymbol{u}) = \exp\left[-j < \boldsymbol{r},\boldsymbol{c}(\boldsymbol{u}) >\right]$$

où $< .,. >$ représente un produit scalaire et $\boldsymbol{c}(\boldsymbol{u})$ un ensemble de fonctions telles que les équations $\boldsymbol{u} = \boldsymbol{c}(\boldsymbol{u})$ décrivent un sous-espace dans le domaine de FOURIER de la fonction $f(\boldsymbol{r})$. À titre d'exemple, prenons le cas où $\boldsymbol{u} = (\Omega,\phi)$, $\boldsymbol{r} = (x,y)$, $c_1(\boldsymbol{u}) = c_1(\Omega,\phi) = \Omega\cos\phi$, $c_2(\boldsymbol{u}) = c_2(\Omega,\phi) = \Omega\sin\phi$ et

$$g(\Omega,\phi) = \iint f(x,y)\exp\left[-j\left[\Omega\cos\phi x + \Omega\sin\phi y\right]\right]\,\mathrm{d}x\,\mathrm{d}y$$

et $u = \Omega\cos\phi$ et $v = \Omega\sin\phi$ décrivent une ligne droite dans l'espace (u,v) faisant un angle ϕ avec l'axe des u.

3.2 Exemples classiques de problèmes inverses linéaires

Parmi les problèmes inverses linéaires que l'on rencontre souvent en traitement du signal et de l'image, nous allons en citer un certain nombre qui sont les plus classiques.

3.2.1 Déconvolution de signaux 1-D

$$\begin{aligned} g(t) &= f(t) * h(t) + b(t) \\ &= \int h(\tau) f(t-\tau)\,\mathrm{d}\tau + b(t) \\ &= \int f(\tau) h(t-\tau)\,\mathrm{d}\tau + b(t) \end{aligned}$$

où $g(t)$ est le signal mesuré, $f(t)$ est le signal recherché, et $h(t)$ est la réponse impulsionnelle du système de mesure. Ici l'opérateur $\mathcal{H}$ est un opérateur de convolution.

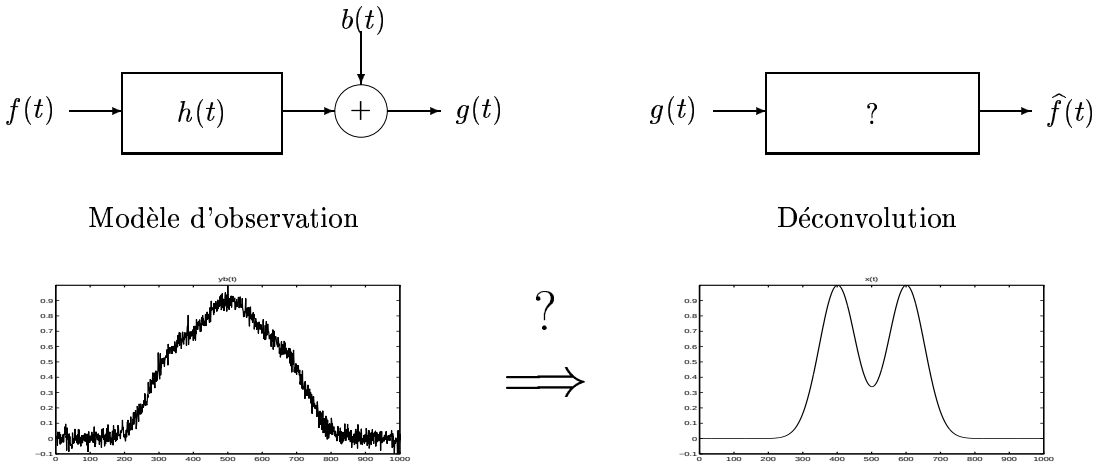


FIG. 3.1 – *Déconvolution de signaux.*

Sous forme discrète cette relation s'écrit :

$$g(m) = \sum_{k=-q}^{p} h(k) f(m-k) + b(m), \quad m = 0, \cdots, M$$

ou encore

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{b}$$

où $\boldsymbol{g} = [g_0, \cdots, g_M]$, $\boldsymbol{f} = [f_{-p}, \cdots, f_{M+q}]$ et où $\boldsymbol{H}$ est une matrice de TOEPLITZ dont les éléments sont déterminés par les échantillons de sa réponse impulsionnelle $\boldsymbol{h} = [h_{-q}, \cdots, 0, \cdots, h_p]$.

Le caractère TOEPLITZ de la matrice $\boldsymbol{H}$ est due à la propriété d'invariance par translation de l'opérateur.

3.2.2 Débruitage ou filtrage

$$g(t) = f(t) + b(t),$$

où $g(t)$ est le signal mesuré, $f(t)$ est le signal recherché, et $b(t)$ est le bruit. Lorsque $g(t)$ est observée aux instants réguliers $t_i = (i-1)\Delta T$,$i = 1, \cdots, n$ on peut considérer ce problème comme un cas particulier de déconvolution où la réponse impulsionnelle $h(t) = \delta(t)$.

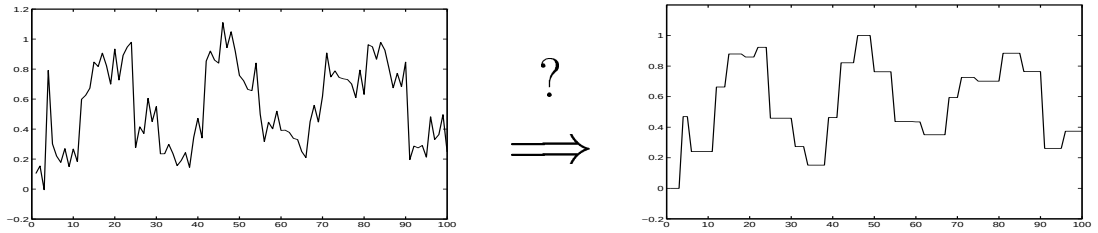


FIG. 3.2 – *Débruitage de signaux.*

Lorsque $g(t)$ est observée aux instants irréguliers t_i il s'agit de problèmes d'interpolation et d'extrapolation des signaux.

De même on peut considérer le débruitage ou le filtrage des images ::

$$g(x,y) = f(x,y) + b(x,y).$$

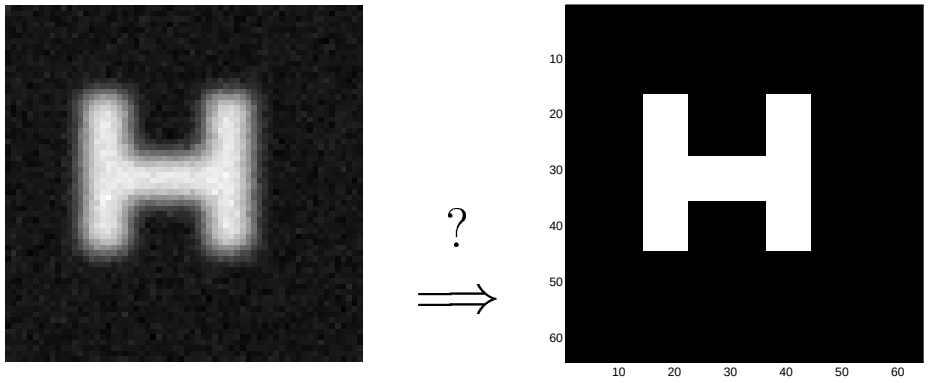


FIG. 3.3 – *Débruitage d'images.*

3.2.3 Restauration d'image

$$\begin{aligned} g(x,y) &= f(x,y) * h(x,y) + b(x,y) \\ &= \iint_D f(x',y')\, h(x-x',y-y')\, \mathrm{d}x'\, \mathrm{d}y' + b(x,y) \\ &= \iint_D h(x',y')\, f(x-x',y-y')\, \mathrm{d}x'\, \mathrm{d}y' + b(x,y), \end{aligned}$$

où $g(x,y)$ est l'image observée, $f(x,y)$ est l'image recherchée, et $h(x,y)$ est la réponse impulsionnelle du système d'imagerie. Ici $\boldsymbol{u} = (x,y)$, $\boldsymbol{r} = (x',y')$, $h(\boldsymbol{r},\boldsymbol{u}) = h(\boldsymbol{r}-\boldsymbol{u})$ et l'opérateur $\mathcal{H}$ est un opérateur de convolution bivariable.

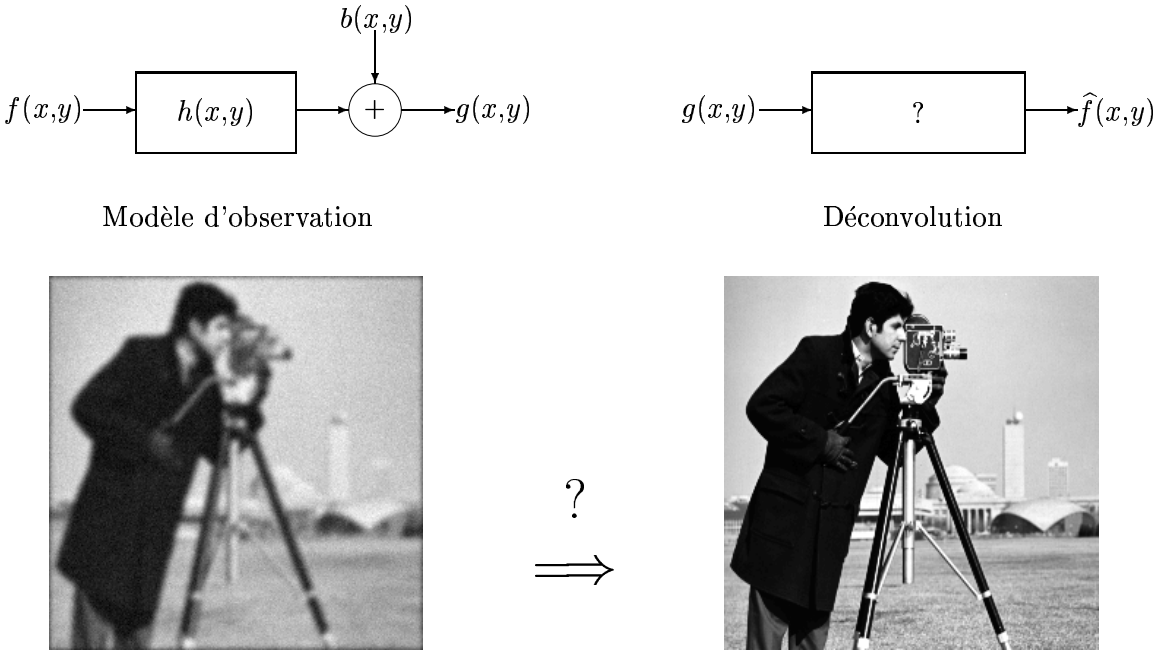


FIG. 3.4 – *Restauration d'image.*

Sous forme discrète cette relation s'écrit :

$$g(m,n) = \sum_{k=-p}^{p'} \sum_{l=-q}^{q'} h(k,l) f(m-k,n-l) + b(m,n)$$

ou encore

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{b}$$

où $\boldsymbol{g}$ et $\boldsymbol{f}$ sont deux vecteurs contenant les lignes (ou les colonnes, dans un ordre lexicographique) de l'image dégradée et de l'image initiale et $\boldsymbol{H}$ est une matrice qui a une structure de TOEPLITZ-BLOC-TOEPLITZ dont les éléments sont déterminés par les échantillons de la réponse impulsionnelle $h(k,l)$.

Le caractère TOEPLITZ-BLOC-TOEPLITZ de la matrice $\boldsymbol{H}$ est propre à la propriété de l'opérateur de convolution.

3.2.4 Reconstruction d'image en tomographie X

$$g(r,\phi) = \iint_D f(x,y)\,\delta(r - x\cos\phi - y\sin\phi)\,\mathrm{d}x\,\mathrm{d}y + b(r,\phi),$$

où $g(r,\phi)$ représente la projection suivant l'angle ϕ, et $f(x,y)$ est l'objet à reconstruire. Ici $\boldsymbol{u} = (r,\phi)$, $\boldsymbol{r} = (x,y)$, $h(\boldsymbol{r},\boldsymbol{u}) = \delta(r - x\cos\phi - y\sin\phi)$ et l'opérateur $\mathcal{H}$ est l'opérateur de la transformée de RADON (TR).

Notons que, théoriquement, cette transformée admet une inversion qui est

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

mais, contrairement à ce qu'on peut imaginer, l'existence d'une expression analytique pour l'inversion de cette transformée ne réduit en rien les difficultés de son inversion dans des situations réelles. Nous reviendrons sur cette remarque dans les chapitres suivants.

La discrétisation de cette équation est plus délicate que celle de la restauration d'image. On peut cependant, imaginer facilement que cette relation sous forme discrète, puisse aussi s'écrire de la forme

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{b}$$

où $\boldsymbol{g}$ est un vecteur concaténant l'ensemble des projections $p(r_i,\phi_j)$ et $\boldsymbol{f}$ est un vecteur contenant les pixels de l'image. La matrice $\boldsymbol{H}$ dont les éléments sont entièrement déterminés par la géométrie de l'instrumentation (la discrétisation de l'objet et les positions de capteurs). Cette matrice n'a pas, en général, de structure particulière, cependant elle est sûrement très creuse. En effet, si on partage l'espace de (x,y) en pixels et si on fait l'hypothèse que $f(x,y)$ est constante à l'intérieure de chaque pixel, alors l'éléments H_{ij} correspond à la longueur du segment du rayons i traversant le pixel j.

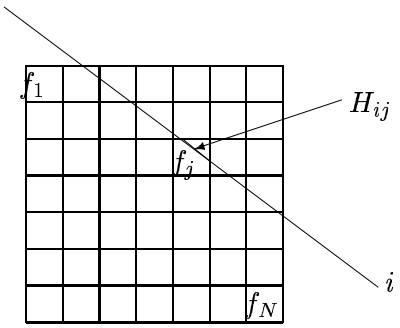


FIG. 3.5 – *Discrétisation de la transformée de* RADON.

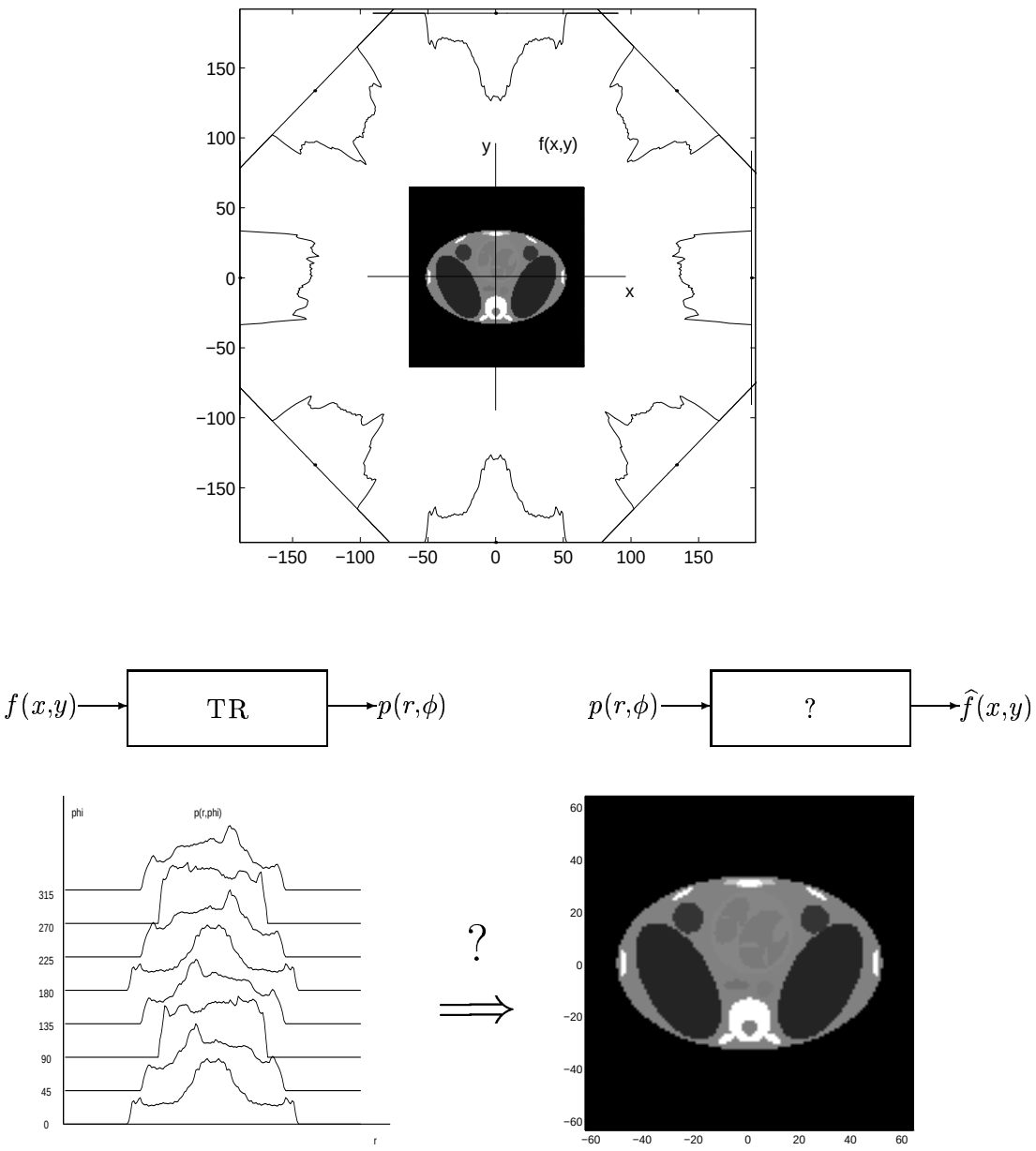


FIG. 3.6 – *Reconstruction d'image en tomographie X.*

3.2.5 Radiographie des objets cylindriques

Dans le problème de reconstruction d'image en tomographie X, lorsque l'objet $f(x,y)$ a une symétrie de révolution, la fonction ne dépend pas de l'angle ϕ et peut être exprimée seulement en fonction de ρ en coordonnées cylindriques: $f(x,y) = f(\rho)$. De même, ses projections $g(r,\phi)$ suivant les différentes angles ϕ sont identiques et ne dépendent pas de ϕ. On peut alors établir une relation entre $f(\rho)$ et $g(r)$ qui est :

$$g(r) = \int_r^R f(\rho) \frac{\rho}{\sqrt{\rho^2 - r^2}} \, \mathrm{d}\rho + b(r),$$

où $g(r)$ est l'intensité des rayons traversés (mesures), $f(\rho)$ est la fonction densité de la matière qui ne dépend que de la coordonnée radiale ρ et R est le rayon extérieur de l'objet. Remarquons que cette équation intégrale correspond à une transformée d'ABEL.

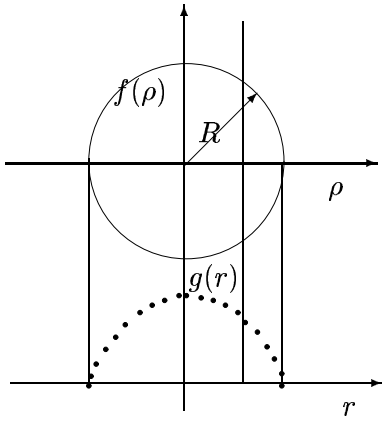


FIG. 3.7 – *Radiographie des objets cylindriques.*

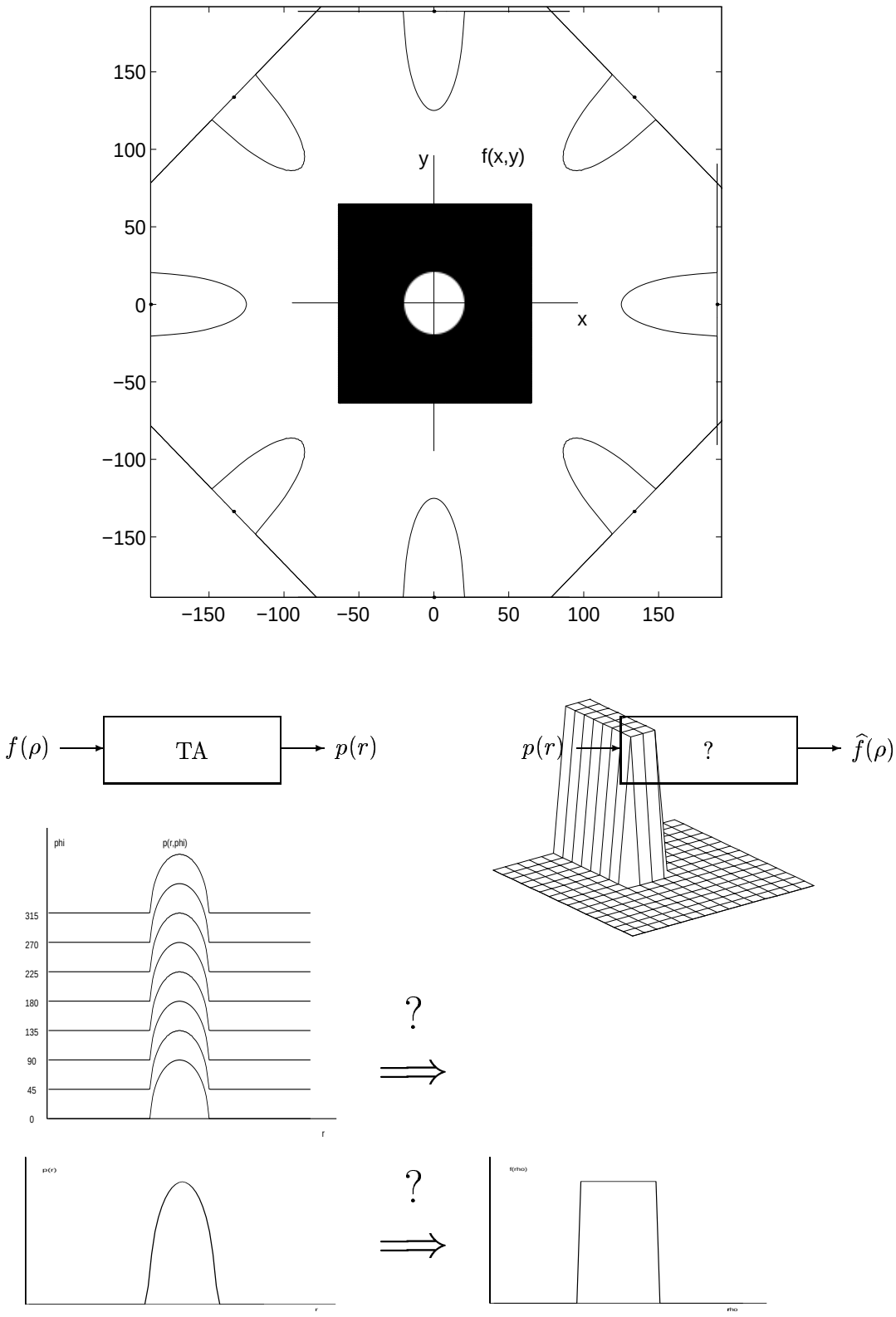


FIG. 3.8 – *Reconstruction d'image en tomographie X pour des objets cylindriques.*

3.2.6 Synthèse d'ouverture en radio astronomie ou imagerie radar

$$g(u,v) = \iint_D f(x,y)\, \exp\left[-j(ux+vy)\right]\, \mathrm{d}x\, \mathrm{d}y + b(u,v),$$

où $f(x,y)$ est l'image à reconstruire, et $g(u,v)$ représente, soit une mesure de la corrélation entre les signaux recueillis par deux antennes en synthèse d'ouverture en radio astronomie, soit la transformée de FOURIER (TF) du signal d'écho en imagerie radar. Dans le cas de la radio astronomie, $f(x,y)$ représente la brillance du ciel tandis que dans le cas de l'imagerie radar elle représente la variation de l'indice de réfraction de l'objet (la cible) par rapport à son environnement.

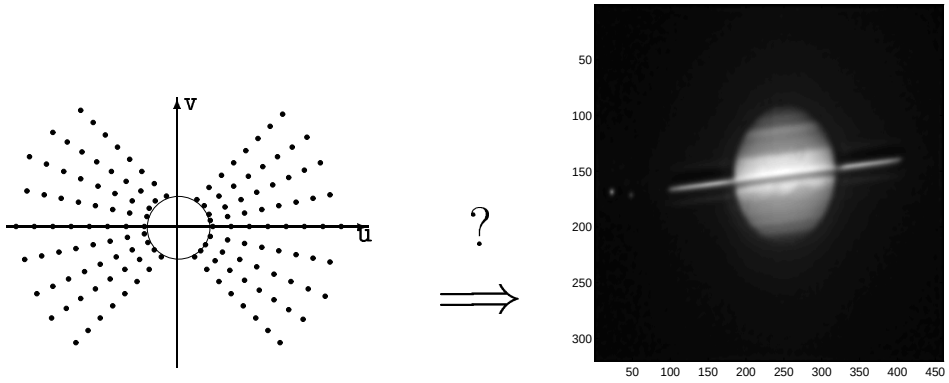


FIG. 3.9 – *Synthèse d'ouverture en radioastronomie.*

Ici $\boldsymbol{u} = (u,v)$, $\boldsymbol{r} = (x,y)$, $h(\boldsymbol{r},\boldsymbol{u}) = \exp\left[-j < \boldsymbol{r},\boldsymbol{u} >\right]$ et l'opérateur $\mathcal{H}$ est l'opérateur de la transformée de FOURIER bivariable.

Notons qu'ici aussi, théoriquement, la TF inverse existe et on a :

$$f(x,y) = \iint g(u,v) \exp\left[+j(ux+vy)\right]\, \mathrm{d}u\, \mathrm{d}v,$$

mais, là encore, l'existence de cette formule ne réduit en rien la difficulté du problème inverse de la synthèse de FOURIER. En effet, si $g(u,v)$ était connue précisement et pour tout $(u,v) \in \mathbf{R}^2$, alors $f(x,y)$ pourrait être calculée à partir de cette intégrale. Mais, en pratique, $g(u,v)$ peut être mesurée sur un nombre fini de points (u_i,v_i) non nécessairement distribués d'une manière uniforme dans l'espace (u,v), et c'est exactement là la difficulté de l'inversion.

3.2.7 Synthèse de Fourier en reconstruction tomographique d'image

$$g(\Omega,\phi) = \iint_D f(x,y)\, \exp\left[-j\left(T_1(\Omega,\phi)x + T_2(\Omega,\phi)y\right)\right]\, \mathrm{d}x\, \mathrm{d}y + b(\Omega,\phi)$$

où

$$\begin{cases} u &= T_1(\Omega,\phi) \\ v &= T_2(\Omega,\phi) \end{cases}$$

définissent un ensemble de contours algébriques dans l'espace de FOURIER (u,v).

Comme nous l'avons vu dans le chapitre précédent, la synthèse de FOURIER se trouve au cœur d'un grand nombre d'applications en imagerie. Parmi celles-ci, on trouve :

- Tomographie à rayons X :

$$\begin{pmatrix} u = T_1(\Omega,\phi) \\ v = T_2(\Omega,\phi) \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} \Omega \\ 0 \end{pmatrix},$$

  où les contours algébriques sont des lignes droites concentriques, et où $g(\Omega,\phi_i)$ représente la TF spatiale de $p(r,\phi_i)$ par rapport à la variable r.
- Imagerie par résonnance magnétique nucléaire (RMN) :
  Cette formulation du problème se trouve aussi dans le cas de l'imagerie RMN. Dans ce cas $g(\Omega,\phi_i) = s(t,\phi_i)$, où $s(t,\phi_i)$ est directement le signal de précession libre mesuré quand le gradient du champ magnétique est dans la direction ϕ_i.
- Tomographie à ondes diffractées :

$$\begin{pmatrix} u = T_1(\Omega,\phi) \\ v = T_2(\Omega,\phi) \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} -k_0 + \sqrt{k_0^2 - \Omega^2} \\ \Omega \end{pmatrix},$$

  où k_0 est la constante de propagation des ondes, et où les $g(\Omega,\phi_i)$ représentent les TF spatiales des champ diffractés mesurés. Ici, les contours algébriques sont des demi-cercles concentriques.

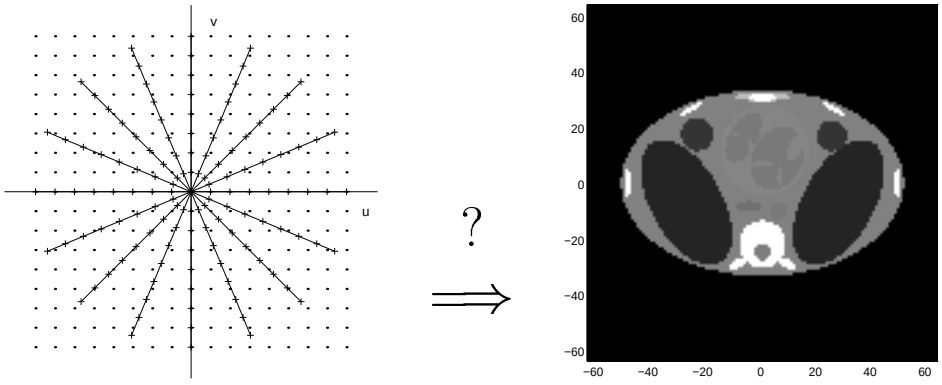


FIG. 3.10 – *Synthèse de* FOURIER *en tomographie X.*

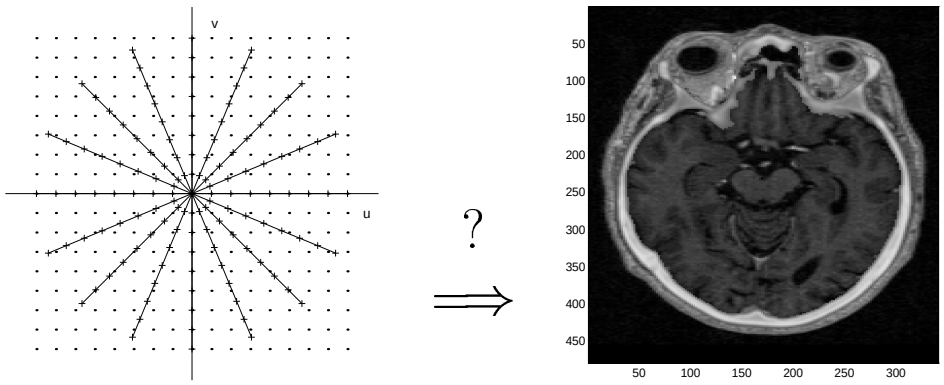


FIG. 3.11 – *Synthèse de* FOURIER *en imagerie RMN.*

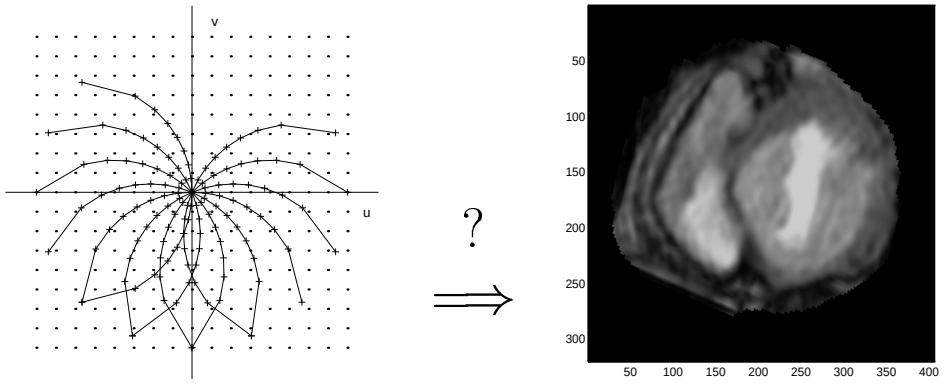


FIG. 3.12 – *Synthèse de* FOURIER *en tomographie par ondes diffractées.*

3.2.8 Synthèse de Fourier-Laplace

$$g(u,s) = \iint_D f(x,y)\, \exp\left[-j(ux + sy)\right]\, \mathrm{d}x\, \mathrm{d}y + b(u,s),$$

où u est une variable réelle, tandis que $s = \gamma + jv$ est une variable complexe.

Cette équation intervient en tomographie de diffraction dans des objets dissipatifs ou bien en imagerie tomographique par courants de FOUCAULT en contrôle non destructif.

(u,v) représente un point dans le domaine de FOURIER spatial de l'objet $f(x,y)$. Ceci correspond au phénomène de propagation dans les deux directions, tandis que γ représente un phénomène de diffusion ou d'atténuation.

3.3 Exemples de problèmes inverses non linéaires

Parmi les problèmes inverses non linéaires classiques nous allons citer deux exemples.

3.3.1 Imagerie par ondes diffractées

Lorsqu'un objet $f(\boldsymbol{r})$ est illuminé par une onde incidente $\phi_0(\boldsymbol{r})$ le champ total $\phi(\boldsymbol{r})$ peut être modélisé par la somme du champ incident et un champ diffracté qui dépend à la fois de l'objet et du champ incident :

$$\phi(\boldsymbol{r}) = \phi_0(\boldsymbol{r}) + \int_D G_o(\boldsymbol{r},\boldsymbol{r}')\phi(\boldsymbol{r}')\, f(\boldsymbol{r}')\, \mathrm{d}\boldsymbol{r}'$$

et le champ diffracté à l'extérieur de l'objet $g(\boldsymbol{u})$ que l'on mesure est une fonction de l'objet et du champ total à l'intérieur de l'objet :

$$g(\boldsymbol{u}) = \int_D G_m(\boldsymbol{r},\boldsymbol{u})\, f(\boldsymbol{r})\, \phi(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} + b(\boldsymbol{u}).$$

Le problème de l'imagerie dans ce contexte est de fournir une estimation de $f(\boldsymbol{r})$ à partir d'un nombre fini de mesures $g(\boldsymbol{u}_i), i = 1, \cdots, M$.

Le cas de l'approximation de Born étant la situation de faibles perturbations où on peut remplacer $\phi(\boldsymbol{r})$ par $\phi_0(\boldsymbol{r})$ dans cette dernière équation. Nous avons déjà présentés et détaillés ces équations dans le cas linéaire. Ici, l'objectif est de considérer le problème non linéaire.

Travailler dans le cadre fonctionnel ne permet pas de voir facilement les difficultés du problème. C'est pourquoi, nous allons discrétiser ces équations, en utilisant par exemple la méthode des moments. Il n'est alors pas difficile de voir que la forme discrétisée de ces deux équations devient :

$$\begin{aligned} \boldsymbol{\phi} &= \boldsymbol{\phi}_0 + \boldsymbol{G}_o \boldsymbol{F} \boldsymbol{\phi}, \\ \boldsymbol{g} &= \boldsymbol{G}_m \boldsymbol{F} \boldsymbol{\phi} + \boldsymbol{b}, \end{aligned}$$

où

- $\boldsymbol{\phi}_0$ et $\boldsymbol{\phi}$ sont deux vecteurs représentant respectivement le champ incident et le champ total,
- $\boldsymbol{G}_m$ et $\boldsymbol{G}_o$ sont deux matrices reliées aux fonctions de Green,
- $\boldsymbol{F}$ est une matrice diagonale dont les éléments diagonaux sont formés par les éléments du vecteur $\boldsymbol{f}$ qui représente l'objet (les pixels de l'image à reconstruire concatenés ligne par ligne ou colonne par colonne) et,
- $\boldsymbol{g}$ est un vecteur contenant les mesures et $\boldsymbol{b}$ un vecteur représentant les erreurs de mesures.

Écrite sous cette forme, pour mieux illustrer le caractère non linéaire du problème, il est maintenant possible d'expliciter ' la relation liant les mesures $\boldsymbol{g}$ aux inconnus $\boldsymbol{f}$, en éliminant les variables intermédiaires $\boldsymbol{\phi}$. En effet, calculant $\boldsymbol{\phi}$ à l'aide de la première équation et la remplaçant dans la deuxième on obtient :

$$\boldsymbol{g} = \boldsymbol{G}_m \boldsymbol{F} \left(\boldsymbol{I} - \boldsymbol{G}_o \boldsymbol{F}\right)^{-1} \boldsymbol{\phi}_0 + \boldsymbol{b} = \boldsymbol{H}(\boldsymbol{f}) + \boldsymbol{b},$$

ce qui permet de montrer d'une manière plus explicite le caractère non linéaire de l'opérateur $\boldsymbol{H}$ liant l'objet aux mesures.

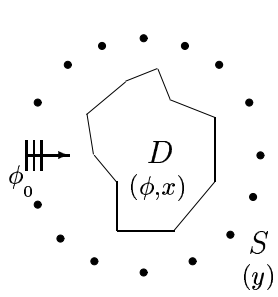


FIG. 3.13 – *Géométrie de l'imagerie par ondes diffractées.*

Evidemment, il n'existe pas de relation explicite pour l'inversion de ces équations et l'obtention d'une solution ne peut se faire que d'une manière itérative, ceci quelque soit le cadre choisi qu'il soit en dimension infinie ou finie.

Dans ce problème non linéaire aussi, la résolution du problème directe, c'est à dire étant donnée $f(\boldsymbol{r})$ calculer $g(\boldsymbol{r})$, ne pose pas de grandes difficultés. En effet, on peut montrer qu'avec un choix judicieux pour les fonctions de bases et la discrétisation de ces équations, on peut calculer $g(\boldsymbol{r})$ aussi précisément que l'on souhaite si on connait l'objet $f(\boldsymbol{r})$. Par contre, la résolution du problème inverse, c'est à dire le calcule de $f(\boldsymbol{r})$ à partir de $g(\boldsymbol{r})$ est plus délicat. En effet, même lorsqu'on suppose pouvoir mesurer $g(\boldsymbol{r})$ partout autour de l'objet avec une précision infinie, l'existence et l'unicité de la solution ne sont pas garanties pour n'importe quel objet $f(\boldsymbol{r})$. De plus, même lorsqu'on impose à l'objet $f(\boldsymbol{r})$ dans un ensemble de fonctions telles que ces deux propriétés soient garanties, la solution du problème reste très sensible aux erreurs de mesures.

A COMPLETER

3.3.2 Tomographie d'impédance

$$p(\boldsymbol{u}_i) = \int_V \sigma(\boldsymbol{r}) |\nabla \phi_i(\boldsymbol{r})|^2 \, \mathrm{d}\boldsymbol{r} + b_i$$

où $p(\boldsymbol{u}_i)$ est la puissance dissipée mesurée entre le point de coordonnée $(\boldsymbol{u}_i)$ et un point de référence, $\sigma(\boldsymbol{r})$ est la distribution des conductivitée électriques dans l'objet, $\phi_i(\boldsymbol{r})$ est la distribution du potentiel électrique dans l'objet. Notons que $\phi_i(\boldsymbol{r})$ dépend du $\sigma(\boldsymbol{r})$, ce qui rend cette équation non linéaire.

A COMPLETER

Chapitre 4

Problèmes mal-posés

Pour mieux comprendre les difficultés de la résolution des problèmes inverses, un cadre classique est l'étude des problèmes inverses linéaires dans des espaces de Hilbert où l'on définit deux ensembles F et G et un opérateur $H : F \mapsto G$ tel que :

$$\boldsymbol{Hf} = \boldsymbol{g} \quad \boldsymbol{f} \in F \quad \boldsymbol{g} \in G \quad \text{espaces de Hilbert}$$

Souvent l'opérateur H est borné et compact et nous verrons que c'est exactement ces caractères qui sont à l'origine des principales difficultés de la résolution des problèmes inverses.

Ce cadre topologique et abstrait nous permettra de définir les notions de problèmes bien posés et mal posés. C'est pourquoi, avant de donner ces définitions un rappel sur la terminologie de l'analyse fonctionnelle est nécessaire. Après ce rappel les notions de problèmes *bien posés* et *mal posés* seront définies et un certain nombre d'exemples détailleront ces notions.

Finalement, une analyse plus détaillée du problème de la déconvolution nous permettra de comprendre d'une manière plus concrète le caractère mal posé du problème et les difficultés de sa résolution.

4.1 Rappels d'analyse fonctionnelle

Considérons deux ensembles F et G et un opérateur $H : F \mapsto G$.

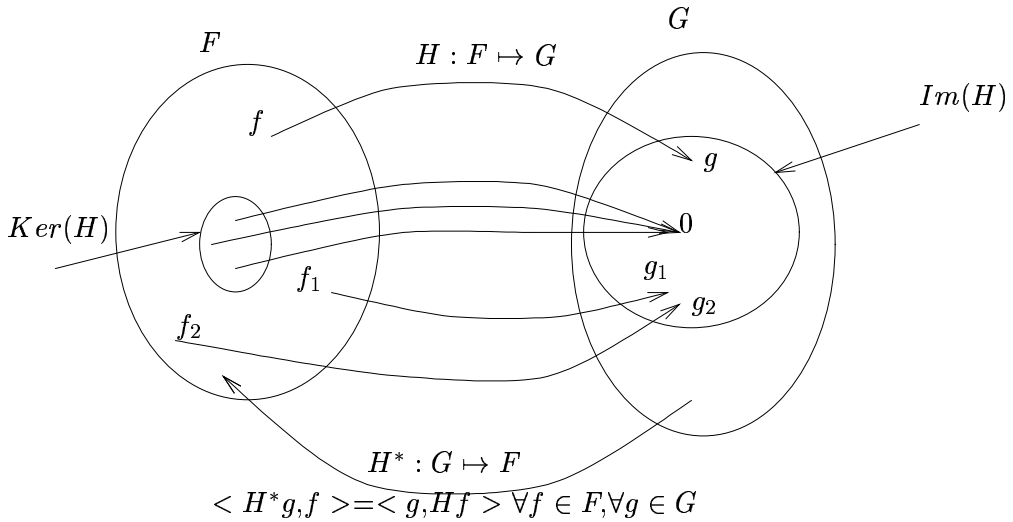


FIG. 4.1 – *Analyse fonctionnelle et topolgie*

- Topologie sur un ensemble F:
  Un système de sous-ensembles de F, dits **ouverts**, tels que la réunion d'ensembles ouverts est un ouvert, et l'intersection d'un nombre fini d'ensembles ouverts est un ouvert.
- $F_1 \subset F$ est dit **fermé** si son complémentaire est un ouvert.
- Pour un ouvert M, on définit :
  - sa fermeture : $\overline{M}$ intersection de tous les fermés qui contiennent M
  - sa frontière : $\overline{M} - M$
- $\mathcal{I}\mathrm{m}\,(\boldsymbol{H}) = \bigcup\{\boldsymbol{Hf}\} \subset G$ est l'image de $\boldsymbol{H}$
- $\mathcal{K}\mathrm{er}\,(\boldsymbol{H}) = \{\boldsymbol{f}, \boldsymbol{Hf} = \boldsymbol{0}\} \subset F$ est le noyau de $\boldsymbol{H}$
- $\boldsymbol{H}^* : G \mapsto F$ est l'opérateur adjoint de $\boldsymbol{H}$

$$\langle \boldsymbol{H}^*\boldsymbol{g}, \boldsymbol{f}\rangle = \langle \boldsymbol{g}, \boldsymbol{Hf}\rangle \quad \forall \boldsymbol{f} \in F, \forall \boldsymbol{g} \in G$$

- $\mathcal{K}\mathrm{er}\,(\boldsymbol{H}) = \mathcal{I}\mathrm{m}\,(\boldsymbol{H}^*)^{\perp}$
- $\mathcal{K}\mathrm{er}\,(\boldsymbol{H}^*) = \mathcal{I}\mathrm{m}\,(\boldsymbol{H})^{\perp}$
- $\mathcal{K}\mathrm{er}\,(\boldsymbol{H})^{\perp} = \overline{\mathcal{I}\mathrm{m}\,(\boldsymbol{H}^*)}$
- $\mathcal{K}\mathrm{er}\,(\boldsymbol{H}^*)^{\perp} = \overline{\mathcal{I}\mathrm{m}\,(\boldsymbol{H})}$

- $\boldsymbol{H}$ est **borné** s'il existe un réel $c \geq 0$ tel que $\|\boldsymbol{H}\boldsymbol{f}\| \leq c\|\boldsymbol{f}\|$, $\forall \boldsymbol{f} \in F$
- Un ensemble $B \subset \mathbf{R}^n$ est dit **borné** s'il est entièrement situé dans une boule ouverte dont le centre est à l'origine et le rayon est fini.
- Un ensemble fermé et borné est dit **compact**.
- Un opérateur $\boldsymbol{H}$ est dit **compact** s'il applique tout ensemble borné $B \subset F$ dans un ensemble $A(B)$ dont la fermeture $\overline{A(B)}$ est compacte.

Exemple 1:

$$\begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}$$

$$\boldsymbol{H} = \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}, \quad \boldsymbol{H}^* = \begin{pmatrix} 1 & 3 \\ 2 & 6 \end{pmatrix}$$

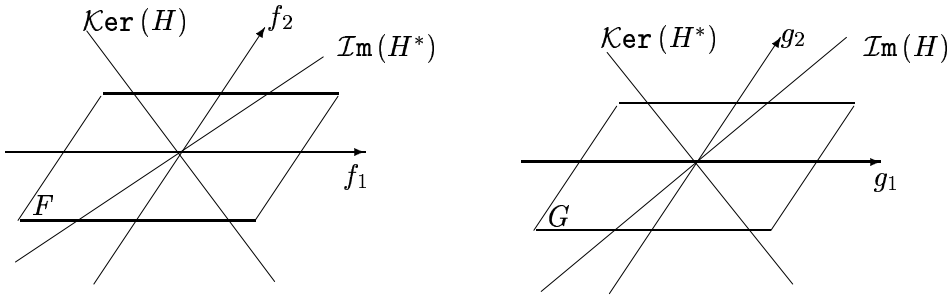


FIG. 4.2 – *Image et noyau d'un opérateur et de son adjoint: $F = G = \mathbb{R}^2$*

$$\begin{aligned}
\mathcal{K}\mathrm{er}\,(\boldsymbol{H}) &= \{(f_1,f_2) \ : f_1 = -2f_2\} \\
\mathcal{I}\mathrm{m}\,(\boldsymbol{H}^*) &= \left\{(f_1,f_2) \ : f_1 = \frac{1}{2}f_2\right\} \\
\mathcal{K}\mathrm{er}\,(\boldsymbol{H}^*) &= \{(g_1,g_2) \ : g_1 = -3g_2\} \\
\mathcal{I}\mathrm{m}\,(\boldsymbol{H}) &= \left\{(g_1,g_2) \ : g_1 = \frac{1}{3}g_2\right\}
\end{aligned}$$

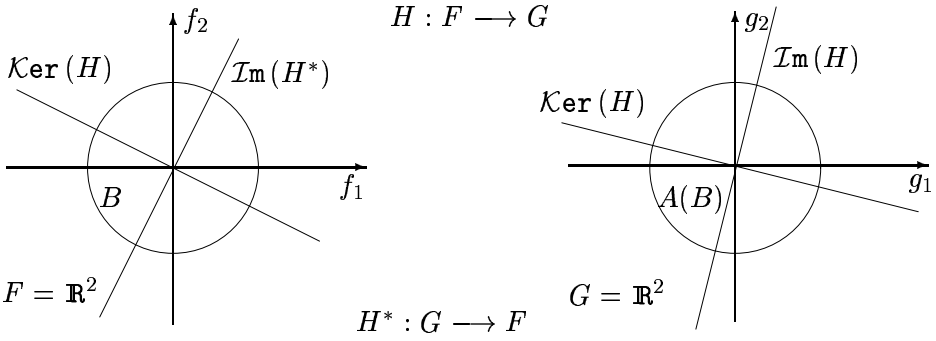


FIG. 4.3 – *Ensemble et opérateur compacts*

Exemple 2:

$$\begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix}$$

$$\boldsymbol{H} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{H}^* = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{H}^*\boldsymbol{H} = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad \boldsymbol{H}\boldsymbol{H}^* = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

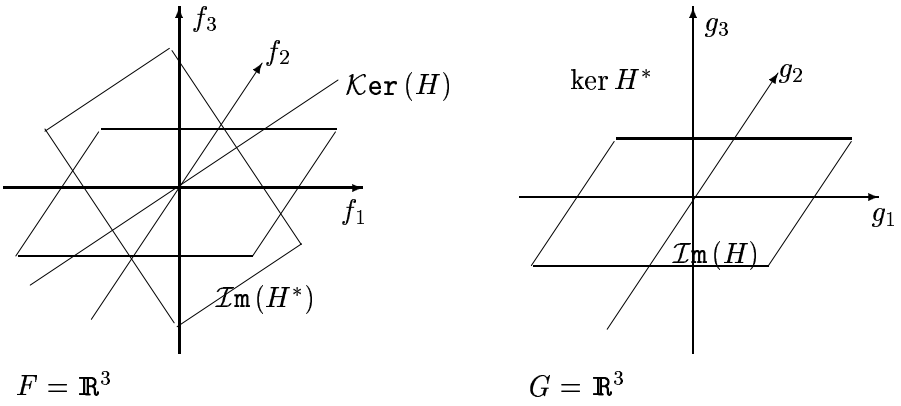


FIG. 4.4 – *Image et noyau d'un opérateur et de son adjoint : $F = G = \mathbb{R}^3$*

$$\begin{aligned}
\mathcal{K}\mathrm{er}\,(\boldsymbol{H}) &= \{(f_1,f_2,f_3) \ : f_1 = f_2, f_3 = 2f_2\} \\
\mathcal{I}\mathrm{m}\,(\boldsymbol{H}^*) &= \{(f_1,f_2,f_3) \ : f_1 + f_2 - 2f_3 = 0\} \\
\mathcal{K}\mathrm{er}\,(\boldsymbol{H}^*) &= \{(g_1,g_2,g_3) \ : g_1 = g_2 = 0\} \\
\mathcal{I}\mathrm{m}\,(\boldsymbol{H}) &= \{(g_1,g_2,g_3) \ : g_3 = 0\}
\end{aligned}$$

Exemple 3:

$$\begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix}$$

$$\boldsymbol{H} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{pmatrix}, \quad \boldsymbol{H}^* = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 0 \end{pmatrix}, \quad \boldsymbol{H}^*\boldsymbol{H} = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad \boldsymbol{H}\boldsymbol{H}^* = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$$

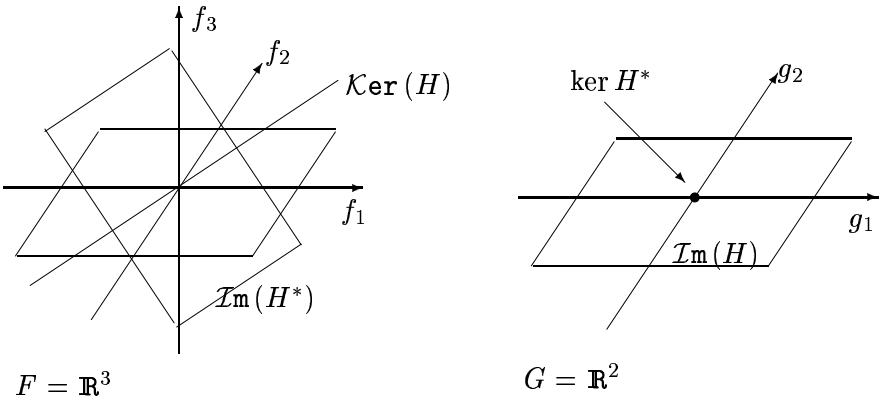


FIG. 4.5 – *Image et noyau d'un opérateur et de son adjoint : $F = \mathbb{R}^3$,$G = \mathbb{R}^2$*

$$\begin{aligned} \mathcal{K}\mathrm{er}\,(\boldsymbol{H}) &= \{(f_1,f_2,f_3) \ : f_1 = f_2, f_3 = 2f_2\} \\ \mathcal{I}\mathrm{m}\,(\boldsymbol{H}^*) &= \{(f_1,f_2,f_3) \ : f_1 + f_2 - 2f_3 = 0\} \\ \mathcal{K}\mathrm{er}\,(\boldsymbol{H}^*) &= \{(g_1,g_2) \ : g_1 = g_2 = 0\} \\ \mathcal{I}\mathrm{m}\,(\boldsymbol{H}) &= \{(g_1,g_2) \ : g_1 = g_2 \in \mathbb{R}\} \end{aligned}$$

4.2 Définitions

D'après HADAMARD, un problème est dit *bien posé* si le problème admet une solution (Existence), si elle unique (Unicité) et si elle est stable (Stabilité). Le problème est dit *mal posé* si au moins une de ces trois conditions n'est pas vérifiée.

- **Existence** $\boldsymbol{g} \in G = \mathcal{I}\mathrm{m}(\boldsymbol{H}) \quad \forall \boldsymbol{g} \in G \exists \boldsymbol{f} : \boldsymbol{H}\boldsymbol{f} = \boldsymbol{g}$

- **Unicité** $\mathcal{K}\mathrm{er}(\boldsymbol{H}) = \{0\} \quad \boldsymbol{H}\boldsymbol{f} = 0 \longrightarrow \boldsymbol{f} = 0$

- **Stabilité** $\mathcal{I}\mathrm{m}(\boldsymbol{H}) = \overline{\mathcal{I}\mathrm{m}(\boldsymbol{H})} = \mathcal{K}\mathrm{er}(\boldsymbol{H}^*)^\perp$
  faible perturbation de $\boldsymbol{g}$ donne
  faible perturbation de la solution $\boldsymbol{f}$

$$F = \mathcal{K}\mathrm{er}(\boldsymbol{H}) + \mathcal{K}\mathrm{er}(\boldsymbol{H})^\perp, \quad G = \mathcal{K}\mathrm{er}(\boldsymbol{H}^*) + \mathcal{K}\mathrm{er}(\boldsymbol{H}^*)^\perp$$

Quatre situations mutuellement exclusives :

1. $\mathcal{I}\mathrm{m}(\boldsymbol{H})$ est dense dans G : $\quad (\mathcal{K}\mathrm{er}(\boldsymbol{H}) = \{0\}$ et $\boldsymbol{g} \in \mathcal{I}\mathrm{m}(\boldsymbol{H}))$
   Il existe une solution au sens classique.
2. $\mathcal{I}\mathrm{m}(\boldsymbol{H})$ n'est pas dense dans G : $\quad (\mathcal{K}\mathrm{er}(\boldsymbol{H}) = \{0\}$ et $\boldsymbol{g} \notin \mathcal{I}\mathrm{m}(\boldsymbol{H}))$
   Il n'existe pas de solution au sens classique.
3. $\overline{\mathcal{I}\mathrm{m}(\boldsymbol{H})} \subset G$ et $\boldsymbol{g} \in \mathcal{I}\mathrm{m}(\boldsymbol{H}) + \mathcal{I}\mathrm{m}(\boldsymbol{H})^\perp$
   On peut définir une solution au sens de l'inversion généralisée.
4. $\overline{\mathcal{I}\mathrm{m}(\boldsymbol{H})} \subset G$ et $\boldsymbol{g} \notin \mathcal{I}\mathrm{m}(\boldsymbol{H}) + \mathcal{I}\mathrm{m}(\boldsymbol{H})^\perp$
   Il n'existe pas de solution au sens classique.

$\boldsymbol{H}$: opérateur linéaire borné et compact $\longrightarrow \mathcal{I}\mathrm{m}(\boldsymbol{H})$ ouvert
La symétrisation de $\boldsymbol{H}$ n'apporte rien.

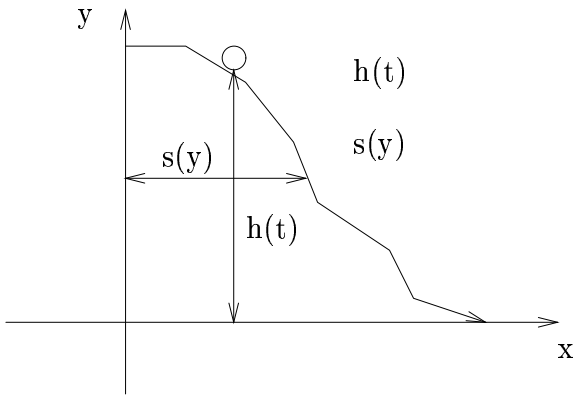
$$\mathcal{I}\mathrm{m}(\boldsymbol{H}^*\boldsymbol{H}) = \mathcal{I}\mathrm{m}(\boldsymbol{H}^*), \text{ si } \mathcal{K}\mathrm{er}(\boldsymbol{H}) = \{0\} \rightarrow \mathcal{K}\mathrm{er}(\boldsymbol{H}\boldsymbol{H}^*) = \mathcal{K}\mathrm{er}(\boldsymbol{H}^*)$$

$$\boldsymbol{H}^*\boldsymbol{g} = \boldsymbol{H}^*\boldsymbol{H}\boldsymbol{f}$$

est aussi mal-posé. Les problèmes directs sont généralement bien posés. Les problèmes inverses sont souvent mal posés.

4.3 Exemples de problèmes mal-posés

Exemple 1: Problème d'Abel



- $s(y)$ profil de trajectoire d'un point pesant, partant de l'altitude h.
- $t(h)$ temps d'arrivée à l'altitude zéro

$$t(h) = \frac{1}{\sqrt{2g}} \int_0^h \frac{s'(y)}{\sqrt{(h-y)}} \, \mathrm{d}y$$

inversement

$$s(y) = -\frac{\sqrt{2g}}{\pi} \int_0^y \frac{t(h)}{\sqrt{(y-h)}} \, \mathrm{d}h$$

- s_i = résultat d'une mesure de $s(y_i)$,
- $d_s = \sup|s_i - s(y_i)|$ mesure "naturelle",
- t_e = valeur de t permettant de coller exactement sur une donnée expérimentale,
- $d_t = \sup|t_e - t|$ mesure "naturelle",
- L'application $s \mapsto t$ n'est pas continue pour ces distances!

Exemple 2: Diffraction inverse ou Synthèse d'ouverture

- $f(r)$ forme d'un objet,
- $g(u)$ champ diffracté,

$$g(u) = \int_{-\infty}^{\infty} f(r) \exp\left[-jur\right] \, \mathrm{d}r$$

$$f(r) = \frac{1}{2\pi} \int_{-\infty}^{\infty} g(u) \exp\left[jur\right] \, \mathrm{d}u$$

$$g_1(u) = g_0(u) + \epsilon \exp\left[-\alpha u^2\right]$$

$$f_1(r) = f_0(r) + \frac{\epsilon}{(4\pi\alpha)^{1/2}} \exp\left[-r^2/4\alpha\right]$$

- norme sup:

$$\max_u |g_1(u) - g_0(u)| \leq \alpha\epsilon$$

$$\max_r |f_1(r) - f_0(r)| \leq \alpha \frac{\epsilon}{(4\pi\alpha)^{1/2}}$$

- norme L^2:

$$\left\{ \int_{-\infty}^{\infty} [g_1(u) - g_0(u)]^2 \, \mathrm{d}u \right\}^{1/2} = \epsilon(\pi/2\alpha)^{1/4}$$

$$\left\{ \int_{-\infty}^{\infty} [f_1(r) - f_0(r)]^2 \, \mathrm{d}r \right\}^{1/2} = \frac{1}{2\pi}\epsilon(\pi/2\alpha)^{1/4}$$

- Un problème mal-posé peut être rendu bien-posé juste en changeant l'espace des solutions.

Exemple 3: Opérateurs linéaires et compacts

$$g(s) = \int_a^b h(s,t) f(t) dt \quad f \in F, \quad g \in G, \quad F = G = L^2[a,b],$$

et

$$\int_a^b \int_a^b |h(s,t)|^2 \, \mathrm{d}s \, \mathrm{d}t < \infty$$

$\boldsymbol{H}$ est compact et admet une décomposition spectrale

$$h(s,t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s) \psi_k(t)$$

$$\int_a^b h(s,t) \psi_k(t) \, \mathrm{d}t = \lambda_k \phi_k(s), \qquad \int_a^b h(s,t) \phi_k(s) \, \mathrm{d}s = \lambda_k \psi_k(t)$$

$\{\psi_k\}$ et $\{\phi_k\}$, fonctions orthonormales singulières de $\boldsymbol{H}$, ou bien fonctions propres de $\boldsymbol{H}^*\boldsymbol{H}$ et $\boldsymbol{H}\boldsymbol{H}^*$, forment une base de $L^2[a,b]$.

$$g(s) = \sum_{k=1}^{\infty} a_k \phi_k(s)$$

$$f(t) = \sum_{k=1}^{\infty} b_k \psi_k(t) \quad \text{avec} \quad b_k = \frac{< g, \phi_k >}{\lambda_k}$$

Pour que cette solution appartienne à $L^2[a,b]$, il faut imposer

$$\|f\|^2 = \sum_{k=1}^{\infty} b_k^2 < \infty$$

$$\sum_{k=1}^{\infty} b_k^2 = \sum_{k=1}^{\infty} \frac{1}{\lambda_k^2} a_k^2 \longrightarrow 0$$

G doit être réduit par rapport à $L^2[a,b]$
sinon, F peut contenir des solutions qui peuvent être inacceptables.

4.4 Déconvolution : un problème mal-posé

Pour montrer que la déconvolution est un problème mal-posé nous pouvons utiliser les outils suivants :

- La transformée de FOURIER qui permet de transformer l'opérateur de la convolution en une multiplication ;
- La discrétisation du problème et l'utilisation de l'algèbre matricielle ;
- L'analyse fonctionnelle et la décomposition spectrale des opérateurs dans l'espace de Hilbert.

Mais avant d'essayer d'expliquer pourquoi regarder sur la figure qui suit

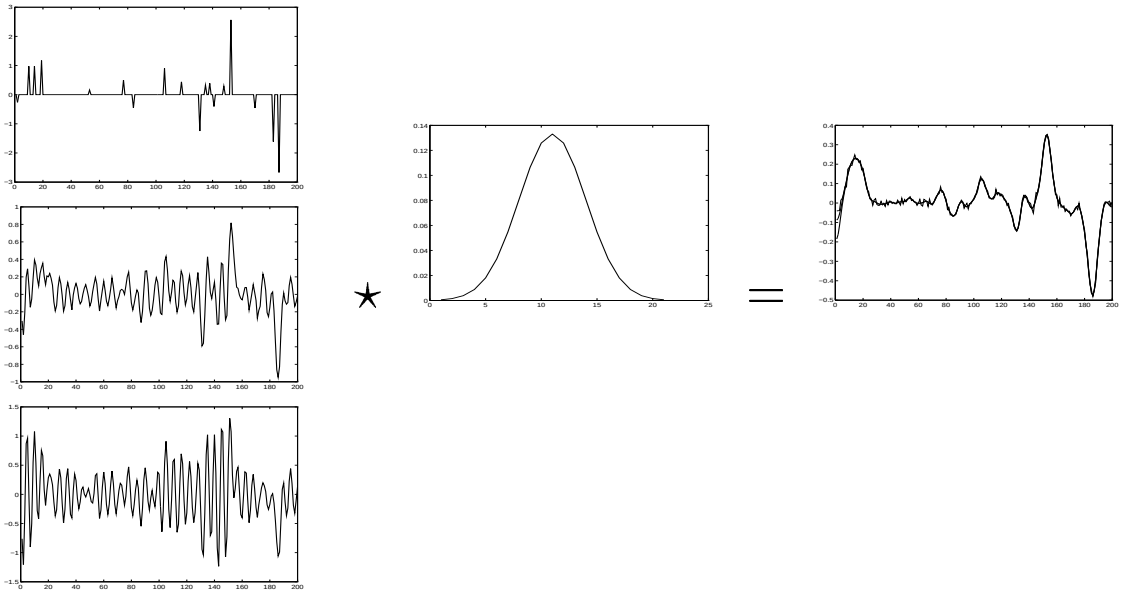


FIG. 4.6 – *Déconvolution : un problème mal-posé.*

On constate que trois signaux $f_1(t)$, $f_2(t)$ et $f_3(t)$ très différents à l'entrée d'un instrument ayant pour réponse impulsionnelle $h(t)$, ont fourni pratiquement les mêmes signaux $g_1(t)$, $g_2(t)$ et $g_3(t)$ de sortie (les trois signaux sont tracés sur un seul graphe). Ceci signifie qu'une méthode d'inversion se basant seulement sur une distance de sortie ne peut distinguer ces trois signaux.

4.4.1 Passage dans le domaine de Fourier

Le schéma suivant résume l'essentiel de cette approche :

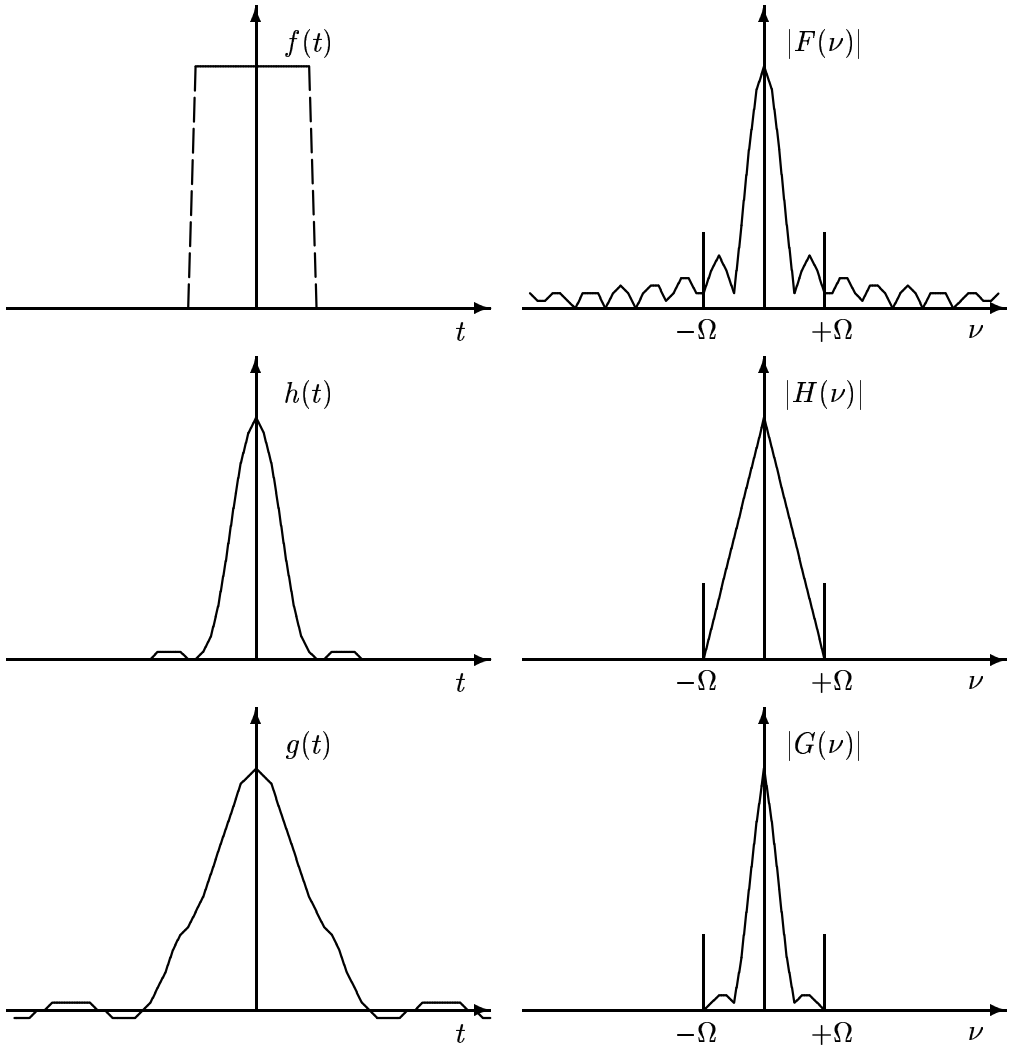


FIG. 4.7 – *Déconvolution : un problème mal posé*

En effet, on peut constater que tout signal $f(t)$ qui a même contenu spectral dans la bande passante de l'opérateur (l'intervalle $[-\Omega,\Omega]$) fournit la même sortie $g(t)$.

La déconvolution est ainsi un problème inverse mal posé. Un moyen pour le rendre bien posé est de restreindre l'espace des solutions aux fonctions à spectre limité dans la bande $[-\Omega,\Omega]$:

$$F(\omega) = 0 \quad \forall \omega \notin [-\Omega, +\Omega]$$

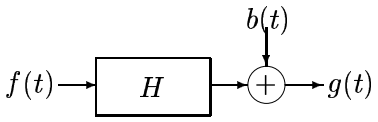
En faisant ceci, on retrouve le critère de Rayleigh classique, on assure l'unicité de la solution mais on n'a pas tout résolu. Une autre difficulté qui persiste est la sensibilité de la solution au bruit. En effet, lorsque les signaux mesurés sont bruités on a :

$$G(\omega) = H(\omega)F(\omega) + B(\omega) \longrightarrow F(\omega) = \frac{G(\omega)}{H(\omega)} + \frac{B(\omega)}{H(\omega)}$$

et il faut étudier le comportement de $B(\omega)/H(\omega)$ qui conditionne cette sensibilité.

4.4.2 Discrétisation

Considérons le modèle :



$$g(t) = \int f(t')h(t-t')\,\mathrm{d}t' + b(t) = \int h(t')f(t-t')\,\mathrm{d}t' + b(t)$$

et faisons les hypothèses suivantes :

- on discrétise les signaux $f(t)$, $g(t)$, $h(t)$ avec le même pas de d'échantillonnage $\Delta T = 1$,
- la réponse impulsionnelle est à support limité : $h(t) = 0, \forall t < -q\Delta T$, et $\forall t > p\Delta T$.

On peut alors écrire :

$$g(m) = \sum_{k=-q}^{p} h(k)f(m-k) + b(m), \quad m = 0, \cdots, M$$

ou sous forme matricielle :

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{b} + \boldsymbol{b}$$

$$\begin{pmatrix} g(0) \\ g(1) \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ g(M) \end{pmatrix} = \begin{pmatrix} h(p) & \cdots & h(0) & \cdots & h(-q) & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & & \ddots & & \ddots & & & & \vdots \\ \vdots & & & & & & & & & \vdots \\ \vdots & & & h(p) & \cdots & h(0) & \cdots & h(-q) & & \vdots \\ \vdots & & & & & & & & & \vdots \\ \vdots & & & & & \ddots & & & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & h(p) & \cdots & h(0) & \cdots & h(-q) \end{pmatrix} \begin{pmatrix} f(-p) \\ \vdots \\ f(0) \\ f(1) \\ \vdots \\ f(M) \\ f(M+1) \\ \vdots \\ f(M+q) \end{pmatrix}$$

Notons que $\boldsymbol{g}$ est de dimension $M+1$, $\boldsymbol{f}$ de dimension $M+p+q+1$, $\boldsymbol{h} = [h(p), \cdots, h(0), \cdots, h(-q)]$ de dimension $(p+q+1)$ et la matrice $\boldsymbol{H}$ de dimensions $(M+1) \times (M+p+q+1)$.

Maintenant si on fait l'hypothèse que le système est causal, on a $q = 0$ et

$$\begin{pmatrix} g(0) \\ g(1) \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ g(M) \end{pmatrix} = \begin{pmatrix} h(p) & \cdots & h(0) & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & & & & & & & \vdots \\ \vdots & & & & & & & \vdots \\ \vdots & & h(p) & \cdots & h(0) & & & \vdots \\ \vdots & & & & & & & \vdots \\ \vdots & & & & & & & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & h(p) & \cdots & h(0) \end{pmatrix} \begin{pmatrix} f(-p) \\ \vdots \\ f(0) \\ f(1) \\ \vdots \\ \vdots \\ f(M) \end{pmatrix}$$

et si de plus, on fait l'hypothèse que le signal d'entrée est aussi causal, on obtient :

$$\begin{pmatrix} g(0) \\ g(1) \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ g(M) \end{pmatrix} = \begin{pmatrix} h(0) & & & & & \\ h(1) & \ddots & & & & \\ \vdots & & & & & \\ h(p) & \cdots & & h(0) & & \\ 0 & \ddots & & & \ddots & \\ \vdots & & & & & \\ 0 & \cdots & 0 & h(p) & \cdots & h(0) \end{pmatrix} \begin{pmatrix} f(0) \\ f(1) \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ f(M) \end{pmatrix}$$

et, finalement, si $p = M$, on obtient :

$$\begin{pmatrix} g(0) \\ g(1) \\ \vdots \\ g(M) \end{pmatrix} = \begin{pmatrix} h(0) & & & \\ h(1) & h(0) & & \\ \vdots & & \ddots & \\ h(M) & \cdots & h(1) & h(0) \end{pmatrix} \begin{pmatrix} f(0) \\ f(1) \\ \vdots \\ f(M) \end{pmatrix}$$

On peut remarquer que dans tous les cas, la matrice $\boldsymbol{H}$ est TOEPLITZ et, sauf dans ces deux derniers cas, elle est aussi circulante.

Dans le cas où la matrice n'est pas circulante, il est possible de la rendre artificiellement circulante en ajoutant des zéros à la suite du vecteur $\boldsymbol{f}$ (bourrage de zéros). En effet on peut toujours écrire :

$$\begin{pmatrix} g(0) \\ g(1) \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ g(M) \end{pmatrix} = \begin{pmatrix} h(0) & & & & & & h(p) & \cdots & h(1) \\ h(1) & \ddots & & & & & & \ddots & \vdots \\ \vdots & & & & & & & & h(p) \\ h(p) & \cdots & & h(0) & & & & & 0 \\ 0 & \ddots & & & \ddots & & & & \vdots \\ \vdots & & & & & & & & \\ 0 & \cdots & 0 & h(p) & \cdots & h(0) & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} f(0) \\ f(1) \\ \vdots \\ \vdots \\ \vdots \\ f(M) \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Mettons nous ensuite dans la situation idéale sans bruit. On pourrait alors s'imaginer qu'il suffit de choisir $p = M$ pour que la matrice $\boldsymbol{H}$ soit carrée et définir la solution par $\widehat{\boldsymbol{f}} = \boldsymbol{H}^{-1}\boldsymbol{g}$ ou même dans les autres cas, de définir la solution soit par $\widehat{\boldsymbol{f}} = (\boldsymbol{H}^t\boldsymbol{H})^{-1}\boldsymbol{H}^t\boldsymbol{g}$ ou par $\widehat{\boldsymbol{f}} = \boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{H}^t)^{-1}\boldsymbol{g}$. Là encore il ne faut oublier que la matrice $\boldsymbol{H}$ (ou $\boldsymbol{H}^t\boldsymbol{H}$ ou $\boldsymbol{H}\boldsymbol{H}^t$) peut être singulière ou du moins très mal conditionnée. En effet, le conditionnement de ces matrices dépend du pas d'échantillonnage. Pour comprendre cela, il suffit remarquer que les lignes successives de $\boldsymbol{H}$ sont très proches (due au caractère TOEPLITZ), et cela d'autant plus que le pas d'échantillonnage est petit. On peut aussi expliquer ceci à l'aide des propriétés des valeurs propres et des vecteurs propres des matrices de TOEPLITZ:

A COMPLETER

4.4.3 Décomposition spectrale

- $\boldsymbol{H} : F \mapsto G$ linéaire borné et Hilbertien
  $\{\psi_k,\phi_k;\lambda_k^2\}$ système singulier de $\boldsymbol{H}$:

$$\boldsymbol{H}\psi_k = \lambda_k\phi_k, \quad \boldsymbol{H}^*\phi_k = \lambda_k\psi_k$$

$$\boldsymbol{H}\boldsymbol{H}^*\phi_k = \lambda_k^2\phi_k, \quad \boldsymbol{H}^*\boldsymbol{H}\psi_k = \lambda_k^2\psi_k$$

- Résolution de $\boldsymbol{H}^*\boldsymbol{g} = \boldsymbol{H}^*\boldsymbol{H}\boldsymbol{f}$ par projection :
  $\{\psi_k\}$ base de G, $\{\phi_k\}$ base de F

$$x = \sum_{k=1}^{\infty} \frac{1}{\lambda_k} < y,\phi_k > \psi_k$$

- **Théorème de Hilbert-Schmidt :**
  Si $\boldsymbol{H}$ est compact, alors 0 est le seul point d'accumulation de $\{\lambda_k^2\}$:

$$\lim_{k \longrightarrow \infty} \lambda_k^2 = 0 \quad \exists k_0 : \forall k > k_0, \lambda_k = 0$$

$$\Downarrow$$

$$x + \delta x = \sum_{k=1}^{\infty} \frac{1}{\lambda_k} < y,\phi_k > \left[1 + \frac{< \delta y,\phi_k >}{< y,\phi_k >}\right] \psi_k$$

◇ **Changer la notion de la solution**

4.5 Changer la notion de solution

- Bien définir le problème direct $\boldsymbol{H} : F \mapsto G$
- Pour tout $\boldsymbol{g} \in G$, étudier l'existence des sous ensembles non vides de $F : \boldsymbol{Hf} = \boldsymbol{g}$
- Rechercher et définir des applications $G \mapsto F$
- étudier le problème des erreurs après avoir donné à F et G des structures d'espaces métriques.

Les propriétés de la solution ne sont pas entièrement contenues dans $\boldsymbol{g} = \boldsymbol{Hf}$

- Solutions approchées $\quad \widehat{\boldsymbol{f}} : \|\boldsymbol{H}\widehat{\boldsymbol{f}} - \boldsymbol{g}\| \leq \epsilon, \epsilon > 0$ fixé
- Quasi-solutions $\quad \widehat{\boldsymbol{f}} : \|\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}\| \leq \|\boldsymbol{g} - \boldsymbol{Hf}\|, \quad \forall \boldsymbol{f} \in F$
- Changement d'espace ou de topologie
- Opérateurs régularisants
- Concepts probabilistes

Chapitre 5

Moindres carrés et inversion généralisée

Nous avons vu qu'un problème est dit mal-posé s'il ne satisfait pas à une des trois conditions : existence, unicité et stabilité. Lorsque la difficulté vient de la non unicité de la solution l'inversion généralisée permet d'y remédier. En effet, la résolution de l'équation $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ présente trois possibilités :

- Il y a une solution au problème ;
- Il y a une infinité de solutions au problème ;
- Il n'y pas de solution au problème.

Dans le premier cas, il ne reste qu'à calculer l'inverse. Dans le second cas, le problème devient celui d'un choix parmi l'ensemble des solutions suivant un critère (par exemple, la norme minimale). Dans le troisième cas, on souhaiterait aussi pouvoir définir une solution, cela est possible en définissant un critère du type distance dans l'espace G entre $\boldsymbol{g}$ et $\boldsymbol{H}\boldsymbol{f}$, (par exemple la distance euclidienne), et en choisissant comme solution celle qui la minimise. Dans le cas où la distance choisie est une distance euclidienne la solution est dite *au sens des moindres carrés* et nous verrons qu'il y a un lien entre la solution au sens des MC et celle de l'inverse généralisée.

5.1 Définition en dimension infinie

Soit $\boldsymbol{H} : F \mapsto G$,
$\boldsymbol{P}$ le projecteur orthogonal de F sur $\mathcal{K}\mathrm{er}\,(\boldsymbol{H})^{\perp} = \overline{\mathcal{I}\mathrm{m}\,(\boldsymbol{H}^*)}$ et
$\boldsymbol{Q}$ le projecteur orthogonal de G sur $\mathcal{K}\mathrm{er}\,(\boldsymbol{H}^*)^{\perp} = \overline{\mathcal{I}\mathrm{m}\,(\boldsymbol{H})}$. On a alors

$$\boldsymbol{H}\boldsymbol{f} = \boldsymbol{H}\boldsymbol{P}\boldsymbol{f} \quad \boldsymbol{H}^*\boldsymbol{g} = \boldsymbol{H}^*\boldsymbol{Q}\boldsymbol{g}$$

Définition 1 *L'inverse généralisée $\boldsymbol{H}^+$ de $\boldsymbol{H}$ est l'opérateur dont le domaine est $\mathcal{D}(\boldsymbol{H}^+) = \mathcal{I}m\,(\boldsymbol{H}) \oplus \mathcal{I}m\,(\boldsymbol{H})^{\perp}$, tel que $\boldsymbol{H}^+\boldsymbol{g} = \boldsymbol{f}_0 \in F$, où $\boldsymbol{f}_0 \in S = \{\boldsymbol{f} \in F : \boldsymbol{H}\boldsymbol{f} = \boldsymbol{Q}\boldsymbol{g}\}$ et $\|\boldsymbol{f}_0\| \leq \|\boldsymbol{f}\|, \quad \forall \boldsymbol{f} \in S, \boldsymbol{f} \neq \boldsymbol{f}_0$.*

Notons que

$$\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 = \|\boldsymbol{Q}\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \|\boldsymbol{g} - \boldsymbol{Q}\boldsymbol{g}\|^2$$

Si $\boldsymbol{g} \in \mathcal{I}\mathrm{m}\,(\boldsymbol{H}) \oplus \mathcal{I}\mathrm{m}\,(\boldsymbol{H})^{\perp}$, l'ensemble des solutions des MC est non-vide, fermé et convexe. Il existe donc un élément unique de norme minimale, qui est $\boldsymbol{H}^+\boldsymbol{g}$.

Limitations :

- $\forall \boldsymbol{g} \in G$, il existe une solution inverse généralisée ssi $\mathcal{I}\mathrm{m}\,(\boldsymbol{H})$ est fermée.
- En général $\boldsymbol{H}$ est compact, $\mathcal{I}\mathrm{m}\,(\boldsymbol{H})$ n'est pas fermé, sauf si $\boldsymbol{H}$ est de dimension finie.

- Existence : Non $\quad \mathcal{I}\mathrm{m}\,(\boldsymbol{H}) \oplus \mathcal{I}\mathrm{m}\,(\boldsymbol{H})^{\perp} \neq G$
- Unicité : Oui
- Stabilité : Non $\quad \mathcal{I}\mathrm{m}\,(\boldsymbol{H}) \neq \overline{\mathcal{I}\mathrm{m}\,(\boldsymbol{H})}$

5.2 Définition en dimension finie

Considérons l'équation

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

où $\boldsymbol{H}$ est une matrice de dimension $[m,n]$.

Définition 2 (Quasi-solution) *On appelle $\widehat{\boldsymbol{f}}$ une quasi-solution de l'équation $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ une solution qui minimise une distance $\Delta(\boldsymbol{g},\boldsymbol{H}\boldsymbol{f})$ entre $\boldsymbol{g}$ et $\boldsymbol{H}\boldsymbol{f}$:*

$$\widehat{\boldsymbol{f}} = arg\min_{\boldsymbol{f}} \{\Delta(\boldsymbol{g},\boldsymbol{H}\boldsymbol{f})\}$$

Lorsque la distance choisie est une distance euclidienne $\Delta(\boldsymbol{g},\boldsymbol{H}\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2$, la solution ainsi définie est dite "au sens des moindres carrés".

Définition 3 (Moindre carré) *On appelle $\widehat{\boldsymbol{f}}$ une solution au sens des moindres carrés de l'équation $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ une solution qui minimise la norme euclidienne $\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2$:*

$$\widehat{\boldsymbol{f}} = arg\min_{\boldsymbol{f}} \left\{\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2\right\}.$$

Cet ensemble coïncide avec l'ensemble des solutions de l'équation normale

$$[\boldsymbol{H}^t\boldsymbol{H}]\widehat{\boldsymbol{f}} = \boldsymbol{H}^t\boldsymbol{g}.$$

- *Si $M = N$ et $rang\{\boldsymbol{H}\} = N$, alors $\widehat{\boldsymbol{f}} = \boldsymbol{f} = \boldsymbol{H}^{-1}\boldsymbol{g}$.*
- *Si $N \leq M$ et $rang\{\boldsymbol{H}\} = N$, alors $[\boldsymbol{H}^t\boldsymbol{H}]$ est de rang plein et $\widehat{\boldsymbol{f}} = [\boldsymbol{H}^t\boldsymbol{H}]\boldsymbol{H}^t\boldsymbol{g}$.*
- *Si $M < N$, alors $[\boldsymbol{H}^t\boldsymbol{H}]$ est forcement singulière. Elle a un rang au maximum égal à M, et l'équation (1) n'a pas de solution unique (Elle a une infinité de solutions).*

Définition 4 (Inverse généralisée) *On appelle $\boldsymbol{f}^+ = \boldsymbol{H}^+\boldsymbol{g}$ la solution inverse généralisée de l'équation $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ la solution MC de norme minimale de l'équation normale $[\boldsymbol{H}^t\boldsymbol{H}]\widehat{\boldsymbol{f}} = \boldsymbol{H}^t\boldsymbol{g}$. Si $\boldsymbol{g} \in \mathcal{I}m(H)$ alors cette solution existe et elle est unique. Elle est donnée par :*

$$\boldsymbol{f}^+ = \boldsymbol{H}^+\boldsymbol{g}.$$

où $\boldsymbol{H}^+$ est l'inverse généralisée de $\boldsymbol{H}$.
Si $rang\{\boldsymbol{H}\} = N$ alors $\boldsymbol{H}^+ = [\boldsymbol{H}^t\boldsymbol{H}]^{-1}\boldsymbol{H}^t$ et si $rang\{\boldsymbol{H}\} = M$ alors $\boldsymbol{H}^+ = \boldsymbol{H}^t[\boldsymbol{H}^t\boldsymbol{H}]^{-1}$.

Pour être complet nous mentionnons deux autres définitions pour l'inverse généralisée d'une matrice.

Définition 5 (Moore) *Une matrice $\boldsymbol{G}$ de dimension $[N,M]$ est dite inverse généralisée de la matrice $\boldsymbol{H}$ si*

$$\boldsymbol{H}\boldsymbol{G} = \boldsymbol{P}_A \quad et \quad \boldsymbol{G}\boldsymbol{H} = \boldsymbol{P}_G$$

où $\boldsymbol{P}_A$ est l'opérateur (matrice) de projection des vecteurs de l'espace $\mathbb{R}^m$ sur $\mu(\boldsymbol{H})$ et $\boldsymbol{P}_G$ est l'opérateur (matrice) de projection des vecteurs de l'espace $\mathbb{R}^n$ sur $\mu(\boldsymbol{G})$.

Définition 6 (Penrose) *Une matrice $\boldsymbol{G}$ de dimension $[N,M]$ est dite inverse généralisée de la matrice $\boldsymbol{H}$ si*

$$\boldsymbol{H}\boldsymbol{G}\boldsymbol{H} = \boldsymbol{H}, \quad (\boldsymbol{H}\boldsymbol{G})^t = \boldsymbol{H}\boldsymbol{G}, \quad \boldsymbol{G}\boldsymbol{H}\boldsymbol{G} = \boldsymbol{G} \quad et \quad (\boldsymbol{G}\boldsymbol{H})^t = \boldsymbol{G}\boldsymbol{H}$$

Ces définitions sont identiques si les normes associées aux espaces $\mathbb{R}^m$ et $\mathbb{R}^n$ sont du type euclidien $\|\boldsymbol{f}\| = < \boldsymbol{f}, \boldsymbol{f} >^{1/2}$.

Supposons que $\boldsymbol{H}\boldsymbol{f} = \boldsymbol{g}$ et calculons l'erreur quadratique moyenne entre une estimation $\widehat{\boldsymbol{f}}$ linéaire quelconque $\widehat{\boldsymbol{f}} = \boldsymbol{G}^+\boldsymbol{g}$ de la solution et $\boldsymbol{f}$:

$$EQM = [\boldsymbol{f} - \widehat{\boldsymbol{f}}]^t[\boldsymbol{f} - \widehat{\boldsymbol{f}}] = \mathrm{Tr}\left\{[\boldsymbol{f} - \widehat{\boldsymbol{f}}][\boldsymbol{f} - \widehat{\boldsymbol{f}}]^t\right\} = \mathrm{Tr}\left\{\boldsymbol{f}\boldsymbol{f}^t[\boldsymbol{I} - \boldsymbol{H}^+\boldsymbol{H}]^t\right\}$$

Il est alors facile de vérifier que cette erreur est minimale si $\widehat{\boldsymbol{f}} = \boldsymbol{H}^+\boldsymbol{g}$ et qu'elle est donnée par

$$EQM = [\boldsymbol{f} - \widehat{\boldsymbol{f}}]^t[\boldsymbol{f} - \widehat{\boldsymbol{f}}] = \mathrm{Tr}\left\{\boldsymbol{f}\boldsymbol{f}^t[\boldsymbol{I} - \boldsymbol{H}^+\boldsymbol{H}]^t\right\}$$

l'estimation est parfaite si $\boldsymbol{H}^+\boldsymbol{H} = \boldsymbol{I}$.

En résumé, lorsque nous voulons calculer une solution au sens inverse généralisée de l'équation $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ avec $\boldsymbol{H}$ une matrice de dimensions $[M,N]$, quatre situations se présentent :

- si $M = N$ et $\mathrm{rang}\{\boldsymbol{H}\} = N$ alors $\boldsymbol{H}^+ = \boldsymbol{H}^{-1}$
- si $M > N$ et $\mathrm{rang}\{\boldsymbol{H}\} = N$ alors $\boldsymbol{H}^+ = (\boldsymbol{H}^t\boldsymbol{H})^{-1}\boldsymbol{H}^t$
- si $M < N$ et $\mathrm{rang}\{\boldsymbol{H}\} = M$ alors $\boldsymbol{H}^+ = \boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{H}^t)^{-1}$
- si $\mathrm{rang}\{\boldsymbol{H}\} = K < \inf(M,N)$ alors $\left\{\begin{array}{l}\text{Décomp. en valeurs singulières}\\ \text{Méthodes itératives}\end{array}\right.$

5.3 Algorithmes d'inversion généralisée

5.4 Décomposition en valeurs singulières

Considérons l'équation

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

et notons $\{\boldsymbol{v}_j, j = 1,\ldots,K\}$ les vecteurs propres de $\boldsymbol{H}^t\boldsymbol{H}$, $\{\boldsymbol{u}_i, i = 1,\ldots,K\}$ les vecteurs propres de $\boldsymbol{H}\boldsymbol{H}^t$ et λ_j^2 les valeurs propres correspondantes. Nous avons alors:

$$\begin{aligned}
\boldsymbol{H}^t\boldsymbol{H}\boldsymbol{v}_j &= \lambda_j^2\boldsymbol{v}_j, \quad j = 1,\ldots,K \\
\boldsymbol{H}\boldsymbol{H}^t\boldsymbol{u}_i &= \lambda_i^2\boldsymbol{u}_i, \quad i = 1,\ldots,K \\
\boldsymbol{H}\boldsymbol{v}_j &= \lambda_j\boldsymbol{u}_j, \quad j = 1,\ldots,K \\
\boldsymbol{H}^t\boldsymbol{u}_i &= \lambda_i\boldsymbol{v}_i, \quad i = 1,\ldots,K
\end{aligned}$$

$$\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^t$$

$$\boldsymbol{U} = \left(\boldsymbol{u}_1 \vdots \boldsymbol{u}_2 \vdots \cdots \vdots \boldsymbol{u}_K\right), \quad \boldsymbol{V} = \left(\boldsymbol{v}_1 \vdots \boldsymbol{v}_2 \vdots \cdots \vdots \boldsymbol{v}_K\right), \quad \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_K)$$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K, \quad K \leq \inf(M,N).$$

La solution inverse généralisée au sens du Moore Penrose est définie par

$$f^+ = \boldsymbol{H}^+\boldsymbol{g}, \quad \text{avec} \quad \boldsymbol{H}^+ = \boldsymbol{V}\boldsymbol{\Lambda}^+\boldsymbol{U}^t$$

et où

$$\Lambda^+ = \mathrm{diag}\{\alpha_i\}, \quad \begin{cases} \alpha_i = \frac{1}{\lambda_i} & \text{si} \quad \lambda_i \neq 0 \\ \alpha_i = 0 & \text{si} \quad \lambda_i \simeq 0 \end{cases}$$

On a aussi les relations suivantes :

$$\mathcal{K}\mathrm{er}\,(\boldsymbol{H}) = \boldsymbol{k}(\boldsymbol{z}) = \boldsymbol{V}(\boldsymbol{I} - \Lambda^{+}\Lambda)\boldsymbol{z}, \quad \forall \boldsymbol{z} \in \mathbf{R}^{n}$$

Les composantes d'indices 0 à K des vecteurs $\boldsymbol{k}(\boldsymbol{z})$ sont nulles. On peut alors écrire :

$$\boldsymbol{k}(\boldsymbol{z}) = \sum_{k=K+1}^{N} z_k \boldsymbol{v}_k \quad \forall \boldsymbol{z} \in \mathbb{R}^{n}$$

et définir ainsi l'ensemble S des solutions de l'équation $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ par

$$S = \{\boldsymbol{f} \ : \boldsymbol{f} = \boldsymbol{f}^{+} + \boldsymbol{k}(\boldsymbol{z}) \quad \forall \boldsymbol{z} \in \mathbf{R}^{n}\}$$

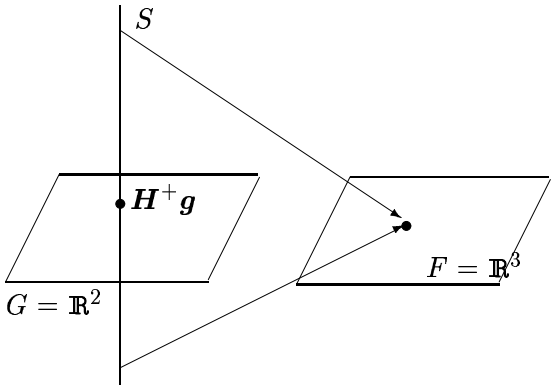


FIG. 5.1 – *Inversion généralisée et espace des solutions.*

5.4.1 Méthode de Bialy

Cette méthode est basée sur le théorème suivant :

Théorème 1 *Soit* $\boldsymbol{H} : H \mapsto H$ *un opérateur borné et non négatif. Soit* $\boldsymbol{P}$ *l'opérateur de projection orthogonale sur* $\mathcal{K}er(\boldsymbol{H})$*. Alors, s'il existe une solution pour* $\boldsymbol{Hf} = \boldsymbol{g}$*, pour tout* $\boldsymbol{g} \in H$,$\boldsymbol{f}_0 \in H$ *la suite des itérations*

$$\boldsymbol{f}_{n+1} = \boldsymbol{f}_n + \alpha(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}_n), \quad \textit{avec } 0 < \alpha < 2/\|\boldsymbol{H}\|$$

converge vers $P\boldsymbol{f}_0 + \boldsymbol{f}_{IG}$*, où* $\boldsymbol{f}_{IG}$ *est la solution IG de* $\boldsymbol{Hf} = \boldsymbol{g}$*.*

- *si* $\boldsymbol{f}_0 = \boldsymbol{0}$*, alors* $\lim_{n\to\infty} \boldsymbol{f}_n = \boldsymbol{f}_{IG}$
- $\boldsymbol{f}_{IG}$ *est la solution de norme minimale de* $\boldsymbol{H}^t\boldsymbol{Hf} = \boldsymbol{H}^t\boldsymbol{g}$*.*

5.4.2 Méthode de Landweber

Cette méthode, qui est une extension de la méthode précédente, est basée sur le théorème suivant :

Théorème 2 *Soient* $\boldsymbol{H} : H_1 \mapsto H_2$ *un opérateur borné,* $\boldsymbol{g} \in \mathcal{I}m(\boldsymbol{H}) + \mathcal{I}m(\boldsymbol{H})^{\perp}$*. Alors la suite des itérations*

$$\begin{aligned} &\boldsymbol{f}_0 = \boldsymbol{0} \\ &\boldsymbol{f}_{n+1} = \boldsymbol{f}_n + \alpha\boldsymbol{H}^t(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}_n), \quad \textit{avec } 0 < \alpha < 2/\|\boldsymbol{H}^t\boldsymbol{H}\| \end{aligned}$$

converge vers la solution IG de $\boldsymbol{Hf} = \boldsymbol{g}$*.*

Notons que cet algorithme peut être considéré comme un algorithme du type gradient pour la minimisation du critère des MC $J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2$. En effet, on peut remarquer que le gradient de ce critère est $\nabla J(\boldsymbol{f}) = -2\boldsymbol{H}^t(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f})$.

5.4.3 Méthode de Van-Cittert

Cette méthode est une extension de la méthode précédente et permet d'apporter des contraintes supplémentaires sur la solution. En effet, si $\boldsymbol{T}$ est un opérateur traduisant ces contraintes on peut envisager l'algorithme suivant :

$$\left\{ \begin{array}{l} \boldsymbol{f}_0 = \boldsymbol{0} \\ \boldsymbol{f}_{n+1} = \boldsymbol{f}_n + \alpha\boldsymbol{H}^t(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{T}\boldsymbol{f}_n), \quad \text{avec } 0 < \alpha < 2/\|\boldsymbol{T}^t\boldsymbol{H}^t\boldsymbol{H}\boldsymbol{T}\| \end{array} \right.$$

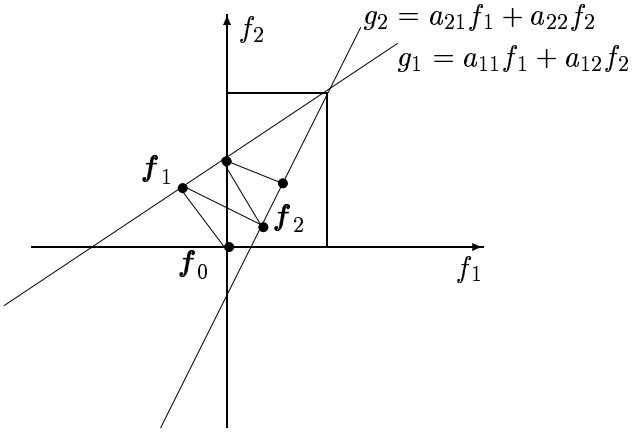
5.4.4 Méthode de Kaczmarz

L'idée de base dans cette méthode est d'utiliser les données g_i une par une. Ainsi, une mise en œuvre en ligne serait possible. Dans une relation linéaire $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ chaque équation $g_i = \sum_{j=1}^{N} h_{ij}\, f_j = \langle \boldsymbol{h}_i, \boldsymbol{f} \rangle$ définit un sous-espace dans l'espace des solutions. Cet algorithme recherche une solution à l'itération $k+1$ en projetant la solution à l'itération k sur le sous-espace créé par l'équation $g_k = \sum_{j=1}^{N} h_{kj}\, f_j$.

$$\begin{cases} \boldsymbol{f}_0 = \boldsymbol{0} \\ \boldsymbol{f}_{n+1} = \boldsymbol{f}_n + \frac{g_i - \langle \boldsymbol{f}_n, \boldsymbol{h}_i \rangle}{\langle \boldsymbol{h}_i, \boldsymbol{h}_i \rangle} \boldsymbol{h}_i \quad i = 1,2,\ldots,M,1,2,\ldots,M,1,2 \end{cases}$$

où $\boldsymbol{h}_i$ est la $i^{\text{ème}}$ ligne de la matrice $\boldsymbol{H}$.

Pour bien comprendre l'essentiel de cet algorithme, considérons le cas d'une matrice $\boldsymbol{H}$ de dimensions (2×2). Dans ce cas, chaque équation décrit une ligne droite dans l'espace des solutions (f_1, f_2) et la figure ci-dessous montre alors le principe de cet algorithme.



Évidemment, lorsqu'on veut imposer d'autres contraintes, comme par exemple la positivité, on peut introduire au cours de ces itérations. Ainsi, si on note par $p_i(\boldsymbol{f}), i = 1, \cdots, m$ les m contraintes supplémentaires que l'on veut imposer à la solution, l'algorithme s'écrit :

$$\begin{cases} \boldsymbol{f}_0 = \boldsymbol{0} \\ \boldsymbol{f}_{n+1} = P(\boldsymbol{f}_n) = p_m(\cdots p_2(p_1(\boldsymbol{f}))\cdots) \\ p_i(\boldsymbol{f}) = \boldsymbol{f}_{(n} + \frac{g_i - \langle \boldsymbol{f}_n, \boldsymbol{h}_i \rangle}{\langle \boldsymbol{h}_i, \boldsymbol{h}_i \rangle} \boldsymbol{h}_i \quad i = 1,2,\ldots,M,1,2,\ldots,M,1,2 \end{cases}$$

Chapitre 6

Régularisation

La régularisation d'un problème **mal-posé** consiste à le transformer en un problème **bien-posé**, c'est-à-dire à définir **une solution unique** pour **toutes les mesures possibles** dans l'espace des observations, et à **assurer la stabilité** de cette solution vis-à-vis des **erreurs** sur ces mesures.

6.1 Définition d'un opérateur de régularisation

Un **régulariseur** de $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ est une famille d'opérateurs $\{\boldsymbol{R}_\lambda; \lambda \in \Lambda \subset \mathbf{R}^+\}$ tels que :

- $\forall \boldsymbol{f} \in F, \lim_{\lambda \to 0} \|\boldsymbol{R}_\lambda \boldsymbol{H}\boldsymbol{f} - \boldsymbol{f}\|^2 = 0$
- $\forall \lambda \in \Lambda, \boldsymbol{R}_\lambda$ est un opérateur continu de G dans F

$$\Downarrow$$

Solution Approchée

$$\boldsymbol{f}_\epsilon = \boldsymbol{R}_\lambda \boldsymbol{g}_\epsilon, \quad \text{avec} \quad \boldsymbol{g}_\epsilon = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{b}$$

$$\|\boldsymbol{f}_\epsilon - \boldsymbol{H}^{-1}\boldsymbol{g}\| \leq \underbrace{\|\boldsymbol{R}_\lambda \boldsymbol{g} - \boldsymbol{H}^{-1}\boldsymbol{g}\|}_{\text{erreur de régularisation}} + \underbrace{\|\boldsymbol{R}_\lambda(\boldsymbol{g}_\epsilon - \boldsymbol{g})\|}_{\text{erreur due au bruit}}$$

- Utilisation d'E.H.N.R.
- Décomposition TSVD
- Régularisation P.T.T.

6.2 Décomposition tronquée en valeurs singulières

Considérons l'équation

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

et notons par $\{\boldsymbol{v}_j, j = 1,\ldots,K\}$ les vecteurs propres de $\boldsymbol{H}^t\boldsymbol{H}$ et par $\{\boldsymbol{u}_i, i = 1,\ldots,K\}$ les vecteurs propres de $\boldsymbol{H}\boldsymbol{H}^t$ et par λ_j^2 les valeurs propres correspondantes. Nous avons alors :

$$\begin{aligned}
\boldsymbol{H}^t\boldsymbol{H}\boldsymbol{v}_j &= \lambda_j^2\boldsymbol{v}_j, \quad j = 1,\ldots,K \\
\boldsymbol{H}\boldsymbol{H}^t\boldsymbol{u}_i &= \lambda_i^2\boldsymbol{u}_i, \quad i = 1,\ldots,K \\
\boldsymbol{H}\boldsymbol{v}_j &= \lambda_j\boldsymbol{u}_j, \quad j = 1,\ldots,K \\
\boldsymbol{H}^t\boldsymbol{u}_i &= \lambda_i\boldsymbol{v}_i, \quad i = 1,\ldots,K
\end{aligned}$$

$$\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^t$$

$$\boldsymbol{U} = \left(\boldsymbol{u}_1 \vdots \boldsymbol{u}_2 \vdots \cdots \vdots \boldsymbol{u}_K\right), \quad \boldsymbol{V} = \left(\boldsymbol{v}_1 \vdots \boldsymbol{v}_2 \vdots \cdots \vdots \boldsymbol{v}_K\right), \quad \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_K)$$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K, \quad K \leq \inf(M,N)$$

- On décide qu'à partir du rang $r < K$, les valeurs singulières ordonnées ne sont plus significatives.
- Dans l'inversion, on utilise la pseudo-inverse $\boldsymbol{H}^+$.

$$\boldsymbol{H}^+ = \boldsymbol{V}\boldsymbol{\Lambda}^+\boldsymbol{U}^t$$

$$\Lambda^+ = \mathrm{diag}\{\alpha_i\}, \quad \begin{cases} \alpha_i = \frac{1}{\lambda_i} & \text{si} \quad \lambda_i \neq 0 \text{ (significatif)} \\ \alpha_i = 0 & \text{si} \quad \lambda_i \simeq 0 \text{ (non-significatif)} \end{cases}$$

$$\boldsymbol{g} + \delta\boldsymbol{g} = H(\boldsymbol{f} + \delta\boldsymbol{f}) \longrightarrow \frac{\|\delta\boldsymbol{f}\|}{\|\boldsymbol{f}\|} \leq \left(\frac{\lambda_{max}}{\lambda_{min}}\right)^2 \frac{\|\delta\boldsymbol{g}\|}{\|\boldsymbol{g}\|}$$

6.3 Régularisation par méthodes itératives

- Méthodes itératives du type inversion généralisée (BIALY, LANDWEBER,...)
- établir une règle d'arrêt fondée sur l'évolution du résidu

$$\|\boldsymbol{H}\widehat{\boldsymbol{f}}^k - \boldsymbol{g}_\epsilon\| > \delta \text{ pour } i < k$$

$$\|\boldsymbol{H}\widehat{\boldsymbol{f}}^k - \boldsymbol{g}_\epsilon\| < \delta \text{ pour } i > k$$

- $\frac{1}{k}$ joue le rôle du paramètre de régularisation.

6.4 Introduction de contraintes supplémentaires

$\boldsymbol{C}$ un opérateur de contrainte

$$\boldsymbol{f} = \boldsymbol{C}\boldsymbol{f} \quad \text{ssi } \boldsymbol{f} \text{ satisfait la contrainte}$$

Exemple : Positivité

$$x_j = \begin{cases} x_j & \text{si } x_j > 0 \\ 0 & \text{sinon} \end{cases}, \quad j = 1, \cdots, n$$

Le problème initial $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ devient alors

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{C}\boldsymbol{f}$$

L'itération générale est modifiée en remplaçant l'opérateur $\boldsymbol{H}$ par $\boldsymbol{HC}$.

6.5 Régularisation avec *a priori* de douceur

Tikhonov (régularisation continue) :

$$g_i = \int f(t)h_i(t)\,\mathrm{d}t + b_i$$

$$\begin{aligned} J(f) &= Q(f) + \lambda\Omega(f) \\ &= \sum_i \left| g_i - \int f(t)h_i(t)\,\mathrm{d}t \right|^2 + \lambda \int \sum_k p_k(t) \left| \frac{\partial^k f}{\partial t^k} \right|^2 \mathrm{d}t, \quad \lambda > 0, \end{aligned}$$

où les $p_k(t)$ sont des fonctions positives, continues, et continûment dérivables.

Philipps et Twomey (régularisation discrète) :

$$J_\lambda(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda\Omega(\boldsymbol{f}) \quad \lambda > 0$$

Exemple : $k = 1$,$p_k(t) = 1$

$$J_\lambda(f) = \sum_i \left| g_i - \int f(t)h_i(t)\,\mathrm{d}t \right|^2 + \lambda \int |f'(t)|^2\,\mathrm{d}t$$

Cas discret :

$$J_\lambda(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{D}_k\boldsymbol{f}\|^2 = [\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}]^t[\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}] + \lambda[\boldsymbol{D}_k\boldsymbol{f}]^t[\boldsymbol{D}_k\boldsymbol{f}]$$

avec

$$\boldsymbol{D}_0 = \boldsymbol{I}, \quad \boldsymbol{D}_1 = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ -1 & 1 & \ddots & & \vdots \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & & & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}, \quad \boldsymbol{D}_2 = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ -2 & 1 & \ddots & & & \vdots \\ 1 & -2 & 1 & \ddots & & \vdots \\ 0 & 1 & -2 & 1 & & \vdots \\ \vdots & \ddots & & & & \vdots \\ & & \ddots & & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}$$

$$\nabla J_\lambda = 2\boldsymbol{H}^t[\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}]^t + 2\lambda\boldsymbol{D}^t\boldsymbol{D}\boldsymbol{f} = 0 \longrightarrow$$

$$[\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{D}^t\boldsymbol{D}]\widehat{\boldsymbol{f}} = \boldsymbol{H}^t\boldsymbol{g} \longrightarrow \widehat{\boldsymbol{f}} = [\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{D}^t\boldsymbol{D}]^{-1}\boldsymbol{H}^t\boldsymbol{g}$$

6.6 Algorithmes de régularisation

$$\text{minimiser} \quad J(\boldsymbol{f}) = Q(\boldsymbol{f}) + \lambda\Omega(\boldsymbol{f})$$

$$\text{avec} \quad Q(\boldsymbol{f}) = [\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}]^t[\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}]$$

$$=$$

$$\text{minimiser } \Omega(\boldsymbol{f}) \text{ sous la contrainte}$$

$$e = [\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}]^t[\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}] < \epsilon$$

Information *a priori*:

- Douceur

$$\Omega(\boldsymbol{f}) = [\boldsymbol{D}\boldsymbol{f}]^t[\boldsymbol{D}\boldsymbol{f}] = \boldsymbol{f}^t\boldsymbol{D}^t\boldsymbol{D}\boldsymbol{f}$$

$$\widehat{\boldsymbol{f}} = [\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{P}]^{-1}\boldsymbol{H}^t\boldsymbol{g} \quad \text{avec} \quad \boldsymbol{P} = \boldsymbol{D}^t\boldsymbol{D}$$

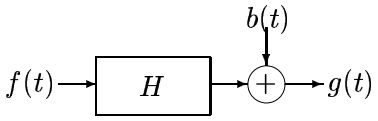
- Positivité $\Omega(\boldsymbol{f})$ = fonction non quadratique de $\boldsymbol{f}$
  Pas de solution explicite

Difficulté : Le calcul de l'expression $[\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{D}^t\boldsymbol{D}]^{-1}$

- approximation circulante
- méthodes récursives
- méthodes itératives

6.6.1 Méthode de Hunt (approximation circulante)

Pour illustrer cette méthode considérons le problème de la déconvolution des signaux :



$$g(t) = \int f(t')h(t-t')\,\mathrm{d}t' + b(t) = \int h(t')f(t-t')\,\mathrm{d}t' + b(t)$$

Nous avons vu que lorsqu'on discrétise cette équation on obtient

$$g(n) = \sum_{i=0}^{p} h(i)f(n-i) + b(n)$$

ou sous forme matricielle :

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{b}$$

où $\boldsymbol{H}$ est une matrice de TOEPLITZ. Notons que si on définit $\boldsymbol{h} = [h_0, \cdots, h_p]$ on peut noter que $\boldsymbol{H}\boldsymbol{f} = \mathrm{conv}(\boldsymbol{h},\boldsymbol{f})$.

La solution régularisée au sens de PTT est définie par

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

avec

$$\begin{aligned} J(\boldsymbol{f}) &= Q(\boldsymbol{f}) + \lambda\Omega(\boldsymbol{f}) \\ Q(\boldsymbol{f}) &= [\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}]^t[\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}] = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 \\ \Omega(\boldsymbol{f}) &= [\boldsymbol{D}\boldsymbol{f}]^t[\boldsymbol{D}\boldsymbol{f}] = \|\boldsymbol{D}\boldsymbol{f}\|^2 \end{aligned}$$

où $\boldsymbol{D}$ est une matrice correspondant à une opération de filtrage linéaire, donc de convolution par une réponse impulsionnelle, par exemple :

$$\boldsymbol{d} = [1, -2, 1], \quad \boldsymbol{D}\boldsymbol{f} = \mathrm{conv}(\boldsymbol{d},\boldsymbol{f})$$

et on a

$$\Omega(\boldsymbol{f}) = \sum_{j=1}^{N} (f(i+1) - 2f(i) + f(i-1))^2 = \boldsymbol{f}^t\boldsymbol{D}^t\boldsymbol{D}\boldsymbol{f}.$$

La solution régularisée a pour expression $\widehat{\boldsymbol{f}} = [\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{D}^t\boldsymbol{D}]^{-1}\boldsymbol{H}^t\boldsymbol{g}$ et la principale difficulté est le coût de calcul de l'inverse de la matrice $[\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{D}^t\boldsymbol{D}]$. Cette matrice est symétrique, mais, dans le cas général n'a pas d'autres propriétés remarquables. Dans le cas particulier où $\boldsymbol{H}$ et $\boldsymbol{D}$ sont des matrices de TOEPLITZ, cette matrice est aussi TOEPLITZ.

L'idée de base est d'étendre les vecteurs $\boldsymbol{f}$, $\boldsymbol{h}$ et $\boldsymbol{g}$ par des zéros :

$$\boldsymbol{f}_e(i) = \begin{cases} f(i) & i = 1, \ldots, Nf \\ 0 & i = Nf+1, \ldots, P \geq Nf + Nh - 1 \end{cases}$$

$$g_e(i) = \begin{cases} g(i) & i = 1, \ldots, Ng \\ 0 & i = Ng+1, \ldots, P \end{cases}$$

$$h_e(i) = \begin{cases} h(i) & i = 1,\ldots,Nh \\ 0 & i = Nh+1,\ldots,P \end{cases}$$

pour obtenir une relation $\boldsymbol{g}_e = \boldsymbol{H}_e \boldsymbol{f}_e$ avec $\boldsymbol{H}_e$ une matrice circulante que l'on peut diagonaliser par transformée de FOURIER discrète (TFD). En effet, pour une matrice circulante $\boldsymbol{H}_e$ on a

$$\boldsymbol{H}_e = \boldsymbol{F}\boldsymbol{\Lambda}\boldsymbol{F}^{-1}, \quad \boldsymbol{\Lambda} = \text{diag}[\lambda_1,\ldots,\lambda_P]$$

avec

$$\boldsymbol{F}[k,l] = \exp\left[j2\pi\frac{kl}{P}\right] \qquad F^{-1} = \frac{1}{P}\exp\left[-j2\pi\frac{kl}{P}\right]$$

$$[\lambda_1,\ldots,\lambda_P] = \text{TFD}\ [h_1,\ldots,h_{Nh},0,\ldots,0]$$

De même, on peut étendre le vecteur $\boldsymbol{d} = [1, -2, 1]$

$$d_e(i) = \begin{cases} d(i) & i = 1,\ldots,3 \\ 0 & i = 4,\ldots,P \end{cases}$$

de telle sorte que l'on puisse écrire les relations suivantes :

$$\widehat{\boldsymbol{f}} = [\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{D}^t\boldsymbol{D}]^{-1}\boldsymbol{H}^t\boldsymbol{g}$$

$$\boldsymbol{F}\widehat{\boldsymbol{f}}_e = [\Lambda_h^t\Lambda_h + \lambda\Lambda_c^t\Lambda_c]^{-1}\Lambda_h^t\boldsymbol{F}\boldsymbol{g}$$

$$\text{TFD}\ \{\boldsymbol{f}_e\} = [\Lambda_h^t\Lambda_h + \lambda\Lambda_c^t\Lambda_c]^{-1}\Lambda_h^t\text{TFD}\ \{\boldsymbol{g}\}$$

$$\widehat{F}(\omega) = \frac{1}{H(\omega)}\frac{|H(\omega)|^2}{|H(\omega)|^2 + \lambda|C(\omega)|^2}G(\omega)$$

Lien avec le filtre de Wiener :

$$C(\omega) = S_{bb}(\omega)/S_{ff}(\omega)$$

Restauration d'image :

$\boldsymbol{D}$ matrice de convolution par une réponse impulsionnelle

$$H_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

$$\begin{aligned} \Omega(\boldsymbol{f}) = \sum\sum \quad & (f(i+1,j) + f(i-1,j) \\ & + f(i,j+1) + f(i,j+1) - 4f(i,j))^2 \end{aligned}$$

$$f_e(k,l) = \begin{cases} f(k,l) & k = 1,\ldots,K \quad l = 1,\ldots,L \\ 0 & k = K+1,\ldots,P \quad l = L+1,\ldots,P \end{cases}$$

$$F(\omega_x,\omega_y) = \frac{1}{H(\omega_x,\omega_y)} \frac{|H(\omega_x,\omega_y)|^2}{|H(\omega_x,\omega_y)|^2 + \lambda |C(\omega_x,\omega_y)|^2)} G(\omega_x,\omega_y)$$

6.6.2 Méthodes itératives

Nous avons vu que la solution régularisée est définie comme l'argument qui minimise un critère $J(\boldsymbol{f})$:

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{J(\boldsymbol{f}) = Q(\boldsymbol{f}) + \lambda\Omega(\boldsymbol{f})\}$$

Lorsque le critère est unimodal, on peut alors utiliser les algorithmes d'optimisation classique du type gradient pour calculer la solution. Notant par :

$$\boldsymbol{g} = \nabla J(\boldsymbol{f}) = \left[\frac{\partial J}{\partial f_1}, \cdots, \frac{\partial J}{\partial f_n}\right]^t \quad \text{le gradient de } J, \quad \boldsymbol{g}_k = \nabla J(\boldsymbol{f}_k)$$

et par

$$\boldsymbol{H} = \nabla^2 J(\boldsymbol{f}) = \left\{\frac{\partial^2 J}{\partial f_i \partial f_j}\right\} \quad \text{la matrice Hessienne de } J, \quad \boldsymbol{H}_k = \nabla^2 J(\boldsymbol{f}_k)$$

La majorité des méthodes d'optimisation locale calcule la solution d'une manière itérative suivante :

$$\begin{aligned} \boldsymbol{f}_{k+1} &= \boldsymbol{f}_k - \alpha_k \boldsymbol{D}_k \nabla J(\boldsymbol{f}_k) \\ &= \boldsymbol{f}_k - \alpha_k \boldsymbol{D}_k \boldsymbol{g}_k \\ &= \boldsymbol{f}_k + \alpha_k \boldsymbol{d}_k \end{aligned}$$

Suivant les différents choix pour α_k et $\boldsymbol{D}_k$ on peut distinguer les algorithmes suivants :

1. **Méthodes de premier ordre ou de gradient**
   - gradient à pas fixe : $\boldsymbol{D}_k = \boldsymbol{I}$, $\alpha_k = \alpha$ et $\boldsymbol{d}_k = -\boldsymbol{g}_k$
   - gradient à pas variable : $\boldsymbol{D}_k = \boldsymbol{I}$ et

$$\alpha_k = -\frac{\boldsymbol{g}^{(k)^t}\boldsymbol{g}_k}{\boldsymbol{g}^{(k)^t}\boldsymbol{H}_k\boldsymbol{g}_k} = \frac{\|\boldsymbol{g}_k\|^2}{\|\boldsymbol{g}_k\|_H^2}$$

   - gradient conjugué :

$$\boldsymbol{f}_{k+1} = \boldsymbol{f}_k + \alpha_k \boldsymbol{d}_k \quad \alpha_k = -\frac{\boldsymbol{d}^{(k)^t}\boldsymbol{g}_k}{\boldsymbol{d}^{(k)^t}\boldsymbol{H}_k\boldsymbol{d}_k}$$

$$\boldsymbol{d}_{k+1} = \boldsymbol{d}_k + \beta_k \boldsymbol{g}_k \quad \beta_k = -\frac{\boldsymbol{g}^{(k)^t}\boldsymbol{g}_k}{\boldsymbol{g}^{(k-1)^t}\boldsymbol{g}^{(k-1)}}$$

2. **Méthodes de second ordre**
   - Méthode de Newton : $\boldsymbol{D}_k = [\boldsymbol{H}_k]^{-1}$

$$\boldsymbol{f}_{k+1} = \boldsymbol{f}_k - \alpha_k [\boldsymbol{H}_k]^{-1} \boldsymbol{g}_k$$

   - Méthode de Newton approchée : $\boldsymbol{D}_k = \text{diag}\{d_{1_k}, \cdots, d_{n_k}\}$ avec

$$d_{i_k} = \boldsymbol{D}_{ii} = \left(\frac{\partial^2 J}{\partial {f_i}^2} \right)^{-1}$$

     les éléments diagonaux de la matrice Hessienne.
   - Méthodes de Newton modifiées :

$$\boldsymbol{D}_k = \boldsymbol{D}_0 = [\boldsymbol{H}_0]^{-1} = \boldsymbol{D}$$

     ou

$$\boldsymbol{D}_k = \boldsymbol{D}_0 = \text{diag}\{d_{1_0}, \cdots, d_{n_0}\} \quad \text{avec} \quad d_{i_0} = \left(\frac{\partial^2 J}{\partial {f_i}^2}(\boldsymbol{f}_0) \right)^{-1}$$

- Avantage : $\Omega(\boldsymbol{f})$ peut être quelconque
- Inconvénient : Coût de calcul

6.6.3 Méthodes récursives

Considérons l'équation $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{b}$ et sa solution régularisée :

$$\widehat{\boldsymbol{f}} = [\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{D}]^{-1}\boldsymbol{H}^t\boldsymbol{g}$$

Notons par $\boldsymbol{H}_i$ la matrice $\boldsymbol{H}$ lorsqu'on a observé les données jusqu'à l'indice i ; $\boldsymbol{g}_i$ et par $\boldsymbol{H}_{i+1}$ la matrice $\boldsymbol{H}$ lorsqu'on a observé les données jusqu'à l'indice i ; $\boldsymbol{g}_{i+1}$. On a alors

$$\boldsymbol{g}_i = \begin{pmatrix} g_1 \\ \vdots \\ g_i \end{pmatrix} = \boldsymbol{H}_i\boldsymbol{f} + \boldsymbol{b},$$

$$\boldsymbol{g}_{i+1} = \begin{pmatrix} g_1 \\ \vdots \\ g_i \\ g_{i+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{g}_i \\ \\ \cdots \\ g_{i+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{H}_i \\ \\ \cdots\cdots \\ \boldsymbol{h}_{i+1}^t \end{pmatrix} \boldsymbol{f} + \boldsymbol{b}$$

L'idée de base consiste à exprimer la solution $\boldsymbol{f}_{i+1}$

$$\begin{aligned} \boldsymbol{f}_{i+1} &= (\boldsymbol{H}_{i+1}^t\boldsymbol{H}_{i+1} + \lambda\boldsymbol{D})^{-1}\boldsymbol{H}_{i+1}^t\boldsymbol{g}_{i+1} \\ &= \left[\begin{pmatrix} \boldsymbol{H}_i^t & \vdots & \boldsymbol{h}_{i+1} \end{pmatrix} \begin{pmatrix} \boldsymbol{H}_i \\ \cdots \\ \boldsymbol{h}_{i+1}^t \end{pmatrix} + \lambda\boldsymbol{D} \right]^{-1} \begin{pmatrix} \boldsymbol{H}_i^t & \vdots & \boldsymbol{h}_{i+1} \end{pmatrix} \begin{pmatrix} \boldsymbol{g}_i \\ \cdots \\ g_{i+1} \end{pmatrix} \end{aligned}$$

en fonction de $\boldsymbol{f}_i$:

$$\boldsymbol{f}_i = (\boldsymbol{H}_i^t\boldsymbol{H}_i + \lambda\boldsymbol{D})^{-1}\boldsymbol{H}_i^t\boldsymbol{g}_i$$

Utilisant la lemme d'inversion des matrices on obtient facilement :

$$\boldsymbol{f}_{i+1} = (\boldsymbol{H}_i^t\boldsymbol{H}_i + \boldsymbol{h}_{i+1}\boldsymbol{h}_{i+1}^t + \lambda\boldsymbol{D})^{-1}(\boldsymbol{H}_i^t\boldsymbol{g}_i + \boldsymbol{h}_{i+1}g_{i+1})$$

qui peut être réécrite d'une manière plus familière en notant :

$$\boldsymbol{P}_i = (\boldsymbol{H}_i^t\boldsymbol{H}_i + \lambda\boldsymbol{D})^{-1}$$

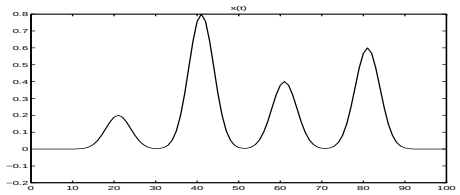
$$\boldsymbol{P}_{i+1}^t = \boldsymbol{P}_i^t + \boldsymbol{H}_{i+1}\boldsymbol{H}_{i+1}^t$$

$$\Downarrow$$

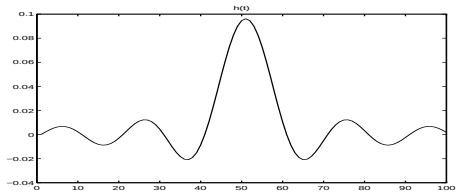
$$\boldsymbol{f}_{i+1} = \boldsymbol{f}_i + \boldsymbol{P}_{i+1}\boldsymbol{H}_{i+1}(\boldsymbol{g}_{i+1} - \boldsymbol{H}_{i+1}\boldsymbol{f}_i)$$

$$\boldsymbol{P}_{i+1} = \boldsymbol{P}_i - \boldsymbol{P}_i\boldsymbol{H}_{i+1}(\boldsymbol{H}_{i+1}^t\boldsymbol{P}_i\boldsymbol{H}_{i+1} + \lambda)^{-1}\boldsymbol{H}_{i+1}^t\boldsymbol{P}_i$$

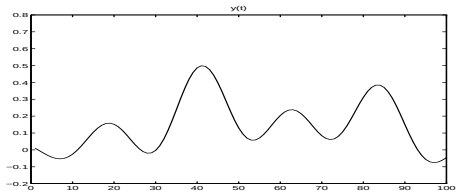
Filtre de Kalman



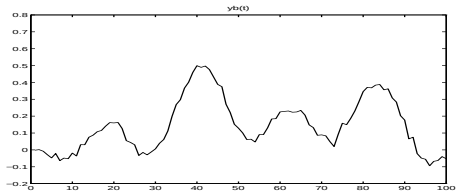
Signal d'entrée $f(t)$



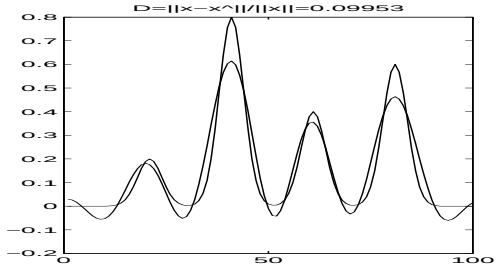
La réponse impulsionnelle $h(t)$



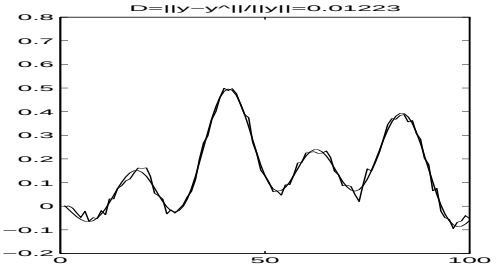
Signal de sortie sans bruit



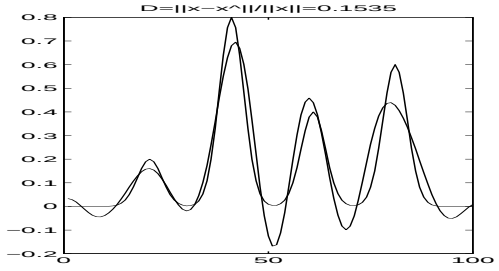
Signal de sortie $g(t)$ avec bruit RSB=20 dB.



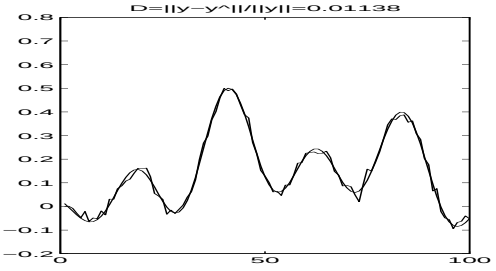
$\lambda = 1 \quad \widehat{f}(t)$



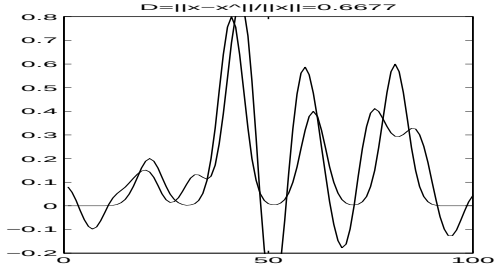
$\widehat{g}(t)$



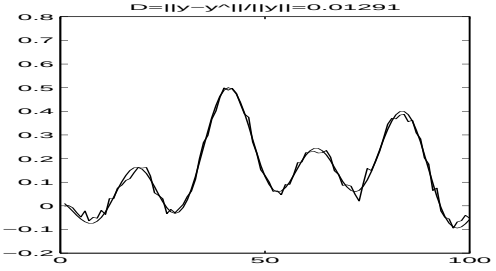
$\lambda = .1 \quad \widehat{f}(t)$



$\widehat{g}(t)$

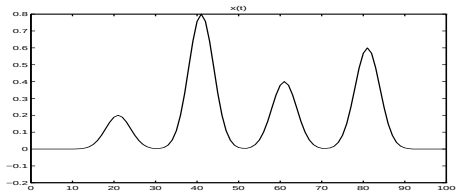


$\lambda = .01 \quad \widehat{f}(t)$

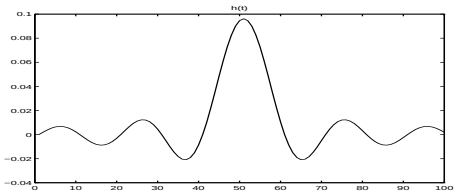


$\widehat{g}(t)$

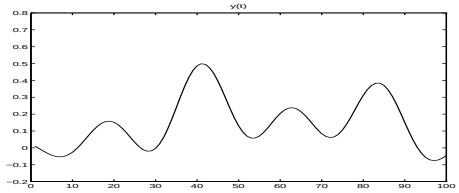
FIG. 6.1 – *Déconvolution par la méthode de Hunt : Exemple 1*



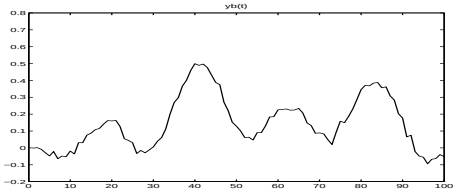
Signal d'entrée $f(t)$



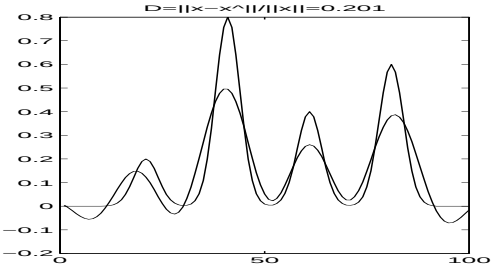
La réponse impulsionnelle $h(t)$



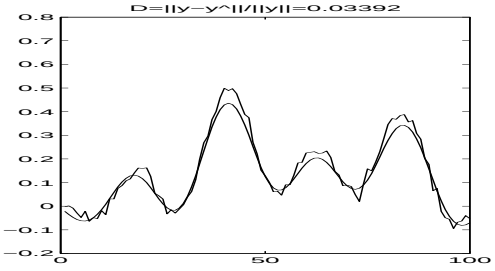
Signal de sortie sans bruit



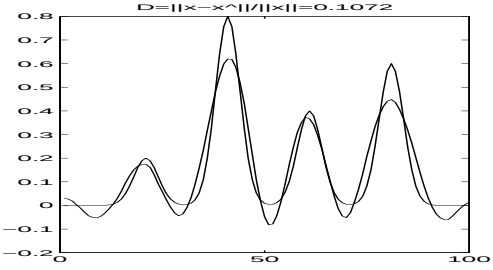
Signal de sortie $g(t)$ avec bruit RSB=20 dB.



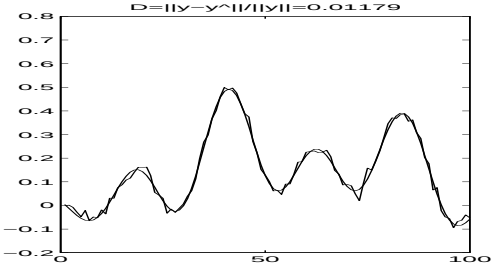
$\lambda = 1 \quad \widehat{f}(t)$



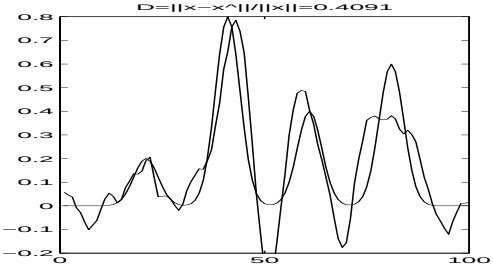
$\widehat{g}(t)$



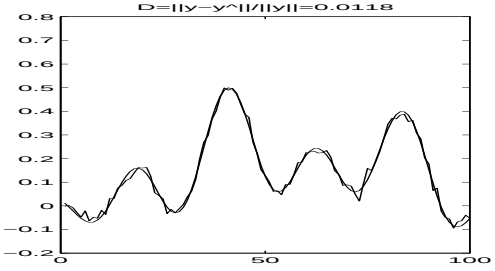
$\lambda = .1 \quad \widehat{f}(t)$



$\widehat{g}(t)$

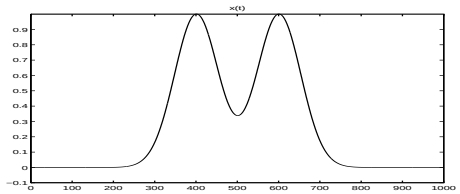


$\lambda = .01 \quad \widehat{f}(t)$

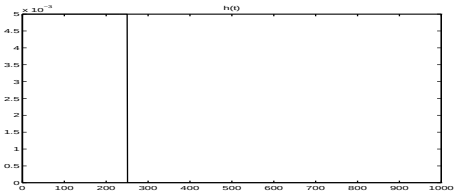


$\widehat{g}(t)$

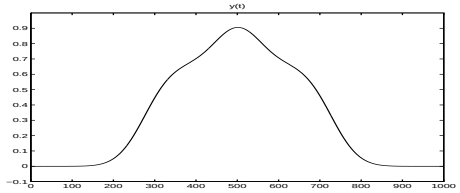
FIG. 6.2 – *Déconvolution par filtre de Wiener : Exemple 1*



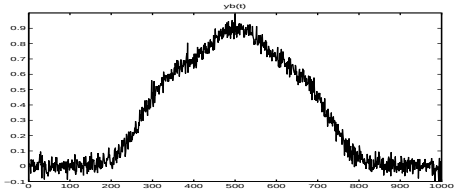
Signal d'entrée $f(t)$



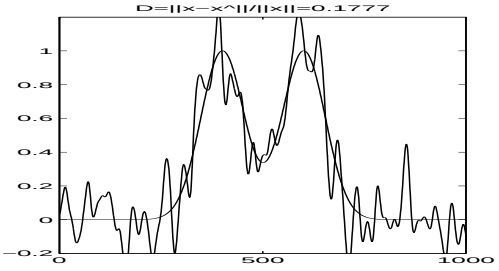
La réponse impulsionnelle $h(t)$



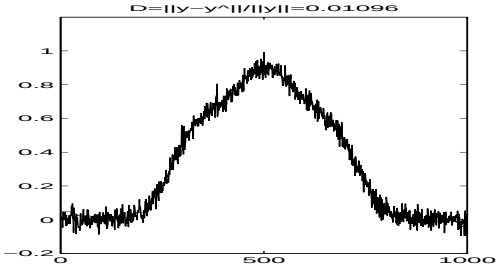
Signal de sortie sans bruit



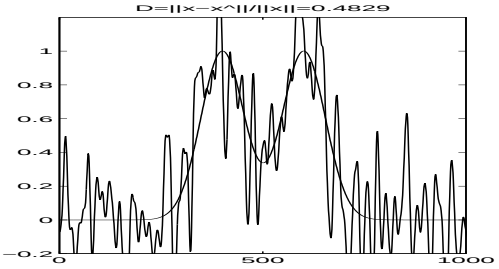
Signal de sortie $g(t)$ avec bruit RSB=20 dB.



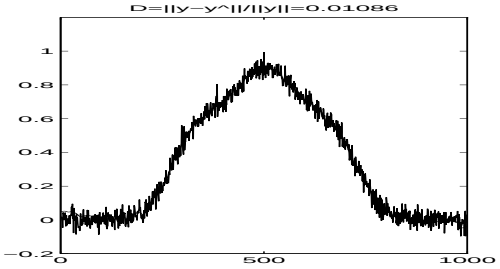
$\lambda = 1 \quad \widehat{f}(t)$



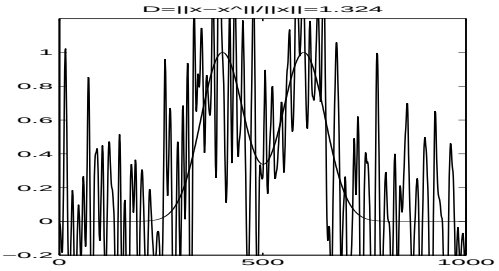
$\widehat{g}(t)$



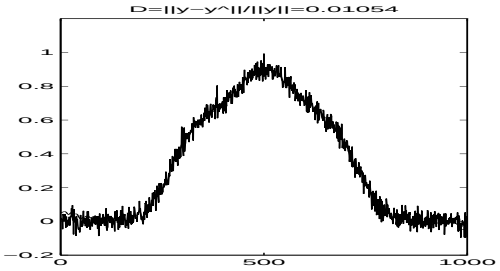
$\lambda = .1 \quad \widehat{f}(t)$



$\widehat{g}(t)$

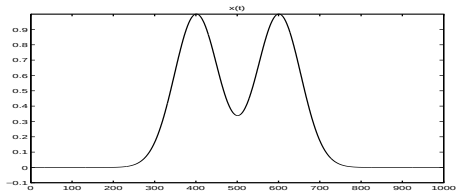


$\lambda = .01 \quad \widehat{f}(t)$

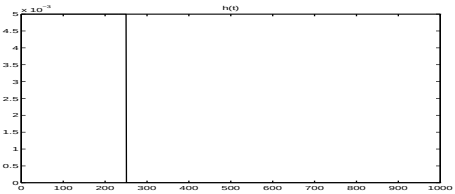


$\widehat{g}(t)$

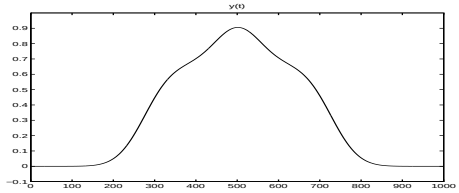
FIG. 6.3 – *Déconvolution par la méthode de Hunt : Exemple 2*



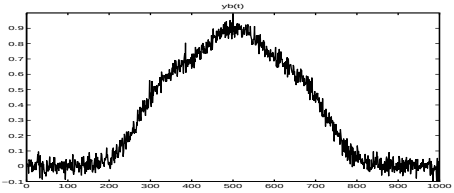
Signal d'entrée $f(t)$



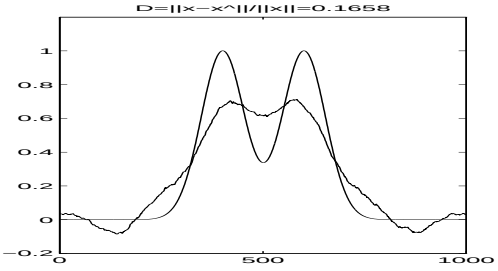
La réponse impulsionnelle $h(t)$



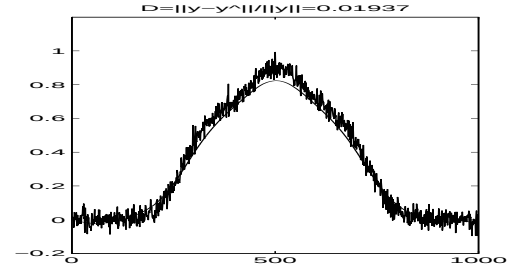
Signal de sortie sans bruit



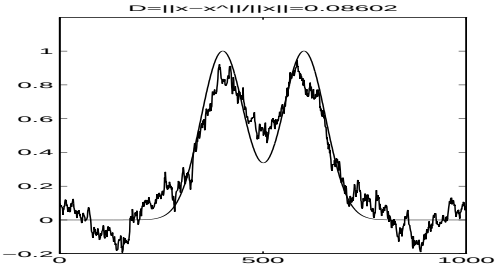
Signal de sortie $g(t)$ avec bruit RSB=20 dB.



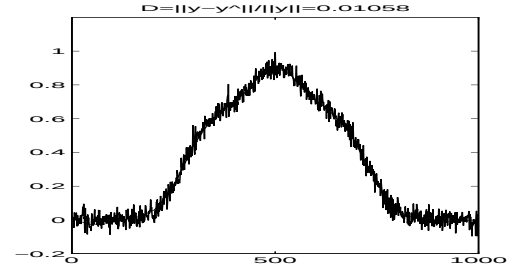
$\lambda = 1 \quad \widehat{f}(t)$



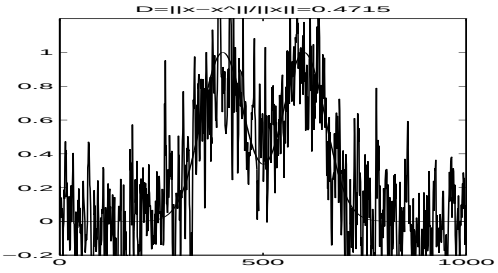
$\widehat{g}(t)$



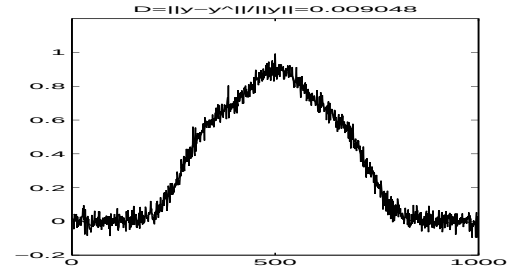
$\lambda = .1 \quad \widehat{f}(t)$



$\widehat{g}(t)$



$\lambda = .01 \quad \widehat{f}(t)$



$\widehat{g}(t)$

FIG. 6.4 – *Déconvolution par filtre de Wiener : Exemple 2*

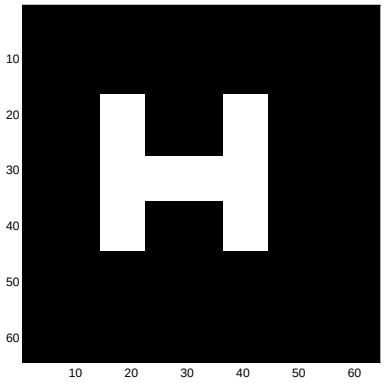
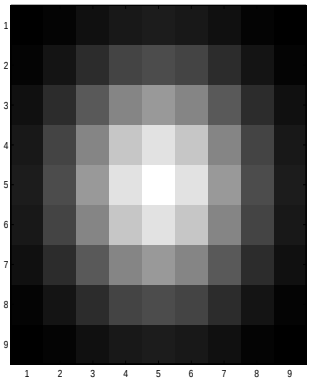


Image d'entrée $f(x,y)$



La réponse impulsionnelle $h(x,y)$

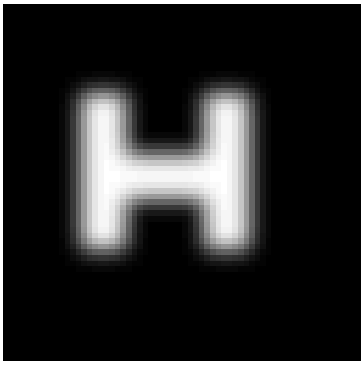


Image de sortie sans bruit

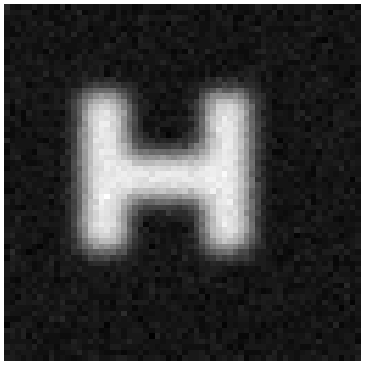
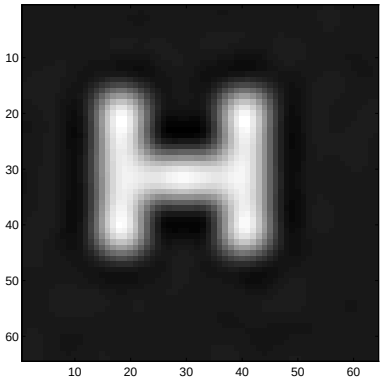
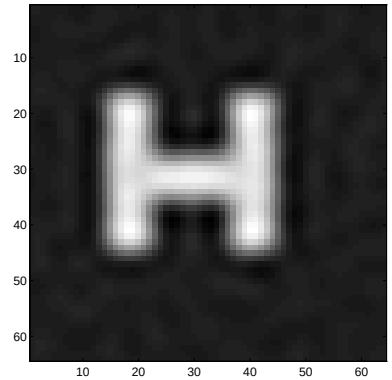


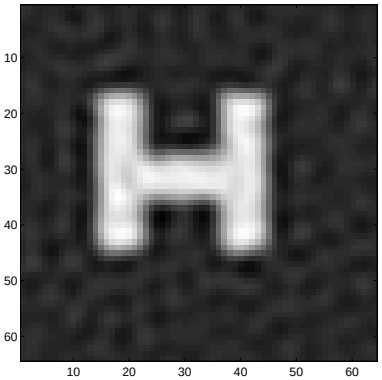
Image de sortie $g(x,y)$ avec bruit RSB=20 dB.



$\lambda = 1 \quad \widehat{f}(x,y)$



$\lambda = .1 \quad \widehat{f}(x,y)$



$\lambda = .01 \quad \widehat{f}(x,y)$

FIG. 6.5 – *Restauration par la méthode de Hunt : Exemple 1*

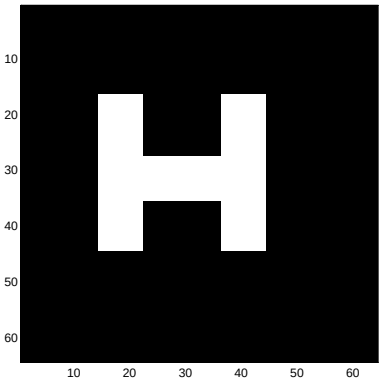
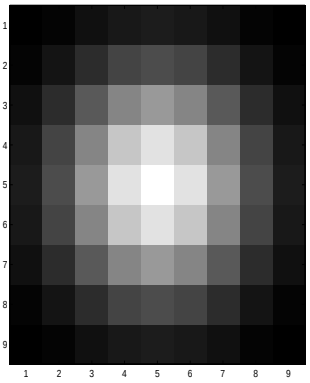


Image d'entrée $f(x,y)$



La réponse impulsionnelle $h(x,y)$

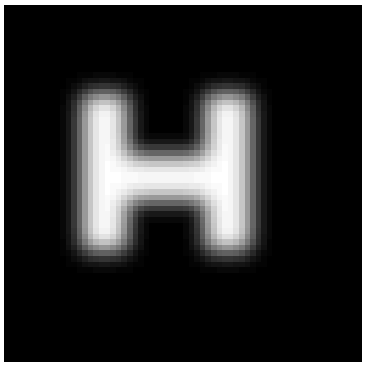


Image de sortie sans bruit

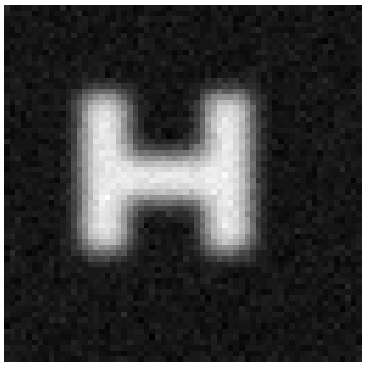
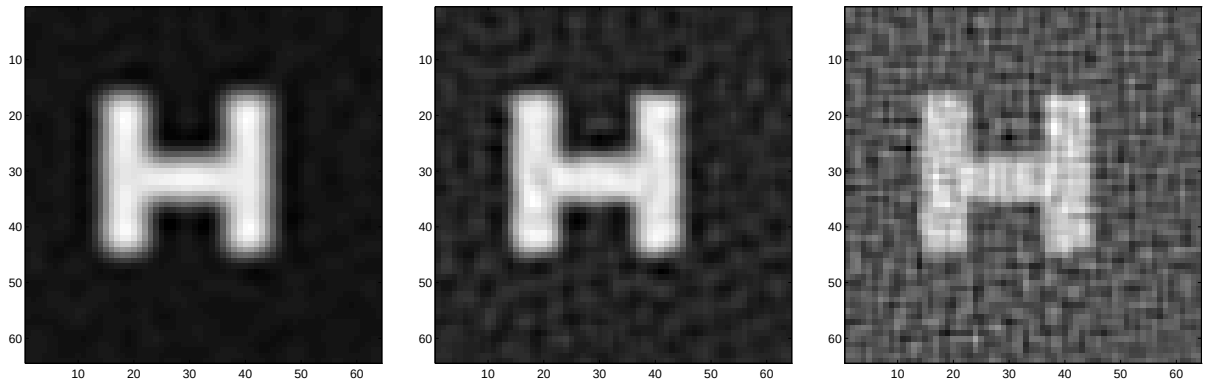


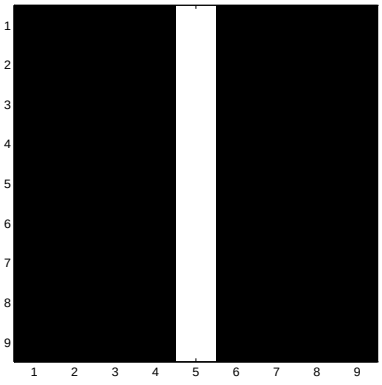
Image de sortie $g(x,y)$ avec bruit RSB=20 dB.



$\lambda = 1$ $\widehat{f}(x,y)$ $\lambda = .1$ $\widehat{f}(x,y)$ $\lambda = .01$ $\widehat{f}(x,y)$

FIG. 6.6 – *Restauration par filtre de Wiener : Exemple 1*

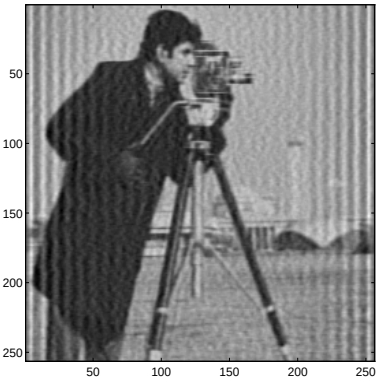
Image d'entrée $f(x,y)$



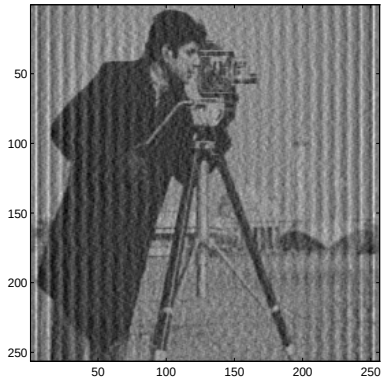
La réponse impulsionnelle $h(x,y)$

Image de sortie sans bruit

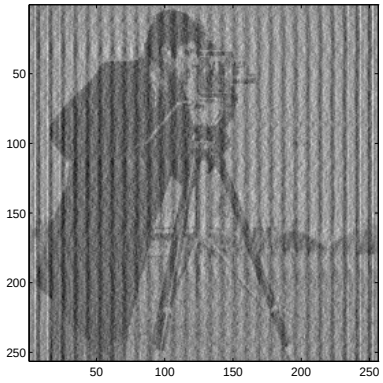
Image de sortie $g(x,y)$ avec bruit RSB=20 dB.



$\lambda = 1 \quad \widehat{f}(x,y)$



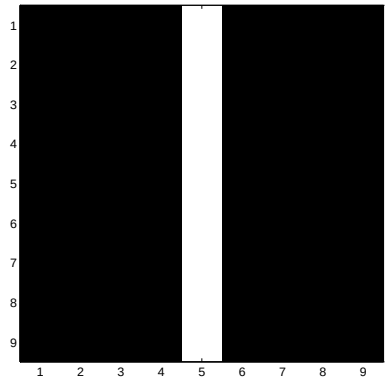
$\lambda = .1 \quad \widehat{f}(x,y)$



$\lambda = .01 \quad \widehat{f}(x,y)$

FIG. 6.7 – *Restauration par la méthode de Hunt : Exemple 2*

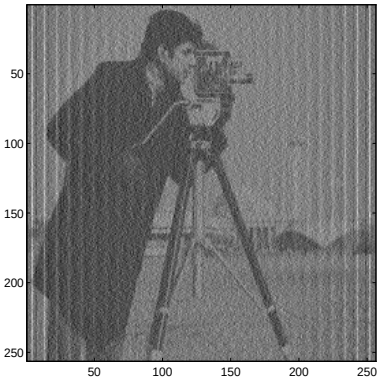
Image d'entrée $f(x,y)$



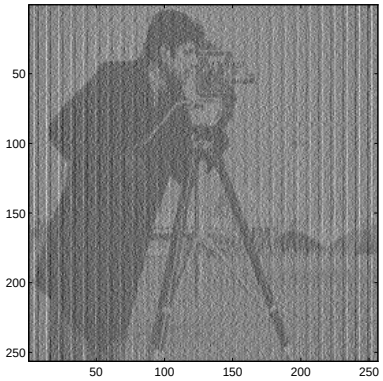
La réponse impulsionnelle $h(x,y)$

Image de sortie sans bruit

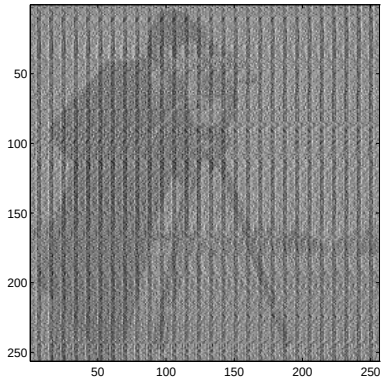
Image de sortie $g(x,y)$ avec bruit RSB=20 dB.



$\lambda = 1$ $\widehat{f}(x,y)$



$\lambda = .1$ $\widehat{f}(x,y)$



$\lambda = .01$ $\widehat{f}(x,y)$

FIG. 6.8 – *Restauration par filtre de Wiener : Exemple 2*

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 quatres 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 utilisé 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 $\mathcal{R}$:

$$p(r,\phi) = \iint f(x,y)\delta(r - x\cos\phi - y\sin\phi)\,\mathrm{d}x\,\mathrm{d}y$$

alors on a une expression analytique pour son inverse :

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

qui est classiquement décomposée en trois opérateurs :

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

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 ϕ alors cela pourrait fournir une solution satisfaisante. Mais, en pratique on connaît $p(r,\phi)$ pour des valeurs discrètes de r et de ϕ 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 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 ϕ 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 Ω 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 :

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

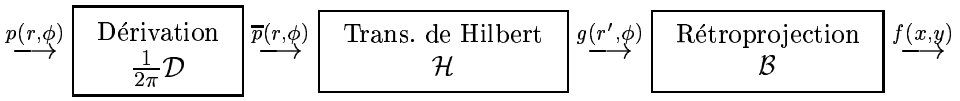
On peut alors résumer les différents algorithmes qui peuvent être utilisés pour le calcul de la solution de la TR inverse :

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 grande 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.

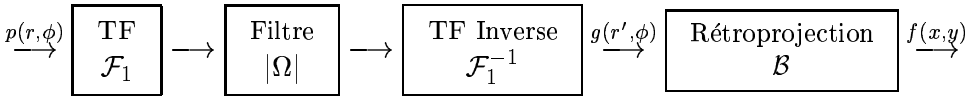
Les figures de la page suivante illustrent ces difficultés.

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.

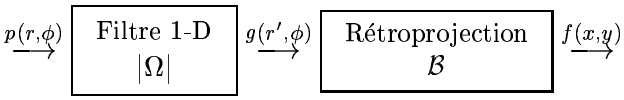
- Inversion directe :



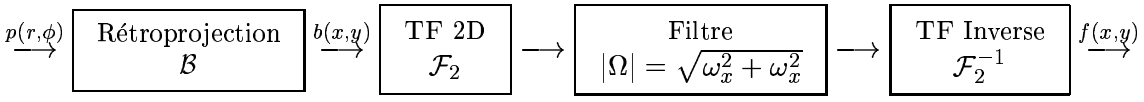
- Rétroprojection des projections filtrées :



- Filtrage par convolution et rétroprojection :



- Rétroprojection suivie de filtrage en 2D :



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

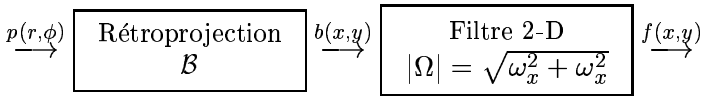


FIG. 7.1 – *Méthodes analytiques pour l'inversion de la* TR.

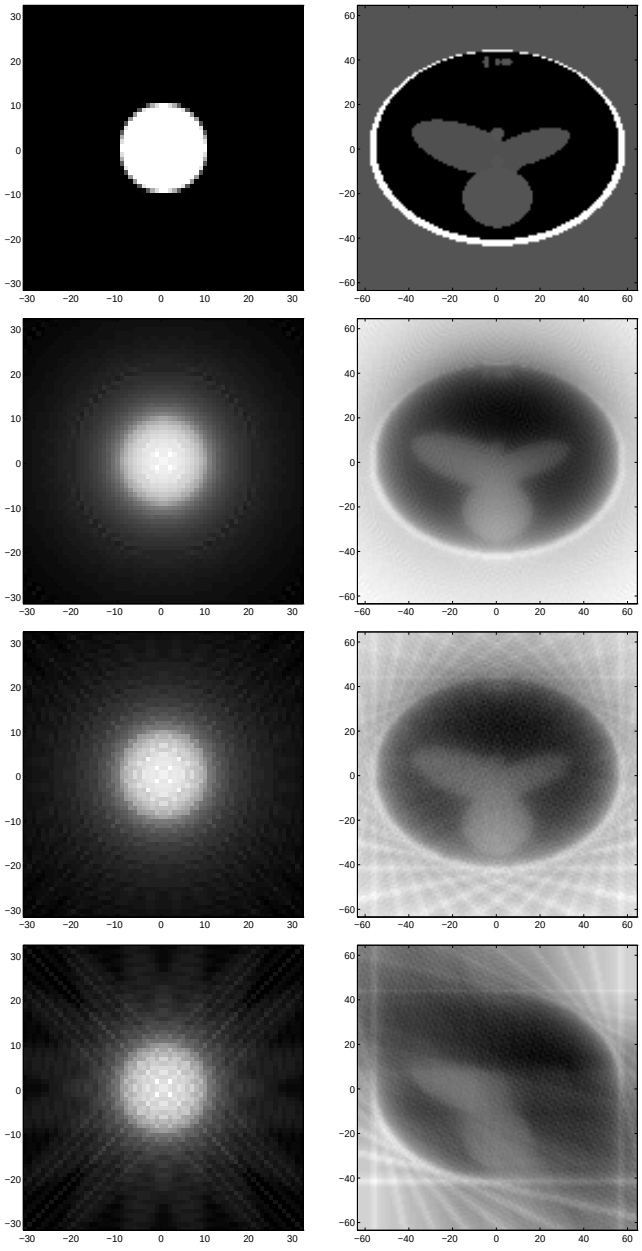


FIG. 7.2 – *Reconstruction par rétroprojection filtrée : Trois cas sont considérés : 64 projections, 16 projections réparties entre 0 et π et 8 projections réparties entre 0 et $\pi/2$.*

7.2 Méthodes de développement en série

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

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

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 $g(\boldsymbol{u}_i), i = 1, \cdots, K$ on détermine K coefficients $g_k, k = 1, \cdots, K$ et on en déduit les K coefficients $f_k, k = 1, \cdots K$. Finalement, on reconstruit $\widehat{f}(\boldsymbol{r})$ par

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

Le choix des $b_k(\boldsymbol{r})$ 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:

$$g(r) = \int f(r')h(r,r')\,\mathrm{d}r'$$

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:

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

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

$$H_k(\omega_0) = \int h(r) \exp\left[+jk\omega_0\right]\,\mathrm{d}r$$

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

$$g_k = \frac{1}{2\pi} \int g(r) \exp\left[-jk\omega_0 r\right]\,\mathrm{d}r, \quad k = 1 \cdots, K$$

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

$$f(r) = \sum_{k}^{K} f_k \exp\left[-jk\omega_0 r\right]$$

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 (ρ,θ) :

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

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

$$p(r,\phi) = \sum_{k} p_k(r) \exp\left[-jk\phi\right], \qquad f(\rho,\theta) = \sum_{k} f_k(\rho) \exp\left[-jk\theta\right],$$

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

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

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

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

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

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

3. On calcule la solution par

$$f(\rho,\theta) = \sum_{k}^{K} f_k(\rho) \exp\left[-jk\theta\right]$$

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 $b_k(\boldsymbol{r})$ 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 $b_j(\boldsymbol{r})$:

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

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

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

on puisse calculer analytiquement les coefficients H_{ij}.

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

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

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

Ensuite, deux approches sont courantes :

1. Choisir $N = M$ et espérer inverser directement la matrice $\boldsymbol{H}$ 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 :

$$F_c(q) = 4\pi \int_0^{\infty} r^2\, J_0(qr)\rho(r)\, \mathrm{d}r \tag{7.1}$$

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 $\boldsymbol{q} = \{q_1,\cdots,q_M\}$ 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 :

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

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

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

il n'est pas difficile de montrer que :

$$\boldsymbol{F}_c = \boldsymbol{A}\boldsymbol{f} \tag{7.4}$$

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

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

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

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

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 $\boldsymbol{A}$ 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

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

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 $\boldsymbol{A}$ n'est plus diagonale. De plus, il n'y a aucune raison de choisir $M = N$, car on peut toujours essayer d'estimer $\boldsymbol{f}$ au moins par MC :

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

ou par MC de norme minimale (MCNM) :

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

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 $\boldsymbol{A}$ 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.

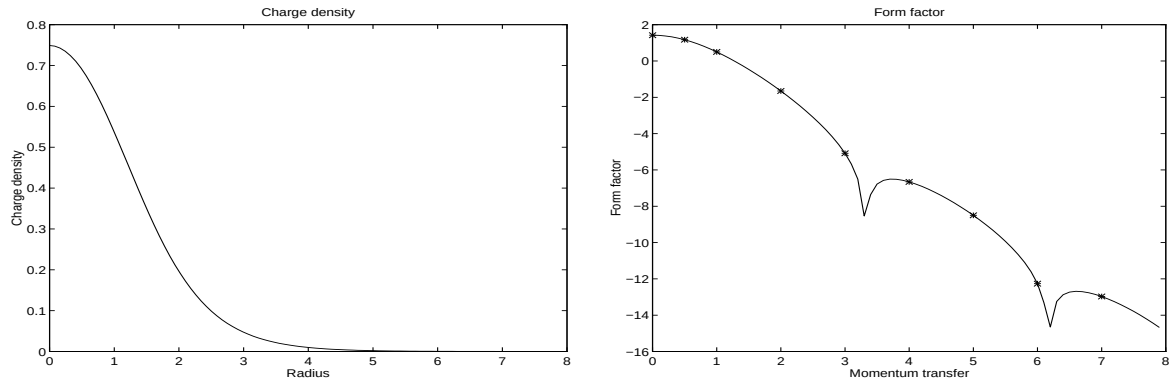


Figure a : La densité de charge théorique $\rho(r)$ [gauche], le facteur de forme théorique $\log \|F_c(q)\|$ et les données utilisées pour les simulations [droite].

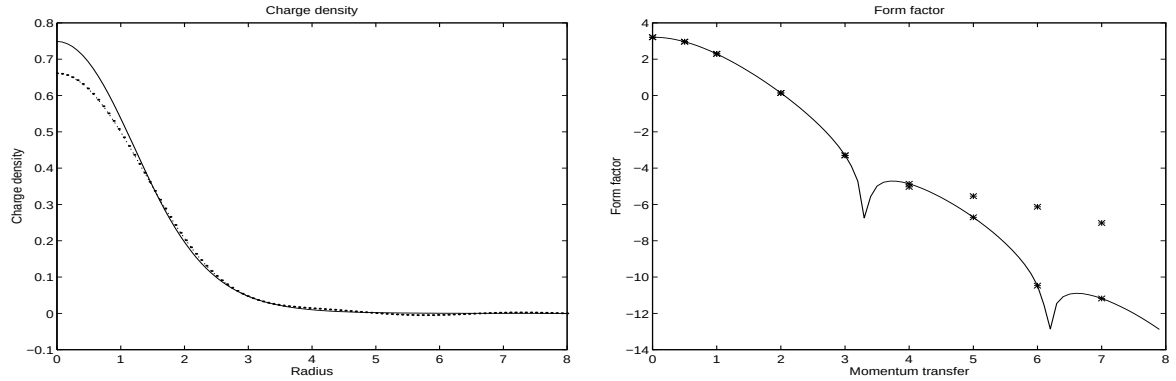


Figure b. Résultat de reconstruction de $\widehat{\rho}(r)$ obtenus par MC (points) et par MCNM (tirets) pour $N = 5$ [gauche] et les facteurs de formes correspondants $\log \|\widehat{F}_c(q)\|$ [droite]. Dans ce cas les deux solutions sont pratiquement identiques car la matrice $\boldsymbol{A}^t\boldsymbol{A}$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

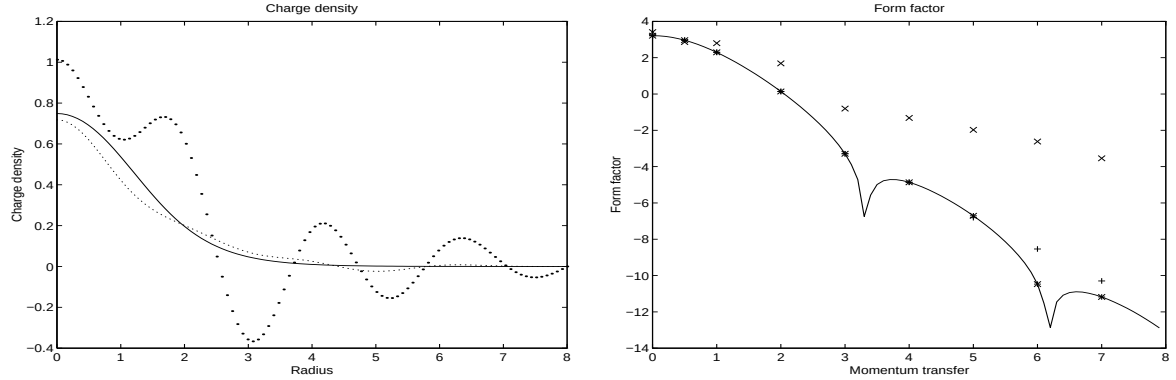


Figure c. Résultat de reconstruction de $\widehat{\rho}(r)$ obtenus par MC (points) et par MCNM (tirets) pour $N = 10$ [gauche] et les facteurs de formes correspondants $\log \|\widehat{F}_c(q)\|$ [droite]. Dans ce cas la matrice $\boldsymbol{A}\boldsymbol{A}^t$ est mal conditionnée, alors que la matrice $\boldsymbol{A}\boldsymbol{A}^t + \lambda \boldsymbol{I}$ 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 $f(\boldsymbol{r})$ 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 $f(\boldsymbol{r})$ sur une base :

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

sauf qu'ici les fonctions de base sont choisies parmi

$$b_n(\boldsymbol{r}) = \delta(\boldsymbol{r} - n\boldsymbol{\Delta}) \tag{7.7}$$

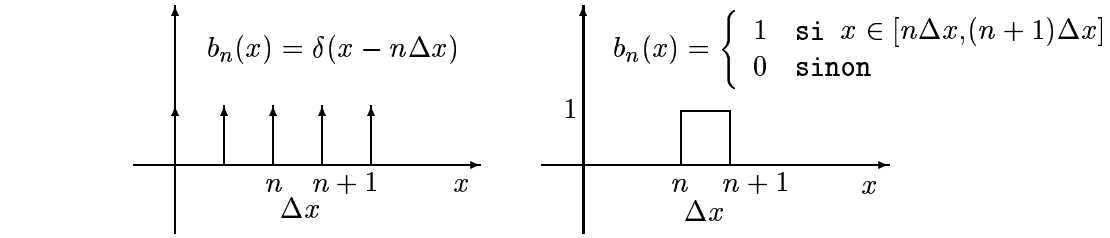
ou

$$b_n(\boldsymbol{r}) = \begin{cases} 1 & \text{si } \boldsymbol{r} \in [(n-1)\boldsymbol{\Delta}, n\boldsymbol{\Delta}] \\ 0 & \text{a}illeurs \end{cases} \tag{7.8}$$

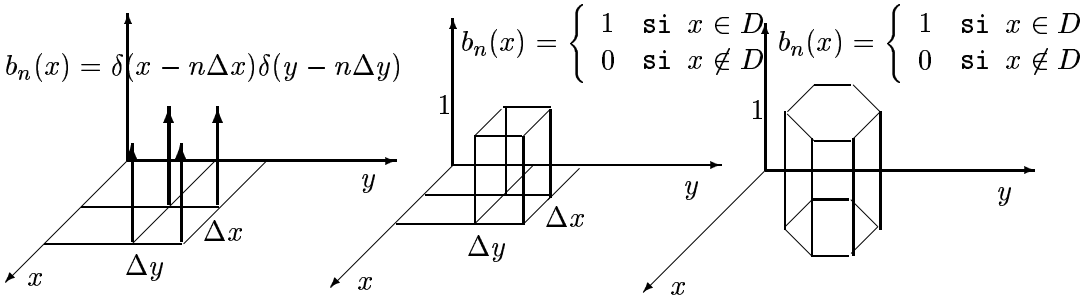
ou encore (pour un échantillonnage non régulier)

$$b_n(\boldsymbol{r}) = \begin{cases} 1 & \text{si } \boldsymbol{r} \in D_n \\ 0 & \text{a}illeurs \end{cases} \tag{7.9}$$

avec D_n un ensemble de régions dans le domaine de la fonction telle que $D_i \cap D_j = 0, \forall i \neq j$ et $\cup D_j = S$.



Cas 1D



Cas 2D

FIG. 7.4 – *chantillonnage et fonctions bases.*

Les pas d'échantillonnage $\boldsymbol{\Delta} = \{\Delta_1, \Delta_2, \cdots\}$ 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 :

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

où $\boldsymbol{f}$ 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 $[(n-1)\boldsymbol{\Delta},n\Delta]$ 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 $\boldsymbol{H}$ 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 $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ avec $\boldsymbol{H}$ une matrice carrée et inversible et vouloir proposer la solution $\widehat{\boldsymbol{f}} = \boldsymbol{H}^{-1}\boldsymbol{g}$.

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

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{\Delta(\boldsymbol{g},\boldsymbol{H}\boldsymbol{f})\}$$

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

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \left\{ \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 \right\}$$

qui satisfont l'équation normale

$$\boldsymbol{H}^t\boldsymbol{H}\widehat{\boldsymbol{f}} = \boldsymbol{H}^t\boldsymbol{g}$$

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 $\boldsymbol{b}$. 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 $J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2$, 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

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

où $H(\boldsymbol{f})$ est une fonctionnelle régularisante et λ 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

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

où Δ_1 et Δ_2 sont deux distances et $\boldsymbol{f}_0$ 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 $H(\boldsymbol{f})$ est assez arbitraire et on manque d'arguments pour le choix du paramètre de régularisation λ ;

2. le bruit $\boldsymbol{b}$ 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 $p(\boldsymbol{g}|\boldsymbol{f})$. On considère ensuite $p(\boldsymbol{g}|\boldsymbol{f})$ comme une fonction de $\boldsymbol{f}$ que l'on nomme la fonction *vraisemblance* $V(\boldsymbol{f})$. On définit alors un critère basé sur cette fonction, par exemple le maximum de vraisemblance (MV), qui permet alors d'estimer $\boldsymbol{f}$.
   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 $\boldsymbol{f}$ 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* $p_0(\boldsymbol{f})$ 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* $p(\boldsymbol{f})$, mais ces mesures n'étant pas suffisantes pour déterminer d'une manière unique $p(\boldsymbol{f})$, on choisirait la loi $p(\boldsymbol{f})$ qui est la plus proche de $p_0(\boldsymbol{f})$ au sens da 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.

Chapitre 8

Approches probabilistes

Nous avons distingué quatre approches probabilistes différentes pour la résolution des problèmes inverses. L'objet de ce chapitre est de détailler un peu plus ces approches.

Dans ce chapitre, nous aborderons les trois premières approches, en montrant par des exemples simples leur limites. Nous réservons une place privilégiée à l'approche bayésienne qui sera détaillée dans un chapitre séparé.

8.1 Approche n'utilisant que les moments

Dans cette approche, on considère que les grandeurs observées et inconnues sont bien des fonctions aléatoires, on se limite à caractériser les lois de probabilités de ces grandeurs que par leurs moments, souvent jusqu'à l'ordre deux. Ensuite on établit un lien entre les moments des grandeurs observées et ceux des grandeurs inconnues puis on cherche à l'inverser.

Mais, un point qui est souvent laissé pour compte est : l'estimation des moments des grandeurs observées. Souvent on fait l'hypothèse que l'on peut estimer ces moments par les moyennes empiriques. Ceci peut être raisonnable lorsqu'on se limite aux moments d'ordre un et deux, mais cela devient inefficace pour les moments d'ordre supérieur. C'est d'ailleurs dans le premier cas que ces méthodes trouvent particulièrement leur intérêt : c'est l'estimation en moyenne quadratique.

Pour illustrer cette approche, nous allons considérer le cas du filtrage optimal de Wiener pour la résolution du problème de déconvolution de signaux.

Filtrage optimal de Wiener :

Considérons le problème de la déconvolution des signaux:

$$g(t) = h(t) * f(t) + b(t)$$

supposons que $f(t)$, $b(t)$ et $g(t)$ sont des fonctions aléatoires et cherchons à lier les moments d'ordre un :

$$\mathrm{E}\left[g(t)\right], \quad \mathrm{E}\left[b(t)\right] \quad \text{et} \quad \mathrm{E}\left[f(t)\right]$$

et les moments d'ordre deux :

$$\begin{aligned} R_{gg}(\tau) &= \mathrm{E}\left[g(t)g(t+\tau)\right], \\ R_{ff}(\tau) &= \mathrm{E}\left[f(t)f(t+\tau)\right], \\ R_{bf}(\tau) &= R_{fb}(-\tau) = \mathrm{E}\left[b(t)f(t+\tau)\right] \\ R_{gf}(\tau) &= R_{fg}(-\tau) = \mathrm{E}\left[g(t)f(t+\tau)\right] \end{aligned}$$

de ces grandeurs. Faisant l'hypothèse que $f(t)$ et $b(t)$ sont indépendantes. On obtient :

$$\begin{aligned} \mathrm{E}\left[g(t)\right] &= h(t) * \mathrm{E}\left[f(t)\right] + \mathrm{E}\left[b(t)\right] \\ R_{gg}(\tau) &= h(t) * h(t) * R_{ff}(\tau) + R_{bb}(\tau) \\ R_{gf}(\tau) &= h(t) * R_{ff}(\tau) \end{aligned}$$

Passant dans le domaine de FOURIER on a :

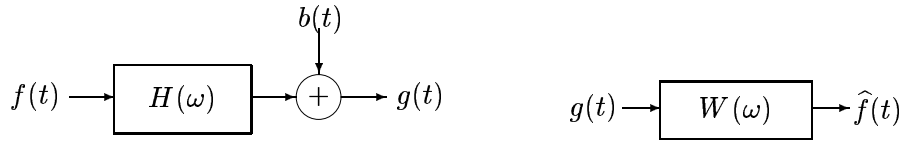
$$\begin{aligned} S_{gg}(\omega) &= |H(\omega)|^2 S_{ff}(\omega) + R_{bb}(\omega) \\ S_{gf}(\omega) &= H(\omega) S_{ff}(\omega), \\ S_{fg}(\omega) &= H^*(\omega) S_{ff}(\omega), \end{aligned}$$

L'objectif du filtrage optimal est de fournir une solution $\widehat{f}(t)$ qui s'obtient par un filtrage linéaire de $g(t)$:

$$\widehat{f}(t) = w(t) * g(t)$$

et la réponse impulsionnelle du filtre $w(t)$ est telle que l'erreur quadratique moyenne

$$EQM = \mathrm{E}\left[\left(f(t) - \widehat{f}(t)\right)^2\right]$$



Modèle d'observation Déconvolution par filtre de Wiener.

FIG. 8.1 – *Filtrage de* WIENER.

soit minimale.

En utilisant le principe d'orthogonalité :

$$\mathrm{E}\left[(f(t) - w(t) * g(t))\ g(t-\tau)\right] = 0 \quad \forall t,\tau \longrightarrow R_{fg}(\tau) = w(t) * R_{gg}(\tau).$$

Passant dans le domaine de FOURIER on a :

$$W(\omega) = \frac{S_{fg}(\omega)}{S_{gg}(\omega)} = \frac{H^*(\omega) S_{ff}(\omega)}{|H(\omega)|^2 S_{ff}(\omega) + S_{bb}(\omega)}.$$

Remplaçant pour les expressions de $S_{fg}(\omega)$ et $S_{gg}(\omega)$ on obtient :

$$W(\omega) = \frac{H^*(\omega) S_{ff}(\omega)}{|H(\omega)|^2 S_{ff}(\omega) + S_{bb}(\omega)} = \frac{1}{H(\omega)} \frac{|H(\omega)|^2}{|H(\omega)|^2 + \frac{S_{bb}(\omega)}{S_{ff}(\omega)}}$$

La difficulté dans cette approche commence lors de la mise en œuvre réelle de la méthode. En effet, il faut pouvoir estimer $S_{gg}(\omega)$ et $S_{fg}(\omega)$ ce qui nécessite la connaissance *a priori* de $S_{ff}(\omega)$ et de $S_{bb}(\omega)$.

Souvent, en pratique on fait l'hypothèse que

$$\frac{S_{bb}(\omega)}{S_{ff}(\omega)} = r,$$

avec r une constante représentant l'inverse du rapport signal à bruit. Avec cette hypothèse on a

$$W(\omega) = \frac{1}{H(\omega)} \frac{|H(\omega)|^2}{|H(\omega)|^2 + r}.$$

Cette hypothèse n'est cependant pas très réaliste car elle considère que le signal et le bruit ont la même forme de spectre, ce qui est rarement le cas. En effet, on peut souvent considére le bruit blanc, mais le signal que l'on cherche n'a pas le même spectre.

Notons que lorsque $r \mapsto 0$ on retrouve le filtrage inverse.

Une autre hypothèse plus réaliste consiste à supposer

$$\frac{S_{bb}(\omega)}{S_{ff}(\omega)} = |D(\omega)|^2,$$

où $D(\omega)$ représente la fonction de transfert d'un filtre passe-haut. Avec cette hypothèse on a

$$W(\omega) = \frac{1}{H(\omega)} \frac{|H(\omega)|^2}{|H(\omega)|^2 + |D(\omega)|^2}.$$

Nous verrons dans le chapitre suivant qu'il existe un lien entre cette approche et celle de l'estimation bayésienne dans le cas linéaire avec des lois gaussiennes.

8.2 Estimation au sens du maximum de vraisemblance

Dans cette approche on prend en compte explicitement le caractère incertain des mesures en les caractérisant par une loi de probabilité $p(\boldsymbol{g}|\boldsymbol{f})$.

L'idée de base ensuite est de considérer $p(\boldsymbol{g}|\boldsymbol{f})$ comme une fonction $V(\boldsymbol{f}) = p(\boldsymbol{g}|\boldsymbol{f})$ appelée fonction *vraisemblance*. On définit alors la solution $\widehat{\boldsymbol{f}}$ du problème inverse comme l'argument qui maximise cette fonction :

$$\widehat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \{p(\boldsymbol{g}|\boldsymbol{f})\} = \arg\min_{\boldsymbol{f}} \{-\ln p(\boldsymbol{g}|\boldsymbol{f})\}$$

Exemple 1: Loi Gaussienne :
Considérons le modèle $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{b}$ et supposons que $\boldsymbol{b}$ puisse être modélisé par un vecteur aléatoire centrée, blanc et gaussien :

$$b_i \sim \mathcal{N}(0,\sigma_b^2) \longrightarrow \boldsymbol{b} \sim \mathcal{N}(\boldsymbol{0},\sigma_b^2\boldsymbol{I}) \longrightarrow \boldsymbol{g}|\boldsymbol{f} \sim \mathcal{N}(\boldsymbol{H}\boldsymbol{f},\sigma_b^2\boldsymbol{I}).$$

On en déduit alors :

$$p(\boldsymbol{g}|\boldsymbol{f}) = K \exp\left[\frac{-1}{2\sigma_b^2}[\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}]^t[\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}]\right].$$

L'estimé au sens du MV devient alors :

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{-\ln p(\boldsymbol{g}|\boldsymbol{f})\} = \arg\min_{\boldsymbol{f}} \left\{\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2\right\}$$

et on retrouve l'estimation au sens des moindres carrés.

Si nous connaissons un peu plus sur les caractéristiques des b_i, par exemple si les b_i ont des variances différentes, alors

$$b_i \sim \mathcal{N}(0,\sigma_i^2)$$

et il n'est pas difficile de montrer que

$$p(\boldsymbol{g}|\boldsymbol{f}) = K \exp\left[\frac{-1}{2}(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f})^t\boldsymbol{W}^{-1}(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f})\right]$$

avec $\boldsymbol{W} = \text{diag}\{\sigma_1^2, \cdots, \sigma_m^2\}$. Alors,

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{-\ln p(\boldsymbol{g}|\boldsymbol{f})\} = \arg\min_{\boldsymbol{f}} \left\{\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2_{\boldsymbol{W}}\right\}$$

et on retrouve l'estimation au sens des moindres carrés généralisés.

L'extension au cas d'un modèle non linéaire se fait aussi très facilement.

Exemple 2: Loi Gaussienne généralisée :
Supposons maintenant que les b_i suivent une loi gaussienne généralisée :

$$b_i \sim p(b_i) \propto \exp\left[-\beta|b_i|^p\right], \quad \beta > 0, \quad 1 < p \leq 2.$$

Si on fait l'hypothèse que $\boldsymbol{b}$ est blanc alors on a :

$$\boldsymbol{b} \sim p(\boldsymbol{b}) \propto \exp\left[-\beta\sum_{i=1}^{M}|b_i|^p\right] \longrightarrow -\ln p(\boldsymbol{b}) = K + \beta\sum_{i=1}^{M}|b_i|^p$$

On en déduit alors :

$$-\ln p(\boldsymbol{g}|\boldsymbol{f}) = K + \beta \sum_{i=1}^{M} |g_i - [\boldsymbol{H}\boldsymbol{f}]_i|^p.$$

L'estimé au sens du MV devient alors :

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{-\ln p(\boldsymbol{g}|\boldsymbol{f})\} = \arg\min_{\boldsymbol{f}} \left\{ \sum_{i=1}^{M} |g_i - [\boldsymbol{H}\boldsymbol{f}]_i|^p \right\} = \arg\min_{\boldsymbol{f}} \{\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^p\}$$

et on retrouve l'estimation au sens des moindres carrés.

Exemple 3: Loi Gamma :
Supposons maintenant que les b_i suivent une loi de gamma :

$$b_i \sim \mathcal{G}(\alpha,\beta) \longrightarrow p(b_i) = K_i b_i^{-\alpha} \exp[-\beta b_i]$$

et si on fait l'hypothèse que $\boldsymbol{b}$ est blanc (b_i et b_j) indépendantes $\forall i,j, i \neq j$, on a

$$p(\boldsymbol{b}) = \prod_{i=1}^{M} p(b_i) \longrightarrow \ln p(\boldsymbol{b}) = K + \sum_{i=1}^{M} \alpha \log(b_i) + \beta b_i.$$

On en déduit :

$$\ln p(\boldsymbol{g}|\boldsymbol{f}) = K + \sum_{i=1}^{M} \alpha \log(g_i - [\boldsymbol{H}\boldsymbol{f}]_i) + \beta(g_i - [\boldsymbol{H}\boldsymbol{f}]_i)$$

et

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{-\ln p(\boldsymbol{g}|\boldsymbol{f})\}.$$

Il n'y a pas d'expression analytique pour cette solution, mais elle peut être calculé numériquement par un algorithme itératif.

En conclusion :

- L'approche MV permet de mieux prendre en compte la nature du bruit en lui attribuant une loi de probabilité $p(\boldsymbol{b})$, ou, d'une manière plus générale, l'incertaine sur les mesures en leur attribuant une loi de probabilité $p(\boldsymbol{g}|\boldsymbol{f})$;
- Dans le cas gaussien on retrouve l'estimation au sens des moindres carrés ;
- Cette approche peut fournir des résultats satisfaisants lorsqu'elle est utilisée en combinaison avec une modélisation paramétrique des inconnues avec un nombre de paramètres très inférieure au nombre de données indépendantes. Malheureusement, cette approche donne rarement des résultats satisfaisants pour la résolution des problèmes inverses dans le cadre algébrique où le nombre d'inconnues est du mêmes ordre de grandeur voir plus grand que le nombre des mesures. C'est pourquoi, certains auteurs ont proposé une approche dite *Maximum de vraisemblance pénalisée* qui consiste à définir la solution par :

  $$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{-\ln p(\boldsymbol{g}|\boldsymbol{f}) + \phi(\boldsymbol{f})\}$$

  où $\phi(\boldsymbol{f})$ est une fonction de pénalisation choisie d'une manière adhoc pour permettre d'obtenir une solution satisfaisante. Nous verrons alors que cette approche peut mieux être interprétée dans le cadre de l'inférence bayésienne où $\phi(\boldsymbol{f}) = -\ln p(\boldsymbol{f})$ avec $p(\boldsymbol{f})$ la loi *a priori*.

8.3 Approche du maximum d'entropie

Avant de voir comment le principe du maximum d'entropie peut être utilisé pour la résolution des problèmes inverses, nous allons rappeler un certain nombre de définitions et de notions qui seront essentielles pour la compréhension de cette approche.

8.3.1 Définition de l'entropie

Le principe du ME peut être approché de différentes manières. L'approche de la théorie de l'information [SW48] est sans doute la mieux adaptée à notre problème. JAYNES ([Jay68, Jay82, Jay85]) est parmi les premiers auteurs modernes à avoir introduit le formalisme du ME par l'approche de la théorie de l'information. Cette notion d'entropie est introduite de la manière suivante : considérons une variable aléatoire discrète X produisant des réalisations $\boldsymbol{x} = \{x_1,x_2,\cdots,x_n\}$ et attribuons les probabilités $\boldsymbol{p} = \{p_1,p_2,\cdots,p_n\}$ à ces réalisations pour représenter notre information partielle sur cette variable. On définit la quantité $I_i = \ln(1/p_i)$ comme la quantité d'information apportée par la réalisation x_i .

Le raisonnement intuitif conduisant à cette expression est le suivant : plus un événement est rare, plus le gain d'information obtenu par sa réalisation est grand. L'utilisation du logarithme rend additif le gain total d'information obtenu par la réalisation de plusieur événements indépendants. On définit alors l'entropie d'un processus par la somme pondérée des informations individuelles de chaque réalisation. C'est la définition de l'entropie donnée par SHANNON :

$$S(\boldsymbol{p}) = \sum_{i=1}^{n} p_i \ln \frac{1}{p_i} = -\sum_{i=1}^{n} p_i \ln p_i. \tag{8.1}$$

L'entropie S est une mesure d'incertitude de la distribution $\boldsymbol{p} = \{p_1,p_2,\cdots,p_n\}$, déterminée uniquement par certaines règles élémentaires de cohérence logique et d'additivité ([Jay82, Jay85, Bal82]).

Généralisant ce concept, on définit $-\ln(q_i/p_i)$ comme le gain d'information, sur une probabilité *a priori* p_i, apporté par la connaissance de la probabilité q_i de réalisation de l'événement x_i. On définit alors :

$$S(\boldsymbol{q},\boldsymbol{p}) = -\sum_{i=1}^{n} q_i \ln \frac{p_i}{q_i} = \sum_{i=1}^{n} q_i \ln \frac{q_i}{p_i}, \tag{8.2}$$

appelée entropie croisée ou entropie relative de la distribution q_i par rapport à la distribution p_i. Notons ici qu'il y a un changement de signe pour des raisons historiques et il est clair que la minimisation de l'entropie croisée $S(\boldsymbol{q},\boldsymbol{p})$ se réduit à la maximisation de l'entropie $S(\boldsymbol{q})$ si l'*a priori* $\boldsymbol{p}$ est uniforme.

Sous réserve de quelques précautions, on peut généraliser ce qui précède au cas de distributions continues, et définir l'entropie [SW48] par :

$$S(p) = -\int p(x) \ln p(x)\, \mathrm{d}x, \tag{8.3}$$

et l'entropie croisée par :

$$S(q,p) = -\int q(x) \ln \frac{p(x)}{q(x)}\, \mathrm{d}x. \tag{8.4}$$

8.3.2 Lois à maximum d'entropie

Voyons maintenant ce que signifie le choix d'une distribution de probabilité à maximum d'entropie contenant une information *a priori*, ou qui soit compatible avec des contraintes connues sur cette distribution. Pour cela, précisons tout d'abord la signification de l'information contenue dans une distribution de probabilité $\{p_1,p_2,\cdots,p_n\}$. À l'évidence il faut que l'on puisse extraire cette information de cette distribution. Considérons une variable aléatoire discrète X qui peut prendre des valeurs $\boldsymbol{x} = \{x_1,x_2,\cdots,x_n\}$ avec une distribution de probabilité $\boldsymbol{p} = \{p_1,p_2,\cdots,p_n\}$. Maintenant si on nous demande quelle est la meilleure estimée $\widehat{\phi}$ d'une fonction $\phi(X)$ au sens du minimum de l'erreur quadratique moyenne. La solution est immédiate :

$$\widehat{\phi} = \mathrm{E}\,[\phi] = \sum_{i=1}^{n} p_i\,\phi(x_i).$$

C'est un problème direct et bien-posé. Inversement, si l'on se pose la question d'ajuster une distribution $\boldsymbol{p}$ pour incorporer une information donnée sur la fonction $\phi(X)$, il faut se poser la question suivante :

> connaissant une règle d'estimation précise, par exemple la minimisation de l'erreur quadratique moyenne, comment choisir $\boldsymbol{p}$ pour que l'on ait $\widehat{\phi} = \mathrm{E}\,[\phi]$?

Le problème est qu'il existe, en général, beaucoup de distributions qui satisfont cette contrainte. Il s'agit d'un problème inverse mal posé au sens où la solution n'est pas unique. Le principe du maximum d'entropie nous permet alors de choisir une solution.

Soit maintenant le cas plus général où nous considérons les m fonctions $\{\phi_1(X),\phi_2(X),\cdots,\phi_m(X)\}$ pour lesquelles nous avons un ensemble de données $\{d_1,d_2,\cdots,d_m\}$ pouvant s'exprimer sous la forme des m contraintes simultanées suivantes :

$$\mathrm{E}\,[\phi_k(X)] = \sum_{i=1}^{n} p_i\,\phi_k(x_i) = d_k, \quad k = 1,\cdots,m.$$

Dans chaque problème, ces données $\{d_1,d_2,\cdots,d_m\}$ peuvent avoir des interprétations physiques différentes et la difficulté consiste à incorporer ces données (contraintes) dans notre distribution de probabilité. Nous voulons ajuster la distribution de probabilité $\{p_1,p_2,\cdots,p_n\}$ à nos données. En général, le nombre des contraintes m est inférieur à n et il y a une infinité de solutions à ce problème. En d'autres termes, on peut trouver une infinité de distributions de probabilité $\{p_1,p_2,\cdots,p_n\}$ qui satisfont ces contraintes. C'est ici que le principe du maximum d'entropie est mis en œuvre pour choisir, parmi ces solutions possibles, celle qui a l'entropie maximale, c'est-à-dire la loi qui satisfait toutes les contraintes (toute l'information connue) et qui est la moins compromettante vis-à-vis de toute autre information inconnue. Le mot *information* devant être pris au sens de la définition de l'information moyenne ou de l'entropie de Shannon (équation 3).

Le problème se formule ainsi :

Problème P1 :

$$\text{maximiser} \quad S(\boldsymbol{p}) = -\sum_{i=1}^{n} p_i\,\ln p_i,$$

$$\text{sous les contraintes} \quad \sum_{i=1}^{n} p_i\,\phi_k(x_i) = d_k, \quad k = 1,\cdots,m.$$

La solution est obtenue par une technique variationnelle de multiplicateurs de Lagrange et est donnée par :

$$p_i = \frac{1}{Z(\lambda_1,\cdots,\lambda_m)} \exp\left[-\sum_{k=1}^{m} \lambda_k \phi_k(x_i)\right], \quad i = 1,\cdots,n, \tag{8.5}$$

où

$$Z(\lambda_1,\cdots,\lambda_m) = \sum_{i=1}^{n} \exp\left[-\sum_{k=1}^{m} \lambda_k \phi_k(x_i)\right] \tag{8.6}$$

est la fonction de partition, et les $\{\lambda_1,\lambda_2,\cdots,\lambda_n\}$ sont déterminés par le système d'équations :

$$-\frac{\partial \ln Z(\lambda_1,\cdots,\lambda_m)}{\partial \lambda_k} = d_k, \quad k = 1,\cdots,m, \tag{8.7}$$

avec m données d_k et m inconnues λ_k. Notons que ce système d'équations n'est autre que le système d'équations des contraintes constituées par les données.

La valeur maximale de l'entropie est :

$$S_{\max} = \ln Z + \sum_{k=1}^{m} \lambda_k d_k.$$

Si on note $\lambda_0 = \ln Z$ on peut vérifier facilement les propriétés suivantes :

$$-\frac{\partial \lambda_0}{\partial \lambda_k} = \mathrm{E}\left[\phi_k(x_i)\right] = d_k, \tag{8.8}$$

$$-\frac{\partial^2 \lambda_0}{\partial {\lambda_k}^2} = \mathrm{E}\left[\phi_k^2(x_i)\right] - d_k^2 = \mathrm{Var}\left[\phi_k(x_i)\right], \tag{8.9}$$

$$-\frac{\partial^2 \lambda_0}{\partial \lambda_k \partial \lambda_l} = \mathrm{E}\left[\phi_k(x_i)\phi_l(x_i)\right] - d_k d_l = \mathrm{Cov}\left\{\phi_k(x_i),\phi_l(x_i)\right\}. \tag{8.10}$$

8.3.3 Lois à minimum d'entropie relative

L'extension de ce qui précède au cas de l'entropie croisée se formule ainsi :

Problème P2 :
Étant donné une distribution de probabilité *a priori* $\boldsymbol{p} = \{p_1,p_2,\cdots,p_n\}$, déterminer la distribution de probabilité *a posteriori* $\boldsymbol{q} = \{q_1,q_2,\cdots,q_n\}$ qui minimise

$$S(\boldsymbol{q},\boldsymbol{p}) = -\sum_i q_i \ln \frac{p_i}{q_i} = \sum_i q_i \ln \frac{q_i}{p_i},$$

et qui satisfait les m contraintes

$$\sum_{i=1}^{n} q_i \, \phi_k(x_i) = d_k, \quad k = 1,\cdots,m.$$

La solution est obtenue, là aussi, par une technique variationnelle de multiplicateurs de Lagrange et donnée par :

$$q_i = \frac{1}{Z} p_i \exp\left[-\sum_{k=1}^{m} \lambda_k \phi_k(x_i)\right], \quad i = 1,\cdots,n, \tag{8.11}$$

où

$$Z(\lambda_1,\cdots,\lambda_m) = \sum_{i=1}^{n} p_i \, \exp\left[-\sum_{k=1}^{m} \lambda_k \phi_k(x_i)\right] \tag{8.12}$$

est la fonction de partition, et les $\{\lambda_1,\lambda_2,\cdots,\lambda_n\}$ sont déterminés par le système d'équations (9).

8.3.4 Extension au cas continu

Dans le cas continu, les définitions (8.3) et (8.4) conduisent à :

Problème P3 :

$$\text{maximiser} \quad S(p) = -\int p(x) \, \ln p(x) \, \mathrm{d}x,$$

$$\text{sous les contraintes} \quad \int \phi_k(x) \, p(x) \, \mathrm{d}x = d_k, \quad k = 1,\cdots,m.$$

La solution est

$$p(x) = \frac{1}{Z} \, \exp\left[-\sum_{k=1}^{m} \lambda_k \phi_k(x)\right], \tag{8.13}$$

où

$$Z(\lambda_1,\cdots,\lambda_m) = \int \exp\left[-\sum_{k=1}^{m} \lambda_k \phi_k(x)\right] \mathrm{d}x \tag{8.14}$$

est la fonction de partition, et les $\{\lambda_1,\lambda_2,\cdots,\lambda_n\}$ sont déterminés par le système d'équations (9).

Problème P4 :
Étant donné une densité de probabilité *a priori* $p(x)$, déterminer la densité de probabilité *a posteriori* $q(x)$ qui minimise l'entropie croisée

$$S(q,p) = -\int q(x) \, \ln \frac{p(x)}{q(x)} \, \mathrm{d}x,$$

et qui satisfait les m contraintes

$$\int \phi_k(x) \, q(x) \, \mathrm{d}x = d_k, \quad k = 1,\cdots,m.$$

La solution est

$$q(x) = \frac{1}{Z} \, p(x) \, \exp\left[-\sum_{k=1}^{m} \lambda_k \phi_k(x)\right], \tag{8.15}$$

où

$$Z(\lambda_1,\cdots,\lambda_m) = \int p(x) \, \exp\left[-\sum_{k=1}^{m} \lambda_k \phi_k(x)\right] \mathrm{d}x \tag{8.16}$$

est la fonction de partition, et les $\{\lambda_1,\lambda_2,\cdots,\lambda_n\}$ sont déterminés par (9).

Notons également que le vecteur $\widehat{\boldsymbol{\lambda}} = (\lambda_1,\cdots,\lambda_m)$ dans tous ces problèmes peut être considéré comme la solution du problème d'optimisation suivant :

Problème dual des problèmes P1-P4 :

$$\widehat{\boldsymbol{\lambda}} = \arg\min_{\boldsymbol{\lambda}} \Big\{ D(\boldsymbol{\lambda}) = \boldsymbol{\lambda}^t \boldsymbol{d} + \ln Z(\boldsymbol{\lambda}) \Big\} \tag{8.17}$$

où $\boldsymbol{d} = (d_1, \cdots, d_m)$. Ce problème d'optimisation est appelé *problème dual* des problèmes P1-P4. Notons que l'expression de $Z(\boldsymbol{\lambda})$ est différente suivant le problème.

Une présentation légèrement différente de ces relations peut être obtenue si on définit

$$\phi_0(x) = 1, \quad \text{et} \quad d_0 = 1,$$

ce qui permet d'inclure la contrainte de normalisation de $q(x)$ et de formuler le problème précédent de la manière suivante :

$$\text{minimiser} \quad S(q,p) = -\int q(x)\, \ln \frac{p(x)}{q(x)}\, \mathrm{d}x,$$

$$\text{sous les contraintes} \quad \int \phi_k(x)\, q(x)\, \mathrm{d}x = d_k, \quad k = 0, \cdots, m.$$

La solution peut être écrite sous la forme

$$q(x) = p(x)\, \exp\left[-\sum_{k=0}^{m} \lambda_k \phi_k(x)\right], \tag{8.18}$$

et les $\{\lambda_0, \lambda_1, \lambda_2, \cdots, \lambda_n\}$ sont déterminés par :

$$G(\lambda_0, \lambda_1, \lambda_2, \cdots, \lambda_n) = -\int \phi_k(x)\, p(x)\, \exp\left[-\sum_{l=0}^{m} \lambda_l \phi_l(x)\right] \mathrm{d}x = d_k, \quad k = 0, \cdots, m. \tag{8.19}$$

On remarque que λ_0 est relié à la fonction de partition Z par $Z = -\exp[\lambda_0]$, et on a :

$$\lambda_0 = -\ln Z = -\ln \int p(x) \exp\left[-\sum_{k=1}^{m} \lambda_k \phi_k(x)\right] \mathrm{d}x. \tag{8.20}$$

Les autres coefficients λ_k sont déterminés par le système d'équations :

$$-\frac{\partial}{\partial \lambda_k} \lambda_0(\lambda_1, \cdots, \lambda_n) = d_k, \quad k = 1, \cdots, m. \tag{8.21}$$

Notons aussi que la valeur minimale $S_{\min}(q,p)$ peut être exprimée en fonction des λ_k par :

$$S_{\min}(q,p) = -\lambda_0 - \sum_{k=1}^{m} \lambda_k \phi_k(x). \tag{8.22}$$

Il n'est en général pas possible d'obtenir une relation explicite pour les coefficients λ_k. On résoud alors le système d'équations (8.19) numériquement par des méthodes itératives. Cependant, dans certaines situations simples on peut résoudre le problème d'une façon analytique [Jay82, JS84]. Le tableau 1 montre quelques exemples de lois, que l'on obtient sous forme analytique, en résolvant le problème P3 :

$$\text{maximiser} \quad S(p) = -\int p(x)\, \ln p(x)\, \mathrm{d}x,$$

$$\text{sous les contraintes} \quad \int \phi_k(x)\, p(x)\, \mathrm{d}x = d_k, \quad k = 1, \cdots, m.$$

Tableau 1: Exemples de lois à maximum d'entropie.

contraintes	domaine de x	loi de probabilité
$\phi_1(x) = x$	$x \in \mathbf{R}_+$	$p(x) = \frac{1}{\mu} \exp[-\mu x]$ loi exponentielle
$\phi_1(x) = \lvert x \rvert$	$x \in \mathbf{R}$	$p(x) = \frac{1}{2\mu} \exp[-\mu \lvert x \rvert]$ loi de LAPLACE
$\begin{cases} \phi_1(x) = x \\ \phi_2(x) = x^2 \end{cases}$	$x \in \mathbf{R}$	$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]$ loi de Gausse
$\begin{cases} \phi_1(x) = x \\ \phi_2(x) = \ln x \end{cases}$	$x \in \mathbf{R}_+$	$p(x) \propto \exp[-\lambda \ln x - \mu x]$ $\propto x^{-\lambda} \exp[-\mu x]$ loi Gamma
$\begin{cases} \phi_1(x) = \ln x \\ \phi_2(x) = \ln(1-x) \end{cases}$	$x \in]0,1[$	$p(x) \propto \exp[-\lambda \ln x - \mu \ln(1-x)]$ $\propto x^{-\lambda}(1-x)^{-\mu}$ loi Béta
$\begin{cases} \phi_1(x) = \ln x \\ \phi_2(x) = x^2 \end{cases}$	$x > 0$	$p(x) \propto \exp[-\lambda \ln x - \mu x^2)]$ $\propto x^{-\lambda} \exp[-\mu x^2]$ loi de Rayleigh

8.3.5 Extension au cas multivariable

Toutes ces relations peuvent être généralisées au cas multivariable. Par exemple le problème P4 devient :

Problème P5 :
Étant donné une densité de probabilité *a priori* $p(\boldsymbol{x})$, déterminer la densité de probabilité *a posteriori* $q(\boldsymbol{x})$ qui minimise l'entropie croisée

$$S(q,p) = -\int q(\boldsymbol{x}) \ln \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})} \, \mathrm{d}\boldsymbol{x},$$

et qui satisfait les m contraintes

$$\int \phi_k(\boldsymbol{x}) \, q(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = d_k, \quad k = 1, \cdots, m.$$

La solution est

$$q(\boldsymbol{x}) = \frac{1}{Z} p(\boldsymbol{x}) \exp\left[-\sum_{k=1}^{m} \lambda_k \phi_k(\boldsymbol{x})\right], \tag{8.23}$$

où

$$Z(\lambda_1, \cdots, \lambda_m) = \int p(\boldsymbol{x}) \exp\left[-\sum_{k=1}^{m} \lambda_k \phi_k(\boldsymbol{x})\right] \mathrm{d}\boldsymbol{x} \tag{8.24}$$

et les $\{\lambda_1,\lambda_2,\cdots,\lambda_n\}$ sont déterminés par :

$$-\frac{\partial \ln Z(\lambda_1,\cdots,\lambda_m)}{\partial \lambda_k} = d_k, \quad k = 1,\cdots,m. \tag{8.25}$$

Ici aussi, dans certains cas on peut obtenir une relation explicite pour la solution. Par exemple, si $p(\boldsymbol{x})$ est une distribution exponentielle multivariée de la forme séparable

$$p(\boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{\alpha_i} \exp\left[-\frac{x_i}{\alpha_i}\right],$$

et si les contraintes sont des contraintes sur les moments d'ordre un

$$\mathrm{E}\left[X_i\right] = \int x_i\, q(x_i)\, \mathrm{d}x_i = m_i, \quad i = 1,\cdots,n,$$

alors la solution $q(\boldsymbol{x})$ reste une fonction exponentielle multivariée séparable :

$$q(\boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{m_i} \exp\left[-\frac{x_i}{m_i}\right].$$

De même si $p(\boldsymbol{x})$ est une fonction gaussienne multivariée séparable :

$$p(\boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\alpha_i} \exp\left[-\frac{1}{2}\left(\frac{x_i - m_i}{\alpha_i}\right)^2\right],$$

et si les contraintes sont des contraintes du second ordre :

$$\mathrm{E}\left[(X_i - m_i)^2\right] = \int (x_i - m_i)^2\, q(x_i)\, \mathrm{d}x_i = \sigma_i^2,$$

alors la solution $q(\boldsymbol{x})$ reste une fonction gaussienne multivariée séparable :

$$q(\boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{1}{2}\left(\frac{x_i - m_i}{\sigma_i}\right)^2\right].$$

Ce dernier résultat peut être généralisé au cas où $p(\boldsymbol{x})$ est une fonction gaussienne multivariée non dégénérée :

$$p(\boldsymbol{x}) = \frac{|\boldsymbol{\Sigma}|^{-1/2}}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{m})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{m})\right],$$

et si les contraintes sont des contraintes sur les moments d'ordre un et deux non singulières ($\boldsymbol{\Sigma}'$ inversible) :

$$\begin{aligned} \mathrm{E}\left[\boldsymbol{x}\right] &= \int \boldsymbol{x} q(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = \boldsymbol{m}', \\ \mathrm{E}\left[(\boldsymbol{x} - \boldsymbol{m}')(\boldsymbol{x} - \boldsymbol{m}')^t\right] &= \int (\boldsymbol{x} - \boldsymbol{m}')(\boldsymbol{x} - \boldsymbol{m}')^t q(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = \boldsymbol{\Sigma}', \end{aligned}$$

alors la solution $q(\boldsymbol{x})$ reste une fonction gaussienne multivariée non dégénérée :

$$q(\boldsymbol{x}) = \frac{|\boldsymbol{\Sigma}'|^{-1/2}}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{m}')^t \boldsymbol{\Sigma}'^{-1} (\boldsymbol{x} - \boldsymbol{m}')\right].$$

8.4 Utilisation pour les problèmes inverses

Arrivés à ce stade, nous avons les outils nécessaires pour comprendre comment le principe du maximum d'entropie peut être utilisé dans la résolution des problèmes inverses.

Il existe au moins deux catégories de méthodes qui utilisent le principe du maximum d'entropie pour la résolution des problèmes inverses :

8.4.1 Maximum d'entropie classique

Dans cette approche qui ne s'applique qu'aux objets positifs, l'idée de base consiste à assimiler l'objet $\boldsymbol{f}$ à une distribution de probabilité et les données $\boldsymbol{g}$ à des contraintes linéaires sur celle-ci. Sachant qu'en général, ces contraintes ne sont pas suffisantes pour définir une unique solution au problème, on utilise le PME pour choisir une solution parmi l'ensemble des solutions admissibles qui peuvent être définies soit par

$$\{\boldsymbol{f} : \boldsymbol{H}\boldsymbol{f} = \boldsymbol{g}\},$$

ou par

$$\{\boldsymbol{f} : \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 \leq c\}.$$

On choisit alors dans cet ensemble la solution qui maximise l'entropie

$$-\sum_{j=1}^{N} f_j \log f_j,$$

ou d'une manière plus générale celle qui minimise

$$-\sum_{j=1}^{N} \left[f_j \log\left(\frac{f_j}{m_j}\right) + (f_j - m_j) \right],$$

où $\boldsymbol{m}$ est une solution par défaut (*a priori*).

Notons que dans cette approche on assimile $\boldsymbol{f}$ à une distribution de probabilité et les données $\boldsymbol{g}$ à des contraintes linéaires sur celle-ci.

8.4.2 Maximum d'entropie sur la moyenne

Dans cette approche

- On fait l'hypothèse que $\boldsymbol{f} \in \mathcal{C}$ où $\mathcal{C}$ est un ensemble convexe muni d'une mesure de référence $\mu(\boldsymbol{f})$ (ou d'une densité de probabilité *a priori* $\mathrm{d}\mu(\boldsymbol{f}) = p_0(\boldsymbol{f})\ \mathrm{d}\boldsymbol{f}$ et on suppose que

$$\begin{aligned} \boldsymbol{m} &= \int_{\mathcal{C}} \boldsymbol{f}\ \mathrm{d}\mu(\boldsymbol{f}) \\ &= \int_{\mathcal{C}} \boldsymbol{f}\, p_0(\boldsymbol{f})\, \mathrm{d}\boldsymbol{f}. \end{aligned}$$

- On considère $\boldsymbol{f}$ comme la moyenne d'un vecteur aléatoire $\boldsymbol{F}$ pour lequel on définit une loi de probabilité $P(\boldsymbol{f})$ (ou une densité de probabilité $\mathrm{d}P(\boldsymbol{f}) = p(\boldsymbol{f})\ \mathrm{d}\boldsymbol{f}$:

$$\begin{aligned} \boldsymbol{f} = \mathrm{E}\,[\boldsymbol{F}] &= \int_{\mathcal{C}} \boldsymbol{f}\ \mathrm{d}P(\boldsymbol{f}) \\ &= \int_{\mathcal{C}} \boldsymbol{f}\, p(\boldsymbol{f})\, \mathrm{d}\boldsymbol{f}, \end{aligned}$$

et les données $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$ comme des contraintes linéaires sur la loi de probabilité $p(\boldsymbol{f})$:

$$\begin{aligned}\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} = \boldsymbol{H}\mathrm{E}[\boldsymbol{F}] = \mathrm{E}[\boldsymbol{H}\boldsymbol{F}] &= \int_{\mathcal{C}} \boldsymbol{H}\boldsymbol{f}\ \mathrm{d}P(\boldsymbol{f}) \\ &= \int_{\mathcal{C}} \boldsymbol{H}\boldsymbol{f}\, p(\boldsymbol{f})\,\mathrm{d}\boldsymbol{f}.\end{aligned}$$

– On cherche alors parmi l'ensemble des lois de probabilités satisfaisant ces contraintes celle qui minimise la distance de Kulback entre $\mathrm{d}P(\boldsymbol{f})$ et $\mathrm{d}\mu(\boldsymbol{f})$:

$$-\int \log \frac{\mathrm{d}P(\boldsymbol{f})}{\mathrm{d}\mu(\boldsymbol{f})}\,\mathrm{d}P(\boldsymbol{f}),$$

ou, d'une manière équivalente, celle qui minimise l'entropie relative

$$-\int p(\boldsymbol{f}) \log\left(\frac{p(\boldsymbol{f})}{p_0(\boldsymbol{f})}\right)\,\mathrm{d}\boldsymbol{f}.$$

La solution, nous l'avons vu, peut être obtenue via le Lagrangien :

$$\begin{aligned}\mathcal{L}(\boldsymbol{f},\boldsymbol{\lambda}) &= \int_{\mathcal{C}} \left[\log \frac{\mathrm{d}P(\boldsymbol{f})}{\mathrm{d}\mu(\boldsymbol{f})} - \sum_{i=1}^{M} \lambda_i (g_i - [\boldsymbol{H}\boldsymbol{f}]_i)\right]\,\mathrm{d}P(\boldsymbol{f}) \\ &= \int_{\mathcal{C}} \left[\log \frac{\mathrm{d}P(\boldsymbol{f})}{\mathrm{d}\mu(\boldsymbol{f})} - \boldsymbol{\lambda}^t(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f})\right]\,\mathrm{d}P(\boldsymbol{f}) \\ &= \int_{\mathcal{C}} \left[\log \frac{p(\boldsymbol{f})}{p_0(\boldsymbol{f})} - \boldsymbol{\lambda}^t(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f})\right] p(\boldsymbol{f})\,\mathrm{d}\boldsymbol{f}\end{aligned}$$

et donnée par

$$\mathrm{d}P(\boldsymbol{f},\boldsymbol{\lambda}) = \exp\left[\boldsymbol{\lambda}^t[\boldsymbol{H}\boldsymbol{f}] - \log Z(\boldsymbol{\lambda})\right]\,\mathrm{d}\mu(\boldsymbol{f}),$$

ou d'une manière équivalente

$$p(\boldsymbol{f},\boldsymbol{\lambda}) = \exp\left[\boldsymbol{\lambda}^t[\boldsymbol{H}\boldsymbol{f}] - \log Z(\boldsymbol{\lambda})\right] p_0(\boldsymbol{f})\,\mathrm{d}\boldsymbol{f},$$

où

$$Z(\boldsymbol{\lambda}) = \int_{\mathcal{C}} \exp\left[\boldsymbol{\lambda}^t[\boldsymbol{H}\boldsymbol{f}]\right]\,\mathrm{d}\mu(\boldsymbol{f}) = \int_{\mathcal{C}} \exp\left[\boldsymbol{\lambda}^t[\boldsymbol{H}\boldsymbol{f}]\right]\,p_0(\boldsymbol{f})\,\mathrm{d}\boldsymbol{f}.$$

Les paramètres de Lagrange sont obtenus en cherchant la solution du système d'équations non linéaires suivant :

$$\frac{\partial \log Z(\boldsymbol{\lambda})}{\partial \lambda_i} = g_i, \quad i = 1, \cdots, M.$$

– Une fois $p(\boldsymbol{f})$ déterminée on définit la solution du problème par

$$\widehat{\boldsymbol{f}} = \mathrm{E}[\boldsymbol{f}] = \int \boldsymbol{f}\,\mathrm{d}P(\boldsymbol{f}) = \int \boldsymbol{f}\,p(\boldsymbol{f})\,\mathrm{d}\boldsymbol{f}.$$

On peut montrer que, cette solution $\widehat{\boldsymbol{f}}$, qui dépend ainsi de $\boldsymbol{\lambda}$, peut être obtenue directement en minimisant un critère $\Omega(\boldsymbol{f},\boldsymbol{m})$ sous les contraintes $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$.

En effet, en notant

$$\boldsymbol{s} = \boldsymbol{H}^t\boldsymbol{\lambda}, \quad G^*(\boldsymbol{s}) = \log Z(\boldsymbol{s}) = \log \int_{\mathcal{C}} \exp\left[\boldsymbol{s}^t\boldsymbol{f}\right]\,\mathrm{d}\mu(\boldsymbol{f}),$$

et

$$\Omega(\boldsymbol{f}) = \inf_{\boldsymbol{s}}\{\boldsymbol{s}^t\boldsymbol{f} - G^*(\boldsymbol{s})\}, \quad D(\boldsymbol{\lambda}) = \boldsymbol{\lambda}^t\boldsymbol{g} - G^*(\boldsymbol{H}^t\boldsymbol{\lambda}),$$

on peut alors montrer que :

- Le vecteur de $\boldsymbol{\lambda}$ optimal peut être calculé en cherchant l'optimum de $D(\boldsymbol{\lambda})$:

$$\widehat{\boldsymbol{\lambda}} = \arg\max_{\boldsymbol{\lambda}}\{D(\boldsymbol{\lambda})\}$$

  Le critère $D(\boldsymbol{\lambda})$ est appelé *Critère dual*;
- La solution du problème $\widehat{\boldsymbol{f}}$ peut être calculée en cherchant l'optimum d'un critère $\Omega(\boldsymbol{f},\boldsymbol{m})$ appelé *critère primal*:

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}\in\mathcal{C}}\{H(\boldsymbol{f},\boldsymbol{m})\} \quad \text{sous les contraintes} \quad \boldsymbol{g} = \boldsymbol{H}\boldsymbol{f};$$

- Il existe une relation explicite pour calculer la solution en fonction de $G^*(\boldsymbol{s})$:

$$\widehat{\boldsymbol{f}}(\boldsymbol{s}) = \frac{\mathrm{d}G^*(\boldsymbol{s})}{\mathrm{d}\boldsymbol{s}},$$

  et où :
- les fonctions G et Ω dépendent de la mesure de référence $\mu(\boldsymbol{f})$;
- $D(\boldsymbol{\lambda})$ est appelé le *critère dual* et dépend des données $\boldsymbol{g}$ et de la fonction G ;
- $\Omega(\boldsymbol{f},\boldsymbol{m})$ est appelé le *critère primal* et peut être considérer comme une mesure de distance entre $\boldsymbol{f}$ et $\boldsymbol{m}$, ce qui signifie qu'elle satisfait les conditions suivantes :
  - $\Omega(\boldsymbol{f},\boldsymbol{m}) \geq 0, \quad$ et $\quad \Omega(\boldsymbol{f},\boldsymbol{m}) = 0 \quad$ ssi $\quad \boldsymbol{f} = \boldsymbol{m}$;
  - $\Omega(\boldsymbol{f},\boldsymbol{m})$ est continue et convexe sur $\mathcal{C}$;
  - $\Omega(\boldsymbol{f},\boldsymbol{m}) = \infty \quad$ si $\boldsymbol{f} \notin \mathcal{C}$.

  Lorsque la mesure de référence est séparable

$$\mu(\boldsymbol{f}) = \prod_{j=1}^{N} \mu_j(f_j),$$

on a aussi :

$$\mathrm{d}P(\boldsymbol{f},\boldsymbol{\lambda}) = \prod_{j=1}^{N} \mathrm{d}P_j(f_j,\boldsymbol{\lambda}),$$

et

$$G(\boldsymbol{s}) = \sum_j g_j\left(s_j\right), \quad \Omega(\boldsymbol{f},\boldsymbol{m}) = \sum_j h_j(f_j,m_j), \quad \widehat{x}_j = g_j'(s_j).$$

En remplaçant $\boldsymbol{s} = \boldsymbol{H}^t\boldsymbol{\lambda}$ on obtient :

$$G(\boldsymbol{\lambda}) = \sum_j g_j\left([\boldsymbol{H}^t\boldsymbol{\lambda}]_j\right), \quad \Omega(\boldsymbol{f},\boldsymbol{m}) = \sum_j h_j(f_j,m_j), \quad \widehat{x}_j = g_j'\left([\boldsymbol{H}^t\widehat{\boldsymbol{\lambda}}]_j\right).$$

On peut voir que $\Omega(\boldsymbol{f},\boldsymbol{m})$ est aussi séparable et que h_j et g_j dépendent de la mesure de référence $\mu_j(f_j)$:

- g_j est la transformée de Cramer (log de la transformée de LAPLACE) de μ_j:

$$g_j(s) = \log\int \exp\left[sf_j\right]\,\mathrm{d}\mu_j(f_j);$$

– h_j est la convexe conjugée de g_j: $\quad h_j(f) = \inf_s\{sf - g_j(s)\}$.

Prenons quelques exemples:

	$\mu_j(f)$	$g_j(s)$	$h_j(f,m)$
Gaussien:	$\exp\left[-\frac{1}{2}(f-m)^2\right]$	$\frac{1}{2}(s-m)^2$	$\frac{1}{2}(f-m)^2$
Poisson:	$\frac{m^f}{f!}\exp[-m]$	$\exp[m-s]$	$-f\log\frac{f}{m}+m-f$
Gamma:	$f^{\alpha-1}\exp\left[-\frac{f}{m}\right]$	$\log(s-m)$	$-\log\frac{f}{m}+\frac{f}{m}-1$

Ainsi, si on considère le critère primal:

$$\text{maximiser} \quad \Omega(\boldsymbol{f},\boldsymbol{m}) = \sum_{j=1}^{N} h_j(f_j,m_j) \quad \text{sous contraintes} \quad \boldsymbol{H}\boldsymbol{f} = \boldsymbol{g},$$

on peut par exemple remarquer que lorsque la mesure de référence est une mesure poissonnienne on a $h_j(f_j,m_j) = -f_j \ln f_j/m_j + (m_j - f_j)$. On peut ainsi faire le parallèle avec la méthode du ME classique.

Cependant, lorsque la mesure de référence n'est pas séparable ou, lorsqu'il y a de l'incertitude sur les données $\boldsymbol{g}$, il n'est plus aussi facile d'utiliser cette approche. Mentionnons quand même que cette approche est récente, et des travaux tout à fait récents ont essayé de prendre en compte le bruit. Ici nous en présentons une approche qui consiste à remplacer $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{n}$ par

$$\boldsymbol{g} = [\boldsymbol{H}|\boldsymbol{I}]\begin{bmatrix} \boldsymbol{f} \\ \boldsymbol{n} \end{bmatrix} \longrightarrow \boldsymbol{g} = \tilde{\boldsymbol{H}}\,\tilde{\boldsymbol{f}}$$

et faire l'hypothèse d'indépendance entre $\boldsymbol{b}$ et $\boldsymbol{f}$:

$$\mu(\tilde{\boldsymbol{f}}) = \mu_x(\boldsymbol{f})\mu_n(\boldsymbol{n})$$

Ensuite, suivant l'approche, on montre que la solution de l'optimisation du critère primal devient:

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}\in\mathcal{C}}\{\mathcal{Q}(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}) + \alpha\Omega(\boldsymbol{f},\boldsymbol{m})\}$$

avec

$$\Omega(\boldsymbol{f},\boldsymbol{m}) = \sum_{j=1}^{N} h_j(f_j,m_j), \qquad \text{et} \qquad \mathcal{Q}(\boldsymbol{z}) = \sum_{i=1}^{M} q_i(z_j).$$

Là encore, les fonctions $h_j(f_j)$ et $q_i(z_i)$ dépendent des mesures de références $\mu_f(\boldsymbol{f})$ et $\mu_n(\boldsymbol{f})$.

Chapitre 9

Approche bayésienne

L'approche bayésienne est une approche cohérente pour la résolution des problèmes inverses car elle permet de prendre en compte et de traiter de la même manière l'information *a priori* sur la solution et celle sur les données.

En effet, cette approche permet non seulement de prendre en compte le caractère incertain des erreurs au travers de l'attribution d'une loi de probabilité aux mesures, mais aussi l'information *a priori* au travers d'une loi de probabilité *a priori* que l'on attribue aux inconnues du problème.

La fusion de ces deux sources d'information s'effectue par la règle de Bayes. On peut alors dire que cette approche généralise la régularisation déterministe.

9.1 Introduction

L'approche bayésienne correspond formellement au cheminement suivant :

1. On explicite un ensemble d'hypothèses $\mathcal{H}$ sur le problème concernant le modèle d'observation, les connaissances *a priori* sur les inconnues $\boldsymbol{f}$ et sur le bruit $\boldsymbol{b}$.
2. On attribue une loi de probabilité *a priori* $p(\boldsymbol{f}|\boldsymbol{\theta}_1,\mathcal{H})$ aux inconnues du problème pour traduire la connaissance initiale sur ces inconnues. Cette loi peut dépendre d'un certain nombre de paramètres $\boldsymbol{\theta}_1$.
3. On attribue une loi de probabilité $p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta}_2;\mathcal{H})$ aux grandeurs mesurées pour traduire l'incertitude (due au bruit, aux erreurs de discrétisation et de quantification, au manque de précision de l'appareil de mesure, etc.) sur ces données. Cette loi aussi peut dépendre d'un certain nombre de paramètres $\boldsymbol{\theta}_2$. L'ensemble des paramètres $\boldsymbol{\theta} = (\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ sont appelés les hyperparamètres du problème.
4. On utilise la règle de Bayes pour combiner l'information contenue dans les données et celle contenue dans la loi *a priori* et calculer ainsi la loi de probabilité *a posteriori*

$$p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta};\mathcal{H}) = \frac{p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta}_2;\mathcal{H})\, p(\boldsymbol{f}|,\boldsymbol{\theta}_1,\mathcal{H})}{p(\boldsymbol{g}|\boldsymbol{\theta};\mathcal{H})}$$

où

$$p(\boldsymbol{g}|\boldsymbol{\theta};\mathcal{H}) = \int p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta}_2;\mathcal{H})\, p(\boldsymbol{f}|\boldsymbol{\theta}_1,\mathcal{H})\, \mathrm{d}\boldsymbol{f}.$$

Cette loi contient toute l'information disponible sur les inconnues $\boldsymbol{f}$ après cette fusion.
5. On peut alors calculer n'importe quelle quantité. Par exemple, la moyenne *a posteriori* de $\boldsymbol{f}$:

$$\mathrm{E}\,[\boldsymbol{f}] = \int \boldsymbol{f}\, p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta};\mathcal{H})\, \mathrm{d}\boldsymbol{f},$$

ou la moyenne de n'importe quelle autre fonction $h(\boldsymbol{f})$:

$$\mathrm{E}\,[h(\boldsymbol{f})] = \int h(\boldsymbol{f})\, p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta};\mathcal{H})\, \mathrm{d}\boldsymbol{f},$$

ou encore calculer la probabilité pour que $\underline{\boldsymbol{f}} < \boldsymbol{f} \leq \overline{\boldsymbol{f}}$:

$$P(\underline{\boldsymbol{f}} < \boldsymbol{f} \leq \overline{\boldsymbol{f}}) = \int_{\underline{\boldsymbol{f}}}^{\overline{\boldsymbol{f}}} p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta};\mathcal{H})\, \mathrm{d}\boldsymbol{f}$$

ou seulement s'intéresser à une des composantes f_j de $\boldsymbol{f}$ et calculer

$$P(\underline{f}_j < f_j \leq \overline{f}_j) = \int_{\underline{f}_j}^{\overline{f}_j} p(f_j|\boldsymbol{g},\boldsymbol{\theta};\mathcal{H})\, \mathrm{d}f_j,$$

où

$$p(f_j|\boldsymbol{g},\boldsymbol{\theta};\mathcal{H}) = \int \cdots \int p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta};\mathcal{H})\, \mathrm{d}f_1 \cdots \mathrm{d}f_{j-1}\, \mathrm{d}f_{j+1} \cdots \mathrm{d}f_n$$

est la loi *a posteriori marginale*.
Lorsque $\boldsymbol{f}$ est un scalaire ou un vecteur avec seulement deux composantes, on peut même tracer $p(f|\boldsymbol{g},\boldsymbol{\theta};\mathcal{H})$ ou $p(f_1,f_2|\boldsymbol{g},\boldsymbol{\theta};\mathcal{H})$, ce qui permettra d'avoir une idée plus précise sur la forme de ces distribution.

Mais en pratique, la manipulation d'une densité de probabilité n'est pas très commode dès lors qu'on a plus de deux variables. C'est pourquoi, on se contente souvent de fournir une description sommaire de la loi *a posteriori* en indiquant

– son mode :

$$\mathrm{M}[\boldsymbol{f}] = \arg\max_{\boldsymbol{f}} \{p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta};\mathcal{H})\},$$

– sa moyenne :

$$\mathrm{E}[\boldsymbol{f}] = \int \boldsymbol{f}\, p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta};\mathcal{H})\,\mathrm{d}\boldsymbol{f},$$

– les modes de ses marginales :

$$\mathrm{M}[f_j] = \arg\max_{f_j} \{p(f_j|\boldsymbol{g},\boldsymbol{\theta};\mathcal{H})\},$$

– les α-quantiles de ses marginales :

$$Q_\alpha(f_j) \ : P(f_j \leq Q_\alpha(f_j)) = \alpha,$$

– les médianes de ses marginales :

$$\mathrm{Méd}[f_j] = Q_{\frac{1}{2}}(f_j) \ : \ P\left(f_j \leq \mathrm{Méd}[f_j]\right) = \frac{1}{2},$$

– les régions de plus haute probabilité ou les supports α–interquantiles :

$$[a,b] = \left[Q_{(1-\alpha)/2}(f_j), Q_{(1+\alpha)/2}(f_j)\right] \ : P(a \leq f_j \leq b) = 1-\alpha,$$

– les variances :

$$\mathrm{Var}[f_j] = \int (f_j - \bar{f}_j)^2\, p(f_j|\boldsymbol{g},\boldsymbol{\theta};\mathcal{H}),$$

où

$$\bar{f}_j = \mathrm{E}[f_j] = \int f_j\, p(f_j|\boldsymbol{g},\boldsymbol{\theta};\mathcal{H}),$$

– les covariances :

$$\mathrm{Cov}[f_j,f_k] = \iint (f_j - \bar{f}_j)(f_k - \bar{f}_k)\, p(f_j,f_k|\boldsymbol{g},\boldsymbol{\theta};\mathcal{H})\,\mathrm{d}f_j\,\mathrm{d}f_k.$$

6. Une autre approche fondée sur la théorie de la décision consiste en la définition d'un estimateur $\widehat{\boldsymbol{f}}$ qui minimiserait la moyenne d'une fonction de coût $C(\widehat{\boldsymbol{f}},\boldsymbol{f})$. Nous reviendront sur cette approche et les choix de ces fonctions coût, ainsi que sur les liens qu'existe entre ces estimateurs et les caractéristiques de la loi *a posteriori*.
7. Finalement, une fois une solution choisie, il est impératif d'y attribuer un degré de confiance ainsi que de déterminer la sensibilité de cette solution vis-à-vis des erreurs de mesures et celles de la modélisation. Heureusement, en principe, cette approche bayésienne nous fourni tout ce qui est nécessaire pour pouvoir satisfaire cette exigence. Car disposant de la loi *a posteriori* on peut calculer, par exemple, la covariance de l'erreur de l'estimation, ce qui nous permet de mettre des barres d'erreurs sur les solutions estimées.

Voyons maintenant quelles sont les difficultés de l'utilisation de cette approche dans un problème réel et quelles sont les solutions existantes pour les résoudre. On peut classer ces difficultés dans les rubriques qui suivent :

1. choix de la forme de la loi *a priori* $p(\boldsymbol{f}|\boldsymbol{\theta}_1;\mathcal{H})$;

2. choix ou estimation de ses paramètres $\boldsymbol{\theta}_1$;
3. choix de la forme de la loi $p(\boldsymbol{g}|\boldsymbol{\beta},\boldsymbol{\theta}_2;\mathcal{H})$;
4. choix ou estimation de ses paramètres $\boldsymbol{\theta}_2$;
5. calcul effectif de la loi *a posteriori* $p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta};\mathcal{H})$;
6. choix d'une fonction de coût $C(\widehat{\boldsymbol{f}},\boldsymbol{f})$ et calcul effectif de l'estimateur qu'en découle ;
7. détermination ou estimation *a posteriori* des hyperparamètres du problème, *i.e.*les paramètres $\boldsymbol{\theta} = (\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ des lois de probabilités directes $p(\boldsymbol{f}|\boldsymbol{\theta}_1;\mathcal{H})$ et $p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta}_2;\mathcal{H})$, à partir des données, c'est-à-dire en même temps que la solution ;

L'objectif principal de ce chapitre est de faire état de l'art concernant les solutions proposées pour résoudre ces problèmes et de fournir quelques exemples concrets (étude de cas) pour illustrer les difficultés et les performances des solutions proposées.

Mais, avant d'aller plus loin prenons un exemple, désormais classique, qui est le cas linéaire gaussien qui nous permettra de comprendre l'essentiel de ces différentes étapes.

9.2 Cas linéaire gaussien

S'il y a un cas où les calculs peuvent se faire facilement jusqu'au bout et où l'on peut facilement comprendre l'approche bayésienne est le cas où le modèle reliant les inconnues $\boldsymbol{f}$ et les données $\boldsymbol{g}$ est linéaire et où on peut attribuer des lois de probabilités gaussiennes à $\boldsymbol{f}$ et aux mesures $\boldsymbol{g}$.

Nous allons suivre les étapes présentées en introduction pour la résolution du problème linéaire

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{b} \tag{9.1}$$

- Information *a priori* sur $\boldsymbol{f}$:
  Faisons l'hypothèse que nous puissions *a priori* connaître seulement la moyenne $\mathrm{E}[\boldsymbol{f}] = \boldsymbol{f}_0$ et la matrice de covariance $\mathrm{E}\left[(\boldsymbol{f}-\boldsymbol{f}_0)(\boldsymbol{f}-\boldsymbol{f}_0)^t\right] = \boldsymbol{R}_f = \sigma_f^2\boldsymbol{P}_0$ de $\boldsymbol{f}$. Nous pouvons alors faire l'hypothèse que $p(\boldsymbol{f}) = \mathcal{N}(\boldsymbol{f}_0,\sigma_f^2\boldsymbol{R}_f)$:

$$p(\boldsymbol{f}) = A\exp\left[\frac{-1}{2}(\boldsymbol{f}-\boldsymbol{f}_0)^t\boldsymbol{R}_f^{-1}(\boldsymbol{f}-\boldsymbol{f}_0)\right] = A\exp\left[\frac{-1}{2\sigma_f^2}(\boldsymbol{f}-\boldsymbol{f}_0)^t\boldsymbol{P}_0^{-1}(\boldsymbol{f}-\boldsymbol{f}_0)\right] \tag{9.2}$$

  avec $A = (2\pi)^{-n/2}\,|\boldsymbol{R}_f|^{-1/2} = (2\pi\sigma_f^2)^{-n/2}\,|\boldsymbol{P}_0|^{-1/2}$.
  D'ailleurs, cette hypothèse peut être validée par le principe du maximum d'entropie.
- Information *a priori* sur $\boldsymbol{b}$:
  Faisons l'hypothèse que $\mathrm{E}[\boldsymbol{b}] = \boldsymbol{0}$ et que sa matrice de covariance $\mathrm{E}\left[\boldsymbol{b}\boldsymbol{b}^t\right] = \boldsymbol{R}_b = \sigma_b^2\boldsymbol{I}$. Cette hypothèse est bien raisonnable. En effet, supposer que $\mathrm{E}[\boldsymbol{b}] = \boldsymbol{0}$ signifie que nous faisons l'hypothèse qu'il n'y a pas d'erreur systématique et l'hypothèse $\mathrm{E}\left[\boldsymbol{b}\boldsymbol{b}^t\right] = \sigma_b^2\boldsymbol{I}$ signifie que le bruit est supposé blanc (non corrélé).
  Avec les mêmes arguments que dans le cas précédent on fait alors l'hypothèse que $p(\boldsymbol{b}) = \mathcal{N}(\boldsymbol{0},\sigma_b^2\boldsymbol{R}_f)$. Maintenant, partant du modèle (9.1) et faisant l'hypothèse que le bruit $\boldsymbol{b}$ est aussi indépendant du $\boldsymbol{f}$ on peut facilement déduire

$$p(\boldsymbol{g}|\boldsymbol{f}) = B\exp\left[\frac{-1}{2}(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f})^t\boldsymbol{R}_b^{-1}(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f})\right] = B\exp\left[\frac{-1}{2\sigma_b^2}(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f})^t(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f})\right] \tag{9.3}$$

  avec $B = (2\pi)^{-m/2}\,|\boldsymbol{R}_b|^{-1/2} = (2\pi\sigma_b^2)^{-m/2}$.
- Règle de Bayes :
  Utilisant la règle de Bayes, il est facile de voir que

$$p(\boldsymbol{f}|\boldsymbol{g}) = C\exp\left[\frac{-1}{2}J(\boldsymbol{f})\right], \tag{9.4}$$

  avec

$$\begin{aligned} J(\boldsymbol{f}) &= (\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f})^t\boldsymbol{R}_b^{-1}(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}) + (\boldsymbol{f}-\boldsymbol{f}_0)^t\boldsymbol{R}_f^{-1}(\boldsymbol{f}-\boldsymbol{f}_0) \\ &= \frac{1}{\sigma_b^2}\left[(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f})^t(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}) + \lambda(\boldsymbol{f}-\boldsymbol{f}_0)^t\boldsymbol{P}_0^{-1}(\boldsymbol{f}-\boldsymbol{f}_0)\right], \end{aligned}$$

  où $\lambda = \frac{\sigma_b^2}{\sigma_f^2}$.
  Avec un peu de calcul il n'est pas difficile de montrer que cette loi *a posteriori* est aussi gaussienne et on a

$$p(\boldsymbol{f}|\boldsymbol{g}) = \mathcal{N}(\widehat{\boldsymbol{f}},\widehat{\boldsymbol{P}})$$

où

$$\left\{\begin{array}{rcl}\widehat{\boldsymbol{f}} &=& \boldsymbol{R}_f\boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{R}_f\boldsymbol{H}^t+\boldsymbol{R}_b)^{-1}\boldsymbol{g}=\widehat{\boldsymbol{P}}\boldsymbol{H}^t\boldsymbol{R}_b^{-1}\boldsymbol{g},\\ \widehat{\boldsymbol{P}} &=& \boldsymbol{R}_f-\boldsymbol{R}_f\boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{R}_f\boldsymbol{H}^t+\boldsymbol{R}_b)^{-1}\boldsymbol{H}\boldsymbol{R}_f=(\boldsymbol{R}_f^{-1}+\boldsymbol{H}^t\boldsymbol{R}_b^{-1}\boldsymbol{H})^{-1},\end{array}\right.$$

ou encore

$$\left\{\begin{array}{rcl}\widehat{\boldsymbol{f}} &=& \left(\boldsymbol{H}^t\boldsymbol{H}+\lambda\boldsymbol{P}_0^{-1}\right)^{-1}\boldsymbol{H}^t\boldsymbol{g}`,\\ \widehat{\boldsymbol{P}} &=& \sigma_b^2\left(\boldsymbol{H}^t\boldsymbol{H}+\lambda\boldsymbol{P}_0^{-1}\right)^{-1}.\end{array}\right.$$

- Caractéristique de la solution :
  Nous avons ainsi une expression analytique pour la loi *a posteriori* et nous pouvons facilement en déduire tout ce que l'on souhaite. Par exemple nous savons que la moyenne, la médiane et la mode sont identiques et égaux à $\widehat{\boldsymbol{f}}$.

Ainsi tous ces estimateurs : le maximum *a posteriori* (MAP) ou la moyenne *a posteriori* (MP) sont identiques et égaux à $\widehat{\boldsymbol{f}}$ qui peut aussi être calculé par

$$\widehat{\boldsymbol{f}}=\arg\max_{\boldsymbol{f}}\{p(\boldsymbol{f}|\boldsymbol{g})\}=\arg\min_{\boldsymbol{f}}\{J(\boldsymbol{f})=Q(\boldsymbol{f})+\lambda\Omega(\boldsymbol{f})\}$$

avec

$$Q(\boldsymbol{f})=(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f})^t(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f})$$

et

$$\Omega(\boldsymbol{f})=(\boldsymbol{f}-\boldsymbol{f}_0)^t\boldsymbol{P}_0^{-1}(\boldsymbol{f}-\boldsymbol{f}_0),$$

ou encore, si on note par $\boldsymbol{P}_0^{-1}=\boldsymbol{D}^t\boldsymbol{D}$ on a

$$\begin{array}{rcl}J(\boldsymbol{f}) &=& (\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f})^t(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f})+\lambda(\boldsymbol{f}^t-\boldsymbol{f}_0)^t\boldsymbol{D}^t\boldsymbol{D}(\boldsymbol{f}-\boldsymbol{f}_0)\\ &=& \|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|^2+\lambda\|\boldsymbol{D}(\boldsymbol{f}-\boldsymbol{f}_0)\|^2.\end{array}$$

On retrouve alors la notion de solution régularisée et la régularisation quadratique avec les avantages suivantes :

- Dans l'approche régularisation déterministe le choix des distances $\Delta_1(\boldsymbol{g},\boldsymbol{H}\boldsymbol{f})=\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|^2$ et de $\Delta_2(\boldsymbol{f},\boldsymbol{f}_0)=\|\boldsymbol{D}(\boldsymbol{f}-\boldsymbol{f}_0)\|^2$ est plutôt descriptif et un peu arbitraire, alors que dans l'approche bayésienne, ces termes sont, respectivement, les conséquences des hypothèses faites sur la loi du bruit et sur la loi *a priori*.
- Dans régularisation déterministe, le coefficient de régularisation λ est un paramètre qui est choisi empiriquement, alors qu'ici, $\lambda=\frac{\sigma_b^2}{\sigma_f^2}$. Si l'on connait les deux variances σ_b^2 et σ_f^2, sa valeur est fixée. Mais, bien sûr, en pratique on ne connait pas ces deux valeurs et on est encore amené à les fixer par expérience. Cependant, rien que l'expession de λ nous permet de mieux comprendre son comportement : plus le niveau du bruit est important, plus il faut choisir sa valeur grande pour obtenir un résultat satisfaisant. Nous verrons que l'approche bayésienne nous donne les outils nécessaires pour la déterminer (voir le chapitre sur l'estimation des hyperparamètres).
- Dans le cadre de la régularisation déterministe, si on se demande qu'elle est le degré de confiance que l'on peut avoir dans la solution $\widehat{\boldsymbol{f}}$ proposée, nous n'avons pas d'outil convenable pour répondre. Dans l'approche bayésienne, nous pouvons répondre à cette question, par exemple en calculant la matrice de covariance *a posteriori* $\boldsymbol{P}=\mathrm{E}\left[(\boldsymbol{f}-\widehat{\boldsymbol{f}})^2\right]$. Sachant que les éléments diagonaux $P_{jj}=\mathrm{E}\left[(f_j-\widehat{f}_j)^2\right]$ sont les variances *a posteriori*, nous pouvons les utiliser pour mettre des barres d'erreur sur la solution. Les éléments non-diagonale $P_{ij}=\mathrm{E}\left[(f_i-\widehat{f}_i)(f_j-\widehat{f}_j)\right]$ peuvent aussi nous renseigner sur le lien (corrélation) qui peut exister entre l'élément f_i et l'élement f_j.

9.3 Calcul de la loi *a posteriori*

9.3.1 Cas gaussien

Nous avons vu que dans le cas d'un modèle linéaire gaussien, la loi *a posteriori* est aussi gaussienne. Sachant qu'une loi gaussienne est entièrement définie par ses deux premiers moments, le problème devient assez simple. Nous disposons des expressions analytiques pour la moyenne et la matrice de covariance *a posteriori*, expressions dont les calculs nécessitent soit l'inversion des matrices soit l'optimisation d'un critère quadratique. Il reste cependant le coût de calcul. Souvent, on peut profiter de la structure des matrices à inverser ou de la forme quadratique du critère à optimiser pour fournir des algorithmes spécifiques pour effectuer ces calculs.

9.3.2 Cas général

Dans le cas général, le calcul complet de la loi *a posteriori* peut devenir plus délicat, et souvent, on se contente, soit de l'approximer par une loi gaussienne et de calculer la moyenne et la matrice de covariance, soit de définir un estimateur ponctuel à partir de cette loi, comme par exemple le *maximum a posteriori* (MAP) :

$$\widehat{\boldsymbol{f}}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{f}} \{p(\boldsymbol{f}|\boldsymbol{g})\} = \arg\min_{\boldsymbol{f}} \{-\ln p(\boldsymbol{f}|\boldsymbol{g})\},$$

ou la *moyenne a posteriori* (en Anglais posterior mean PM) :

$$\widehat{\boldsymbol{f}}_{\mathrm{PM}} = \int \boldsymbol{f}\, p(\boldsymbol{f}|\boldsymbol{g})\, \mathrm{d}\boldsymbol{f},$$

ou encore le *maximum a posteriori marginal* (MAPmarginal) :

$$\widehat{f}_j = \arg\max_{f_j} \{p(f_j|\boldsymbol{g})\},$$

où

$$p(f_j|\boldsymbol{g},\boldsymbol{\theta};\mathcal{H}) = \int \cdots \int p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta};\mathcal{H})\, \mathrm{d}f_1 \cdots \mathrm{d}f_{j-1}\, \mathrm{d}f_{j+1} \cdots \mathrm{d}f_n.$$

Notons aussi que le calcul des estimateurs PM et MAPmarginal nécessite le calcul des intégrales de dimension très élevée, alors que le calcul de l'estimateur MAP ne nécessite qu'une optimisation.

Notons aussi que dans le cas gaussien tous ces estimateurs sont identiques. En revanche, dans les autres cas, ils peuvent fournir des résultats très différents. Nous reviendrons sur les propriétés de ces estimateurs et les outils qui existent pour les calculer dans les chapitres suivants.

9.4 Choix des fonctions de coût

Dans l'approche théorie de la décision nous avons vu que, partant de la loi *a posteriori*, on peut définir une solution au problème, par l'intermédiaire d'une fonction de coût $C(\widehat{\boldsymbol{f}},\boldsymbol{f})$. En effet, dans cette approche on définit l'estimateur optimal $\widehat{\boldsymbol{f}}$ comme l'argument qui minimise le coût moyen :

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{z}} \{\mathrm{E}\left[C(\boldsymbol{z},\boldsymbol{f})\right]\} = \arg\min_{\boldsymbol{z}} \left\{ \int C(\boldsymbol{z},\boldsymbol{f}) p(\boldsymbol{f}|\boldsymbol{g})\, \mathrm{d}\boldsymbol{f} \right\}.$$

Nous allons voir que suivant le choix de la fonction de coût on obtient des estimateurs différents, et en particulier, les estimateurs MAP, PM et MAPmarginal :

- Coût quadratique et estimateur PM :

$$C(\boldsymbol{f},\widehat{\boldsymbol{f}}) = (\boldsymbol{f}-\widehat{\boldsymbol{f}})^t \boldsymbol{Q} (\boldsymbol{f}-\widehat{\boldsymbol{f}})^t \longrightarrow \widehat{\boldsymbol{f}} = \int \boldsymbol{f}\, p(\boldsymbol{f}|\boldsymbol{g})\, \mathrm{d}\boldsymbol{f}.$$

- Coût dirac et estimateur MAP :

$$C(\boldsymbol{f},\widehat{\boldsymbol{f}}) = 1 - \delta(\boldsymbol{f}-\widehat{\boldsymbol{f}}) \longrightarrow \widehat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \{p(\boldsymbol{f}|\boldsymbol{g})\}.$$

- Coût produit de diracs et estimateur MAPmarginal :

$$C(\boldsymbol{f},\widehat{\boldsymbol{f}}) = \prod_j 1 - \delta(x_j - \widehat{x}_j) \longrightarrow \widehat{f}_j = \arg\max_{x_j} \{p(x_j|\boldsymbol{g})\},$$

D'autres fonctions de coût peuvent aussi être utilisées, mais leur usage reste plus spécifique.

9.5 Choix de la loi *a priori*

Le problème de la conversion d'une information *a priori* en une loi de probabilité est un problème difficile et encore largement ouvert.

La principale difficulé réside dans le fait que rarement, en pratique, l'information *a priori* se présente directement sous une forme probabiliste. Par exemple, on nous dit que la grandeur recherchée est *positive ou bornée entre [0, 1]*, de variation *douce, croissante ou décroissante*, *à bande limitée ou à spectre limitée*, de variation *douce ou constante par morceaux ou par région*, de variation *impulsionnelle*, etc. La tâche, ensuite, est de construire des lois de probabilités qui puissent incorporer ces qualificatifs.

Les méthodes existantes peuvent être schématiquement regroupées en trois grandes classes.

Certaines reposent sur la théorie des *groupes de transformation* pour déterminer la mesure de référence "naturelle" pour la grandeur recherchée. On utilise ces méthodes surtout lorsqu'on sait peu de choses sur la grandeur recherchée, c'est à dire lorsque notre information *a priori* se résume à une connaissance qualitative sur la nature de la grandeur recherchée. Par exemple, nous savons que la grandeur recherchée représente un paramètre de *localisation* pouvant prendre une valeur quelconque sur l'axe des réels $\mathbb{R}$ ou un paramètre *d'échelle* qui est par nature positif et pouvant prendre une valeur quelconque sur la demi-droite des réels positifs $\mathbb{R}_+$.

Mais en pratique, cette approche a permis de justifier après coup l'emploi de la mesure de Lebesgue pour les paramètres de localisation (fournissant ainsi une extension au cas continu de la distribution uniforme résultant de l'application du "Principe d'indifférence" de Bernoulli dans le cas discret) et de la mesure de Jeffreys dans le cas des paramètres d'échelle.

D'autres méthodes reposent sur des principes informationnel. Il s'agit principalement des méthodes dites "à maximum d'entropie" dans lesquelles on recherche une distribution qui soit la plus proche (au sens d'une distance de Kullback) d'une distribution de référence (souvent choisie par l'approche précédente) tout en vérifiant une information incomplète connue *a priori* sous la forme des moments (des contraintes linéaires) sur la loi recherchée. Mais là encore, cette approche a surtout permis de justifier après coup certains choix. De plus, elle n'est véritablement praticable que lorsque cette information *a priori* est faite de contraintes linéaires sur la distribution recherchée. On trouve alors la famille des *distributions exponentielles* comme nous l'avons vu dans l'étude du principe du maximum d'entropie.

Il existe enfin une dernière classe très importante, celle des construction faites "à la main". C'est dans cette catégorie qu'entrent par exemple les modèles markoviens qui ont connu un développement spectaculaire depuis 1984 en traitement d'image, et qui permettent d'incorporer dans une distribution *a priori* des propriétés locales essentielles que doit posséder l'objet. La construction de ces modèles demande beaucoup de savoir-faire.

Bien entendu, il est hors de question de vouloir détailler toutes ces méthodes dans le cadre de ce document. Nous nous limiterons, dans la section suivante, à une description succinte des deux premières approches, plutôt à travers de quelques exemples que d'une manière formelle. Mais, en ce qui concerne la modélisation markovienne, nous lui réservons un chapitre tout entier.

9.5.1 Maximum d'entropie

Nous avons vu dans la section (8.3) que le principe du maximum d'entropie peut être utilisé pour attribuer une loi de probabilité à une grandeur traduisant une information *a priori* sous forme de moments. Nous allons illustrer ceci à travers quelques exemples.

Supposons que nous ayons un voltmètre et que nous cherchions à mesurer la tension de la ville $X(t)$ à différents instants. Supposons que notre instrument soit un voltmètre idéal et ce que nous mesurons $Y(t)$ peut être modélisé par

$$Y(t) = X(t) + B(t)$$

où $B(t)$ représent à la fois les erreurs de lecture et le bruit externe. Nous voudrions appliquer l'approche bayésienne pour fournir une estimation $\widehat{X}(t)$ de $X(t)$. Considérons tout d'abord que nous avons fait juste une mesure y_1 à partir de laquelle que nous voudrions estimer X. Nous avons vu que la première étape avant appliquer la régle de Bayes est d'attribuer les lois de probabilité $p(x)$ et $p(y|x)$.

Concernant X, nous savons que sa moyenne est de 220 volts et que sa variation autour de cette moyenne est ± 5 volts (ceci dépend sûrement de pays et de la qualité des installations, mais supposons que nous soyons en France!). Comment traduire alors cette information en une loi de probabilité? C'est exactement dans ce cadre que le principe du ME peut fournir une réponse satisfaisante :

$$\mathrm{E}\,[X] = x_0 = 220 \text{ volts}, \quad \mathrm{Var}\,[X] = \sigma_x^2 = 25 \text{ volts}^2 \quad \stackrel{\mathrm{ME}}{\longrightarrow} \quad X \simeq \mathcal{N}(x_0,\sigma_x^2)$$

Concernant B, il est naturel de faire l'hypothèse que sa moyenne est nulle (il n'y a pas d'erreur systématique) et supposons que l'on puisse avoir une idée de sa puissance. En langage d'ingénieur : "le bruit est d'environ 10 dB". Là aussi, le principe du ME peut fournir une réponse satisfaisante :

$$\mathrm{E}\,[B] = 0, \quad \mathrm{Var}\,[B] = \sigma_b^2 = 1 \text{ volts}^2 \quad \stackrel{\mathrm{ME}}{\longrightarrow} \quad B \simeq \mathcal{N}(0,\sigma_b^2).$$

Nous avons maintenant tous les ingrédients pour calculer la loi *a posteriori* :

$$X|Y \simeq \mathcal{N}(\widehat{x},\widehat{\sigma}^2),$$

et il n'est pas difficile de calculer $\widehat{x}$ et $\widehat{\sigma}^2$:

$$\widehat{x} = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_b^2} y + \frac{\sigma_b^2}{\sigma_x^2 + \sigma_b^2} x_0$$

$$\widehat{\sigma}^2 = \frac{\sigma_x^2 \sigma_b^2}{\sigma_x^2 + \sigma_b^2}.$$

Notons que

si $\quad \sigma_b^2 = 0 \longrightarrow \widehat{x} = y, \quad \widehat{\sigma}^2 = 0 \quad$ et

si $\quad \sigma_x^2 = 0 \longrightarrow \widehat{x} = x_0, \quad \widehat{\sigma}^2 = 0.$

Considérons maintenant que la grandeur X que nous voulons mesurer est l'énergie ou la puissance consommée dans un circuit et l'appareil de mesure est un wattmètre. Mais, avant même de choisir notre instrument nous savons que la grandeur à mesuer est positive et que son ordre de grandeur est d'environ 200 Watts :

$$X \geq 0, \quad \mathrm{E}\,[X] = x_0 = 200 \text{ Watts} \quad \stackrel{\mathrm{ME}}{\longrightarrow} \quad X \simeq \mathcal{E}(x_0)$$

ce qui signifie :

$$p(x) = x_0 \exp\left[-x/x_0\right], \quad x > 0.$$

Si nous gardons les mêmes hypothèses sur B alors nous avons

$$p(y|x) \propto \exp\left[-\frac{1}{2\sigma_b^2}(y-x)^2\right]$$

ce qui nous donne pour la loi *a posteriori* :

$$p(x|y) \propto \exp\left[-\frac{1}{2\sigma_b^2}\left((y-x)^2 + \frac{2\sigma_b^2}{x_0}x\right)\right], \quad x > 0.$$

En réarrangeant le terme en exponentiel, on peut facilement voir que cette loi est une loi gaussienne tronquée :

$$p(x|y) \propto \exp\left[-\frac{1}{2\sigma_b^2}(x-\widehat{x})^2\right], \quad x > 0, \quad \text{avec} \quad \widehat{x} = y - \frac{\sigma_b^2}{x_0}$$

Revenons maintenant sur le premier exemple et considérons que nous avons fait M mesures aux intervalles réguliers $\boldsymbol{Y} = \{y_1, \cdots, y_M\}$ et que nous voulons estimer $\boldsymbol{X} = \{X_1, \cdots, X_M\}$.

Une première hypothèse consiste à dire qu'*a priori*, il n'y a pas de raison de faire l'hypothèse que la loi de X_i soit différente de la loi de X_j et que X_i soit dépendante de X_j (hypothèse dite i.i.d.) et d'attribuer alors une loi de probabilité jointe :

$$p(\boldsymbol{x}) = \prod_{j=1}^{M} p(x_j) \longrightarrow \boldsymbol{X} \simeq \mathcal{N}(\boldsymbol{x}_0, \sigma_x^2 \boldsymbol{I}).$$

La même hypothèse concernant les éléments de $\boldsymbol{B} = \{B_1, \cdots, B_M\}$ nous permet d'écrire :

$$p(\boldsymbol{b}) = \prod_{j=1}^{M} p(b_j) \longrightarrow \boldsymbol{B} \simeq \mathcal{N}(\boldsymbol{0}, \sigma_b^2 \boldsymbol{I}).$$

Il n'est pas difficile de montrer que la loi *a posteriori* est :

$$p(\boldsymbol{x}|\boldsymbol{y}) = \prod_{j=1}^{M} p(x_j|y_j) \longrightarrow \boldsymbol{X}|\boldsymbol{Y} \simeq \mathcal{N}(\boldsymbol{Y}, \sigma^2 \boldsymbol{I})$$

Une deuxième hypothèse plus réaliste est de dire qu'*a priori*, X_j dépend de X_{j-1} (la valeur de tension à un instant ne peut pas être très différente de sa valeur à un pas d'échantillonnage plus loin). Comment peut-on alors traduire cette information? Supposons que nous puissions avoir une idée *a priori* sur le coéfficient de corrélation entre X_j et X_{j-1}. Ainsi, si on résume tout ce que nous connaissons sur $\boldsymbol{X}$:

$$\mathrm{E}\left[X_j\right] = x_0, \quad \mathrm{Var}\left[X_j\right] = \sigma_x^2, \quad \mathrm{E}\left[(X_j - x_0)(X_{j-1} - x_0)\right] = \beta\sigma_x^2.$$

Utilisant de nouveau le principe du ME, nous obtenons :

$$p(\boldsymbol{x}) \propto \exp\left[\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^t \boldsymbol{P}^{-1}(\boldsymbol{x}-\boldsymbol{x}_0)\right], \quad \text{avec} \quad \boldsymbol{P} = \sigma_x^2 \begin{pmatrix} 1 & \beta & & & & \\ \beta & 1 & \beta & & & \\ & \ddots & \ddots & \ddots & & \\ & & & \beta & 1 & \beta \\ & & & & \beta & 1 \end{pmatrix}$$

Ainsi, d'une manière générale, l'utilisation du principe du maximum d'entropie nous conduit vers les lois exponentielles généralisées. En effet, si on suppose que notre information *a priori* est de la forme :

$$\mathrm{E}\left[\phi_k(\boldsymbol{X})\right] = d_k, \quad k = 1, \cdots, K,$$

nous avons vu que la loi à entropie maximale qui satisfait ces contraintes est de la forme :

$$p(\boldsymbol{x}) = \frac{1}{Z} \exp\left[-\sum_{k=1}^{K} \lambda_k \phi_k(\boldsymbol{x})\right],$$

où

$$Z(\lambda_1, \cdots, \lambda_m) = \int p(\boldsymbol{x}) \exp\left[-\sum_{k=1}^{K} \lambda_k \phi_k(\boldsymbol{x})\right] \mathrm{d}\boldsymbol{x}$$

et où les $\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$ sont déterminés par :

$$-\frac{\partial \ln Z(\lambda_1, \cdots, \lambda_m)}{\partial \lambda_k} = d_k, \quad k = 1, \cdots, K.$$

Un cas particulier intéressant est lorsque les fonctions $\phi_k(x)$ sont séparables :

$$\phi_k(\boldsymbol{x}) = \sum_{j=1}^{N} \phi_{k_j}(x_j)$$

on a alors

$$p(\boldsymbol{x}) = \prod_{j=1}^{N} p(x_j), \quad \text{avec} \quad p(x_j) = \frac{1}{Z_j} \exp\left[-\sum_{k=1}^{K} \lambda_k \phi_{k_j}(x_j)\right],$$

ce qui signifie que les x_j sont indépendants, et si de plus $\phi_{k_j} = \phi_k$ on a

$$p(\boldsymbol{x}) = \prod_{j=1}^{N} p(x_j), \quad \text{avec} \quad p(x_j) = \frac{1}{z} \exp\left[-\sum_{k=1}^{K} \lambda_k \phi_k(x_j)\right],$$

ce qui signifie que les x_j sont i.i.d (indépendants et de même loi).

On trouve parfois l'appellation *lois entropiques* pour les lois appartenant à ces deux derniers cas. Quelques cas particuliers de ces lois utilisées fréquemment sont :

– Gaussienne généralisée :

$$\phi(x_j) = |x_j|^r, 1 < r < 2, \quad \longrightarrow \quad p(x_j) \propto \exp\left[-\alpha |x_j|^r\right].$$

La loi conjointe est alors de la forme :

$$p(\boldsymbol{x}) \propto \exp\left[-\alpha \sum_{j=1}^{N} |x_j|^r\right] \quad \longrightarrow \quad -\ln p(\boldsymbol{x}) = cte + \alpha \sum_{j=1}^{N} |x_j|^r.$$

– Loi Gamma :

$$\phi(x_j) = -\alpha \ln(\frac{x_j}{m_j}) + (\frac{x_j}{m_j}), \quad \longrightarrow \quad p(x_j) \propto (\frac{x_j}{m_j})^{-\alpha} \exp\left[\frac{x_j}{m_j}\right].$$

et

$$p(\boldsymbol{x}) = \prod_{j=1}^{N} p(x_j) \propto \left(\prod_{j=1}^{N} (\frac{x_j}{m_j})^{-\alpha} \right) \exp \left[\sum_{j=1}^{N} \frac{x_j}{m_j} \right]$$

et par conséquent

$$-\ln p(\boldsymbol{x}) = cte - \alpha \sum_{j=1}^{N} \ln(\frac{x_j}{m_j}) + \sum_{j=1}^{N} \frac{x_j}{m_j}.$$

– Loi Béta :

$$\phi(x_j) = \alpha \ln(x_j) + \beta \ln(1 - x_j), \quad \longrightarrow \quad p(x_j) \propto x_j^{\alpha}(1 - x_j)^{\beta}.$$

et la loi conjointe est :

$$p(\boldsymbol{x}) \propto \prod_{j=1}^{N} x_j^{\alpha}(1 - x_j)^{\beta} \quad \longrightarrow \quad -\ln p(\boldsymbol{x}) = cte + \alpha \sum_{j=1}^{N} \ln(x_j) + \beta \sum_{j=1}^{N} \ln(1 - x_j).$$

Nous avons vu que dans le cas i.i.d., la forme générale de ces lois est :

$$p(\boldsymbol{x}) \propto \exp \left[-\sum_{k=1}^{K} \lambda_k \phi_k(\boldsymbol{x}) \right], \quad \text{avec} \quad \phi_k(\boldsymbol{x}) = \sum_{j=1}^{N} \phi_k(x_j).$$

On peut se poser alors une question sur le choix des fonctions $\phi_k(x_j)$. Par exemple, on peut se demander s'il existe des familles de fonctions ϕ_k de telle sorte que la lois $p(x_j)$ et par conséquent la loi $p(\boldsymbol{x})$ aient une propriété, par exemple d'invariance par changement d'échelle. Dans une étude récente, certains auteurs se sont posés de telles questions et ont essayé d'y répondre. Ici, nous résumons quelques uns de ces résultats.

L'idée de base se trouve dans la recherche des fonctions ϕ_k de telle sorte que la famille de lois $p(x_j)$ correspondantes soit fermée pour un groupe de transformations comme le changement d'échelle par exemple.

Prenons le cas du changement d'échelle pour illustrer cette idée. Si x_j représente une quantité et $p(x_j)$ sa loi de probabilité, on voudrait que lors d'un changement d'échelle αx_j avec $\alpha > 0$, la loi $p(\alpha x_j)$ reste dans la même famille que $p(x_j)$. Autrement dit :

$$p(x_j) \propto \exp \left[-\sum_{k=1}^{K} \lambda_k \phi_k(x_j) \right] \longrightarrow p(\alpha x_j) \propto \exp \left[-\sum_{k=1}^{K} \lambda'_k \phi_k(x_j) \right]$$

où λ'_k ne doivent dépendre que des λ_k et de facteur d'échelle α.

Il est alors facile de montrer que pour satisfaire à cette propriété il suffit d'avoir :

$$\forall x_j \ \& \ \forall \alpha > 0, \quad \sum_{k=1}^{K} \lambda_k \phi_k(x_j) = cte + \sum_{k=1}^{K} \lambda'_k \phi_k(x_j)$$

L'étude de cette propriété pour $K = 1$ et $K = 2$ donne les familles suivantes :

Lois à un paramètre $K = 1$:

$$\Big\{ \phi(x) \Big\} = \Big\{ x^r, \ln x \Big\}, \quad r > 0.$$

Lois à deux paramètres $K = 2$:

$$\left\{(\phi_1(x),\phi_2(x))\right\} = \left\{(x^{r_1},x^{r_2}),(x^{r_1},\ln x),(x^{r_1},x^{r_1}\ln x),(\ln x,\ln^2 x)\right\},$$

où r, r_1 et r_2 sont des réels positifs.

On retrouve quelques cas particuliers intéressants en choisissant $\phi_2(x) = x$ et pour différents choix de $\phi_1(x)$:

- $\phi_1(x) = x^2$; on retrouve la loi gaussienne :

$$p(x) \propto \exp\left[-\lambda x^2 - \mu x\right] \propto \exp\left[-\lambda\left(x + \frac{\mu}{2\lambda}\right)^2\right] = \mathcal{N}\left(m = \frac{-\mu}{\lambda},\sigma^2 = \frac{1}{2\lambda}\right).$$

- $\phi_1(x) = \ln x$; on retrouve la loi gamma :

$$p(x) \propto \exp\left[-\lambda \ln x - \mu x\right] = x^{-\lambda}\exp\left[-\mu x\right].$$

- $\phi_1(x) = x \ln x$; on retrouve une loi dite de la forme entropie de Shannon :

$$p(x) \propto \exp\left[-\lambda x \ln x - \mu x\right].$$

9.5.2 Modèles markoviens

La modélisation markovienne a pris une importance sans égale en traitement d'image depuis les travaux de Geman et Geman. L'objet de ce document n'est pas de donner une présentation complète de la modélisation markovienne, mais plutôt une présentation différente.

Nous avons vu dans la section précédente que, d'une manière générale, l'utilisation du principe du maximum d'entropie avec une information *a priori* de la forme:

$$\mathrm{E}\left[\phi_k(\boldsymbol{x})\right] = d_k, \quad k = 1, \cdots, K,$$

nous conduit aux lois de la forme:

$$p(\boldsymbol{x}) = \frac{1}{Z} \exp\left[-\sum_{k=1}^{K} \lambda_k \phi_k(\boldsymbol{x})\right].$$

Considérons maintenant le cas où les fonctions $\phi_k(\boldsymbol{x})$ sont de la forme

$$\phi_k(\boldsymbol{x}) = \sum_j \sum_{i \in V_j} \phi_k(x_j, x_i)$$

où V_j signifie les (échantillons ou les pixels) voisins de j. Les fonctions ϕ_k sont alors appelées *potentiels* et

$$U(\boldsymbol{x}) = \sum_j \sum_{i \in V_j} \phi_k(x_j, x_i)$$

l'énergie.

Un tel choix pour ces fonctions signifie que nous faisons l'hypothèse qu'il y a une corrélation locale entre les valeurs du signal aux instants successifs ou de l'image sur des pixels voisins.

Un cas particulier très courant est lorsque $k = 1$ et $\phi_k(x_j, x_i) = \phi_k(x_j - x_i)$. On a alors

$$p(\boldsymbol{x}) \propto \exp\left[-\lambda \sum_j \sum_{i \in V_j} \phi(x_j - x_i)\right]$$

Le choix de la fonction potentiel ϕ est crucial: les fonctions convexes permettront de modéliser des signaux et des champs continus et les fonctions non convexes permettrons de modéliser des signaux et des champs qui contiennent des discontinuitées. Une liste de ces fonctions classiquement utilisées en traitement du signal et des images est fournie en annexe D.

9.6 Quelques exemples de construction à la main

Dans cette section, nous allons montrer comment on peut choisir une loi de probabilité *a priori* à la main et avec du bon sens. La figure qui suit montre un ensemble de signaux très différents.
Nous allons, pour chacune de ces formes de signaux proposer une famille de lois convenable.

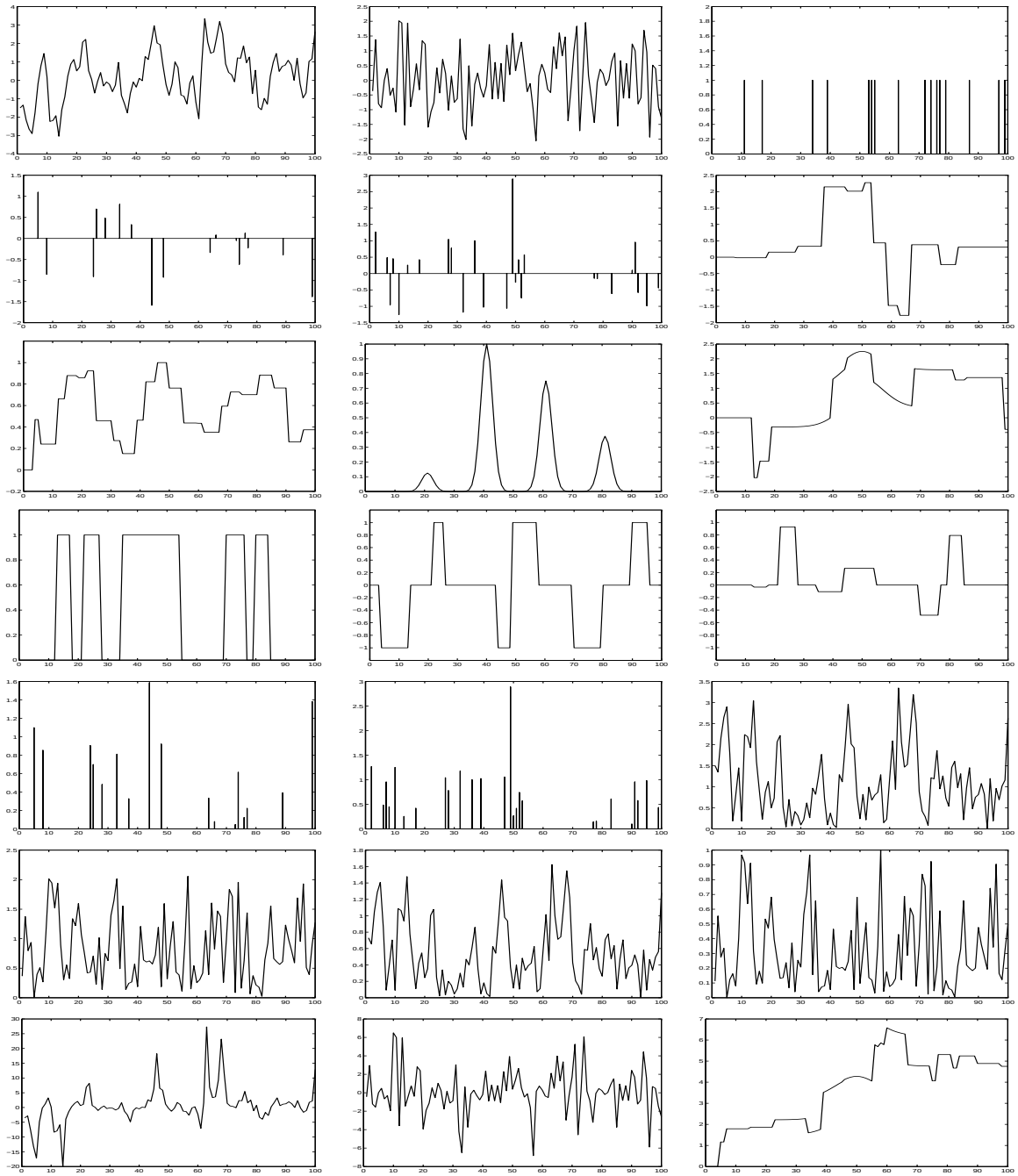
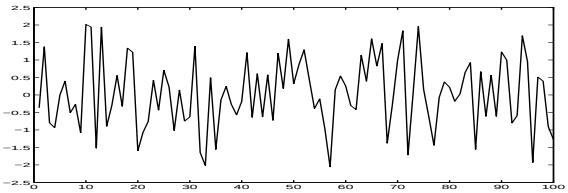


FIG. 9.1 – *Différents types de signaux.*

1. **Modèle gaussien blanc pour des signaux continus à variation rapide**



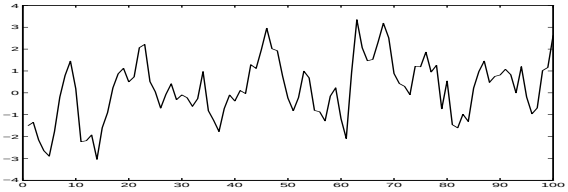
En regardant la forme de ce signal et en constatant que les échantillons du signal sont réparties d'une manière symétrique autours de zéro, on peut proposer une loi gaussienne centrée :

$$\boldsymbol{x} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{R}_x = \sigma_x^2 \boldsymbol{I}\right) \longrightarrow p(\boldsymbol{x}) \propto \exp\left[-\frac{1}{2\sigma_x^2}\|\boldsymbol{x}\|^2\right]$$

ou encore

$$p(\boldsymbol{x}) \propto \exp\left[-\frac{1}{2\sigma_x^2}\sum_j x_j^2\right].$$

2. **Modèle gaussien coloré pour des signaux continus à variation lente :**



En regardant la forme de ce signal, et en la comparant au cas précédent, on peut proposer une loi centré, gaussienne mais colorée :

$$\boldsymbol{x} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{R}_x = \sigma_x^2 \boldsymbol{P}_0\right) \longrightarrow p(\boldsymbol{x}) \propto \exp\left[-\frac{1}{2\sigma_x^2}\boldsymbol{x}^t \boldsymbol{P}_0^{-1} \boldsymbol{x}\right],$$

avec $\boldsymbol{P}_0$ une matrice symétrique et définie positive, ce qui permet d'écrire $\boldsymbol{P}_0^{-1} = \boldsymbol{D}^t\boldsymbol{D}$ et on a alors

$$p(\boldsymbol{x}) \propto \exp\left[-\frac{1}{2\sigma_x^2}\|\boldsymbol{D}\boldsymbol{x}\|^2\right].$$

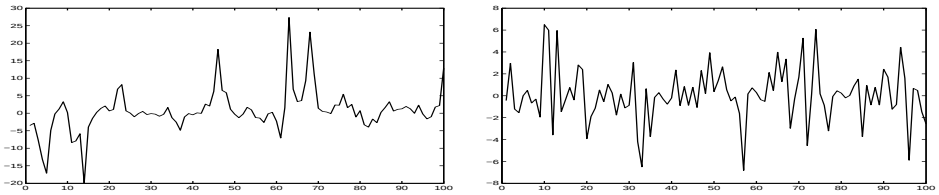
Un choix possible pour $\boldsymbol{D}$ est la matrice de différences finies d'ordre un :

$$\boldsymbol{D}_1 = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ -1 & 1 & \ddots & & \vdots \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & & & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}$$

ce qui permet d'écrire :

$$p(\boldsymbol{x}) \propto \exp\left[-\frac{1}{2\sigma_x^2}\sum_j (x_j - x_{j-1})^2\right]$$

3. **Modèle gaussien généralisé pour des signaux continus plutôt impulsionnels**



En regardant la forme de ce signal, et en la comparant au cas précédent, on peut constater que la proportion des échantillons ayant des valeurs proches de zéro est plus importante que celle des échantillons ayant des valeurs plus grandes et que cette proportion semble décroître très rapidement. On peut alors proposer une loi gaussienne généralisée :

$$p(\boldsymbol{x}) \propto \exp\left[-\beta \|\boldsymbol{D}_k \boldsymbol{x}\|^p\right], \quad 1 < p \leq 2,$$

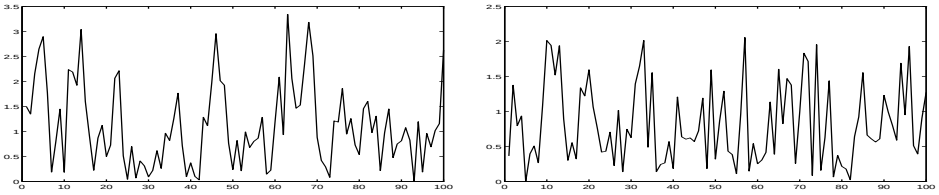
avec $\boldsymbol{D}_0 = \boldsymbol{I}$ et $\boldsymbol{D}_k$ la matrice des différences finies d'ordre k.
Dans le cas $\boldsymbol{D}_0 = \boldsymbol{I}$ on a

$$p(\boldsymbol{x}) \propto \exp\left[-\beta \sum_j |x_j|^p\right]$$

et dans le cas $\boldsymbol{D}_1$ on a

$$p(\boldsymbol{x}) \propto \exp\left[-\beta \sum_j |x_j - x_{j-1}|^p\right].$$

4. **Modèle des signaux positifs**



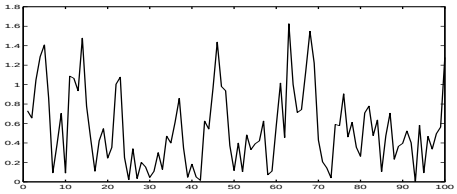
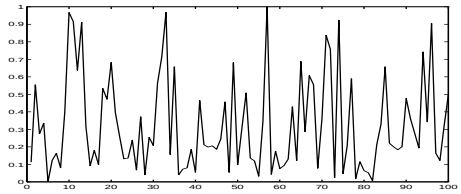
Loi gaussienne généralisée tronquée :

$$p(\boldsymbol{x}) \propto \exp\left[-\beta \sum_j x_j^p\right], \quad x_j > 0.$$

Loi Gamma :

$$p(\boldsymbol{x}) \propto \exp\left[-\alpha \sum_j \ln x_j - \beta \sum_j x_j\right], \quad x_j > 0.$$

5. **Modèle des signaux continus mais bornés entre 0 et 1 : Loi Béta**

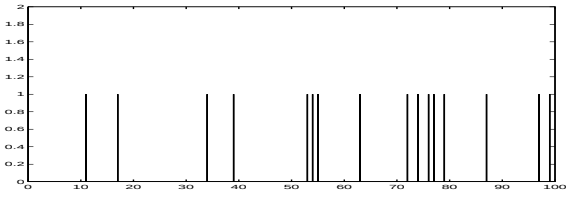
$$p(\boldsymbol{x}) = \prod_{j=1}^{N} p(x_j),$$

avec

$$\begin{aligned} p(x_j) &\propto x_j^{-\alpha} - (1-x_j)^{-\beta} \\ &\propto \exp\left[-\alpha \ln x_j - \beta \ln(1-x_j)\right] \end{aligned}$$

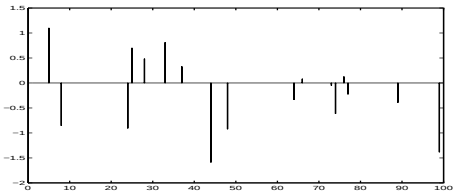
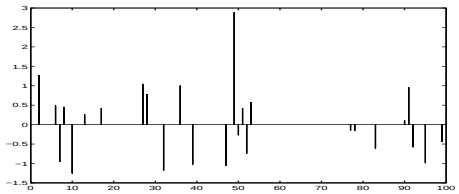
$$-\ln p(\boldsymbol{x}) = cte + \alpha \ln x_j + \beta \ln(1-x_j)$$

6. **Modèle de Bernoulli :**



$$\begin{cases} P(X_j = 1) = \alpha, \\ P(X_j = 0) = 1 - \alpha \end{cases} \longrightarrow P(\boldsymbol{X} = \boldsymbol{x}) = \alpha^s (1-\alpha)^{1-s} \text{ avec } s = \sum(x_j)$$

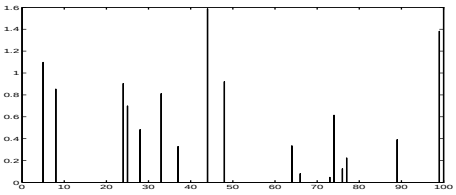
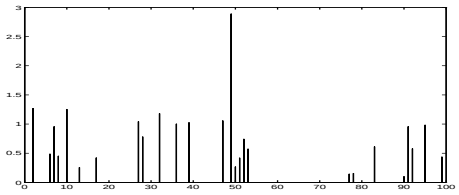
7. **Modèle Bernoulli-Gaussien :**

$$p(x_j|q_j) = \begin{cases} 0 & \text{si } q_j = 0, \\ \mathcal{N}(0,\sigma_x^2) & \text{si } q_j = 1 \end{cases} \quad \text{et} \quad \begin{cases} P(q_j = 1) = \lambda, \\ P(q_j = 0) = 1 - \lambda \end{cases}$$

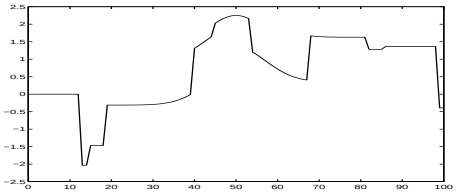
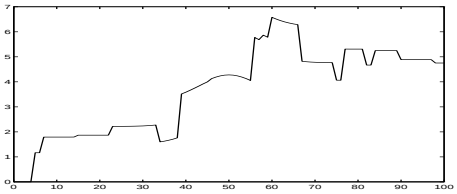
$$p(\boldsymbol{x}|\boldsymbol{q}) \propto \exp\left[\frac{-1}{2q_j\sigma_x^2}\sum_j q_j x_j^2\right]$$

8. **Modèle Bernoulli-Gamma :**

$$p(x_j|q_j) = \begin{cases} 0 & \text{si } q_j = 0, \\ \text{Gamma}(\alpha,\beta) & \text{si } q_j = 1 \end{cases} \quad \text{et} \quad \begin{cases} P(q_j = 1) = \lambda, \\ P(q_j = 0) = 1 - \lambda \end{cases}$$

9. **Modèles pour les signaux continus par morceaux**

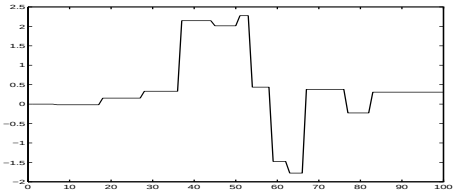
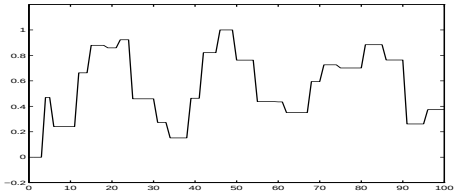



$$p(x_j|x_{j-1},q_j) = \begin{cases} \mathcal{N}(x_{j-1},\sigma_1^2) & \text{si } q_j = 0 \\ \mathcal{N}(0,\sigma_2^2) & \text{si } q_j = 1 \end{cases}$$

ou

$$p(\boldsymbol{x}|\boldsymbol{q}) \propto \exp\left[-\frac{1}{\sigma_1^2}\sum_j q_j(x_j - x_{j-1})\right]$$

10. **Modèles pour les signaux constants par morceaux**

$$p(x_j|x_{j-1},q_j) = \begin{cases} \delta(x_j - x_{j-1}) & \text{si } q_j = 0 \\ \mathcal{N}(0,\sigma_2^2) & \text{si } q_j = 1 \end{cases}$$

9.7 Estimation des paramètres d'une loi à partir des observations directes

Nous avons vu que le PME peut nous permettre d'attribuer une loi de probabilité $p(x)$ à une quantité X pour traduire une information incomplète sur X si cette information porte sur des moments $\mathrm{E}[\phi_k(X)] = d_k$. Malheureusement, en pratique, rarement nous connaissons les valeurs numériques de d_k. C'est pourquoi, en général, l'attribution d'une loi de probabilité se fait en deux étapes :

1. Choix d'une famille de loi qui est équivalent au choix des fonctions ϕ_k ;
2. Détermination des valeurs des paramètres à partir d'un certain nombres d'observations directes ou indirectes du X.

Dans cette section, nous allons nous limiter au cas où nous avons la possibilité d'observer directement X. Supposons maintenant que l'étape 1 est faite et que nous avons choisi une loi $p(x;\boldsymbol{\theta})$ dépendant des paramètres $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_K]$. Considérons alors le cas où nous avons N échantillons $\{x_1, \cdots, x_N\}$. Nous cherchons à déterminer $\boldsymbol{\theta}$ à partir de ces échantillons. Deux méthodes classiquement utilisées sont : la méthode des moments (MM) et la méthode du maximum de vraisemblance (MV).

9.7.1 Méthode des moments

L'idée de base dans cette méthode est de déterminer $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_K)$ en utilisant un système d'équations reliant les moments théoriques $\mathrm{E}\left[X^k\right]$ et les moments empiriques $\overline{X^k} = \frac{1}{N}\sum_{j=1}^{N} x_j^k$:

$$\mathrm{E}\left[X^k\right] = \int x^k\, p(x;\boldsymbol{\theta})\, \mathrm{d}x = \frac{1}{N}\sum_{j=1}^{N} x_j^k, \quad k = 1, \cdots, M$$

avec $M \geq K$ et en espérant qu'il existe une solution unique à ce système d'équations (non linéaires en $\boldsymbol{\theta}$).

Pour illustrer cette méthode considérons les deux exemples suivants :

Exemple 1 :
Cas d'une loi gaussienne à deux paramètres $\boldsymbol{\theta} = [m, \sigma^2]$

$$p(x; m, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-m)^2\right].$$

Si on forme le système d'équations :

$$\begin{aligned} \mathrm{E}[X] &= m = \frac{1}{N}\sum_{j=1}^{N} x_j \\ \mathrm{E}\left[X^2\right] &= \sigma^2 + m^2 = \frac{1}{N}\sum_{j=1}^{N} x_j^2 \end{aligned}$$

il est alors facile de voir qu'il y a une solution qui est :

$$\begin{aligned} m &= \overline{X} = \frac{1}{N}\sum_{j=1}^{N} x_j \\ \sigma^2 &= \overline{(X-m)^2} = \frac{1}{N}\sum_{j=1}^{N} (x_j - m)^2. \end{aligned}$$

Exemple 2 :
Cas d'une loi gamma à deux paramètres $\boldsymbol{\theta} = [\alpha,\beta]$

$$p(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma\alpha} x^{\alpha-1} \exp[-\beta x].$$

Si on forme le système d'équations :

$$\begin{aligned} \mathrm{E}[X] &= \frac{\alpha}{\beta} = \frac{1}{N}\sum_{j=1}^{N} x_j, \\ \mathrm{E}\left[X^2\right] &= \frac{\alpha(1+\alpha)}{\beta^2} = \frac{1}{N}\sum_{j=1}^{N} x_j^2, \end{aligned}$$

et si on note par $m = \overline{X} = \frac{1}{N}\sum_{j=1}^{N} x_j$ et par $v = \overline{(X-m)^2} = \frac{1}{N}\sum_{j=1}^{N}(x_j - m)^2$, il est alors facile de voir qu'il y a une solution qui est :

$$\begin{aligned} \alpha &= \frac{2v - m^2}{v}, \\ \beta &= \frac{m}{v}. \end{aligned}$$

Les inconvenients majeurs de cette méthode sont :

- l'absence de base théorique pour cette méthode ;
- la sensibilité au nombre de données N de ces estimateurs.

9.7.2 Méthode du maximum de vraisemblance

Estimation au sens du MV consiste à définir la fonction de vraisemblance $V(\boldsymbol{\theta}) = p(x_1,\cdots,x_n;\boldsymbol{\theta})$ ou plutôt la fonction log-vraisemblance $L(\boldsymbol{\theta}) = -\ln V(\boldsymbol{\theta})$ et à calculer l'argument $\widehat{\boldsymbol{\theta}}$ qui la maximise :

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \{V(\boldsymbol{\theta})\} = \arg\min_{\boldsymbol{\theta}} \{L(\boldsymbol{\theta})\}$$

Le principal avantage de cet estimateur est qu'il est efficace.

Pour illustrer cette méthode et la comparer à la méthode des moments nous allons considérer les deux exemples de la section précédente.

Exemple 1 :
Cas d'une loi gaussienne à deux paramètres $\boldsymbol{\theta} = [m,\sigma^2]$

$$p(x;m,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-m)^2\right]$$

La fonction du log-vraisemblance s'écrit :

$$L(m,\sigma^2) = \frac{N}{2}\ln(2\pi) + \frac{N}{2}\ln\sigma^2 + \frac{1}{2\sigma^2}\sum_{j=1}^{N}(x_j - m)^2.$$

Pour calculer l'argument qui la minimise, il faut annuler simultanément la dérivée par rapport à m et la dérivée par rapport à σ^2, ce qui permet d'obtenir un résulte bien connu :

$$\begin{aligned} m &= \frac{1}{N}\sum_{j=1}^{N} x_j, \\ \sigma^2 &= \frac{1}{N-1}\sum_{j=1}^{N}(x_j - m)^2. \end{aligned}$$

Notons la seule différence par rapport aux résultats de la méthode des moments apparaît dans l'estimation du σ^2.

Exemple 2 :
Cas d'une loi gamma à deux paramètres $\boldsymbol{\theta} = [\alpha,\beta]$

$$p(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp\left[-\beta x\right].$$

La fonction du log-vraisemblance s'écrit :

$$L(\alpha,\beta) = N[\alpha \ln\beta - \ln\Gamma(\alpha)] - (\alpha-1)\sum_{j=1}^{N} \ln x_j - \beta \sum_{j=1}^{N} x_j.$$

Pour calculer l'argument qui la minimise, il faut annuler simultanément la dérivée par rapport à α et la dérivée par rapport à β :

$$\begin{aligned} \frac{\partial L}{\partial \alpha} &= N\left[\ln\beta - \frac{\partial \ln\Gamma(\alpha)}{\partial\alpha}\right] - \sum_{j=1}^{N} \ln x_j = 0 \\ \frac{\partial L}{\partial \beta} &= \frac{\alpha}{\beta} - \sum_{j=1}^{N} x_j = 0 \end{aligned}$$

Comme on peut le constater, il n'y a pas de solution analytique à ces deux équations, mais on peut la calculer numériquement.

Pour montrer cependant qu'il peut y avoir un lien entre ces deux méthodes, on peut reécrire les relations différemment.

Supposons que la loi $p(x;\boldsymbol{\theta})$ soit de la forme générale :

$$p(x;\boldsymbol{\theta}) \propto \exp\left[-\sum_{k=1}^{K} \theta_k \phi_k(x)\right].$$

On a alors

$$L(\boldsymbol{\theta}) = -\ln p(x_1,\cdots,x_N;\boldsymbol{\theta}) = cte + \sum_{j=1}^{N}\sum_{k=1}^{K} \theta_k \phi_k(x_j).$$

On peut alors montrer facilement que, dans la méthode des moments nous avons à résoudre

$$\int x^k\, p(x;\boldsymbol{\theta})\, \mathrm{d}x = \frac{1}{N}\sum_{j=1}^{N} x_j^k, \quad k = 0,\cdots,K,$$

et dans la méthode du MV nous avons à résoudre

$$\int \phi_k(x)\, p(x;\boldsymbol{\theta})\, \mathrm{d}x = \frac{1}{N}\sum_{j=1}^{N} \phi_k(x_j), \quad k = 0, \cdots, K,$$

où nous avons noté $\phi_0(x) = 1$.
On peut alors remarquer que les deux méthodes sont équivalentess pour la famille des lois exponentielles où $\phi_k(x) = x^k$.

9.8 Estimation des hyperparamètres

Il existe assez peu de méthodes de détermination des hyperparamètres. Dans le cas où ceux-ci se limitent au seul coefficient de régularisation, et où le régulariseur est quadratique, les méthodes de *validation croisée* fournissent des solutions acceptables. Mais il s'agit de méthodes déterministes par essence, qui minimisent un critère de risque dépendant de l'objet inconnu $\boldsymbol{f}$, ce qui ne peut être effectué qu'asymptotiquement puisque cet objet est évidemment inconnu.

Les méthodes bayésienne, qui attribuent une distribution de probabilité *a priori* à l'objet, ne présentent pas cette limitation. Les hyperparamètres θ constituent un second niveau de description du problème, indispensable pour "rigidifier" le premier niveau constitué par les paramètres eux-mêmes— c'est-à-dire l'objet $\boldsymbol{f}$. Dans un problème mal-posé, la valeur des hyperparamètres est importante pour obtenir une solution acceptable, mais ne présente pas d'intérêt en soi. Dans une approche bayésienne, on peut donc distinguer deux niveaux d'inférence. Le premier infère sur $\boldsymbol{f}$, pour une valeur donnée de θ, au travers de la distribution *a posteriori*. Le second infère sur θ grâce à une relation analogue :

$$p(\theta \mid \boldsymbol{g},\boldsymbol{H}) = \frac{p(\theta \mid \boldsymbol{H})\, p(\boldsymbol{g} \mid \theta,\boldsymbol{H})}{p(\boldsymbol{g} \mid \boldsymbol{H})}\,. \tag{9.5}$$

On retrouve là une caractéristique de l'utilisation de la règle de Bayes : la vraisemblance $p(\boldsymbol{g} \mid \theta,\boldsymbol{H})$ attachée aux données dans le second niveau est le coefficient de normalisation dans le premier.

Si, comme cela est souvent le cas, ce terme est suffisamment "piqué", l'influence de la distribution *a priori* $p(\theta \mid \boldsymbol{H})$ est négligeable, et le second niveau d'inférence peut être résolu par maximisation de cette vraisemblance. Mais il faut pour cela résoudre un problème de marginalisation :

$$p(\boldsymbol{g} \mid \theta,\boldsymbol{H}) = \int p(\boldsymbol{f},\boldsymbol{g} \mid \theta,\boldsymbol{H})\, \mathrm{d}\boldsymbol{f} = \int p(\boldsymbol{g} \mid \boldsymbol{f},\theta,\boldsymbol{H})\, p(\boldsymbol{f} \mid \theta)\, \mathrm{d}\boldsymbol{f}\,. \tag{9.6}$$

Une telle intégrale conduit très rarement à une forme explicite.

Pour contourner cette difficulté, on peut introduire des "variables cachées" $\boldsymbol{z}$ qui viennent compléter les observations $\boldsymbol{g}$ de manière à ce que la nouvelle vraisemblance $p(\boldsymbol{g},\boldsymbol{z} \mid \theta,\boldsymbol{H})$ soit plus simple à calculer. On est alors conduit à maximiser des espérances conditionnelles par des techniques itératives, déterministes ou stochastiques (algorithmes EM et SEM).

On peut aussi remarquer que la *vraisemblance généralisée*

$$p(\boldsymbol{g},\boldsymbol{f} \mid \theta,\boldsymbol{H}) = p(\boldsymbol{f} \mid \boldsymbol{g},\theta,\boldsymbol{H})\, p(\boldsymbol{g} \mid \theta,\boldsymbol{H}) = p(\boldsymbol{g} \mid \boldsymbol{f},\theta,\boldsymbol{H})\, p(\boldsymbol{f} \mid \theta) \tag{9.7}$$

résume toute l'information propre au premier niveau d'inférence, et vouloir en faire la maximisation conjointe par rapport à $\boldsymbol{f}$ et θ. Le problème d'intégration soulevé par (9.6) est évidemment évacué. A θ fixé, le maximum de vraisemblance généralisée (MVG) coïncide avec le MAP. Par contre, à $\boldsymbol{f}$ fixé, la situation est beaucoup moins favorable : la vraisemblance généralisée n'est pas, en général, majorée sur son domaine de définition, et il n'existe même pas toujours un maximum local.

A COMPLETER

Chapitre 10

Modélisation markovienne

Dans ce chapitre nous donnons un apérçu général de la modélisation markovienne des signaux et des images. Les modèles markoviens et particulèrement les champs de Markov et les champs de Gibbs sont devenus des outils indispensables pour la traduction d'une information *a priori* plus élaborée que la douceuer ou la positivité en une loi de probabilité. En effet, la modélisation markovienne permet de construire des modèles qui peuvent prendre en compte les discontinuités dans un signal ou une image. Cette outil prend particulièrement son importance en traitement d'image où l'on peut construire les modèles composites (bords, contours, intensité) pour les images et les uitliser lors de la segmentation, restauration ou reconstruction d'image.

10.1 Introduction et notations

10.1.1 Site, voisinage et cliques

Considérons un ensembles de sites $\mathcal{S} = \{s_1, \ldots, s_n\}$ et un système de voisinage $\mathcal{V} = V_s, s \in \mathcal{S}$ sur cet ensemble.

– On dit que le site r est voisin du site s si

$$r \in V_s \Longleftrightarrow s \in V_r, \quad s \notin V_s$$

$\{\mathcal{S},\mathcal{V}\}$ est un graphe.

– $c \subset \mathcal{S}$ est une clique de $\{\mathcal{S},\mathcal{V}\}$ si :

$$|c| = 1 \text{ ou si } |c| > 1 \text{ et } \forall s_1, s_2 \in c \longrightarrow s_1 \text{ et } s_2 \text{ sont voisins}$$

Notons par $\mathcal{C} = \{c\}$ l'ensemble de cliques associées au graphe $\{\mathcal{S},\mathcal{V}\}$.

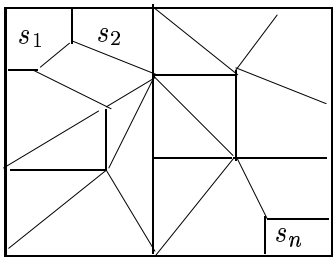
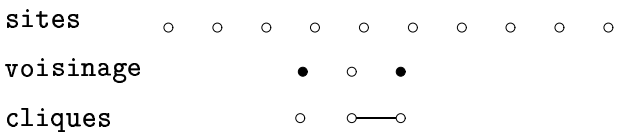


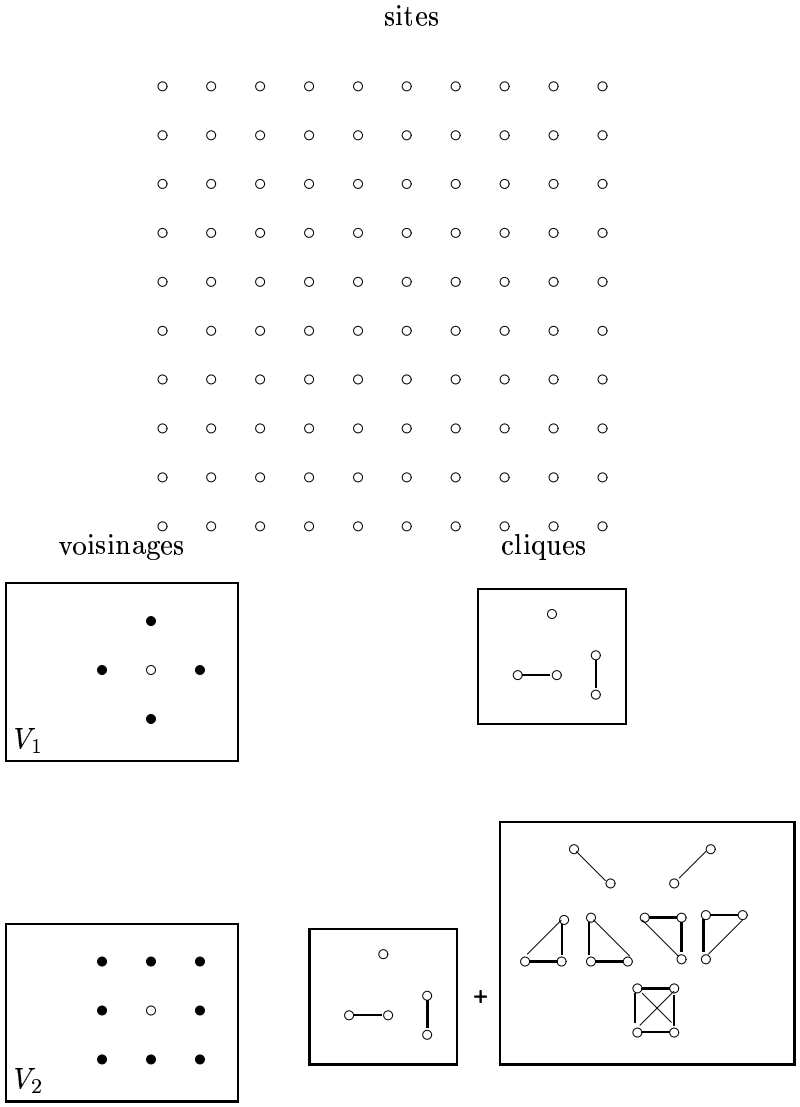
FIG. 10.1 – *Sites, voisinage et champ.*

Notons aussi que les sites ne sont pas obligés être tous identiques, mais en traitement du signal ces sites peuvent correspondre aux positions des échantillons. En traitement d'image on a l'habitude de parler de *pixels*. Ci-dessous deux exemples

Exemple 1 : $\mathcal{S} = \mathbf{Z}$



Exemple 2 : $\mathcal{S} = \mathbf{Z}^2$



10.1.2 Champ aléatoire

Ensembles des variables aléatoires $\mathcal{X} = \{X_s, s \in \mathcal{S}\}$ est appelé un champ aléatoire sur les sites $\mathcal{S}$. La variable aléatoire X_s prend sa valeur dans l'ensemble $\mathcal{G} = \{g_1, \ldots, g_k\}$ qui appelé l'espace des phases (niveau de gris, couleurs, etc.).

Une réalisation du champ aléatoire $\mathcal{X}$ est notée :

$$\{\boldsymbol{X} = \boldsymbol{x}\} = \{X_{s_1} = x_{s_1}, \ldots, X_{s_N} = x_{s_N}\}$$

et l'ensemble des réalisations (ou configurations) possibles est noté :

$$\Omega = \{\boldsymbol{X} = (X_{s_1}, \ldots, X_{s_N}) \ : X_{s_n} \in \mathcal{G}, n = 1, \ldots, N\}$$

Lorsque l'espace des phases est à dimension finie, (cas discret) le champ est caractérisé par la loi de probabilité conjointe $P(\boldsymbol{X} = \boldsymbol{x})$, et lorsque l'espace des phases est à dimension infinie, (cas continu) le champ est caractérisé par sa densité de probabilité conjointe $p(\boldsymbol{x})$.

On peut aussi définir les lois de probabilités conditionnelles :

$$P(X_s = x_s | X_r = x_r), \text{ cas discret} \qquad \text{ou } p(x_s | X_r = x_r), \text{ cas continu}$$

10.1.3 Champ aléatoire markovien

- $\mathcal{X}$ champ aléatoire sur $\mathcal{S}$ à valeurs dans $\mathcal{G}$ est dit markovien relativement au système de voisinage $\mathcal{V}$ si
  - $P(\boldsymbol{X} = \boldsymbol{x}) > 0, \quad \forall \boldsymbol{x} \in \Omega, \quad$ et
  - $\forall X_s \in \mathcal{G}, \forall s \in \mathcal{S}, \quad P(X_s = x_s | X_r = x_r, r \neq s) = P(X_s = x_s | X_r = x_r, r \in V_s)$
- Les fonctions : $\pi_s(x_s | x_{V_s}) = P(X_s = x_s | X_{V_s} = x_{V_s}), s \in \mathcal{S}$
  sont les caractéristiques locales du champ markovien $\mathcal{X}$

Deux questions :

Etant donné un système $\{\pi_s(x_s | x_{V_s}), s \in \mathcal{S}\} = \mathcal{P}$,

1. existe-t-il un champ markovien avec $\mathcal{P}$ comme caractéristiques locales ?
2. est-il unique ?

10.2 Champs de Gibbs et équivalence Gibbs-Markov

10.2.1 Energie et Potentiel

- $\{V_a, a \subseteq \mathcal{S}\} : \Omega \longrightarrow \mathbf{R}^+$ telle que $V_a(\boldsymbol{x})$ ne dépend que de $\boldsymbol{x}_a$
- $V_c, c \in \mathcal{C}$ est un potentiel de voisinage si $V_c = 0$ lorsque c n'est pas une clique
- $U(\boldsymbol{x}) = \sum_{a \subseteq \mathcal{S}} V_a(\boldsymbol{x})$ est l'énergie associée au potentiels V_a

Lorsque U est l'énergie associée au potentiel de voisinage $V_c, c \in \mathcal{C}$ on a

$$U(\boldsymbol{x}) = \sum_{c \in \mathcal{C}} V_c(\boldsymbol{x})$$

10.2.2 Mesure de Gibbs

– la mesure de Gibbs associée à l'énergie $U(\boldsymbol{x})$ et à la température T est la mesure de probabilité sur Ω :

$$\pi_T(\boldsymbol{x}) = \frac{1}{Z_T} \exp\left[-\frac{U(\boldsymbol{x})}{T}\right]$$

– la constante

$$Z_T = \sum_{\boldsymbol{x} \in \Omega} \exp\left[-\frac{U(\boldsymbol{x})}{T}\right]$$

est la fonction de partition.

– T est la température

– Un champ aléatoire $\mathcal{X}$ défini sur $\mathcal{S}$ est un champ de Gibbs si :

$$P(\boldsymbol{X} = \boldsymbol{x}) = \frac{1}{Z} \exp\left[-\frac{U(\boldsymbol{x})}{T}\right] \quad \text{avec} \quad U(\boldsymbol{x}) = \sum_{C \in \mathcal{C}} V_c(\boldsymbol{x})$$

10.2.3 Équivalence Gibbs-Markov

Théorème : Soit $\mathcal{V}$ un système de voisinage sur $\mathcal{S}$. Alors, $\mathcal{X}$ est un champ markovien sur $\mathcal{S}$ à valeurs dans $\mathcal{G}$ relativement au système de voisinage $\mathcal{V}$ ssi $p(\boldsymbol{x}) = P(\boldsymbol{X} = \boldsymbol{x})$ est une distribution de Gibbs relativement à $\mathcal{V}$.

(Hammersley	&	Clifford)
Gibbs	$\Longleftrightarrow$	Markov
Potentiels	$\Longleftrightarrow$	caracteristiques locales

Gibbs $\longrightarrow$ Markov :

$$P(X_s = x_s | X_{\text{-S}} = x_{\text{-S}}) = \frac{\exp\left[-1/T \sum_{c \in \mathcal{C}} V_c(x_s))\right]}{\sum_{y_s \in \mathcal{G}} \exp\left[-1/T \sum_{c \in \mathcal{C}} V_c(y_s, x_{\text{-S}})\right]}$$

Markov $\longrightarrow$ Gibbs : plus délicat. Potentiels canoniques

10.3 Exemples de modèles classiques

10.3.1 Auto-modèles

$$U(\boldsymbol{x}) = \sum_r x_r g_r(x_r) + \sum_r \sum_s \beta_{rs} x_r x_s, \quad r,s \in \mathcal{S}$$

$g_r(.)$ fonctions arbitraires
β_{rs} coefficients d'interaction entre les sites r et s
$\beta_{rs} \neq 0$ ssi les sites r et s sonts voisines: $|r-s|^2 \leq d$

10.3.2 Auto-logistique

$\mathcal{G} = \{0,1\}$ ou $\{-1,+1\}$

$$U(\boldsymbol{x}) = \sum_r \alpha_r x_r + \sum_r \sum_s \beta_{rs} x_r x_s \quad r,s \in \mathcal{C}$$

Si $\alpha = 0$, et $\beta_{rs} = \begin{cases} \beta & \text{si } |r-s|^2 \leq d \\ 0 & \text{sinon} \end{cases}$, U mesure la longueur d'un contour.

10.3.3 Modèle d'Ising

Auto-logistique avec voisinage d'ordre 1
Deux exemples spécifiques:

- Chaîne de Markov binaire: $\mathcal{G} = \{0,1\}$, $\mathcal{S} = \mathbf{Z}$,
  Matrice de ransitions $\begin{pmatrix} 0-0 & 0-1 \\ 1-0 & 1-1 \end{pmatrix}$ $P = \begin{pmatrix} s & 1-s \\ 1-t & t \end{pmatrix}$

$$\pi(\boldsymbol{x}) = \frac{1}{Z} \exp\left[\overline{V}_0 \sum_{i=0,n} x_i + V_0 \sum_{0<i<n} x_i + V_1 \sum_{|i-j|=1} x_i x_j \right]$$

$$\overline{V}_0 = \log \frac{1-s}{s}, \quad V_0 = \log \frac{(1-s)(1-t)}{s^2}, \quad V_1 = \log \frac{st}{(1-s)(1-t)}$$

- Images binaires et voisinage d'ordre 1:

  $\mathcal{G} = \{0,1\}$, $\mathcal{S} = \mathbf{Z}^2$, $V_1 = \left\{ \begin{array}{ccc} & \bullet & \\ \bullet & \circ & \bullet \\ & \bullet & \end{array} \right\}$

$$U(\boldsymbol{x}) = \alpha \sum x_{ij} + \beta \sum x_{i,j} v_{i,j}$$

avec

$$v_{i,j} = x_{i,j-1} + x_{i-1,j} + x_{i,j+1} + x_{i+1,j}$$

10.3.4 Multi-Level Logistic Model (MLL)

ou Ising généralisé ou encore Champ de labels

$\mathcal{G} = g_1, \ldots, g_M$ niveaux des gris dans chaque région

$$U(\boldsymbol{x}) = \sum_k V_k(x) + \sum_c V_c(\boldsymbol{x})$$

– pour les cliques simples

$$V_c(x) = \begin{cases} \alpha_k & \text{si } x_{ij} = g_k \\ 0 & \text{sinon} \end{cases}$$

– pour les cliques multiples

$$V_c(\boldsymbol{x}) = \begin{cases} +\beta & \text{si } x_{ij} \text{ dans } C \text{ ont les mêmes valeurs} \\ -\beta & \text{sinon} \end{cases}$$

10.4 Tentative de classification

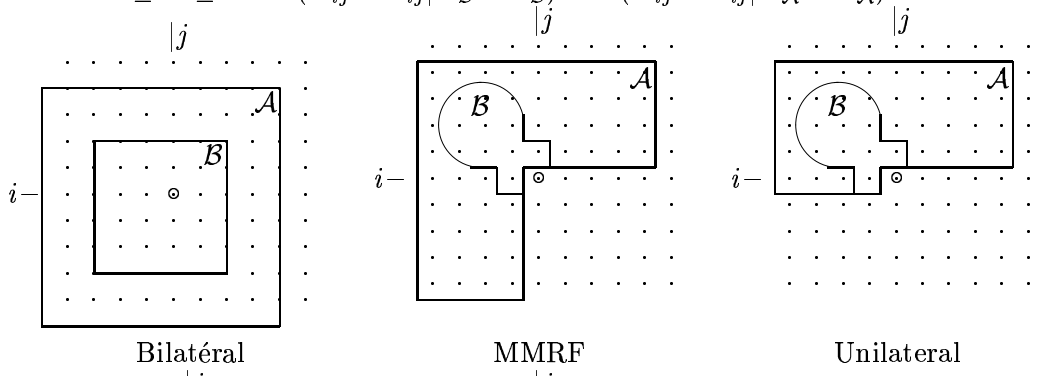
- **Champ de markov au sens strict (SSM)** :
  (Définition via les probabilités conditionnelles)
  Soient $\mathcal{B} \subseteq \mathcal{A} \subseteq \mathcal{S} \quad P(X_{ij} = x_{ij} | X_{\mathcal{B}} = x_{\mathcal{B}}) = P(X_{ij} = x_{ij} | X_{\mathcal{A}} = x_{\mathcal{A}})$
- **Champ de markov au sens large (WSM)**
  (Définition via la régression linéaire)
  (Estimation linéaire en moyenne quadratique ELMQ)
  $X = \{X_{ij}\}, \quad (i,j) \in \mathbf{Z}^2, \quad \mathcal{B} \subseteq \mathcal{A} \subseteq \mathcal{S}$

$$ELMQ\{X_{ij} | X_{\mathcal{B}} = x_{\mathcal{B}}\} = ELMQ\{X_{ij} | X_{\mathcal{A}} = x_{\mathcal{A}}\}$$

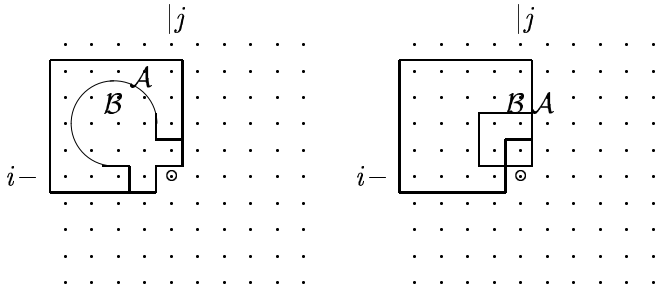
  - Propriétés au sens large = Propriétés au deuxième ordre
  - Représentation équivalente par un modèle AR : $X_s = \sum_r a_r X_r + Z_s$

Champ de markov au sens strict (SSM)

Soient $\mathcal{B} \subseteq \mathcal{A} \subseteq \mathcal{S}$ $\quad P(X_{ij} = x_{ij} | X_{\mathcal{B}} = x_{\mathcal{B}}) = P(X_{ij} = x_{ij} | X_{\mathcal{A}} = x_{\mathcal{A}})$



Bilatéral MMRF Unilateral



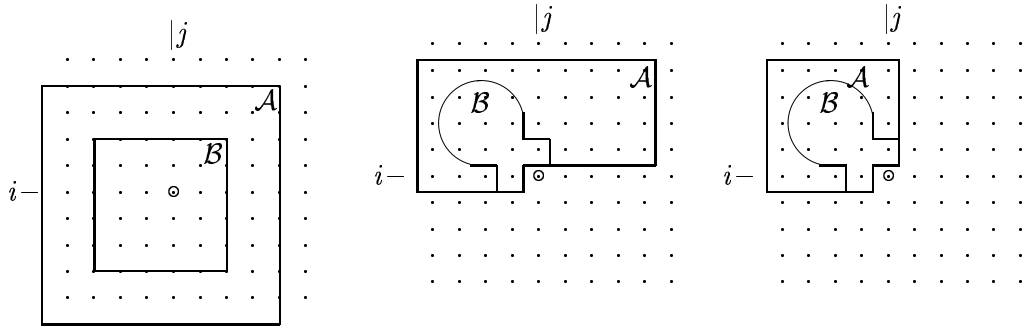
Causal Pickard

Pickard ⊆ Causal
Causal ⊆ Unilatéral
Unilatéral ⊆ MMRF
MMRF ⊆ Bilatéral

Champ de markov au sens large (WSM)

(Définition via la régression linéaire)
(Estimation linéaire en moyenne quadratique ELMQ)
$X = \{X_{ij}\}, \quad (i,j) \in \mathbf{Z}^2, \quad \mathcal{B} \subseteq \mathcal{A} \subseteq \mathcal{S}$

$$ELMQ\{X_{ij} | X_{\mathcal{B}} = x_{\mathcal{B}}\} = ELMQ\{X_{ij} | X_{\mathcal{A}} = x_{\mathcal{A}}\}$$



Bilatéral Unilateral Causal

Champ de markov gaussien : sens large = sens stricte

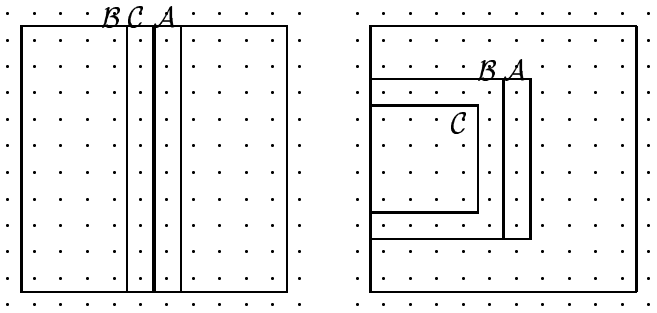
Un exemple classique utilisé en traitement d'image :
$\mathcal{S} = \mathbf{Z}^2_{mn} = \{(i,j) \ : 1 \leq i \leq m, 1 \leq j \leq n\}, \mathcal{X} = \{X_{i,j}, (i,j) \in \mathcal{S}\}$

$$\mathcal{V}_1 = \left\{ \begin{array}{ccc} & \bullet & \\ \bullet & \circ & \bullet \\ & \bullet & \end{array} \right\} \quad \mathcal{V}_2 = \left\{ \begin{array}{ccc} \bullet & \bullet & \bullet \\ \bullet & \circ & \bullet \\ \bullet & \bullet & \bullet \end{array} \right\}$$

$$P(X_{ij} = x_{ij} | X_{kl} = x_{kl}, (k,l) \in \mathcal{S} - \{i,j\}) = P(X_{ij} = x_{ij} | X_{kl} = x_{kl}, (k,l) \in V_{ij})$$

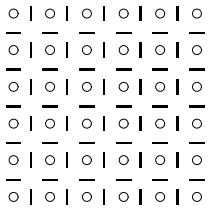
Quelques propriétés :

$$P(x_a, a \in \mathcal{A} | x_b, b \in \mathcal{B}) = P(x_a, a \in \mathcal{A} | x_b, b \in \mathcal{C})$$

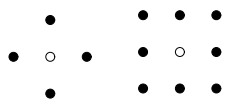


10.5 Modèles couplés ou hiérarchiques

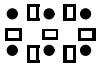
- $\mathbf{Z}_m$ réseau des entiers
- $\mathcal{S}_1 = \mathbf{Z}_m$ sites pixels $\circ$
- $\mathcal{S}_2 = \mathbf{Z}_m$ sites contours $-$
- $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2$ sites image



- $\mathcal{F} = \{F_{i,j}, (i,j) \in \mathcal{S}_1\}$ champ intensité
- $\mathcal{L} = \{L_{\alpha,\beta}, (\alpha,\beta) \in \mathcal{S}_2\}$ champ des lignes
- $\mathcal{X} = (\mathcal{F},\mathcal{L})$ image



Voisinage pixels



voisinage contours

Modèle hiérarchique :

$$P(\boldsymbol{f},\boldsymbol{l}) = P(\boldsymbol{f}|\boldsymbol{l})P(\boldsymbol{l})$$

$$P(\boldsymbol{l}) = \frac{1}{Z_L} \exp\left[-U_L(\boldsymbol{l})\right]$$

$$P(\boldsymbol{f}|\boldsymbol{l}) = \frac{1}{Z_{F|L}} \exp\left[-U_{F|L}(\boldsymbol{f}|\boldsymbol{l})\right]$$

$$\longrightarrow \quad P(\boldsymbol{f},\boldsymbol{l}) = \frac{1}{Z_{F,L}} \exp\left[-U_{F,L}(\boldsymbol{f},\boldsymbol{l})\right],$$

avec

$$U_{F,L}(\boldsymbol{f},\boldsymbol{l}) = U_{F|L}(\boldsymbol{f}|\boldsymbol{l}) + U_L(\boldsymbol{l})$$

10.6 Application en segmentation d'image

- $\mathcal{S} = \{(i,j) \ : 1 \leq i \leq m, 1 \leq j \leq n\}$ sites pixels $\circ$
- $\mathcal{X} = \{X_{i,j}, (i,j) \in \mathcal{S}\}$ image
- $\mathcal{Y} = \{Y_{i,j}, (i,j) \in \mathcal{S}\}$ image dégradée
- $\mathcal{N} = \{N_{i,j}, (i,j) \in \mathcal{S}\}$ bruit

Modèle de dégradation :

$$\mathcal{Y} = \mathcal{X} + \mathcal{N}$$

$X_{i,j} = g_m$, si $X_{i,j} = m$ avec $\mathcal{G} = \{g_1, \ldots, g_M\}$ niveaux des gris dans chaque région

Hypothèses statistiques :

- $\mathcal{X}$ un champ de Markov à valeur dans $\mathcal{G} = \{g_1, \ldots, g_M\}$
- $\mathcal{N}$ un champ gaussien blanc
- $N_{i,j}$ v.a. indépendantes et de lois $\mathcal{N}(0, \sigma^2)$
- $\mathcal{N}$ et $\mathcal{X}$ indépendants

Modèle a posteriori et choix de l'estimateur

– $\mathcal{X}$ est un champ de Gibbs d'énergie $U(\boldsymbol{x})$:

$$P(\boldsymbol{X} = \boldsymbol{x}) = \frac{1}{Z} \exp\left[-U(\boldsymbol{x})\right]$$

$$\begin{aligned} U(\boldsymbol{x}) &= \sum_j \alpha_i x_j \\ &+ \beta_1 \sum_{(i,j)0} \sum x_i x_j + \beta_2 \sum_{(i,j)90} \sum x_i x_j + \beta_3 \sum_{(i,j)-45} \sum x_i x_j + \beta_4 \sum_{(i,j)45} \sum x_i x_j \\ &+ \gamma_1 \sum_{(i,j,k)} x_i x_j x_k + \gamma_2 \sum_{(i,j)90} \sum x_i x_j x_k + \ldots \\ &+ \xi_1 \sum_{(i,j,k,l)} x_i x_j x_k x_l \end{aligned}$$

avec M paramètres $\alpha_m, m = 1 = \cdots, M$, 4 paramètres $\beta_1, \cdots, \beta_4$, 4 paramètres $\gamma_1, \cdots, \gamma_4$ et un paramètre ξ_1.

- $\mathcal{N}$ est un champ gaussien, blanc et indépendant de $\mathcal{X}$,

$$P(\boldsymbol{Y} = \boldsymbol{y} | \boldsymbol{X} = \boldsymbol{x}) = K \exp\left[-\frac{1}{2\sigma_b^2} \|\boldsymbol{y} - \boldsymbol{x}\|^2\right]$$

- Règle de Bayes :

$$P(\boldsymbol{X} = \boldsymbol{x} | \boldsymbol{Y} = \boldsymbol{y}) = \frac{P(\boldsymbol{Y} = \boldsymbol{y} | \boldsymbol{X} = \boldsymbol{x}) P(\boldsymbol{X} = \boldsymbol{x})}{P(\boldsymbol{Y} = \boldsymbol{y})}$$

- Énergie a posteriori : $U^p(\boldsymbol{x}) = \frac{1}{2\sigma_b^2} \|\boldsymbol{y} - \boldsymbol{x}\|^2 + U(\boldsymbol{x})$

Deux types de problèmes :

1. **Classification ou segmentation supervisée :**
   On connait $\boldsymbol{\theta} = \{\{g_1,\ldots,g_M\},\{\beta_1,\cdots,\beta_4\},\{\gamma_1,\cdots,\gamma_4\},\xi_1\}$ et $\boldsymbol{y} \longrightarrow \widehat{\boldsymbol{x}}$.
2. **Segmentation non supervisée :** $\boldsymbol{y} \longrightarrow \widehat{\boldsymbol{x}},\widehat{\boldsymbol{\theta}}$
   - Apprentissage supervisée : $\boldsymbol{x},\boldsymbol{y} \longrightarrow \widehat{\boldsymbol{\theta}}$
   - Apprentissage non supervisée : $\boldsymbol{y} \longrightarrow \widehat{\boldsymbol{\theta}}$

Quelques estimateurs possibles pour le problème 1

- MAP (Maximum A Posteriori) : $\widehat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \{P(\boldsymbol{x}|\boldsymbol{y})\}$
- MPM (Maximum of the Posterior Marginal) :

$$\widehat{\boldsymbol{x}} = \{\widehat{x}_{s\in S}\} \quad \text{avec} \quad \widehat{x}_s = \arg\max_{x_s} \{p_s(X_s = x_s|\boldsymbol{y})\}$$

- PM (Posterior Mean) Moyenne *a posteriori* :

$$\widehat{x} = \{\widehat{x}_{s\in S}\} \quad \text{avec} \quad \widehat{x}_s = \sum_{x_i\in\mathcal{G}} x_s\, p_s(X_s = x_s|\boldsymbol{y})$$

Deux algorithmes stochastiques :

- **Le recuit simulé ou relaxation stochastique**
  permet d'estimer le MAP
- **L'échantillonneur de Gibbs**
  permet de simuler des réalisations de $p(x_s|\boldsymbol{y},x_{j\neq s})$, et de $p(\boldsymbol{x}|\boldsymbol{y})$ et donc d'estimer le MPM ou la moyenne *a posteriori*.

Quelques estimateurs possibles pour les problèmes 2

- Apprentissage supervisée : $\boldsymbol{x},\boldsymbol{y} \longrightarrow \widehat{\boldsymbol{\theta}}$ MAP, PM, MPM
- Apprentissage non supervisée : $\boldsymbol{y} \longrightarrow \widehat{\boldsymbol{\theta}}$ MV (EM)
- Segmentation non supervisée : $\boldsymbol{y} \longrightarrow \widehat{\boldsymbol{x}},\widehat{\boldsymbol{\theta}}$ MVG

Lorsque $\boldsymbol{x}$ est modélisé par un champ unilatéral ou par une chaîne de Markov :

- $\boldsymbol{y},\boldsymbol{\theta} \longrightarrow \widehat{\boldsymbol{x}}_{\text{MAP}}$ Viterbi (Programmation dynamique)
- $\boldsymbol{y} \longrightarrow \widehat{\boldsymbol{\theta}}_{\text{MV}}$ EM, (Forward-Backward)
- $\boldsymbol{y} \longrightarrow \widehat{\boldsymbol{x}}_{\text{MAP}},\widehat{\boldsymbol{\theta}}_{\text{MV}}$: Algorithme itératif, par ex. MVG:

$$\widehat{\boldsymbol{\theta}}^{(0)},\boldsymbol{y} \longrightarrow \widehat{\boldsymbol{x}}^{(1)}_{\text{MAP}},\boldsymbol{y} \longrightarrow \widehat{\boldsymbol{\theta}}^{(1)}_{\text{MV}},\boldsymbol{y} \longrightarrow \widehat{\boldsymbol{x}}^{(1)}_{\text{MAP}},\boldsymbol{y} \longrightarrow \cdots$$

10.7 Application en restauration d'image

Modèle de Geman-Geman : modèle couplé pixels-contours

- $\mathcal{F} = \{F_{i,j},(i,j) \in \mathbf{Z}_m\}$ champ intensité $U(\boldsymbol{f}|\boldsymbol{l})$
- $\mathcal{L} = \{L_{\alpha,\beta},(\alpha,\beta) \in \mathcal{D}_m\}$ champ des lignes $U(\boldsymbol{l})$
- $\mathcal{X} = (\mathcal{F},\mathcal{L})$ image $U(\boldsymbol{f},\boldsymbol{l}) = U(\boldsymbol{f}|\boldsymbol{l}) + U(\boldsymbol{l})$
- $\mathcal{G} = \{G_{i,j},(i,j) \in \mathbf{Z}_m\}$ image dégradée

$$G_{i,j} = \sum_{(k,l)} H_{i-k,j-l} F_{k,l} + N_{i,j} \quad \text{ou} \quad \boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{n}$$

Remarque : le champ des ligne n'est pas dégradée

- $\mathcal{N} = \{N_{i,j},(i,j) \in \mathbf{Z}_m\}$ champ bruit d'intensité (gaussien, blanc)
- $\boldsymbol{H} = \{H_{i,j},(i,j) \in \mathbf{Z}_m\}$ matrice de flou

Estimateur du maximum a posteriori (MAP)

$$\begin{array}{ll} \mathcal{X} = (\mathcal{F},\mathcal{L}), & \boldsymbol{x} = (\boldsymbol{f},\boldsymbol{l}) \\ \mathcal{Y} = (\mathcal{G},\mathcal{L}), & \boldsymbol{y} = (\boldsymbol{g},\boldsymbol{l}) \end{array}$$

– $\mathcal{X}$ est un champ de Markov d'énergie $U(\boldsymbol{f},\boldsymbol{l})$:

Voisinage pixels et voisinage contours

– $\mathcal{N}$ est gaussien, blanc et indépendant de $\mathcal{F}$,

– $\boldsymbol{H}$ est invariant par translation :

$$P(\boldsymbol{G} = \boldsymbol{g}|\boldsymbol{X} = \boldsymbol{x}) = K \exp\left[-\frac{1}{2\sigma_b^2}\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2\right]$$

– Règle de Bayes : $P(\boldsymbol{X} = \boldsymbol{x}|\boldsymbol{G} = \boldsymbol{g}) = A \exp\left[-U^p(\boldsymbol{f},\boldsymbol{l})\right]$ avec

$$U^p(\boldsymbol{f},\boldsymbol{l}) = \frac{1}{2\sigma_b^2}\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + U(\boldsymbol{f},\boldsymbol{l})$$

– Loi *a posteriori* est aussi une mesure de Gibbs. Il reste à déterminer le système de voisinage de cette loi.

– Le champ des lignes $\mathcal{L}$ n'est pas affecté par la dégradation

– Le voisinage du champ d'intensité $\mathcal{X}$ est élargi par la réponse impulsionnelle de la dégradation.

10.8 Outils de simulation et d'optimisation

10.8.1 Echantillonneur de Gibbs

- Soit $P(\boldsymbol{X} = \boldsymbol{x})$ une distribution de probabilité.
- Soit (s_k),$k \in \mathcal{N}$ une suite de visite des sites telle que, pour tout $s \in \mathcal{S}$, s est visité un nombre infini de fois.

$$X_s^{(k+1)} = \begin{cases} X_s^{(k+1)} & \text{si} \quad s \neq s_k \\ \xi & \text{si} \quad s = s_k \end{cases}$$

ξ est une v.a. tirée suivant la loi conditionnelle

$$p_{s_k}(\xi = \lambda) = p_{s_k}(\xi = \lambda | x_j = x_j(k), j \neq s_k)$$

Si l'ensemble des états est fini ou si X est gaussien on a :

Théorème :
Quelle que soit la configuration initiale x_0

$$\lim_{k \longrightarrow \infty} P(\boldsymbol{X}(k) = \boldsymbol{x} | \boldsymbol{X}(0) = \boldsymbol{x}(0)) = P(\boldsymbol{X} = \boldsymbol{x})$$

Remarques :

- Si X est un champ de Markov, la loi conditionnelle s'exprime simplement en fonction de la configuration sur un voisinage de s_k, ce qui rend cette simulation raisonable en temps de calcul.
- Pour la segmentation $X|Y$ a la même taille de voisinage que X, mais

- Pour la restauration la taille de voisinage *a posteriori* dépend à la fois de la taille de voisinage a priori et du support de l'opérateur de flou.

10.8.2 Recuit simulé

$$P(\boldsymbol{X} = \boldsymbol{x}|\boldsymbol{Y} = \boldsymbol{y}) = \frac{1}{Z}\exp\left[-U(\boldsymbol{x})\right] = \pi(\boldsymbol{x})$$

- **Physique statistique :** niveau d'intensité d'un pixel=état d'un particule élémentaire
- distribution de Gibbs : $\pi_T(\boldsymbol{x}) = \frac{1}{Z_T}\exp\left[-\frac{1}{T}U(\boldsymbol{x})\right]$
- chaque réalisation d'image représente l'état de l'ensemble des particule auquel on associe une énergie U($\boldsymbol{x}$)
  $T \longrightarrow \infty \quad \pi_T(\boldsymbol{x}) \longrightarrow$ loi uniforme
  $T = 1 \quad \pi_T(\boldsymbol{x}) \longrightarrow \pi(\boldsymbol{x})$
  $T \longrightarrow 0 \quad \pi_T(\boldsymbol{x}) \longrightarrow$ loi concentrée sur l'estimée MAP

Algorithme de recuit simulé :
(avec l'echantillonneur de Gibss)

$$X_s(k+1) = \begin{cases} X_s^{(k+1)} & \text{si} \quad s \neq s_k \\ \xi & \text{si} \quad s = s_k \end{cases}$$

ξ est une v.a. tirée suivant la loi conditionnelle

$$p_{s_k,T_k}(X = \xi) = \frac{1}{Z_k}\exp\left[-\frac{1}{T_k}U(X_{s_k} = x|x_j, j \neq s_k)\right]$$

Théorème : Si $\lim_{k\longrightarrow\infty} T_k \ln k \geq K$ où K est une constante assez grande, alors quelle que soit la configuration initiale $X(0) = x_0$, $P(X(k) = x|X(0) = x0)$ converge en loi vers une distribution uniforme sur les minima globaux de U

10.8.3 Algorithme d'ICM (Iterated Conditional Modes)

(BESAG)
Il s'agit en effet d'une adaptation déterministe de l'algorithme précédent.

$$X_s^{(k+1)} = \begin{cases} X_s^{(k+1)} & \text{si} \quad s \neq s_k \\ \arg\max_X \{p_i(X_{s_k} = x|x_j, j \neq s_k)\} & \text{si} \quad s = s_k \end{cases}$$

Cet algorithme est rapide mais le résultat dépend du point initial et fournit un minimum local de U.

10.9 Algorithmes de relaxations déterministes

Lors de calcul de la solution au sens du MAP, deux situations peuvent se présenter:

1. La loi *a posterioi* est unimodale.
   Alors il n'y a qu'une seule solution et, en général, l'optimisation du critère MAP (l'énérgie *a posterioi*) ne pose pas de grandes difficultées et on peut utiliser les algorithmes de descentes du type gradient ou autre pour effectuer cette optimisation.
2. La loi *a posterioi* est multimodale.
   Dans ce cas le problème devient plus difficile et il faut aller à la recherche de techniques d'optimisation globale. L'algorithme du recuit simulé (RS) peut alors être utilisé, mais son coût de calcul devient très vite prohibitif, surtout lorsque la taille de voisinage de la loi *a posterioi*) devient plus grande. En effet, la taille de voisinage de la loi *a posterioi*) est directement liée à la taille de voisinage de la loi *a priori*), mais aussi et surtout aux caractéristiques de la l'opérateur liant la grandeur observable à la grandeur inconnue. Par exemple, lorsqu'il s'agit de la ségmentation ou la restauration d'image avec un noyau de taille très réduite, alors on peut utiliser la RS. Mais dans d'autres applicatiuons du genre la reconstruction d'image en tomographie X ou par ondes diffractées où le support de l'opérateur devient important, on ne peut plus envisager l'utilisation de tel algorithme. On peut alors faire recours aux techniques de relaxations déterministes qui cherchent à faire un compromis entre l'optimalité au sens de la recheche de l'optimum global et le coût de calcul. Parmi ces techniques on peut citer la technique de non convexité graduelle (Graduated Non Convexity GNC) que nous allons la décrire ici dans un cas particulier.

10.9.1 Principe de base du GNC

Soit $J(\boldsymbol{f})$ un critère multimodal dont nous cherchons à trouver l'argument $\boldsymbol{f}$ correspondant à son minimum global. La principale idée du GNC est de construire une suite de critères $J_{c_k}(\boldsymbol{f})$ telles que:
– $J_{c_0}(\boldsymbol{f})$ soit convexe et
– $\forall \boldsymbol{f}\ \lim_{c \longrightarrow c_\infty} J_c(\boldsymbol{f}) = J(\boldsymbol{f})$.
Ensuite, on minimise $J_{c_0}(\boldsymbol{f})$ pour obtenir $\boldsymbol{f}_0$, et finalement,
pour une suite de $\{c_1, \cdots, c_\infty\}$, on minimise localement $J_{c_k}(\boldsymbol{f})$ au voisinage de la solution obtenue à l'itération précédente.

On epère ainsi que la suite des solutions $\widehat{\boldsymbol{f}}_k$ ainsi obtenues converge vers le minimum global du critère. Notons que, la convexité du J_{c_0} assure l'unicité de la solution, mais, il n'y a aucune guarantie que l'algorithme converge vers le minimum global. Parmis les points importants concernant cet algorithme, il faut noter que son coût, contrairement aux algorithmes à base du recuits simulé, ne dépends pas de la taille de voisinage de la loi *a posteriori* dans le cas d'un estimateur au sens du MAP.

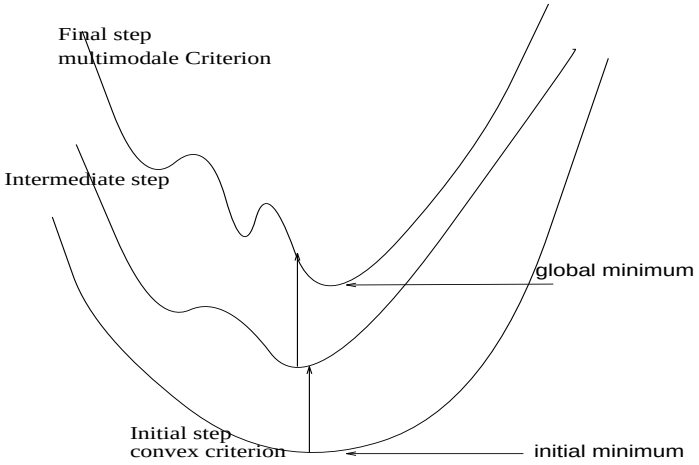


FIG. 10.2 – *Principe de base de la technique du GNC.*

Pour illustrer la mise en œuvre de cette technique nous allons considérer deux cas :
– le cas du calcul de la solution au sens du MAP d'un problème linéaire avec une loi *a priori* markovienne dont le terme d'énérgie est nonconvexe ;
– le cas d'un problème non linéaire où la multimodalité du critère MAP vient plutôt de la non linéarité du problème que du choix de la loi *a priori*.

Cas des problèmes linéaires :

Considérons le cas d'un problème inverse linéaire $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{b}$ en une dimension et supposons que le signal d'entrée est un signal continu par morceau et notons par $\boldsymbol{f}$ le vecteur contenant les valeurs du signal et par $\boldsymbol{l}$ le vecteur contenant les positions des discontinuités ($l_j = 0$, pas de discontinuité, $l_j = 1$, il y a une discontinuité). L'objectif de la reconstruction est alors d'estimer à la fois $\boldsymbol{f}$ et $\boldsymbol{l}$.

– Modèles avec un processus de ligne binaire

$$p(\boldsymbol{f},\boldsymbol{l}|\boldsymbol{g}) = p(\boldsymbol{g}|\boldsymbol{f})\, p(\boldsymbol{f}|\boldsymbol{l})\, p(\boldsymbol{l}) \longrightarrow$$

$$U^p(\boldsymbol{f},\boldsymbol{l}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + U(\boldsymbol{f},\boldsymbol{l}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + U(\boldsymbol{f}|\boldsymbol{l}) + U(\boldsymbol{l})$$

– Si le processus de ligne est non-interactif, ce qui signifie $p(\boldsymbol{l}) = \prod_j p(l_j)$ ou encore $U(\boldsymbol{l}) = \sum_j \phi_1(l_j)$, alors on peut montrer que la solution

$$(\widehat{\boldsymbol{l}},\widehat{\boldsymbol{f}}) = \arg\min_{(\boldsymbol{f},\boldsymbol{l})} \{U^p(\boldsymbol{f},\boldsymbol{l}|\boldsymbol{g})\}$$

peut être calculés en deux étapes :
– Calculer $\widehat{\boldsymbol{f}}$ en minimisant un critère ne dépendant que de $\boldsymbol{f}$:

$$U^p(\boldsymbol{f}|\boldsymbol{g}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda \sum_{s\in\mathcal{S}} \phi([\boldsymbol{D}\boldsymbol{f}]_s)$$

avec $\boldsymbol{D}\boldsymbol{f}$ un opérateur de différences finies et ϕ une fonction non-convexe et ;
– Déduire $\widehat{\boldsymbol{l}}$ par une opération directe à partir de $\widehat{\boldsymbol{f}}$.

– **Exemple :**

$$U(\boldsymbol{f}|\boldsymbol{l}) = \lambda^2 \sum_j (f_j - f_{j-1})^2 (1 - l_j), \quad U(\boldsymbol{l}) = \mu \sum_j l_j,$$

Si

$$(\widehat{\boldsymbol{l}},\widehat{\boldsymbol{f}}) = \arg\min_{(\boldsymbol{f},\boldsymbol{l})} \{U^p(\boldsymbol{f},\boldsymbol{l}|\boldsymbol{g})\}$$

alors

$$\widehat{\boldsymbol{f}} = \arg\min_{(\boldsymbol{f})} \{U^p(\boldsymbol{f}|\boldsymbol{g})\} \quad \text{avec} \quad U^p(\boldsymbol{f}|\boldsymbol{g}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \sum_j \phi(x_j - x_{j-1})$$

avec

$$\phi(t) = \begin{cases} (\lambda t)^2 & \text{si} \quad |t| \leq T \\ \alpha = (\lambda T)^2 & \text{si} \quad |t| > T \end{cases}$$

et

$$\widehat{l}_j = \begin{cases} 1 & \text{si} \quad |x_j - x_{j-1}| \leq T \\ 0 & \text{si} \quad |x_j - x_{j-1}| > T \end{cases}$$

- $U^p(\boldsymbol{f}|\boldsymbol{g})$ est alors multimodale et la recherche de la solution MAP ne peut se faire par une méthode d'optimisation locale.
- Recuit simulé trop coûteux lorsque l'opérateur $\boldsymbol{H}$ a un support large. En effet le coût de calcul de cet algorithme croit d'une manère exponentielle avec la taille de son support.
- GNC est une technique d'optimisation non-locale dont le coût de calcul n'est pas dépendant du support de l'opérateur $\boldsymbol{H}$.

Non convexité vient du choix du terme $U^p(\boldsymbol{f}|\boldsymbol{g})$ qui est une conséquence du choix de la loi *a priori* . C'est pourquoi pour mettre en œuvre l'algorithme du GNC dans ce cas, on peut envisager de définir une suite de critères $J_c(\boldsymbol{f})$:

$$J_c(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \sum_j \phi_c(x_j - x_{j-1})$$

avec

$$\phi_c(t) = \begin{cases} (\lambda t)^2 & \text{si} \quad |t| < q_c \\ \alpha - \frac{1}{2}c(|t| - r_c)^2 & \text{si} \quad q_c > |t| > r_c \\ \alpha = (\lambda T)^2 & \text{si} \quad |t| > r_c \end{cases}$$

où

$$\begin{cases} q_c = T(1 + 2\lambda^2/c)^{-1/2} \\ r_c = T(1 + 2\lambda^2/c)^{1/2} \end{cases}$$

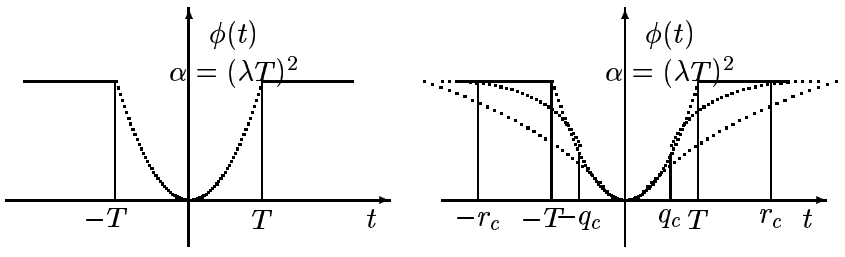


FIG. 10.3 – *Fonction quadratique tronquée et sa relaxation*

La principale difficulté est ensuite de trouver un c_0 telle que $J_{c_k}(\boldsymbol{f})$ soit convexe et ensuite choisir la suite $\{c_0,c_1,\cdots\}$. La solution est alors calculée par

$$\widehat{\boldsymbol{f}}_{c_k} = \arg \min_{\boldsymbol{f} \in V(\widehat{\boldsymbol{f}}_{c_{k-1}})} \{J_{c_k}(\boldsymbol{f})\}$$

- [Blake et Zisserman] :
  - Segmentation d'image $\boldsymbol{A} = \boldsymbol{I} \longrightarrow$ Problème bien posé
  $J_{c_k}(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{f}\|^2 + \Phi_{c_k}(\boldsymbol{f}), \quad \exists\, c_0$ tel que J_{c_0} soit convexe
- [Nikolova, Djafari, Idier] :
  - Extension pour les problème mal posés $\boldsymbol{A}$
  $\nexists\, c_0$ tel que J_{c_0} soit convexe $\longrightarrow$

$$J_{c_k}(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}\|^2 + \Phi_{c_k}(\boldsymbol{f}) + \Psi_{a_k}(\boldsymbol{f})$$

  Double relaxations : $a_k \mapsto 0$ et $c_0 \mapsto c_\infty$
  - Extension pour d'autres modèles que celui de la chaîne ou membrane faible

[GG84, Gem90, GR92, Bes74, Bes86, Bes89, BG93, BZ87, NMD96a]

A COMPLETER

Cas d'un problème inverse non linéaire : Dans le cas d'un problème inverse non linéaire $\boldsymbol{g} = \boldsymbol{H}(\boldsymbol{f}) + \boldsymbol{b}$ si on choisit de définir la solution comme le minimum d'un critère du type :

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \left\{ J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}(\boldsymbol{f})\|^2 + \lambda\Omega(\boldsymbol{f}) \right\},$$

même si on choisit $\Omega(\boldsymbol{f})$ convexe, $J(\boldsymbol{f})$ est en général non convexe et multimodale à cause du terme de l'adéquation aux données $\|\boldsymbol{g} - \boldsymbol{H}(\boldsymbol{f})\|^2$.

Ici, pour illustrer l'usage de l'algorithme du GNC, il faut bien entendu préciser la nature de l'opérateur nonlinéaire $\boldsymbol{H}(\boldsymbol{f})$. Pour ceci, nous allons considérer le cas d'un problème de l'imagerie à ondes diffractées :

$$\begin{aligned} g(\boldsymbol{r}_i) &= \iint_D G_m(\boldsymbol{r}_i,\boldsymbol{r}')\phi(\boldsymbol{r}')f(\boldsymbol{r}')\,\mathrm{d}\boldsymbol{r}', \quad \boldsymbol{r}_i \in S \\ \phi(\boldsymbol{r}) &= \phi_0(\boldsymbol{r}) + \iint_D G_o(\boldsymbol{r},\boldsymbol{r}')\phi(\boldsymbol{r}')f(\boldsymbol{r}')\,\mathrm{d}\boldsymbol{r}', \quad \boldsymbol{r} \in D \end{aligned}$$

La version discrétisée de ces deux équations :

$$\begin{aligned} \boldsymbol{g} &= \boldsymbol{G}_m\boldsymbol{F}\boldsymbol{\phi} + \boldsymbol{b} \\ \boldsymbol{\phi} &= \boldsymbol{\phi}_0 + \boldsymbol{G}_o\boldsymbol{F}\boldsymbol{\phi} \end{aligned}$$

Remplaçant la deuxième équation dans la première on a :

$$\boldsymbol{g} = \boldsymbol{H}(\boldsymbol{f}) + \boldsymbol{b} \quad \text{avec} \quad \boldsymbol{H}(\boldsymbol{f}) = \boldsymbol{G}_m\boldsymbol{F}\left(\boldsymbol{I} - \boldsymbol{G}_o\boldsymbol{F}\right)^{-1}\boldsymbol{\phi}_0$$

$$J(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{x})\|^2 + \Omega(\boldsymbol{x}) \tag{10.1}$$

$$\boldsymbol{A}(\boldsymbol{x}) = \boldsymbol{G}_m\boldsymbol{X}\left(\boldsymbol{I} - \boldsymbol{G}_o\boldsymbol{X}\right)^{-1}\boldsymbol{\phi}_0 \tag{10.2}$$

Un choix possible pour créer une suite de critères relaxés est :

$$\boldsymbol{H}_{c_k}(\boldsymbol{f}) = \boldsymbol{G}_m\boldsymbol{F}(\boldsymbol{I} - c_k\boldsymbol{G}_o\boldsymbol{F})^{-1}\boldsymbol{\phi}_0 \tag{10.3}$$

avec $c_0 = 0$, and $\lim_{k\to\infty} c_k = 1$.

Notons que pour $c_0 = 0$ on a :

$$\boldsymbol{H}_0(\boldsymbol{f}) = \boldsymbol{G}_m\boldsymbol{F}\boldsymbol{\phi}_0 \longrightarrow \quad J_0(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{G}_m\boldsymbol{F}\boldsymbol{\phi}_0\|^2 + \Omega(\boldsymbol{f}).$$

et on retrouve l'approximation linéaire de BORN.

Pour $c_k = 1$, effectivement on retrouve :

$$\boldsymbol{H}_1(\boldsymbol{f}) = \boldsymbol{H}(\boldsymbol{f}) \longrightarrow \quad J_1(\boldsymbol{f}) = J(\boldsymbol{f})$$

A COMPLETER

10.10 Problèmes et questions ouverts

Problème essentiel :

Coût de calcul très élevé du MAP

- potentiels convexes/non convexes $\longrightarrow$ optimisation locale/globale
- Optimisation stochastique ou déterministe

Solution proposée :
Relaxation déterministe : GNC
Non-convexité graduelle (GNC): Technique d'optimisation non-locale

A COMPLETER

Chapitre 11

Choix des hyperparamètres

Dans ce chapitre nous aborderons le problème de l'estimation des hyperparamètres, et en particulier, celui de la détermination du paramètre de la régularisation. Dans un prmier temps, nous décrirons brièvement les méthodes classiques du type : adéquation au données, la courbe en L, la validation croisée, et la validation croisée généralisée. Ensuite, nous présenterons le problème général de l'estimation des hyperparamètres dans le cadre bayésien.

11.1 Position du problème

Dans les chapitres précédents, nous avons vu que la résolution d'un problème inverse ne peut se faire d'une manière satisfaisante que qu'avec l'apport d'une information *a priori* appropriée. L'idée de combiner l'information contenue dans les données avec celle de l'*a priori* est à la base de la notion de la régularisation.

Par exemple dans l'approche classique de la régularisation (au sens de Phillips, Towmay et Tikhonov), nous avons vu que la solution est définie par :

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \left\{ \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda \|\boldsymbol{D}\boldsymbol{f}\|^2 \right\}$$

où λ est le paramètre de régularisation qui a pour role de faire une compromis entre la fidélité aux données et l'information *a priori* qui est ici la variation douce de la grandeur inconnue.

D'une manière plus générale on peut définir une solution régularisée par :

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \left\{ \Delta_1(\boldsymbol{g},\boldsymbol{H}\boldsymbol{f}) + \lambda \Delta_2(\boldsymbol{f},\widehat{\boldsymbol{f}}_\infty) \right\}$$

où Δ_1 et Δ_2 sont deux distances, la première dans l'espace des mesures $\boldsymbol{g}$ et la deuxième dans l'espace des inconnues $\boldsymbol{f}$.

Notant, la solution pour $\lambda = 0$ par $\widehat{\boldsymbol{f}}_0$ et la solution pour $\lambda = \infty$ par $\widehat{\boldsymbol{f}}_\infty$ on peut encore généraliser (ou reécrire la définition de la solution régularisée par :

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \left\{ \Delta_1(\boldsymbol{f},\widehat{\boldsymbol{f}}_0) + \lambda \Delta_2(\boldsymbol{f},\widehat{\boldsymbol{f}}_\infty) \right\}$$

où Δ_1 et Δ_2 sont deux distances euclidiennes, $\widehat{\boldsymbol{f}}_\infty$ est la solution que l'on obtient lorsque $\lambda \longrightarrow \infty$ (la solution a priori) et $\widehat{\boldsymbol{f}}_0$ est la solution que l'on obtient lorsque $\lambda \longrightarrow 0$ (la solution inverse généralisée par exemple).

Arrivé à ce stade deux questions se sonts posées :

- Régulariser de quelle façon ou comment choisir Δ_1 et Δ_2 ?
- Régulariser jusqu'à quel point ou comment choisir le paramètre de la régularisation ?

Les figures de la page suivante montrent, sur un exemple du problème de déconvolution, l'effet du paramètre λ sur la solution.

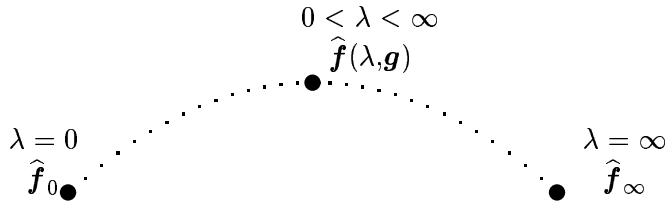
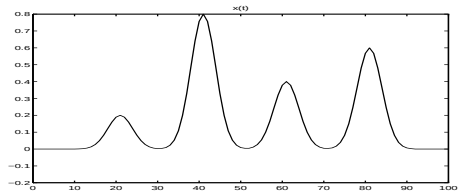
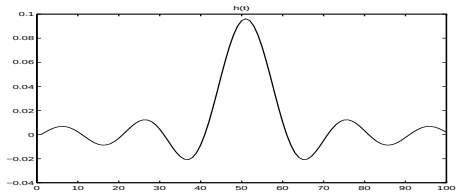


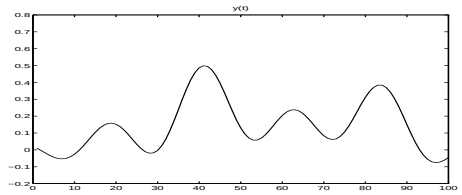
FIG. 11.1 – *Choix du paramètre de régularisation*



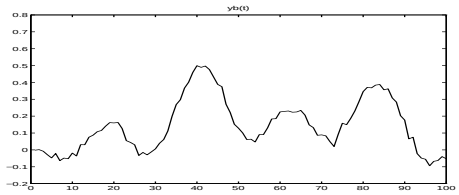
Signal d'entrée $f(t)$



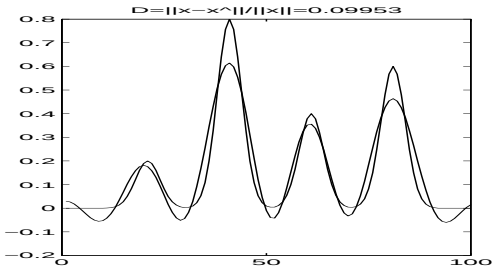
La réponse impulsionnelle $h(t)$



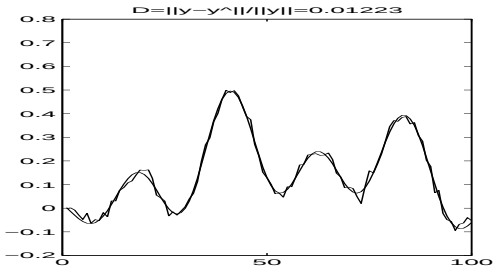
Signal de sortie sans bruit



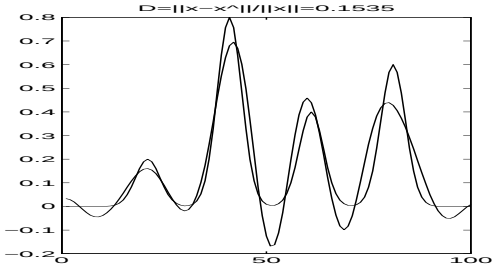
Signal de sortie $g(t)$ avec bruit RSB=20 dB.



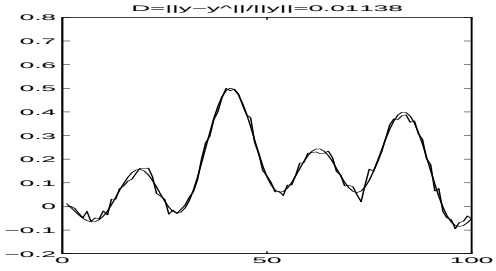
$\lambda = 1 \quad \widehat{f}(t)$



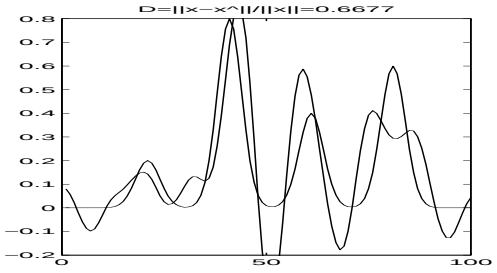
$\widehat{g}(t)$



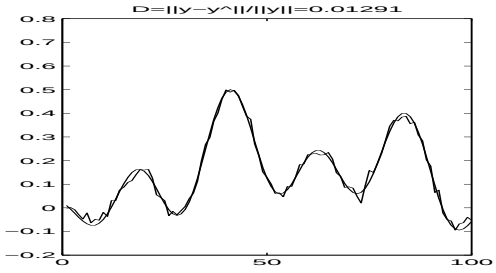
$\lambda = .1 \quad \widehat{f}(t)$



$\widehat{g}(t)$



$\lambda = .01 \quad \widehat{f}(t)$



$\widehat{g}(t)$

FIG. 11.2 – *Effet du coéfficients de la régularisation sur le résultat d'une déconvolution.*

Nous avons vu ensuite que les approches probabilistes ont essayé, d'abord, répondre à la première question. Mais, la réponse à la deuxième question est resté à ce jour encore un problème ouvert.

Dans ce chapitre, d'abord, nous allons considérer le problème de la détermination du paramètre de la régularisation, ensuite, nous allons étendre la question dans l'approche bayésienne à la question plus générale de la détermination des hyperparamètres.

11.2 Choix du coefficient de régularisation

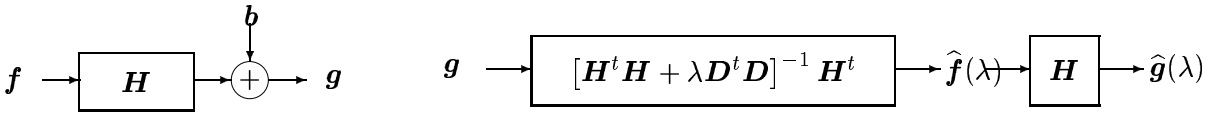
Dans cette section, nous allons considérer le problème de la détermination du paramètre de régularisation dans le cas de la régularisation quadratique :

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \{J(\boldsymbol{f})\} \quad \text{avec} \quad J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{D}\boldsymbol{f}\|^2$$

En effet, dans ce cas simple, nous avons une expression analytique pour la solution régularisée :

$$\widehat{\boldsymbol{f}}(\lambda) = \left[\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{D}^t\boldsymbol{D}\right]^{-1}\boldsymbol{H}^t\boldsymbol{g}$$

ce qui facilite la compréhension des principales méthodes de la détermination du paramètre de régularisation.



Modèle d'observation Inversion

FIG. 11.3 – *Modèle d'observation et d'inversion en régularisation quadratique.*

Notant par :

$$\boldsymbol{A}(\lambda) = \boldsymbol{H}\left[\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{D}^t\boldsymbol{D}\right]^{-1}\boldsymbol{H}^t$$

on a les relations suivantes :

$$\begin{aligned}\left\|\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}(\lambda)\right\|^2 &= \left\|\boldsymbol{g} - \boldsymbol{H}\left[\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{D}^t\boldsymbol{D}\right]^{-1}\boldsymbol{H}^t\boldsymbol{g}\right\|^2 \\ &= \|[\boldsymbol{I} - \boldsymbol{A}(\lambda)]\boldsymbol{g}\|^2\end{aligned}$$

$$\begin{aligned}\|\boldsymbol{H}\boldsymbol{f} - \widehat{\boldsymbol{g}}(\lambda)\|^2 &= \left\|\boldsymbol{H}\boldsymbol{f} - \boldsymbol{H}\left[\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{D}^t\boldsymbol{D}\right]^{-1}\boldsymbol{H}^t\boldsymbol{g}\right\|^2 \\ &= \|\boldsymbol{H}\boldsymbol{f} - \boldsymbol{A}(\lambda)\boldsymbol{g}\|^2\end{aligned}$$

$$\|\boldsymbol{f} - \widehat{\boldsymbol{f}}(\lambda)\|^2 = \|\boldsymbol{f} - \left[\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{D}^t\boldsymbol{D}\right]^{-1}\boldsymbol{H}^t\boldsymbol{g}\|^2$$

Utilisant ces relations, on peut envisager un certain nombre de méthodes pour la détermination du λ :

- Adéquation aux données $\quad \text{RSS}(\lambda) \stackrel{\text{def}}{=} \left\|\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}(\lambda)\right\|^2$
- Risque moyen $\quad \text{E}\left[\|\boldsymbol{f} - \widehat{\boldsymbol{f}}(\lambda)\|^2\right]$
- Validation croisée $\quad \text{E}\left[\|\boldsymbol{g} - \widehat{\boldsymbol{g}}(\lambda)\|^2\right]$

11.2.1 Adéquation aux données

Le problème de la minimisation d'un critère composite

$$\min_{\boldsymbol{f}} \Delta_1(\boldsymbol{g},\boldsymbol{H}\boldsymbol{f}) + \lambda\Delta_2(\boldsymbol{f},\widehat{\boldsymbol{f}}_\infty)$$

peut être interprété comme un problème d'optimisation sous contrainte

$$\min_{\boldsymbol{f}} \Delta_2(\boldsymbol{f},\widehat{\boldsymbol{f}}_\infty) \quad \text{s.c.} \quad \Delta_1(\boldsymbol{g},\boldsymbol{H}\boldsymbol{f}) \leq c_1$$

ou encore

$$\min_{\boldsymbol{f}} \Delta_1(\boldsymbol{g},\boldsymbol{H}\boldsymbol{f}) \quad \text{s.c.} \quad \Delta_2(\boldsymbol{f},\widehat{\boldsymbol{f}}_\infty) \leq c_2$$

où c_1 et c_2 sont deux constantes qu'il faudra déterminer.

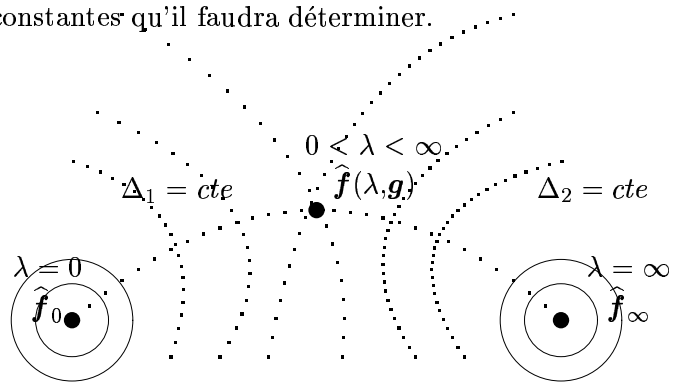


FIG. 11.4 – *Choix du paramètre de régularisation par adéquation aux données*

Pour mieu voir la limitation de cette approche, considérons le deuxième cas. c_2 peut alors être considéré comme une statistique dont sa loi découle de la loi du bruit ou d'une façon plus générale de la loi $p(\boldsymbol{g}|\boldsymbol{f})$.

Supposons par exemple que $b_i \quad v.a.i.\mathcal{N}(0,\sigma_b^2)$. Alors Δ_1 devient une distance quadratique. Supposons aussi que Δ_2 est choisi quadratqiue et égale à $\|\boldsymbol{D}\boldsymbol{f}\|^2$. On a alors :

$$\widehat{\boldsymbol{f}}(\lambda) = \arg\min_{\boldsymbol{f}} \left\{ \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{D}\boldsymbol{f}\|^2 \right\} = \left[\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{D}^t\boldsymbol{D}\right]^{-1}\boldsymbol{H}^t\boldsymbol{g}$$

$$\text{RSS}(\lambda) = \left[\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}(\lambda)\right]^t \left[\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}(\lambda)\right] = \left\|\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}(\lambda)\right\|^2$$

c : $\chi^2(N) \quad \longrightarrow c = N$ ou $N + 2\sqrt{N}$

$$\text{RSS}(\lambda) = N\sigma_b^2 \quad \longrightarrow \quad \lambda_\chi$$

- Il faut connaître σ_b^2
- $\left[\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}(\lambda)\right] \neq [\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}]$
- sur-régularisation

11.2.2 Risque moyen

$$\mathrm{RM} \stackrel{\mathrm{def}}{=} \mathrm{E}\left[\|\boldsymbol{f} - \widehat{\boldsymbol{f}}(\lambda,\boldsymbol{g})\|^2\right] = \int \left[\boldsymbol{f} - \widehat{\boldsymbol{f}}(\lambda,\boldsymbol{g})\right]^2 p(\boldsymbol{g}|\boldsymbol{f},\lambda)\,\mathrm{d}\boldsymbol{g}$$

$$\widehat{\boldsymbol{f}}(\lambda) = \left[\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{D}^t\boldsymbol{D}\right]^{-1}\boldsymbol{H}^t\boldsymbol{g}$$

$$\begin{aligned}\hat{\lambda} &= \arg\min_{\lambda}\left\{\mathrm{E}\left[\|\boldsymbol{f} - \widehat{\boldsymbol{f}}(\lambda)\|^2\right]\right\} \\ &= \|[\boldsymbol{I} - \boldsymbol{A}(\lambda)]\boldsymbol{H}\boldsymbol{f}\|^2 + \sigma_b^2\mathrm{Tr}\left\{\boldsymbol{A}(\lambda)\right\}^2\end{aligned}$$

où

$$\boldsymbol{A}(\lambda) = \boldsymbol{H}\left[\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{D}^t\boldsymbol{D}\right]^{-1}\boldsymbol{H}^t$$

et si $\boldsymbol{H}$ n'est pas singulière

$$\boldsymbol{A}(\lambda) = [\boldsymbol{I} - \lambda\boldsymbol{Q}]^{-1} \quad \text{avec} \quad \boldsymbol{Q} = [\boldsymbol{H}^t]^{-1}\boldsymbol{D}^t\boldsymbol{D}\boldsymbol{H}^{-1}$$

$$\lambda_{\mathrm{RM}} = \arg\min_{\lambda}\left\{\mathrm{RM}(\lambda,\boldsymbol{f})\right\}$$

- Il faut connaître σ_b^2
- $\mathrm{RM}(\lambda,\boldsymbol{f})$ dépend du vrai $\boldsymbol{f}$ qui est inconnu

11.2.3 Validation croisée

L'idée de base dans les techniques de validations croisées est de partionner les données en deux sous ensembles. On utiliserais alors les données $\boldsymbol{g}_1$ dans le première sous ensemble pour calculer la solution $\widehat{\boldsymbol{f}}(\lambda)$ et en déduire (ou prédire) celles du deuxième ; $\widehat{\boldsymbol{g}}_2 = \boldsymbol{H}\widehat{\boldsymbol{f}}(\lambda)$. On peut alors déterminer la valeur optimale (au sens de la VC) du λ par :

$$\widehat{\lambda} = \arg\min_{\lambda} \left\{ \mathrm{E}\left[\| \boldsymbol{g}_2 - \widehat{\boldsymbol{g}}_2(\lambda) \| \right]^2 \right\}$$

Une partition la plus simple possible est d'éliminer à chaque étape une seule donnée g_k, c'est à dire :

$$P_1 = \{g_1, \ldots, g_{k-1}, g_{k+1}, \ldots, g_M\}, \quad P_2 = \{g_k\}$$

Notons par $\boldsymbol{g}^{(-k)}$ $[g_1, \ldots, g_{k-1}, g_{k+1}, \ldots, g_N]$. Le critère de VC peut alors s'écrire :

$$\mathrm{VC}(\lambda) \stackrel{\mathrm{def}}{=} \sum_{k=1}^{M} \left[g_k - \boldsymbol{h}_k^t \widehat{\boldsymbol{f}}\left(\lambda, \boldsymbol{g}^{(-k)}\right) \right]^2$$

et

$$\widehat{\lambda}_{VC} = \arg\min_{\lambda} \{\mathrm{VC}(\lambda)\}$$

Pour aller un peu plus dans le détail, considérons le cas de la régularisation quadratique :

$$\widehat{\boldsymbol{f}}\left(\lambda, \boldsymbol{g}^{(-k)}\right) = \left[\boldsymbol{H}^t\boldsymbol{H} - \boldsymbol{h}_k\boldsymbol{h}_k^t + \lambda \boldsymbol{P}_0^{-1} \right]^{-1} \left[\boldsymbol{H}^t\boldsymbol{g} - \boldsymbol{h}_k \boldsymbol{g}^{(-k)} \right]$$

$$+$$

lemme d'inversion des matrices

$$\Downarrow$$

$$\mathrm{VC}(\lambda) = [\boldsymbol{B}(\lambda)[\boldsymbol{I} - \boldsymbol{A}(\lambda)]\boldsymbol{g}]^2$$

- $\boldsymbol{B}(\lambda) = \mathrm{diag}\left\{(1 - a_{kk})^{-1}\right\}$
- Le calcul de $\mathrm{VC}(\lambda)$ se ramène à celui des éléments diagonaux de $\boldsymbol{A}(\lambda)$
- Si $\boldsymbol{A}(\lambda)$ diagonale, $\mathrm{VC}(\lambda)$ n'a pas de minimum

11.2.4 Validation croisée généralisée

$$\mathrm{VCG}(\lambda) = \frac{\mathrm{VC}(\lambda)}{\mathrm{Tr}\left\{\boldsymbol{I} - \boldsymbol{A}(\lambda)\right\}^2}$$

$$\mathrm{VCG}(\lambda) \stackrel{\mathrm{def}}{=} \sum_{k=1}^{M} w_k(\lambda)^2 \left[g_k - \boldsymbol{h}_k^t \widehat{\boldsymbol{f}}\left(\lambda, \boldsymbol{g}^{(-k)}\right)\right]^2$$

coeffs de pondération : $w_k(\lambda) = (1 - a_{kk})(1 - \mathrm{Tr}\left\{\boldsymbol{I} - \boldsymbol{A}(\lambda)\right\})^{-1}$

$$\widehat{\lambda}_{VCG} = \arg\min_{\lambda} \left\{\mathrm{VCG}(\lambda)\right\}$$

Calcul de VCG(λ)

Ici aussi, comme dans le cas précédent, en utilisant la lemme d'inversion des matrices, il est possible d'exprimer le critère en fonctions des termes calculable d'une manière récursive :

$$\mathrm{VCG}(\lambda) = \sum_{k=1}^{M} \left(\frac{u_k}{v_k}\right)^2$$

- $u_k = g_k - \boldsymbol{h}_k^t \widehat{\boldsymbol{f}}\left(\lambda, \boldsymbol{g}^{(-k)}\right)$ résidu
- $v_k = a_{kk} - 1$
- Auxiliaire de calcul :

$$\begin{aligned} \boldsymbol{P}_{M|M} &= (\boldsymbol{H}^t\boldsymbol{H} + \lambda \boldsymbol{P}_0^t)^{-1} \quad \text{Covariance de } (\widehat{\boldsymbol{f}} - \boldsymbol{f}) \\ \boldsymbol{A}(\lambda) &= \boldsymbol{H}\boldsymbol{P}_{M|M}\boldsymbol{H}^t \end{aligned}$$

11.3 Un autre regard

$$\boldsymbol{f} \sim \mathcal{N}(0,\sigma_x^2 \boldsymbol{P}_0)$$

$$\boldsymbol{b} \sim \mathcal{N}(0,\sigma_b^2 \boldsymbol{I})$$

$$\boldsymbol{g}|\boldsymbol{f} \sim \mathcal{N}(\boldsymbol{H}\boldsymbol{f},\sigma_b^2 \boldsymbol{I})$$

Problème 1 :
Étant donné $\boldsymbol{H}$, σ_x^2,$\boldsymbol{P}_0$,σ_b^2 et $\boldsymbol{g}$, estimer $\boldsymbol{f}$

Solution :

$$\boldsymbol{f}|\boldsymbol{g} \sim \mathcal{N}(\widehat{\boldsymbol{f}},\boldsymbol{P}) \quad \left\{ \begin{array}{rcl} \widehat{\boldsymbol{f}} & = & \left(\boldsymbol{H}^t\boldsymbol{H} + \lambda \boldsymbol{P}_0^{-1}\right)^{-1} \boldsymbol{H}^t \boldsymbol{g} \\ \boldsymbol{P} & = & \sigma_b^2 \left(\boldsymbol{H}^t\boldsymbol{H} + \lambda \boldsymbol{P}_0^{-1}\right)^{-1} \end{array} \right. , \quad \lambda = \frac{\sigma_b^2}{\sigma_x^2}$$

Problème 2 :
Étant donné $\boldsymbol{H}$, $\boldsymbol{P}_0$,σ_b^2 et $\boldsymbol{g}$, estimer σ_x^2 et $\boldsymbol{f}$

ou d'une manière équivalente :

Étant donné $\boldsymbol{H}$, $\boldsymbol{P}_0$,σ_b^2, et $\boldsymbol{g}$, estimer λ et $\boldsymbol{f}$

Problème 3 :
Étant donné $\boldsymbol{H}$, $\boldsymbol{P}_0$ et $\boldsymbol{g}$, estimer σ_x^2,σ_b^2 et $\boldsymbol{f}$

ou d'une manière équivalente :

Étant donné $\boldsymbol{H}$, $\boldsymbol{P}_0$ et $\boldsymbol{g}$, estimer λ,σ_b^2 et $\boldsymbol{f}$

Notons : $\boldsymbol{\theta} = [\sigma_x^2,\sigma_b^2]$ ou $\boldsymbol{\theta} = [\lambda,\sigma_b^2]$

Adéquation aux données

$$Z = \frac{1}{\sigma_b^2}\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 \sim \text{loi de } \chi^2(N)$$

$$\mathrm{E}\left[\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2\right] = N\sigma_b^2$$

- Première approximation :

$$\mathrm{E}\left[\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2\right] \approx \mathrm{RSS}(\lambda) = \|\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}(\lambda)\|^2$$

$$\|\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}(\lambda)\|^2 = N\sigma_b^2 \quad \longrightarrow \quad \widehat{\lambda} \tag{11.1}$$

- Deuxième approximation :

$$\mathrm{E}\left[\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2\right] \approx \frac{\mathrm{RSS}(\lambda)}{\mathrm{Tr}\left\{\boldsymbol{I} - \boldsymbol{A}(\lambda)\right\}^2} = \frac{\|\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}(\lambda)\|^2}{\mathrm{Tr}\left\{\boldsymbol{I} - \boldsymbol{A}(\lambda)\right\}^2}$$

où

$$\boldsymbol{A}(\lambda) = \boldsymbol{H}\left[\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{D}^t\boldsymbol{D}\right]^{-1}\boldsymbol{H}^t$$

$$\frac{\|\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}(\lambda)\|^2}{\mathrm{Tr}\left\{\boldsymbol{I} - \boldsymbol{A}(\lambda)\right\}^2} = N\sigma_b^2 \quad \longrightarrow \quad \widehat{\lambda} \tag{11.2}$$

- Les deux équations (11.1) et (11.2) sont non linéaires
- Dans le cadre des approximations circulantes des matrices, il est possible d'obtenir des relation explicites pour leurs solutions.

Risque moyen

$$\mathrm{E}\left[\|\boldsymbol{f} - \widehat{\boldsymbol{f}}(\boldsymbol{\theta},\boldsymbol{g})\|^2\right] = \int \left[\boldsymbol{f} - \widehat{\boldsymbol{f}}(\boldsymbol{\theta},\boldsymbol{g})\right]^2 p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{g}$$

Validation croisée

$$\mathrm{E}\left[\|\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}(\boldsymbol{\theta},\boldsymbol{g})\|^2\right] = \int \left[\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}(\boldsymbol{\theta},\boldsymbol{g})\right]^2 p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{g}$$

Validation croisée généralisée

$$\mathrm{E}\left[\|\boldsymbol{H}\boldsymbol{f} - \widehat{\boldsymbol{g}}(\boldsymbol{\theta},\boldsymbol{g})\|^2\right] = \int \left[\boldsymbol{H}\boldsymbol{f} - \widehat{\boldsymbol{g}}(\boldsymbol{\theta},\boldsymbol{g})\right]^2 p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{g}$$

Vraisemblance généralisée

$$(\widehat{\boldsymbol{f}},\widehat{\boldsymbol{\theta}}) = \underset{(\boldsymbol{f},\boldsymbol{\theta})}{\arg\max}\,\{p(\boldsymbol{f},\boldsymbol{g};\boldsymbol{\theta})\} = \underset{(\boldsymbol{f},\boldsymbol{\theta})}{\arg\max}\,\{p(\boldsymbol{g}|\boldsymbol{f};\boldsymbol{\theta})p(\boldsymbol{f};\boldsymbol{\theta})\}$$

Vraisemblance marginale

$$\begin{cases} \widehat{\boldsymbol{\theta}} &= \underset{\boldsymbol{\theta}}{\arg\max}\,\{p(\boldsymbol{g};\boldsymbol{\theta})\} = \underset{\boldsymbol{\theta}}{\arg\max}\left\{\int p(\boldsymbol{g}|\boldsymbol{f};\boldsymbol{\theta})p(\boldsymbol{f};\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{f}\right\} \\ \widehat{\boldsymbol{f}} &= \underset{\boldsymbol{f}}{\arg\max}\left\{p(\boldsymbol{f}|\boldsymbol{g};\widehat{\boldsymbol{\theta}})\right\} \end{cases}$$

La principale difficultée réside dans le calcul de $p(\boldsymbol{g};\boldsymbol{\theta})$. C'est pourquoi, en général, on utilise l'algorithme EM pour calculer la solution $\widehat{\boldsymbol{\theta}}$.

Pseudo-vraisemblance

Lorsque $\boldsymbol{\theta}$ constitue les paramètres de la loi *a priori*, $p(\boldsymbol{f};\boldsymbol{\theta})$ et lorsque cette loi est une loi markovienne, on peut envisager d'approximer la vraisemblance par

$$p(\boldsymbol{g};\boldsymbol{\theta}) \approx p(\boldsymbol{g}|\boldsymbol{f};\boldsymbol{\theta}) \sum_i p\left(f_i|f_j;\boldsymbol{\theta};j\in v_i\right)$$

Contexte complètement bayésien

$$(\widehat{\boldsymbol{f}},\widehat{\boldsymbol{\theta}}) = \underset{(\boldsymbol{f},\boldsymbol{\theta})}{\arg\max}\,\{p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g})\} = \underset{(\boldsymbol{f},\boldsymbol{\theta})}{\arg\max}\,\{p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta})p(\boldsymbol{f}|\boldsymbol{\theta})p(\boldsymbol{\theta})\}$$

Ici, la principale difficulté est le choix de la loi $p(\boldsymbol{\theta})$ de telle sorte que $p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g})$ soit unimodale et que la solution puisse exister et soit unique.

11.4 Algorithme EM

(Expectation-Maximization)

Rechercher la solution à maximum de vraisemblance
Exemple :

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \{p(\boldsymbol{g};\boldsymbol{\theta})\} = \arg\max_{\boldsymbol{\theta}} \{\ln p(\boldsymbol{g};\boldsymbol{\theta})\}$$

- vraisemblance $l(\boldsymbol{\theta}) = p(\boldsymbol{g};\boldsymbol{\theta})$
- log-vraisemblance $L(\boldsymbol{\theta}) = \ln p(\boldsymbol{g};\boldsymbol{\theta})$
- Calcul de $p(\boldsymbol{g};\boldsymbol{\theta})$ n'est souvent pas aisé
- Algorithme EM permet d'atteindre le maximum sans calculer explicitement $p(\boldsymbol{g};\boldsymbol{\theta})$:

Le principe de l'algorithme EM consiste à :

- Introduire une variable auxiliaire $\boldsymbol{z}$ reliée à $\boldsymbol{g}$
- $\boldsymbol{z}$ données complètes
- $\boldsymbol{g}$ données incomplètes
- $\{\boldsymbol{z}\} \mapsto \{\boldsymbol{g}\}$ Projection : $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{z}$

$$p(\boldsymbol{z}|\boldsymbol{g};\boldsymbol{\theta}) = \frac{p(\boldsymbol{z},\boldsymbol{g};\boldsymbol{\theta})}{p(\boldsymbol{g};\boldsymbol{\theta})} = \frac{p(\boldsymbol{z};\boldsymbol{\theta})p(\boldsymbol{g}|\boldsymbol{z})}{p(\boldsymbol{g};\boldsymbol{\theta})} = \frac{p(\boldsymbol{z};\boldsymbol{\theta})I_{\boldsymbol{g}}(\boldsymbol{z})}{p(\boldsymbol{g};\boldsymbol{\theta})}$$

où $I_{\boldsymbol{g}}(\boldsymbol{z})$ est la fonction indicatrice : $I_{\boldsymbol{g}}(\boldsymbol{z}) = \begin{cases} 1 & \text{si } \boldsymbol{g} = \boldsymbol{H}\boldsymbol{z} \\ 0 & \text{sinon} \end{cases}$

$$p(\boldsymbol{g};\boldsymbol{\theta}) = \int_{I_{\boldsymbol{g}}(\boldsymbol{z})} p(\boldsymbol{z};\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{z}$$

- Algorithme iteratif :

$$\begin{cases} \text{(E)} & \text{Calculer} & Q(\boldsymbol{\theta};\widehat{\boldsymbol{\theta}}^{(k)}) & = \mathrm{E}_{Z|Y;\widehat{\boldsymbol{\theta}}^{(k)}}\{\ln p(\boldsymbol{z};\boldsymbol{\theta})\} \\ \text{(M)} & \text{Choisir} & \widehat{\boldsymbol{\theta}}^{(k+1)} & = \arg\max_{\boldsymbol{\theta}} \left\{ Q(\boldsymbol{\theta};\widehat{\boldsymbol{\theta}}^{(k)}) \right\} \end{cases}$$

- $\widehat{\boldsymbol{\theta}}^{k}$ converge vers $\widehat{\boldsymbol{\theta}}$

$$\begin{aligned} Q(\boldsymbol{\theta};\widehat{\boldsymbol{\theta}}^{(k)}) &= \mathrm{E}\left[\ln p(\boldsymbol{z};\boldsymbol{\theta})|\boldsymbol{g};\widehat{\boldsymbol{\theta}}^{(k)}\right] \\ &= \mathrm{E}_{Z|Y;\widehat{\boldsymbol{\theta}}^{(k)}}\{\ln p(\boldsymbol{z};\boldsymbol{\theta})\} \\ &= \int_{I_{\boldsymbol{g}}(\boldsymbol{z})} \ln p(\boldsymbol{z};\boldsymbol{\theta})\, p(\boldsymbol{z}|\boldsymbol{g};\widehat{\boldsymbol{\theta}}^{(k)})\,\mathrm{d}\boldsymbol{z} \end{aligned}$$

11.5 Exemples et exercises :

Exercise 1 :

$$y = x + b, \qquad x \sim \mathcal{N}\left(0,\sigma_x^2\right), \qquad b \sim \mathcal{N}\left(0,\sigma_b^2\right)$$

Vérifier les relations suivantes :

$$\begin{aligned}
y|x &\sim \mathcal{N}\left(x,\sigma_b^2\right), \\
y &\sim \mathcal{N}\left(0,\sigma_x^2+\sigma_b^2\right) \\
\begin{pmatrix} x \\ y \end{pmatrix} &\sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_x^2 \\ \sigma_x^2 & \sigma_x^2+\sigma_b^2 \end{pmatrix}\right) \\
x|y &\sim \mathcal{N}\left(\widehat{x}, \frac{\sigma_x^2\,\sigma_b^2}{\sigma_x^2+\sigma_b^2}\right), \quad \text{avec} \quad \widehat{x} = \frac{\sigma_x^2}{\sigma_x^2+\sigma_b^2} y
\end{aligned}$$

Exercise 2 :

Étant donné y et $\sigma_b^2 = 1$, estimer x et $\theta = \sigma_x^2$

Vérifier les relations suivantes :

$$\begin{aligned}
x \sim \mathcal{N}(0,\theta) &\longrightarrow p(x;\theta) = \frac{1}{\sqrt{2\pi\theta}} \exp\left[-\frac{x^2}{2\theta}\right] \\
b \sim \mathcal{N}(0,1) &\longrightarrow p(b) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{b^2}{2}\right] \\
y|x \sim \mathcal{N}(x,1) &\longrightarrow p(y|x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y-x)^2}{2}\right] \\
y \sim \mathcal{N}(0,\theta+1) &\longrightarrow p(y;\theta) = \frac{1}{\sqrt{2\pi(\theta+1)}} \exp\left[-\frac{y^2}{2(\theta+1)}\right]
\end{aligned}$$

$$\Sigma = \begin{pmatrix} \theta & \theta \\ \theta & \theta+1 \end{pmatrix}, \qquad \det|\Sigma| = \theta$$

$$\Sigma^{-1} = \frac{1}{\theta}\begin{pmatrix} \theta+1 & -\theta \\ -\theta & \theta \end{pmatrix}, \qquad \det\left|\Sigma^{-1}\right| = 1$$

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix},\Sigma\right] \longrightarrow p(x,y;\theta) = \frac{\det|\Sigma|^{-\frac{1}{2}}}{2\pi} \exp\left[-\frac{1}{2}(x,y)\Sigma^{-1}\begin{pmatrix} x \\ y \end{pmatrix}\right]$$

$$x|y \sim \mathcal{N}(\widehat{x}, \frac{\theta}{\theta+1}), \quad \widehat{x} = \frac{\theta}{\theta+1} y$$

$$p(x|y;\theta) = \frac{1}{\sqrt{2\pi\frac{\theta}{\theta+1}}} \exp\left[-\frac{(\theta+1)(x-\widehat{x})^2}{2\theta}\right]$$

Vraisemblance généralisée :

$$\begin{aligned}
\mathrm{VG}(x,\theta|y) &= p(x,y;\theta) = p(y|x)p(x;\theta) \\
&= \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y-x)^2}{2}\right] \frac{1}{\sqrt{2\pi\theta}} \exp\left[-\frac{x^2}{2\theta}\right] \\
&= \frac{1}{2\pi\sqrt{\theta}} \exp\left[-\frac{1}{2}(y^2 + \frac{\theta+1}{\theta}x^2 - 2xy)\right]
\end{aligned}$$

Vraisemblance marginale :

$$\begin{aligned}
\mathrm{VM}(\theta|y) &= \int p(x,y;\theta)\,\mathrm{d}x = p(y;\theta) \\
&= \frac{1}{\sqrt{2\pi(\theta+1)}} \exp\left[-\frac{y^2}{2(\theta+1)}\right]
\end{aligned}$$

Estimation au sens du MAP de x si θ connu :

$$\begin{aligned}
\widehat{x} &= \arg\max_x \{p(x|y;\theta)\} = \arg\max_x \{p(y|x;\theta)p(x;\theta)\} \\
&= \arg\max_x \left\{ \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y-x)^2}{2}\right] \frac{1}{\sqrt{2\pi\theta}} \exp\left[-\frac{x^2}{2\theta}\right] \right\} \\
&= \arg\max_x \left\{ \frac{1}{2\pi\sqrt{\theta}} \exp\left[-\frac{1}{2}(y^2 + \frac{\theta}{\theta+1}x^2 - 2xy)\right] \right\} \\
&= \frac{\theta}{\theta+1}y \quad \longrightarrow \widehat{x} \sim \mathcal{N}\left(\frac{\theta}{\theta+1}y, \frac{\theta}{\theta+1}\right)
\end{aligned}$$

Estimation au sens du MV de x si θ connu :

$$\begin{aligned}
\widehat{x} &= \arg\max_x \{p(y|x;\theta)\} \\
&= \arg\max_x \left\{ \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y-x)^2}{2}\right] \right\} \\
&= y \quad \longrightarrow \widehat{x} \sim \mathcal{N}(y,1)
\end{aligned}$$

Vraisemblance généralisée :

$$(\widehat{x},\widehat{\theta}) = \arg\max_{(x,\theta)} \{p(x,y;\theta)\} = \arg\min_{(x,\theta)} \{-\ln p(x,y;\theta)\}$$

$$\begin{aligned} p(x,y;\theta) &= p(y|x)p(x;\theta) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y-x)^2}{2}\right] \frac{1}{\sqrt{2\pi\theta}} \exp\left[-\frac{x^2}{2\theta}\right] \\ &= \frac{1}{2\pi\sqrt{\theta}} \exp\left[-\frac{1}{2}\left[(y-x)^2 + \frac{x^2}{\theta}\right]\right] \end{aligned}$$

$$L(x,\theta) = -\ln p(x,y;\theta) = \ln(2\pi) + \frac{1}{2}\left[\ln(\theta) + y^2 + \frac{\theta+1}{\theta}x^2 - 2xy\right]$$

$$\left\{ \begin{array}{rcl} \dfrac{\partial L(x,\theta)}{\partial x} &=& \left(\dfrac{\theta+1}{\theta}x - y\right) = 0 \quad \longrightarrow x = \dfrac{\theta}{\theta+1}y \\ \dfrac{\partial L(x,\theta)}{\partial \theta} &=& \dfrac{1}{2}\left[\dfrac{1}{\theta} - \dfrac{x^2}{\theta^2}\right] = 0 \quad \longrightarrow \theta = x^2 \end{array} \right.$$

$$\left\{ \begin{array}{rcl} \widehat{x} &=& \dfrac{\theta}{\theta+1}y \\ \widehat{\theta} &=& \widehat{x}^2 = \left(\dfrac{\widehat{\theta}}{\widehat{\theta}+1}\right)^2 y^2 \end{array} \right.$$

Condition aux limites :

$$\theta = \left(\frac{\widehat{\theta}}{\widehat{\theta}+1}\right)^2 y^2 \longrightarrow \theta^2 + 2\theta + 1 - \theta y^2 = 0 \longrightarrow \Delta = (2-y^2)^2 - 4 >= 0 \longrightarrow 0 > y^2 > 4$$

$$\left\{ \begin{array}{rcl} \widehat{x} &=& \arg\max\limits_{x} \left\{p(x|y;\widehat{\theta})\right\} \\ \widehat{\theta} &=& \arg\max\limits_{\theta} \{p(\widehat{x};\theta)\} \end{array} \right.$$

Vraisemblance marginale :

$$\left\{\begin{array}{rcl} \widehat{\theta} & = & \arg\max\limits_{\theta} \{p(y;\theta)\} \\ \widehat{x} & = & \arg\max\limits_{x} \left\{p(x|y;\widehat{\theta})\right\} \end{array}\right.$$

$$\left\{\begin{array}{rcl} p(y;\theta) & = & \dfrac{1}{\sqrt{2\pi(\theta+1)}} \exp\left[-\dfrac{y^2}{2(\theta+1)}\right] \\ p(x|y;\theta) & = & \dfrac{1}{\sqrt{2\pi\frac{\theta}{\theta+1}}} \exp\left[-\dfrac{(\theta+1)(x-\widehat{x})^2}{2\theta}\right] \end{array}\right.$$

$$\left\{\begin{array}{rcl} \widehat{\theta} & = & y^2-1 \text{ si } y^2>1 \\ \widehat{x} & = & \dfrac{\widehat{\theta}}{\widehat{\theta}+1} y = y - \dfrac{1}{y} \end{array}\right.$$

Contexte complètement bayésien :

$$(\widehat{x},\widehat{\theta}) = \arg\max_{(x,\theta)} \{p(x,\theta|y)\} = \arg\max_{(x,\theta)} \{p(y|x)p(x|\theta)p(\theta)\} = \arg\min_{(x,\theta)} \{-lnp(y,x,\theta)\}$$

$$\begin{array}{rcl} L(x,\theta) & = & \ln p(x,y;\theta) \\ & = & -\ln(2\pi) - \dfrac{1}{2}\left[\ln(\theta) + y^2 + \dfrac{\theta+1}{\theta}x^2 - 2xy\right] + \ln p(\theta) \end{array}$$

$$\left\{\begin{array}{rcll} \dfrac{\mathrm{d}L(x,\theta)}{\mathrm{d}x} & = & -\dfrac{\theta+1}{\theta}x + y = 0 & \longrightarrow x = \dfrac{\theta}{\theta+1}y \\ \dfrac{\mathrm{d}L(x,\theta)}{\mathrm{d}\theta} & = & -\dfrac{1}{2}\left[\dfrac{3}{\theta} - \dfrac{x^2}{\theta^2}\right] = 0 & \longrightarrow \theta = \dfrac{1}{3}x^2 \end{array}\right.$$

Avec un choix de $p(\theta) \propto \frac{1}{\theta}$ on a :

$$p(\theta) \propto \frac{1}{\theta} \quad \longrightarrow \quad \left\{\begin{array}{rcl} \widehat{x} & = & \dfrac{\theta}{\theta+1}y \\ \widehat{\theta} & = & \dfrac{1}{3}\widehat{x}^2 \end{array}\right.$$

$$p(\theta) \propto \frac{1}{\sqrt{\theta}} \quad \longrightarrow \quad \left\{\begin{array}{rcl} \widehat{x} & = & \dfrac{\theta}{\theta+1}y \\ \widehat{\theta} & = & \dfrac{1}{2}\widehat{x}^2 \end{array}\right.$$

Algorithme EM :

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \quad \longrightarrow p(z;\theta) = p(x,y;\theta) = \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma\right)$$

$$\Sigma = \begin{pmatrix} \theta & \theta \\ \theta & \theta + 1 \end{pmatrix}$$

$$-\frac{1}{2}(x,y)\Sigma^{-1}\begin{pmatrix} x \\ y \end{pmatrix} = -\frac{1}{2}\left(\frac{\theta+1}{\theta}x^2 - 2xy + y^2 + \ln\det|\Sigma| + 2\ln 2\pi\right)$$

$$\ln p(z;\theta) \propto -\ln\theta - \frac{\theta+1}{\theta}x^2 + 2xy - y^2$$

$$p(z|y;\widehat{\theta}) = p(x,y|y;\widehat{\theta}) = p(x|y;\widehat{\theta})$$

$$\mathrm{E}\left[\ln p(z;\theta)|y;\widehat{\theta}\right] = \ln\theta - \frac{\theta+1}{\theta}\mathrm{E}\left[x^2|y;\widehat{\theta}\right] + 2y\mathrm{E}\left[x|y;\widehat{\theta}\right] - y^2$$

$$Q(\theta,\widehat{\theta}) = \ln\theta - \frac{\theta+1}{\theta}\mathrm{E}\left[x^2|y;\widehat{\theta}\right] + 2y\mathrm{E}\left[x|y;\widehat{\theta}\right] + \mathrm{cte}$$

$$\begin{aligned} \frac{\partial Q(\theta,\widehat{\theta}^{(k)})}{\partial\theta} = 0 \quad \longrightarrow \widehat{\theta}^{(k+1)} &= \mathrm{E}\left[x^2|y;\widehat{\theta}^{(k)}\right] \\ &= \frac{\theta^{(k)}}{\theta^{(k)}+1} + \left(\frac{\theta^{(k)}}{\theta^{(k)}+1}y\right)^2 \end{aligned}$$

$$\widehat{\theta}^{(k+1)} = -\frac{\theta^{(k)}}{\theta^{(k)}+1} - \left(\frac{\theta^{(k)}}{\theta^{(k)}+1}y\right)^2$$

Exercise 3 :
Extension au cas : $\boldsymbol{g = Hf + b}$

$$\boldsymbol{f} \sim \mathcal{N}(0, \boldsymbol{R}_X = \sigma_x^2 \boldsymbol{P}_0), \qquad \boldsymbol{b} \sim \mathcal{N}(0, \boldsymbol{R}_B = \sigma_b^2 \boldsymbol{I})$$

Vérifier les relations suivantes :

$$\begin{array}{rcl}
\boldsymbol{g}|\boldsymbol{f} & \sim & \mathcal{N}(\boldsymbol{Hf}, \boldsymbol{R}_B = \sigma_b^2 \boldsymbol{I}) \\
\boldsymbol{g} & \sim & \mathcal{N}(0, \boldsymbol{R}_Y) \quad \text{avec} \quad \boldsymbol{R}_Y = \boldsymbol{H}\boldsymbol{R}_X\boldsymbol{H}^t + \boldsymbol{R}_B \\
\begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{g} \end{pmatrix} & \sim & \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Sigma}\right) \quad \text{avec} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{R}_X & \boldsymbol{R}_{XY} \\ \boldsymbol{R}_{YX} & \boldsymbol{R}_Y \end{pmatrix} \\
 & & \text{où} \qquad \boldsymbol{R}_{YX} = \boldsymbol{R}_{XY}^t = \boldsymbol{H}\boldsymbol{R}_X \\
\boldsymbol{f}|\boldsymbol{g} & \sim & \mathcal{N}(\widehat{\boldsymbol{f}}, \boldsymbol{P})
\end{array}$$

avec

$$\left\{ \begin{array}{rcl}
\widehat{\boldsymbol{f}} & = & \boldsymbol{R}_X\boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{R}_X\boldsymbol{H}^t + \boldsymbol{R}_B)^{-1}\boldsymbol{g} = \boldsymbol{P}\boldsymbol{H}^t\boldsymbol{R}_B^{-1}\boldsymbol{g} \\
\boldsymbol{P} & = & \boldsymbol{R}_X - \boldsymbol{R}_X\boldsymbol{H}^t(\boldsymbol{H}\boldsymbol{R}_X\boldsymbol{H}^t + \boldsymbol{R}_B)^{-1}\boldsymbol{H}\boldsymbol{R}_X = (\boldsymbol{R}_X^{-1} + \boldsymbol{H}^t\boldsymbol{R}_B^{-1}\boldsymbol{H})^{-1}
\end{array} \right.$$

Avec $\boldsymbol{R}_B = \sigma_b^2 \boldsymbol{I}$ et $\boldsymbol{R}_X = \sigma_x^2 \boldsymbol{P}_0$ on obtient :

$$\left\{ \begin{array}{rcl}
\widehat{\boldsymbol{f}} & = & \left(\boldsymbol{H}^t\boldsymbol{H} + \boldsymbol{\lambda}\boldsymbol{P}_0^{-1}\right)^{-1}\boldsymbol{H}^t\boldsymbol{g} \quad \text{avec} \quad \lambda = \dfrac{\sigma_b^2}{\sigma_x^2} = \dfrac{1}{\theta} \\
\boldsymbol{P} & = & \sigma_b^2\left(\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{P}_0^{-1}\right)^{-1}
\end{array} \right.$$

Exercise 4 :
Étant donnée $\boldsymbol{H}$,$\boldsymbol{g}$,$\sigma_b^2 = 1$ et $\boldsymbol{P}_0$ estimer $\boldsymbol{f}$ et $\theta = \sigma_x^2$.

Vérifier et compléter les relations suivantes :

$$\begin{aligned} \boldsymbol{f} &\sim \mathcal{N}(0,\theta\boldsymbol{P}_0) \\ \boldsymbol{b} &\sim \mathcal{N}(0,\boldsymbol{I}) \\ \boldsymbol{g}|\boldsymbol{f} &\sim \mathcal{N}(\boldsymbol{H}\boldsymbol{f},\boldsymbol{I}) \\ \boldsymbol{g} &\sim \mathcal{N}(0,\boldsymbol{R}_Y) \end{aligned}$$

$$\boldsymbol{\Sigma} = \theta \begin{pmatrix} \boldsymbol{P}_0 & \boldsymbol{P}_0^t\boldsymbol{H}^t \\ \boldsymbol{H}\boldsymbol{P}_0 & \boldsymbol{H}\boldsymbol{P}_0\boldsymbol{H}^t + \boldsymbol{I} \end{pmatrix},$$

$$\det|\boldsymbol{\Sigma}| = \det\left|\boldsymbol{R}_X - \boldsymbol{R}_{XY}\boldsymbol{R}_Y^{-1}\boldsymbol{R}_{YX}\right| \det|\boldsymbol{R}_Y|$$

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \boldsymbol{P}^{-1} & -\boldsymbol{P}^{-1}\boldsymbol{R}_{XY}\boldsymbol{R}_Y^{-1} \\ -\boldsymbol{R}_Y^{-1}\boldsymbol{R}_{YX}\boldsymbol{P}^{-1} & \boldsymbol{R}_Y^{-1} + \boldsymbol{R}_Y^{-1}\boldsymbol{R}_{YX}\boldsymbol{P}^{-1}\boldsymbol{R}_{XY}\boldsymbol{R}_Y^{-1} \end{pmatrix}$$

$$\det\left|\boldsymbol{\Sigma}^{-1}\right| =$$

$$\begin{aligned} \begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{g} \end{pmatrix} &\sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix},\boldsymbol{\Sigma}\right] \\ &= \frac{\det|\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{(2\pi)^{\frac{N+M}{2}}} \exp\left[-\frac{1}{2}[\boldsymbol{f},\boldsymbol{g}]^t\boldsymbol{\Sigma}^{-1}\begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{g} \end{pmatrix}\right] \end{aligned}$$

$$\boldsymbol{f}|\boldsymbol{g} \sim \mathcal{N}(\widehat{\boldsymbol{f}},\boldsymbol{P}) = (2\pi)^{-\frac{N}{2}}\det|\boldsymbol{P}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left[[\boldsymbol{f}-\widehat{\boldsymbol{f}}]^t\boldsymbol{P}^{-1}[\boldsymbol{f}-\widehat{\boldsymbol{f}}]\right]\right]$$

avec

$$\left\{\begin{array}{rcl} \widehat{\boldsymbol{f}} &=& \left(\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{P}_0^{-1}\right)^{-1}\boldsymbol{H}^t\boldsymbol{g} \quad \text{avec} \quad \lambda = \dfrac{1}{\theta} \\ \boldsymbol{P} &=& (\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{P}_0^{-1})^{-1} \end{array}\right.$$

Vraisemblance généralisée :

$$\mathrm{VG}(\boldsymbol{f},\theta|\boldsymbol{g}) = p(\boldsymbol{f},\boldsymbol{g};\theta) = p(\boldsymbol{g}|\boldsymbol{f})p(\boldsymbol{f};\theta)$$

$$\begin{aligned}
&= (2\pi)^{-\frac{M}{2}} \det|\boldsymbol{R}_B|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left[[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]^t \boldsymbol{R}_B^{-1}[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]\right]\right] \times \\
&\quad \times (2\pi)^{-\frac{N}{2}} \det|\boldsymbol{R}_X|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left[\boldsymbol{f}^t \boldsymbol{R}_X^{-1}\boldsymbol{f}\right]\right] \\
&= (2\pi)^{-\frac{N}{2}} (\theta \det|\boldsymbol{P}_0|)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\theta}\left[\theta[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]^t[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}] + \boldsymbol{f}^t\boldsymbol{P}_0^{-1}\boldsymbol{f}\right]\right]
\end{aligned}$$

Vraisemblance marginale :

$$\begin{aligned}
\mathrm{VM}(\theta|\boldsymbol{g}) &= \int p(\boldsymbol{f},\boldsymbol{g};\theta)\,\mathrm{d}\boldsymbol{f} = p(\boldsymbol{g};\theta) \\
&= (2\pi)^{-\frac{M}{2}} \det|\boldsymbol{R}_Y|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left[\boldsymbol{g}^t\boldsymbol{R}_Y^{-1}\boldsymbol{g}\right]\right]
\end{aligned}$$

$$\boldsymbol{R}_Y = \theta\left(\boldsymbol{H}\boldsymbol{P}_0\boldsymbol{H}^t + \frac{1}{\theta}\right)$$

Estimation au sens du MAP de $\boldsymbol{f}$ si θ connu :

$$\begin{aligned}
\widehat{\boldsymbol{f}} &= \arg\max_{\boldsymbol{f}}\{p(\boldsymbol{f}|\boldsymbol{g};\theta)\} = \arg\max_{\boldsymbol{f}}\{p(\boldsymbol{g}|\boldsymbol{f};\theta)p(\boldsymbol{f};\theta)\} \\
&= \arg\min_{\boldsymbol{f}}\left\{[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]^t\boldsymbol{R}_B^{-1}[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}] + \boldsymbol{f}^t\boldsymbol{R}_X^{-1}\boldsymbol{f}\right\} \\
&= \left(\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{P}_0^{-1}\right)^{-1}\boldsymbol{H}^t\boldsymbol{g}, \quad \lambda = \frac{1}{\theta}
\end{aligned}$$

Estimation au sens du MV de $\boldsymbol{f}$ si θ connu :

$$\begin{aligned}
\widehat{\boldsymbol{f}} &= \arg\max_{\boldsymbol{f}}\{p(\boldsymbol{g}|\boldsymbol{f};\theta)\} \\
&= \arg\max_{\boldsymbol{f}}\left\{(2\pi)^{-\frac{M}{2}} \det|\boldsymbol{R}_B|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left[[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]^t\boldsymbol{R}_B^{-1}[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]\right]\right]\right\} \\
&= \arg\min_{\boldsymbol{f}}\left\{[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]^t\boldsymbol{R}_B^{-1}[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]\right\} \\
&= (\boldsymbol{H}^t\boldsymbol{H})^{-1}\boldsymbol{H}^t\boldsymbol{g}
\end{aligned}$$

Vraisemblance généralisée

$$(\widehat{\boldsymbol{f}},\widehat{\theta}) = \underset{(\boldsymbol{f},\theta)}{\arg\max}\,\{p(\boldsymbol{f},\boldsymbol{g};\theta)\} = \underset{(\boldsymbol{f},\theta)}{\arg\max}\,\{\ln p(\boldsymbol{f},\boldsymbol{g};\theta)\}$$

$$\begin{aligned} p(\boldsymbol{f},\boldsymbol{g};\theta) &= p(\boldsymbol{g}|\boldsymbol{f})p(\boldsymbol{f};\theta) \\ &= (2\pi)^{-\frac{M}{2}}\det|\boldsymbol{R}_B|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left[[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]^t\boldsymbol{R}_B^{-1}[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]\right]\right]\times \\ &\quad \times(2\pi)^{-\frac{N}{2}}\det|\boldsymbol{R}_X|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left[\boldsymbol{f}^t\boldsymbol{R}_X^{-1}\boldsymbol{f}\right]\right] \\ &= K\,\theta^{-\frac{1}{2}}\exp\left[-\frac{1}{2}[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]^t[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]-\frac{1}{2\theta}\boldsymbol{f}^t\boldsymbol{P}_0^{-1}\boldsymbol{f}\right] \end{aligned}$$

$$L(x,\theta) = \ln p(\boldsymbol{f},\boldsymbol{g};\theta) \propto \left[\ln\theta + [\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]^t[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}] + \frac{1}{\theta}\boldsymbol{f}^t\boldsymbol{P}_0^{-1}\boldsymbol{f}\right]$$

$$\begin{cases} \dfrac{\mathrm{d}L(\boldsymbol{f},\theta)}{\mathrm{d}\boldsymbol{f}} &= 0 \quad \longrightarrow \boldsymbol{f} = \left(\boldsymbol{H}^t\boldsymbol{H}+\lambda\boldsymbol{P}_0^{-1}\right)^{-1}\boldsymbol{H}^t\boldsymbol{g} \\ \dfrac{\mathrm{d}L(\boldsymbol{f},\theta)}{\mathrm{d}\theta} &= 0 \quad \longrightarrow \theta = \boldsymbol{f}^t\boldsymbol{P}_0^{-1}\boldsymbol{f} \end{cases}$$

$$\begin{cases} \widehat{\boldsymbol{f}} &= \left(\boldsymbol{H}^t\boldsymbol{H}+\lambda\boldsymbol{P}_0\right)^{-1}\boldsymbol{H}^t\boldsymbol{g} \\ \widehat{\theta} &= \widehat{\boldsymbol{f}}^t\boldsymbol{P}_0^{-1}\widehat{\boldsymbol{f}} \end{cases}$$

$$\begin{cases} \widehat{\boldsymbol{f}} &= \underset{\boldsymbol{f}}{\arg\max}\left\{p(\boldsymbol{f}|\boldsymbol{g};\widehat{\theta})\right\} \\ \widehat{\theta} &= \underset{\theta}{\arg\max}\left\{p(\widehat{\boldsymbol{f}};\theta)\right\} \end{cases}$$

Vraisemblance marginale :

$$\left\{\begin{array}{rcl} \widehat{\theta} & = & \arg\max\limits_{\theta}\left\{p(\boldsymbol{g};\theta)\right\} = (2\pi)^{-\frac{M}{2}}\det\left|\boldsymbol{R}_Y\right|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left[\boldsymbol{g}^t\boldsymbol{R}_Y^{-1}\boldsymbol{g}\right]\right] \\ \widehat{\boldsymbol{f}} & = & \arg\max\limits_{\boldsymbol{f}}\left\{p(\boldsymbol{f}|\boldsymbol{g};\widehat{\theta})\right\} \end{array}\right.$$

$$\boldsymbol{R}_Y = \theta\boldsymbol{H}\boldsymbol{P}_0\boldsymbol{H}^t + \boldsymbol{I}$$

$$\begin{array}{rcl} \widehat{\theta} & = & \arg\min\limits_{\theta}\left\{\ln\det\left|\theta\boldsymbol{H}\boldsymbol{P}_0\boldsymbol{H}^t + \boldsymbol{I}\right| - \boldsymbol{g}^t\left(\theta\boldsymbol{H}\boldsymbol{P}_0\boldsymbol{H}^t + \boldsymbol{I}\right)^{-1}\boldsymbol{g}\right\} \\ \widehat{\boldsymbol{f}} & = & \arg\max\limits_{\boldsymbol{f}}\left\{p(\boldsymbol{f}|\boldsymbol{g};\widehat{\theta}\right\} \end{array}$$

Contexte complètement bayésien :

$$(\widehat{\boldsymbol{f}},\widehat{\theta}) = \arg\max_{(\boldsymbol{f},\theta)}\left\{p(\boldsymbol{f},\theta|\boldsymbol{g})\right\} = \arg\max_{(\boldsymbol{f},\theta)}\left\{p(\boldsymbol{g}|\boldsymbol{f})p(x|\theta)p(\theta)\right\}$$

$$p(\theta) \propto \frac{1}{\theta}$$

$$L(x,\theta) = \ln p(\boldsymbol{f},\boldsymbol{g};\theta) \propto \left[\ln\theta + [\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}]^t[\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}] + \frac{1}{\theta}\boldsymbol{f}^t\boldsymbol{P}_0^{-1}\boldsymbol{f}\right] - \ln\theta$$

$$\left\{\begin{array}{rcll} \dfrac{\mathrm{d}L(\boldsymbol{f},\theta)}{\mathrm{d}\boldsymbol{f}} & = & 0 & \longrightarrow \boldsymbol{f} = \left(\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{P}_0\right)^{-1}\boldsymbol{H}^t\boldsymbol{g} \\ \dfrac{\mathrm{d}L(\boldsymbol{f},\theta)}{\mathrm{d}\theta} & = & 0 & \longrightarrow \theta = \boldsymbol{f}^t\boldsymbol{P}_0^{-1}\boldsymbol{f} \end{array}\right.$$

$$\left\{\begin{array}{rcl} \widehat{\boldsymbol{f}} & = & \left(\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{P}_0\right)^{-1}\boldsymbol{H}^t\boldsymbol{g} \\ \widehat{\theta} & = & \widehat{\boldsymbol{f}}^t\boldsymbol{P}_0^{-1}\widehat{\boldsymbol{f}} \end{array}\right.$$

Algorithme EM

$$\boldsymbol{z} = \begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{g} \end{pmatrix} \quad \longrightarrow p(\boldsymbol{z};\theta) = p(\boldsymbol{f},\boldsymbol{g};\theta) = \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Sigma}\right)$$

$$\ln p(\boldsymbol{z};\theta) \propto \ln\det|\boldsymbol{\Sigma}| - [\boldsymbol{f},\boldsymbol{g}]\boldsymbol{\Sigma}^{-1}\begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{g} \end{pmatrix}$$

$$p(\boldsymbol{z}|\boldsymbol{g};\widehat{\theta}) = p(\boldsymbol{f},\boldsymbol{g}|\boldsymbol{g};\widehat{\theta}) = p(\boldsymbol{f}|\boldsymbol{g};\widehat{\theta})$$

$$\mathrm{E}\left[\ln p(\boldsymbol{z};\theta)|\boldsymbol{g};\widehat{\theta}\right] = \ln\det|\boldsymbol{\Sigma}| - \mathrm{E}\left[[\boldsymbol{f},\boldsymbol{g}]\boldsymbol{\Sigma}^{-1}\begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{g} \end{pmatrix}\right]$$

$$Q(\theta,\widehat{\theta}) = \ln\theta - \frac{\theta+1}{\theta}\mathrm{E}\left[\boldsymbol{f}^2|\boldsymbol{g};\widehat{\theta}\right] + 2\boldsymbol{g}\mathrm{E}\left[\boldsymbol{f}|\boldsymbol{g};\widehat{\theta}\right] + \mathrm{cte}$$

$$\frac{\mathrm{d}Q(\theta,\widehat{\theta}^{(k)})}{\mathrm{d}\theta} = 0 \quad \longrightarrow \widehat{\theta}^{(k+1)} = \mathrm{E}\left[\boldsymbol{f}^2|\boldsymbol{g};\widehat{\theta}^{(k)}\right] = \frac{\theta^{(k)}}{\theta^{(k)}+1} + \left(\frac{\theta^{(k)}}{\theta^{(k)}+1}\boldsymbol{g}\right)^2$$

Chapitre 12

Étude de cas et exemples d'applications

Dans ce chapitre nous étudierons un certain nombre de problèmes classiques comme la déconvolution des signaux, la restauration d'image, la reconstruction d'images en tomographie X, la reconstruction d'images en tomographie à ondes diffractées, la reconstruction d'images par courants de FOUCAULT en contrôle non destructif, etc.

12.1 Déconvolution des signaux

12.1.1 Un problème d'instrumentation

Voici un exemple de problème où les mesures effectuées par l'instrument ne sont pas satisfaisantes et il est impossible de changer l'instrument. On voudrait cependant corriger les défaults de l'instrument par une méthode numérique de déconvolution.

Dans cet exemple nous allons étudier successivement les problèmes suivants :

- l'estimation de la réponse impulsionnelle d'un système de mesure à partir de sa réponse indicielle;
- l'estimation de la réponse impulsionnelle d'un système de mesure à partir de sa sorite à une entrée quelconque mais connue;
- l'estimation de l'entrée lorsque la réponse impulsionnelle est connue (déconvolution simple);
- l'estimation de l'entrée lorsque la réponse impulsionnelle n'est pas connue (déconvolution aveugle);

Nous examinerons :

- Méthodes naïves
- Filtre de Wienner
- Régularisation quadratique pour l'inversion
- Régularisation quadratique pour l'estimation de la réponse impulsionnelle
- Une méthode de déconvolution aveugle basée sur la régularisation quadratique

12.1.2 Méthodes naïves

Il s'agit d'examiner les résultats que l'on obtient lorsqu'on utilise une méthode naïve du type filtrage inverse pour inversion :

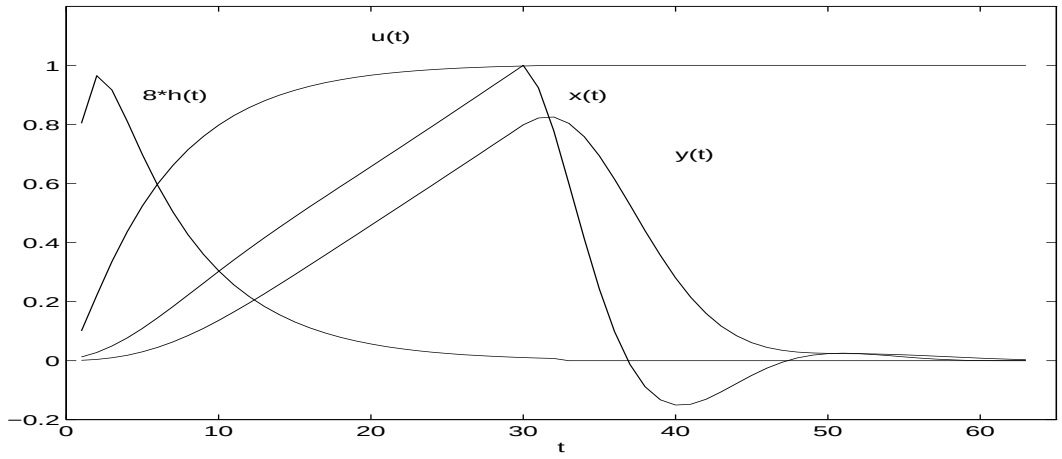
$$y(t) = h(t) * x(t) + b(t) \longrightarrow \text{TF} \longrightarrow Y(\omega) = H(\omega)\, X(\omega) + B(\omega) \longrightarrow$$

$$\widehat{X}(\omega) = \frac{1}{H(\omega)}\, Y(\omega) \longrightarrow \text{TF inverse} \longrightarrow \widehat{x}(t)$$

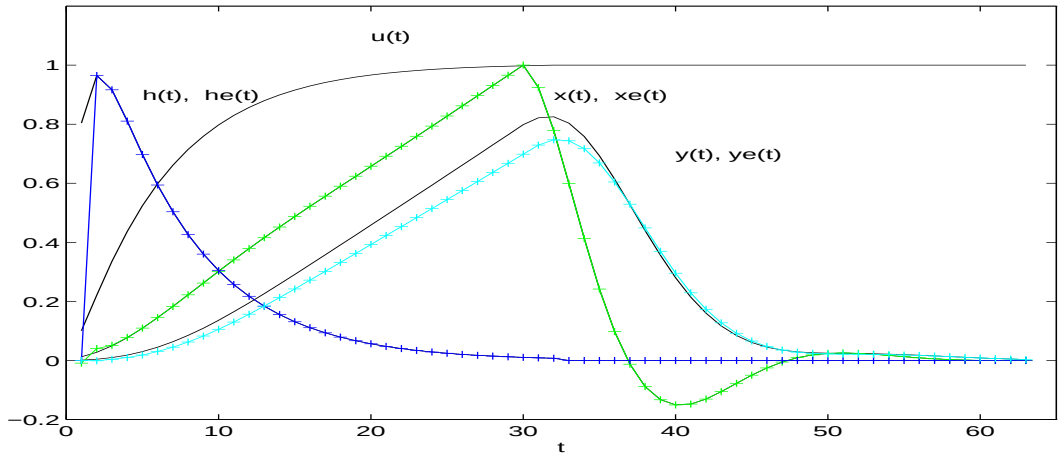
et du type dérivation de la réponse indicielle pour le calcul de la réponse impulsionnelle :

$$u(t) = \int_0^t h(t)\, \mathrm{d}t \longrightarrow h(t) = \frac{\mathrm{d}u(t)}{\mathrm{d}t}$$

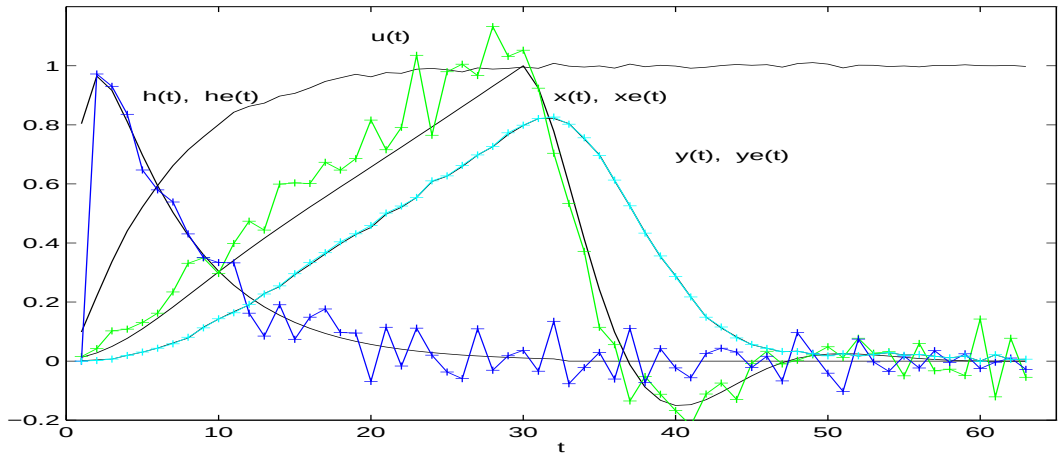
Effectivement, comme le montre les figures qui suivent, ces méthodes peuvent marcher lorsqu'il n'y a pas du bruit. Mais, en présence du bruit, même très très faible, ces méthodes échouent très facilement.



Signaux originaux: $u(t)$, $h(t)$, $x(t)$ et $y(t)$

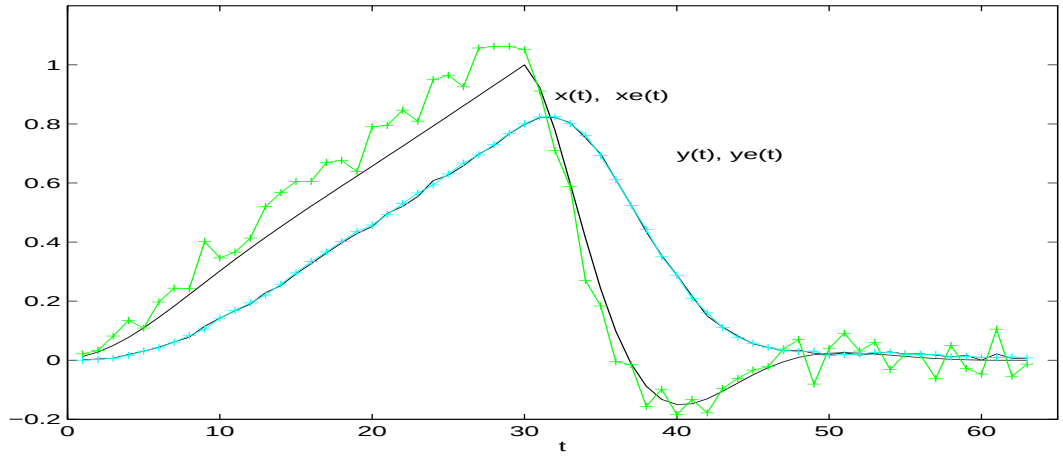


Pas de bruit : tout va bien.

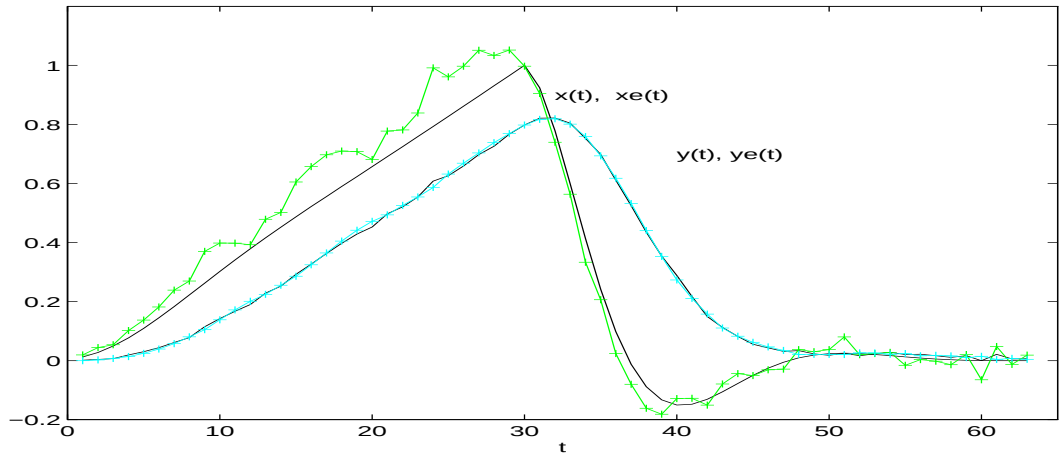


Avec bruit : rien ne va plus

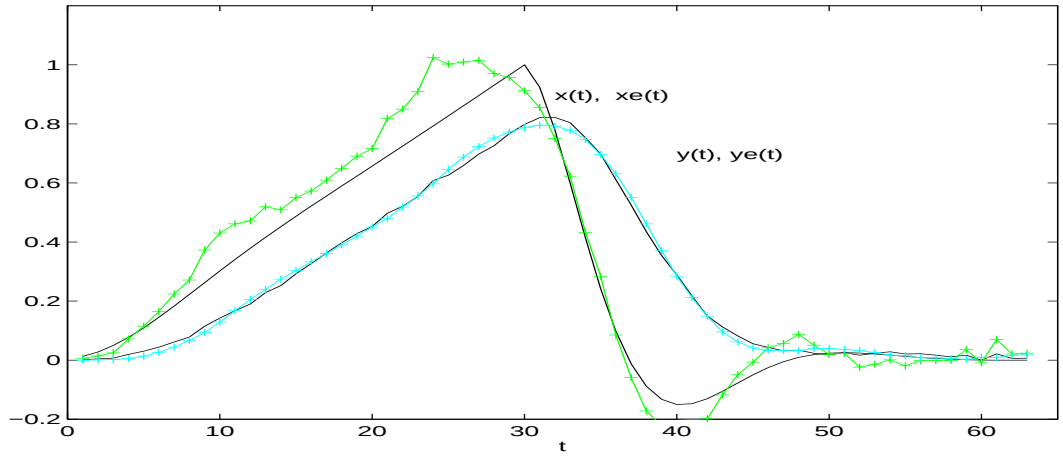
FIG. 12.1 – **Méthodes naïves :**
On calcule $h(t)$ en dérivant $u(t)$ et $x(t)$ par filtrage inverse. Lorsqu'il n'y a pas de bruit, tous va bien, mais lorqu'on rajouté un bruit sur les données les résultats deviennent très vite inexploitables.



$\epsilon = .1$



$\epsilon = .2$



$\epsilon = .3$

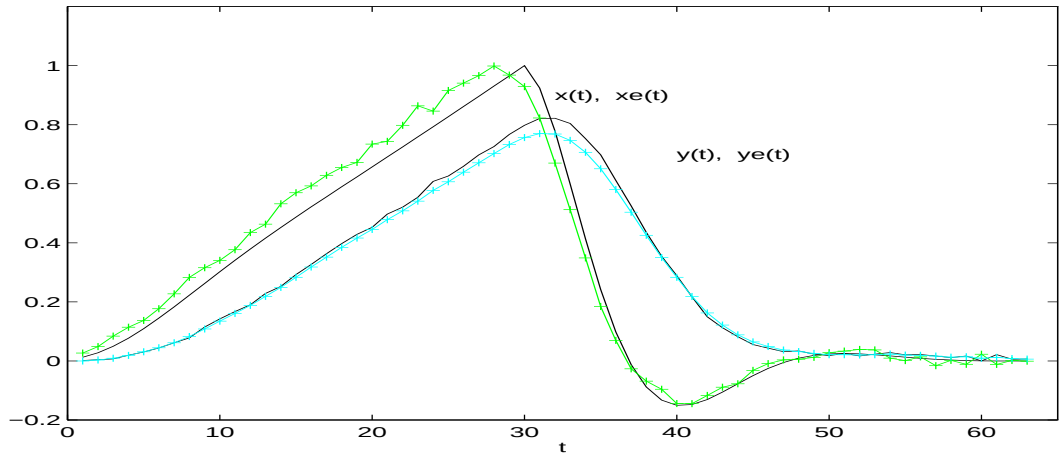
FIG. 12.2 – **Méthodes naïves :** *On bricole en remplaçant tous les valeurs de* $|H(\omega)| < \epsilon$ *par* $|H(\omega)| = \epsilon$ *pour tenter d'éliminer la source d'ennui, mais sans succes. Trois valeurs de* ϵ *sont essayées.*

12.1.3 Filtre de Wiener pour l'inversion

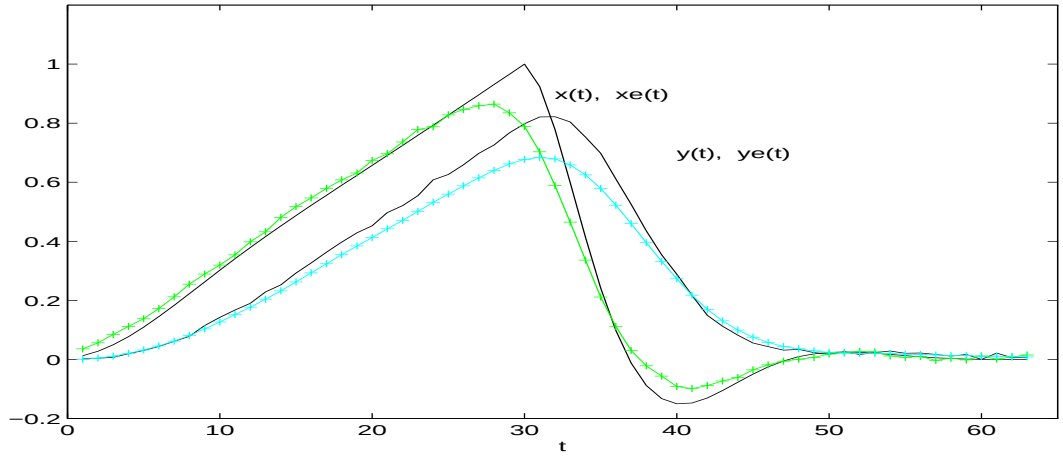
Dans cette étude, nous considérons seulement le problème de la déconvolution simple, en faisant l'hypothèse que la réponse impulsionnelle est connue parfaitement. La solution est calculée par

$$\widehat{X}(\omega) = \frac{H^*(\omega)}{|H(\omega)|^2 + \frac{S_{bb}(\omega)}{S_{xx}(\omega)}} Y(\omega)$$

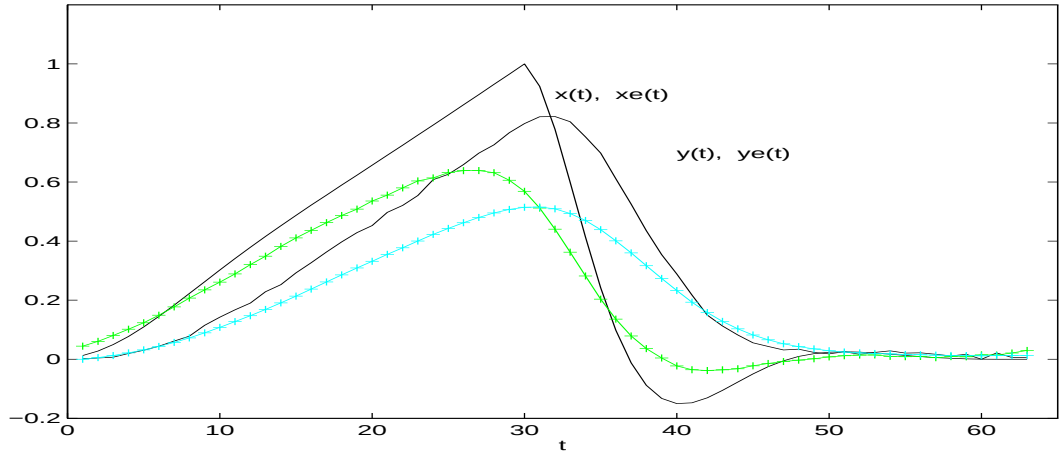
avec l'hypothèse $\frac{S_{bb}(\omega)}{S_{xx}(\omega)} = r$ et différentes valeurs pour r.



$r = 30dB.$



$r = 20dB.$



$r = 10dB.$

FIG. 12.3 – **Filtre de Wiener en déconvolution :**
On essai le filtre de Wiener avec l'hypothèse $\frac{S_{xx}(\omega)}{S_{bb}(\omega)} = r$ *pour différentes valeurs de* r.

12.1.4 Régularisation quadratique pour l'inversion

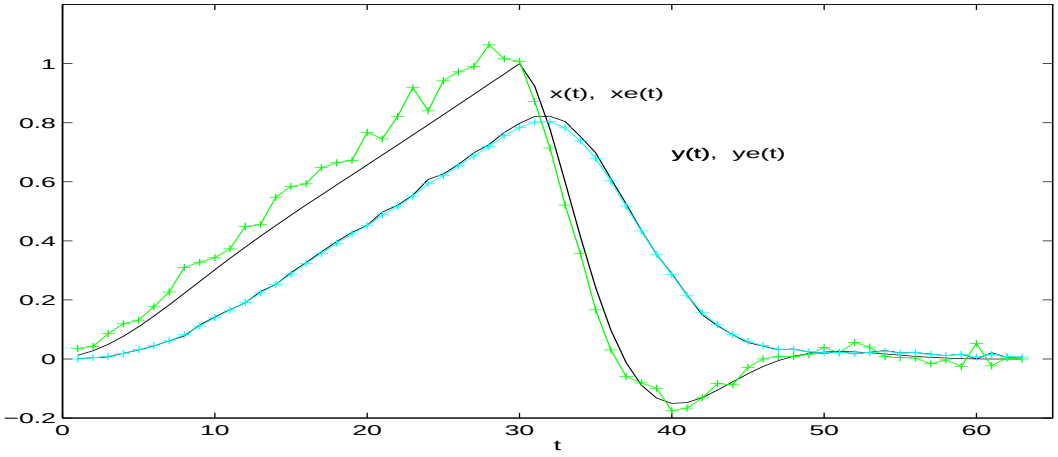
Dans cette étude, nous considérons seulement le problème de la déconvolution simple, en faisant l'hypothèse que la réponse impulsionnelle est connue parfaitement. La solution est calculée par

$$\widehat{\boldsymbol{x}} = \left(\boldsymbol{H}^t\boldsymbol{H} + \lambda \boldsymbol{D}^t\boldsymbol{D}^t\right)^{-1} \boldsymbol{H}^t\boldsymbol{y}$$

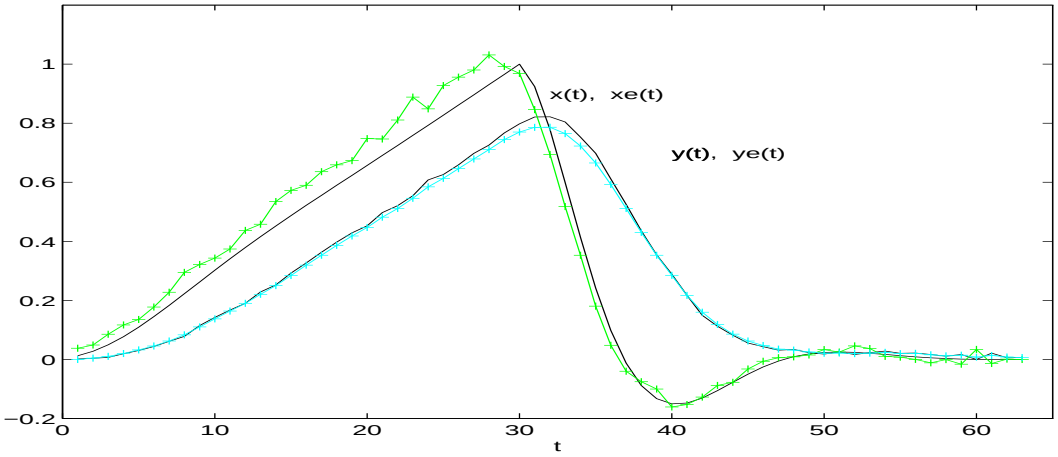
avec différentes valeurs pour le paramètre de régularisation λ.

Deux cas sont étudiés correspondant à deux choix pour la matrice $\boldsymbol{D}$:

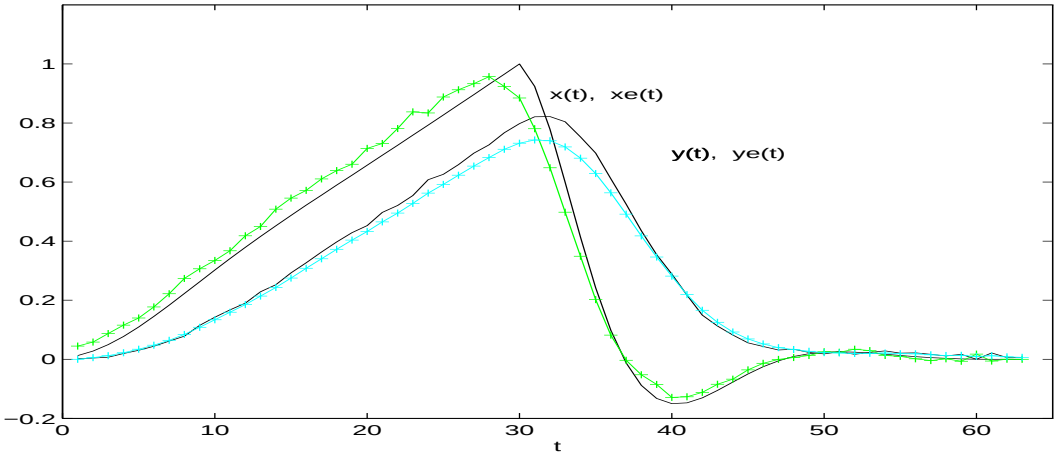
- régularisation d'ordre zéro : $\boldsymbol{D} = \boldsymbol{I}$
- régularisation d'ordre un : $\boldsymbol{D} = \textsc{Toeplitz}[-1,1]$ est une matrice des différences finies d'ordre un.



Régularisation d'ordre zéro. $\lambda = .01$

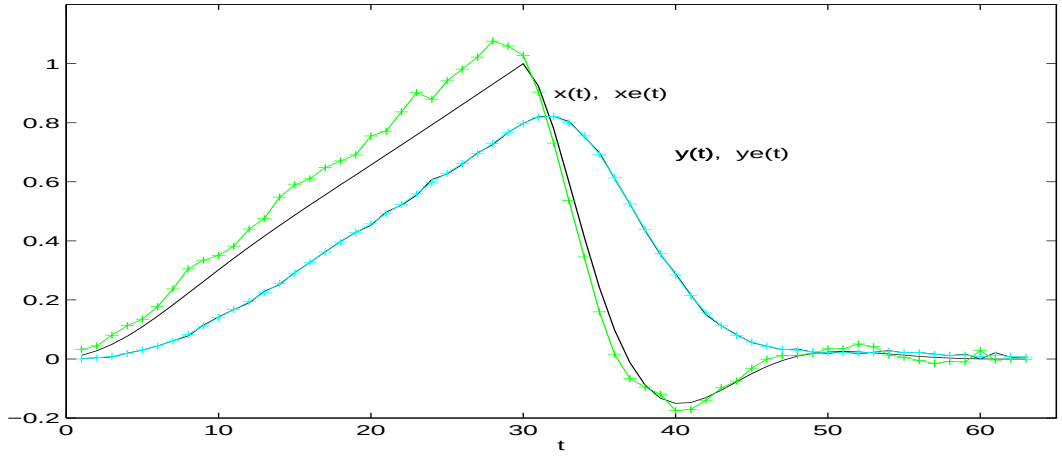


Régularisation d'ordre zéro. $\lambda = .02$

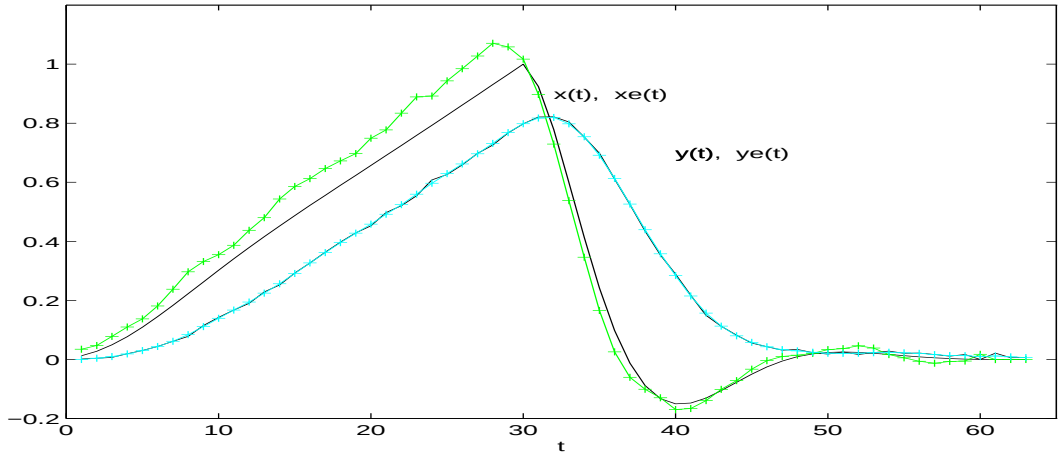


Régularisation d'ordre zéro. $\lambda = .05$

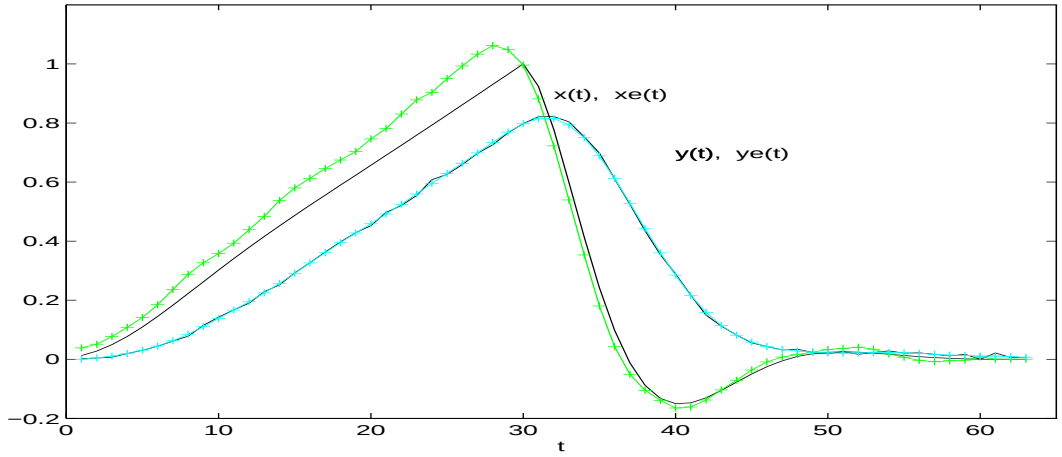
FIG. 12.4 – *Déconvolution par régularisation d'ordre zéro.*



Régularisation d'ordre un. $\lambda = .01$



Régularisation d'ordre un. $\lambda = .02$



Régularisation d'ordre un. $\lambda = .05$

FIG. 12.5 – *Déconvolution par régularisation d'ordre un.*

12.1.5 Régularisation quadratique pour l'estimation de la réponse impulsionnelle

Dans cette étude, nous considérons seulement le problème de l'estimation de la réponse impulsionnelle $h(t)$ à partir de la réponse indicielle $u(t)$. Notons que la relation entre ces deux grandeur s'écrit :

$$u(t) = \int_0^t h(t)\,\mathrm{d}t \longrightarrow h(t) = \frac{\mathrm{d}u(t)}{\mathrm{d}t}$$

Discrétisation :

$$u(m) = \sum_{k=0}^{m} h(k) \longrightarrow \begin{pmatrix} u(0) \\ \vdots \\ u(M) \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & & \vdots \\ \vdots & & \ddots & \\ 1 & \cdots & & 1 \end{pmatrix} \begin{pmatrix} h(0) \\ \vdots \\ h(M) \end{pmatrix}$$

$$\boldsymbol{u} = \boldsymbol{A}\boldsymbol{h}$$

Solution régularisée :

$$\widehat{\boldsymbol{h}} = \left(\boldsymbol{A}^t\boldsymbol{A} + \lambda \boldsymbol{D}^t\boldsymbol{D}\right)^{-1} \boldsymbol{A}^t\boldsymbol{u}$$

Dans le cas plus général où l'entrée est un signal quelconque, on a :

$$y(t) = [h * f](t) + b(t)$$

Discrétisation :

$$y(m) = \sum_{k=0}^{M} x(k)h(m-k) + b(m)$$

$$\begin{pmatrix} y(0) \\ y(1) \\ \vdots \\ y(p) \\ \vdots \\ y(M) \end{pmatrix} = \begin{pmatrix} x(0) & 0 & \cdots & \cdots & \cdots & 0 \\ x(1) & x(0) & & & & \vdots \\ \vdots & & & & & \\ x(p) & & & & & x(0) \\ x(p+1) & x(p) & & & & x(1) \\ \vdots & & & & & \vdots \\ x(M) & \cdots & & & \cdots & x(M-p) \end{pmatrix} \begin{pmatrix} h(0) \\ \vdots \\ h(p) \end{pmatrix}$$

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{h} + \boldsymbol{b} \quad \text{avec} \quad \boldsymbol{X} = \textsc{Toeplitz}(\boldsymbol{x}),$$

Solution régularisée :

$$\widehat{\boldsymbol{h}} = \left(\boldsymbol{X}^t\boldsymbol{X} + \lambda \boldsymbol{D}^t\boldsymbol{D}\right)^{-1} \boldsymbol{X}^t\boldsymbol{y}$$

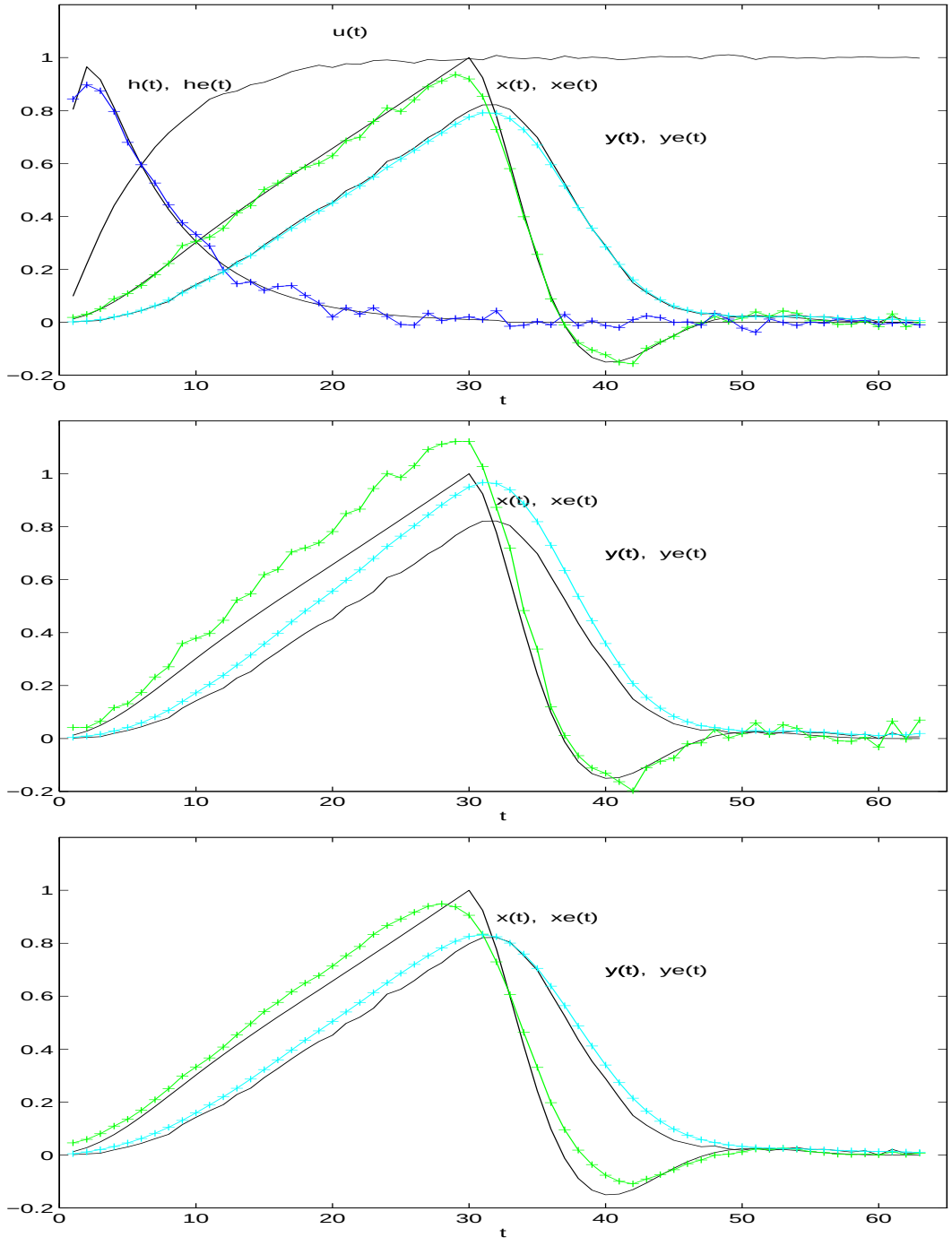


FIG. 12.6 – *Exemple de résultat. On estime d'abord $h(t)$ à partir de $u(t)$ et on l'utilise pour l'estimation de $x(t)$ à partir de $y(t)$*

12.1.6 Déconvolution aveugle

Estimation simultanée de la réponse impulsionnelle $h(t)$ et l'entrée $x(t)$ à partir de la sortie $y(t)$.

$$y(t) = [h * f](t) + b(t)$$

Discrétisation :

$$y(m) = \sum_{k=0}^{p} h(k)x(m-k) + b(m) \quad \text{ou} \quad y(m) = \sum_{k=0}^{M} x(k)h(m-k) + b(m)$$

$$\begin{pmatrix} y(0) \\ y(1) \\ \vdots \\ y(p) \\ \vdots \\ y(M) \end{pmatrix} = \begin{pmatrix} h(0) & 0 & \cdots & \cdots & \cdots & & 0 \\ h(1) & h(0) & & & & & \vdots \\ \vdots & h(1) & & & & & \\ h(p) & & & & & & \\ 0 & h(p) & & & & & \vdots \\ \vdots & & & & & h(0) & 0 \\ 0 & \cdots & 0 & h(p) & \cdots & h(1) & h(0) \end{pmatrix} \begin{pmatrix} x(0) \\ x(1) \\ \vdots \\ x(M-p) \\ \vdots \\ x(M) \end{pmatrix}$$

$$\begin{pmatrix} y(0) \\ y(1) \\ \vdots \\ y(p) \\ \vdots \\ y(M) \end{pmatrix} = \begin{pmatrix} x(0) & 0 & \cdots & \cdots & \cdots & 0 \\ x(1) & x(0) & & & & \vdots \\ \vdots & & & & & \\ x(p) & & & & & x(0) \\ x(p+1) & x(p) & & & & x(1) \\ \vdots & & & & & \vdots \\ x(M) & \cdots & & & \cdots & x(M-p) \end{pmatrix} \begin{pmatrix} h(0) \\ \vdots \\ h(p) \end{pmatrix}$$

$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{b}$ ou $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{h} + \boldsymbol{b}$ avec $\boldsymbol{H} = \textsc{Toeplitz}(\boldsymbol{h})$, $\boldsymbol{X} = \textsc{Toeplitz}(\boldsymbol{x})$,

$$\begin{aligned} p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{h}) &\propto \exp\left[-\frac{1}{2\sigma_b^2}\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2\right] \\ p(\boldsymbol{x}) &\propto \exp\left[-\frac{\alpha}{2}\|\boldsymbol{D}_x\boldsymbol{x}\|^2\right] \\ p(\boldsymbol{h}) &\propto \exp\left[-\frac{\beta}{2}\|\boldsymbol{D}_h\boldsymbol{x}\|^2\right] \\ p(\boldsymbol{x},\boldsymbol{h}|\boldsymbol{y}) &\propto \exp\left[-\frac{1}{2\sigma_b^2}J(\boldsymbol{x},\boldsymbol{h})\right] \\ & J(\boldsymbol{x},\boldsymbol{h}) = \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2 + \lambda_1\|\boldsymbol{D}_h\boldsymbol{h}\|^2 + \lambda_2\|\boldsymbol{D}_x\boldsymbol{x}\|^2 \end{aligned}$$

Estimation au sens du MAP :

$$(\widehat{\boldsymbol{x}},\widehat{\boldsymbol{h}}) = \underset{(\boldsymbol{x},\boldsymbol{h})}{\arg\min}\{J(\boldsymbol{x},\boldsymbol{h})\}$$

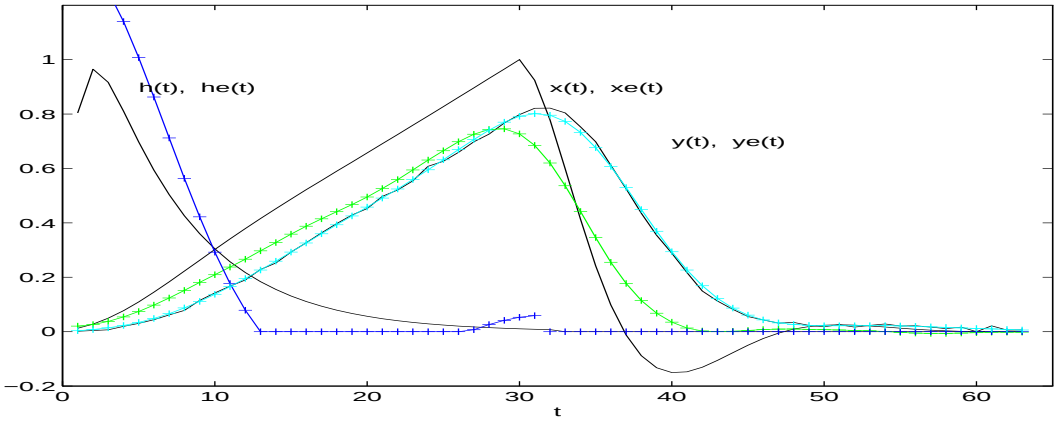
Mise en œuvre :
On constate que J est quadratique en $\boldsymbol{x}$ à $\boldsymbol{h}$ fixé et quadratique en $\boldsymbol{h}$ à $\boldsymbol{x}$ fixé. On peut utiliser cette constatation pour proposer un algorithme itérative :

$$\begin{aligned}\widehat{\boldsymbol{x}}^{(k)} &= \underset{(\boldsymbol{x})}{\arg\min}\left\{J(\boldsymbol{x},\boldsymbol{h}^{(k-1)})\right\}\\ \widehat{\boldsymbol{h}}^{(k)} &= \underset{(\boldsymbol{h})}{\arg\min}\left\{J(\boldsymbol{x}^{(k-1)},\boldsymbol{h})\right\}\end{aligned}$$

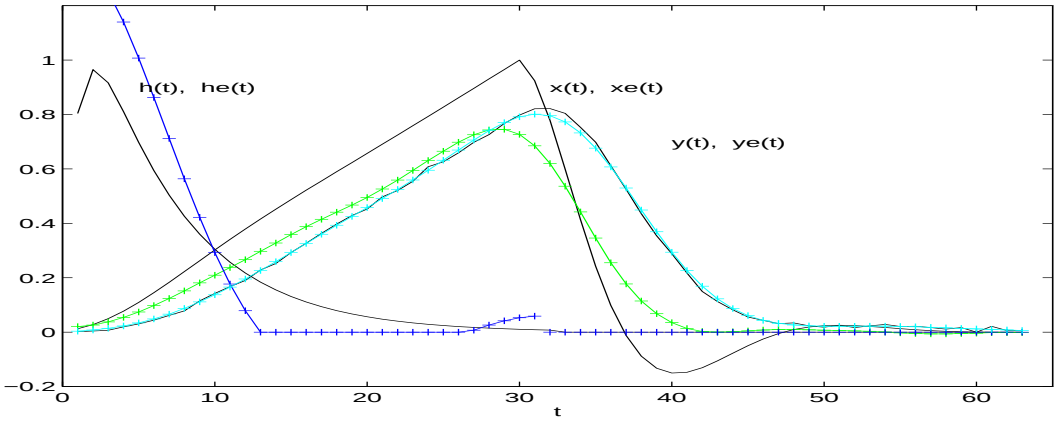
$$\begin{aligned}\widehat{H}(\omega) &= \frac{X^*(\omega)}{|X(\omega)|^2+\lambda_1 D_h(\omega)}Y(\omega)\\ \widehat{X}(\omega) &= \frac{H^*(\omega)}{|H(\omega)|^2+\lambda_2 D_x(\omega)}Y(\omega)\end{aligned}$$

Dans le cas de la restauration d'image :

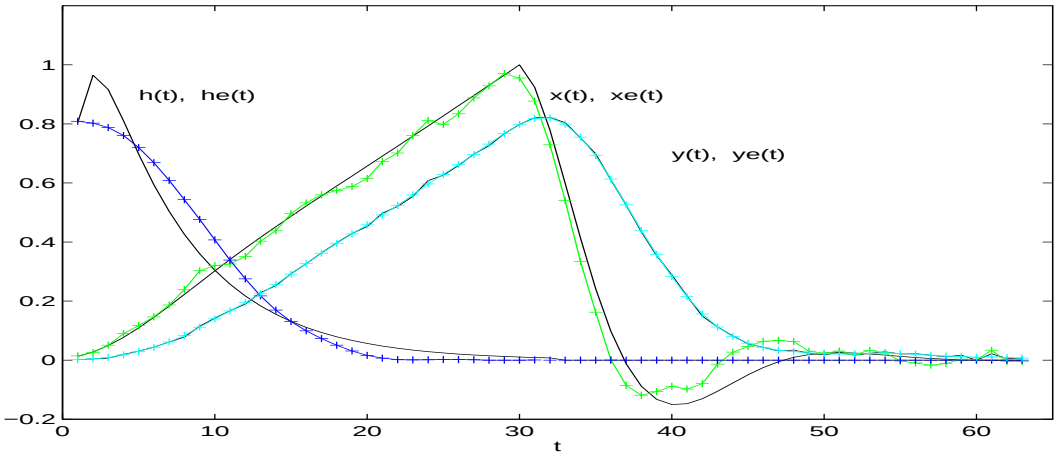
$$\begin{aligned}\widehat{H}(\omega_x,\omega_y) &= \frac{X^*(\omega_x,\omega_y)}{|X(\omega_x,\omega_y)|^2+\lambda_1 D_h(\omega_x,\omega_y)}Y(\omega_x,\omega_y)\\ \widehat{X}(\omega_x,\omega_y) &= \frac{H^*(\omega_x,\omega_y)}{|H(\omega_x,\omega_y)|^2+\lambda_2 D_x(\omega_x,\omega_y)}Y(\omega_x,\omega_y)\end{aligned}$$



Résultat à l'itération 1



Résultat à l'itération 2



Résultat à l'itération finale.

FIG. 12.7 – **Déconvolution aveugle :**
Exemples de résultats de l'estimation simultanée de $h(t)$ et de $x(t)$ à partir de $y(t)$ avec trois jeux de paramètres différents pour $(\lambda_1,\lambda_2) = \{(20,1),(10,1),(20,05)\}$.

Écriture des relations en continu:

Considérons le problème de la déconvolution aveugle des signaux :

$$g(t) = h(t) * f(t) + b(t)$$

et définissons le critère suivant :

$$J(f,h) = \|g - h * f\|^2 + \lambda_1 \|d_1 * f\|^2 + \lambda_2 \|d_2 * h\|^2$$

où $d_1(t)$ et $d_2(t)$ sont deux fonctions connues.

$$\begin{aligned} \frac{\partial J}{\partial f} &= -2h(-t) * [g(t) - h(t) * f(t)] + 2\lambda_1 d_1(-t) * [d_1(t) * f(t)] \\ \frac{\partial J}{\partial h} &= -2f(-t) * [g(t) - h(t) * f(t)] + 2\lambda_2 d_2(-t) * [d_2(t) * h(t)] \end{aligned}$$

$$\begin{aligned} [h(-t) * h(t) + \lambda_1 d_1(-t) * d_1(t)] * f(t) &= h(-t) * g(t) \\ [f(-t) * f(t) + \lambda_2 d_2(-t) * d_2(t)] * h(t) &= f(-t) * g(t) \end{aligned}$$

Passant dans le domaine de FOURIER on obtient :

$$\begin{aligned} \left[|H(\omega)|^2 + \lambda_1 |D_1(\omega)|^2\right] F(\omega) &= H^*(\omega) G(\omega) \\ \left[|F(\omega)|^2 + \lambda_2 |D_2(\omega)|^2\right] H(\omega) &= F^*(\omega) G(\omega) \end{aligned}$$

$$\begin{aligned} F(\omega) &= \frac{H^*(\omega) G(\omega)}{|H(\omega)|^2 + \lambda_1 |D_1(\omega)|^2} \\ H(\omega) &= \frac{F^*(\omega) G(\omega)}{|F(\omega)|^2 + \lambda_2 |D_2(\omega)|^2} \end{aligned}$$

Déconvolution avec une étape de calibration : :

Considérons maintenant le cas où, en plus des données, nous avons eu la possibilité d'effectuer une calibration :

$$\begin{aligned} g_c(t) &= h(t) * f_c(t) + b_1(t) \\ g(t) &= h(t) * f(t) + b_2(t) \end{aligned}$$

et définissons le critère suivant :

$$J(f,h) = \|g_c - h * f_c\|^2 + \|g - h * f\|^2 + \lambda_1 \|d_1 * f\|^2 + \lambda_2 \|d_2 * h\|^2$$

où $d_1(t)$ et $d_2(t)$ sont deux fonctions connues.

$$\begin{aligned} \frac{\partial J}{\partial f} &= -2h(-t) * [g(t) - h(t) * f(t)] + 2\lambda_1 d_1(-t) * [d_1(t) * f(t)] \\ \frac{\partial J}{\partial h} &= -2f_c(-t) * [g_c(t) - h(t) * f_c(t)] - 2f(-t) * [g(t) - h(t) * f(t)] + 2\lambda_2 d_2(-t) * [d_2(t) * h(t)] \end{aligned}$$

$$\begin{aligned} [h(-t) * h(t) + \lambda_1 d_1(-t) * d_1(t)] * f(t) &= h(-t) * g(t) \\ [f(-t) * f(t) + f_c(-t) * f_c(t) + \lambda_2 d_2(-t) * d_2(t)] * h(t) &= f_c(-t) * g_c(t) + f(-t) * g(t) \end{aligned}$$

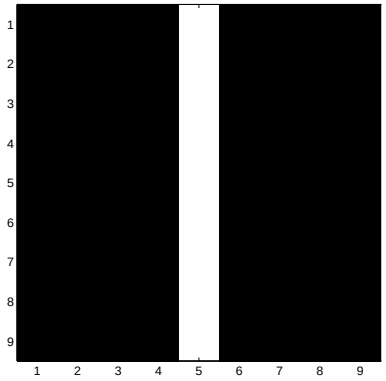
Passant dans le domaine de FOURIER on obtient :

$$\begin{aligned} \left[|H(\omega)|^2 + \lambda_1 |D_1(\omega)|^2\right] F(\omega) &= H^*(\omega) G(\omega) \\ \left[|F_c(\omega)|^2 + |F(\omega)|^2 + \lambda_2 |D_2(\omega)|^2\right] H(\omega) &= F_c^*(\omega) G_c(\omega) + F^*(\omega) G(\omega) \end{aligned}$$

$$\begin{aligned} F(\omega) &= \frac{H^*(\omega) G(\omega)}{[|H(\omega)|^2 + \lambda_1 |D_1(\omega)|^2]} \\ H(\omega) &= \frac{F_c^*(\omega) G_c(\omega) + F^*(\omega) G(\omega)}{|F_c(\omega)|^2 + |F(\omega)|^2 + \lambda_2 |D_2(\omega)|^2} \end{aligned}$$

12.2 Restauration d'image

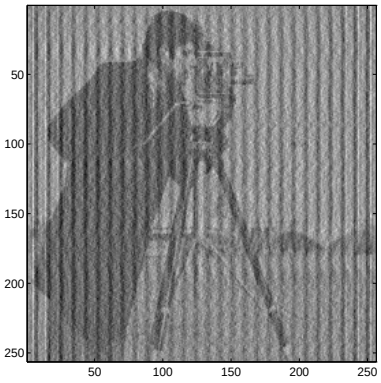
Image d'entrée $f(x,y)$



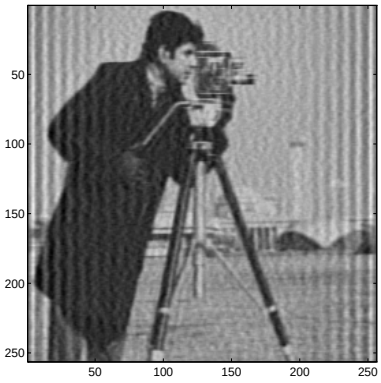
La réponse impulsionnelle $h(x,y)$

Image de sortie sans bruit

Image de sortie $g(x,y)$ avec bruit RSB=20 dB.



Restauration sans régularisation



Restauration avec régularisation

FIG. 12.8 – *Restauration d'image sans et avec régularisation*

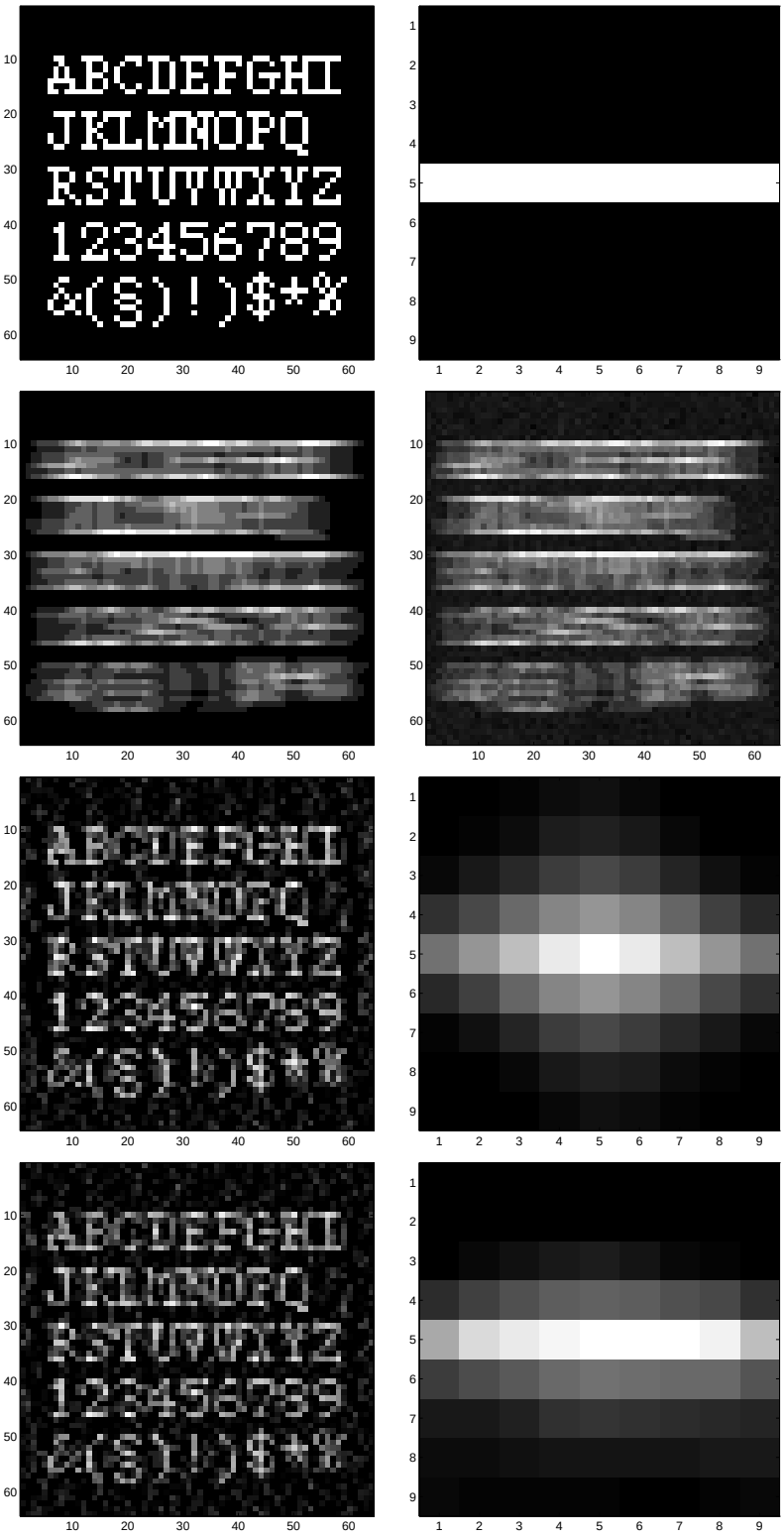


FIG. 12.9 – *Restauration aveugle d'image :*
a) Image originale, b) réponse impulsionnelle
c) Image dégradée sans bruit, d) Image dégradée avec bruit (données)
e) réponse impulsionnelle à initialisation, f) Image estimée à initialisation,
g) réponse impulsionnelle estimée, h) Image estimée

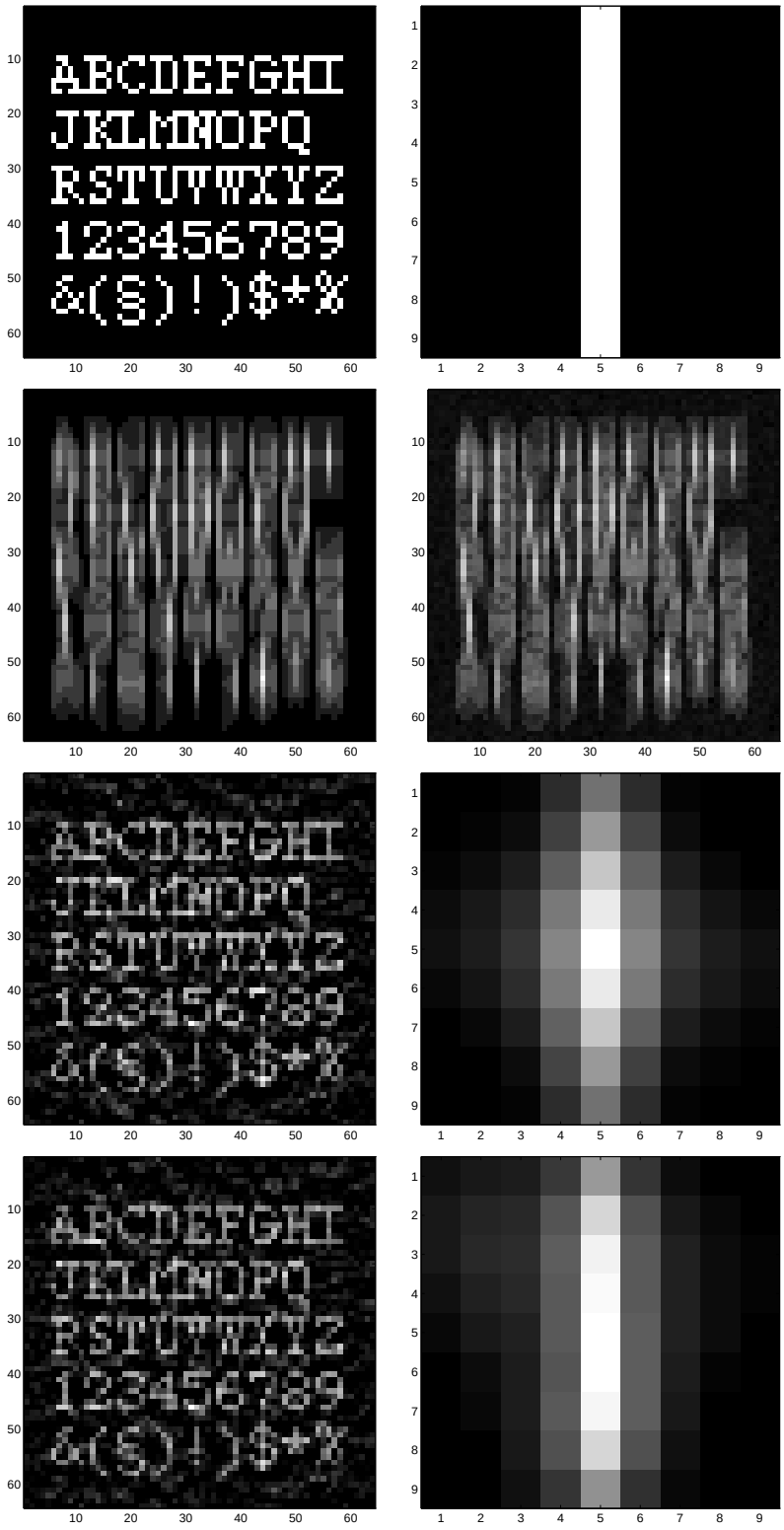


FIG. 12.10 – *Restauration aveugle d'image :*
a) Image originale, b) réponse impulsionnelle
c) Image dégradée sans bruit, d) Image dégradée avec bruit (données)
e) réponse impulsionnelle à initialisation, f) Image estimée à initialisation,
g) réponse impulsionnelle estimée, h) Image estimée

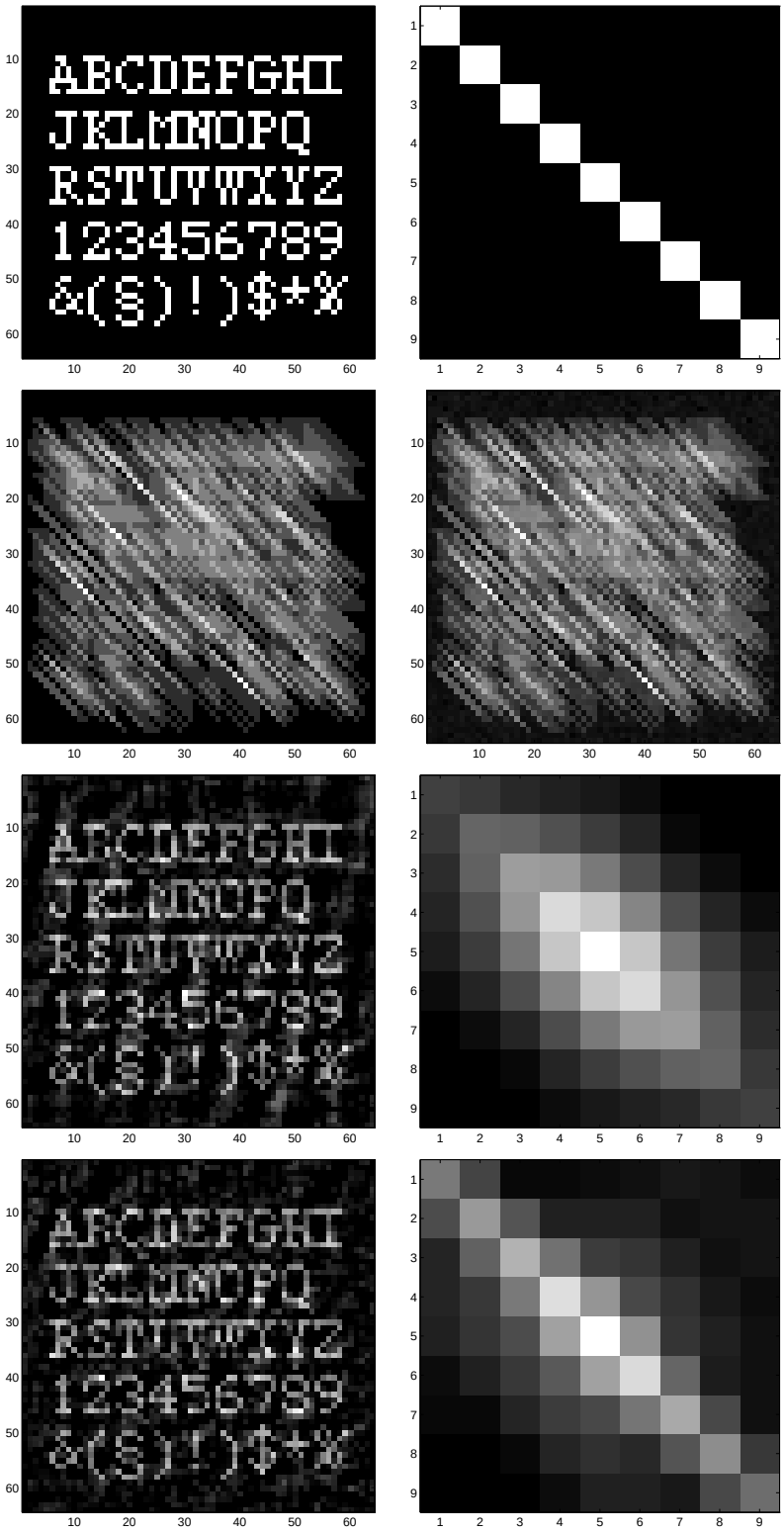


FIG. 12.11 – *Restauration aveugle d'image :*
a) Image originale, b) réponse impulsionnelle
c) Image dégradée sans bruit, d) Image dégradée avec bruit (données)
e) réponse impulsionnelle à initialisation, f) Image estimée à initialisation,
g) réponse impulsionnelle estimée, h) Image estimée

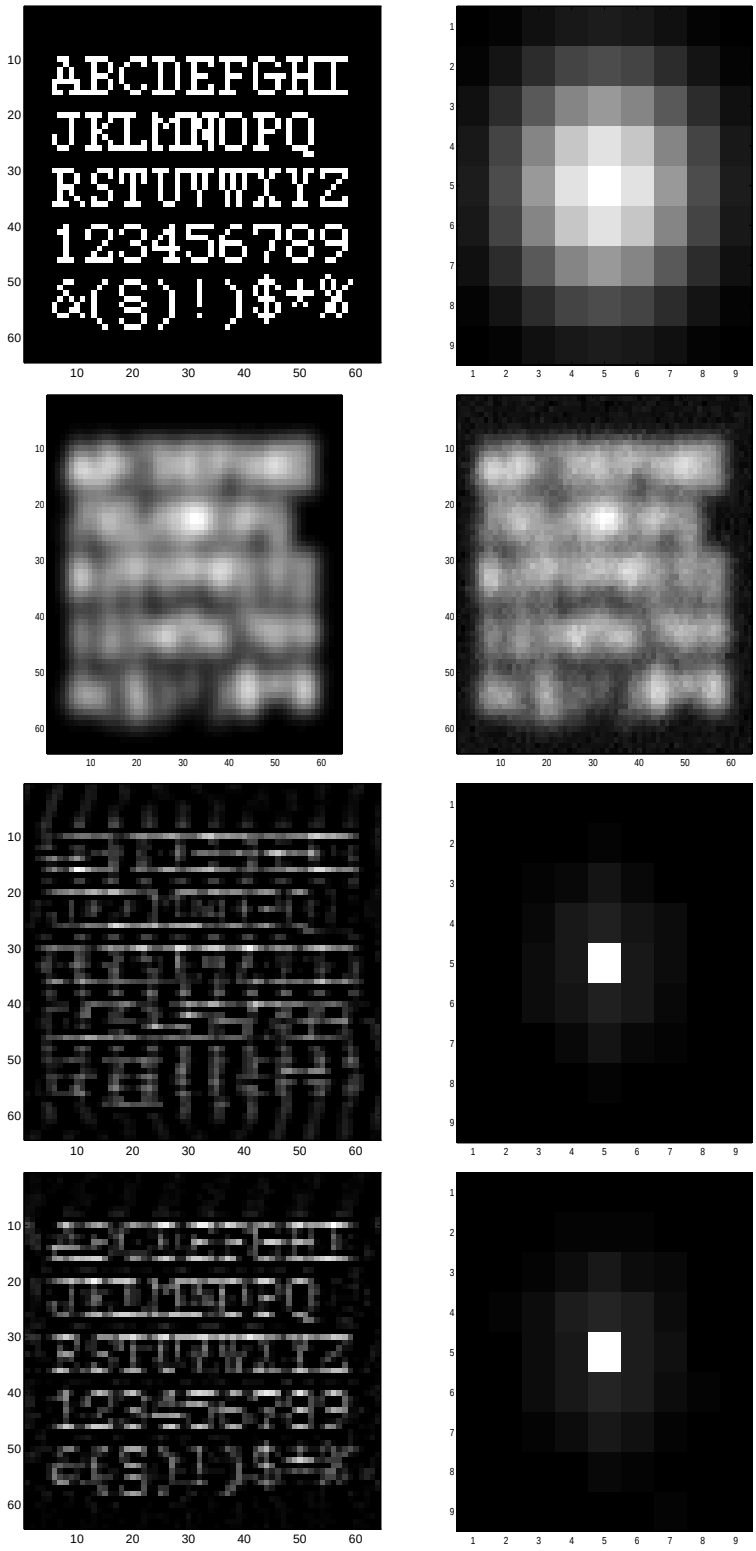


FIG. 12.12 – *Restauration aveugle d'image :*
a) Image originale, b) réponse impulsionnelle
c) Image dégradée sans bruit, d) Image dégradée avec bruit (données)
e) réponse impulsionnelle à initialisation, f) Image estimée à initialisation,
g) réponse impulsionnelle estimée, h) Image estimée

12.3 Reconstruction d'image en tomographie X

12.4 Reconstruction d'image en tomographie microondes

12.5 Reconstruction d'image en CND par courant de Foucault

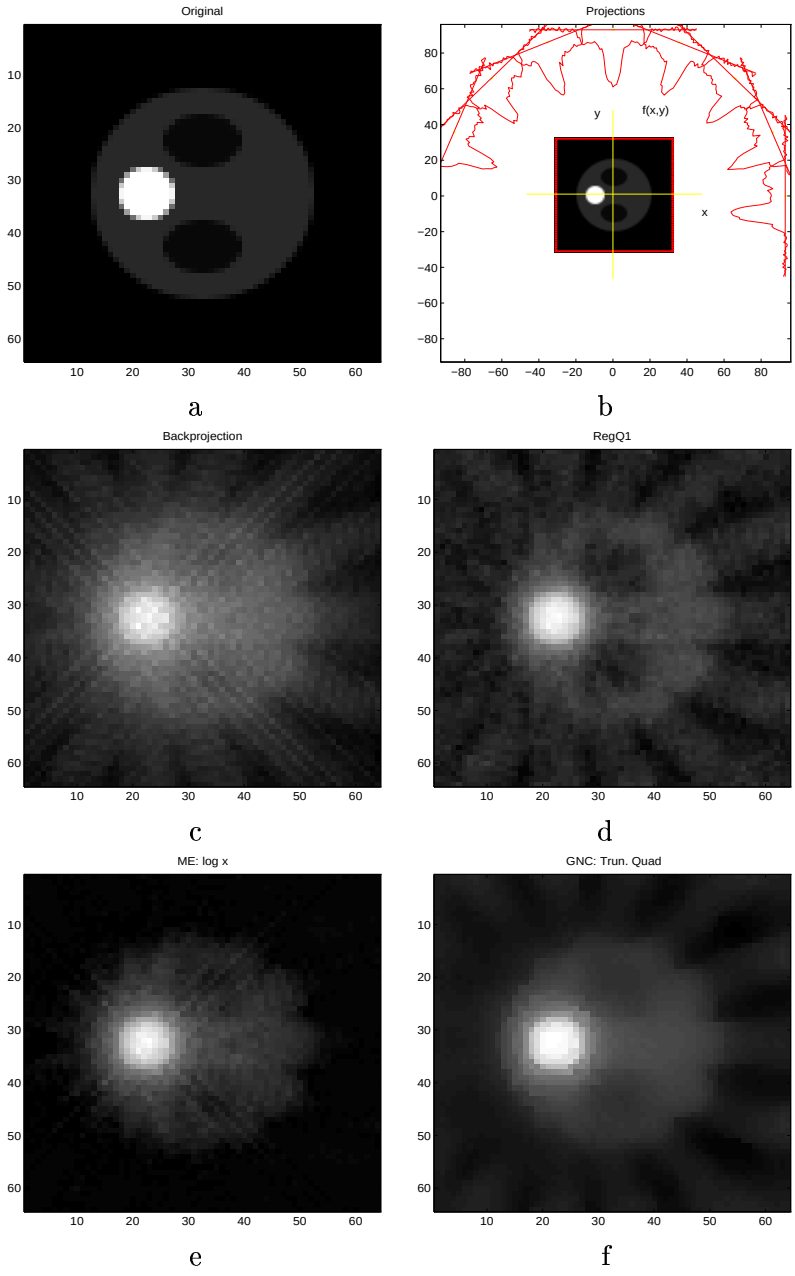


FIG. 12.13 – *Reconstruction d'image en tomographie X :*
a) objet originale, b) projections (données),
c) Reconstruction par rétroprojection, d) Reconstruction par régularisation quadratique,
d) Reconstruction par ME,
e) Reconstruction par modèle markovien et GNC,

Annexe A

Algorithmes d'optimisation pour un critère de moindre carrés

Optimisation d'un critère et en particulier un critère de moindre carrés (MC) se trouve au cœur d'un grand nombre de problèmes. L'objet de cette annexe est de fournir une présentation des différents algorithmes d'optimisation que l'on peut utiliser pour cette tâche, surtout dans les cas d'un système non linéaire. Le cas linéaire s'en déduit, bien entendu, comme un cas particulier.

A.1 Introduction

Considérons tout d'abord le cas général de la minimisation d'un critère $J(\boldsymbol{x})$, c'est à dire le calcul de

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \{J(\boldsymbol{x})\}$$

Lorsque le critère est unimodal, on peut alors utiliser les algorithmes d'optimisation classique du type gradient pour calculer la solution. Notant par :

$$\boldsymbol{g} = \nabla J(\boldsymbol{x}) = \left[\frac{\partial J}{\partial x_1}, \cdots, \frac{\partial J}{\partial x_n}\right]^t \quad \text{le gradient du } J, \quad \boldsymbol{g}^k = \nabla J\left(\boldsymbol{x}^k\right)$$

et par

$$\boldsymbol{H} = \nabla^2 J(\boldsymbol{x}) = \left\{\frac{\partial^2 J}{\partial x_i \partial x_j}\right\} \quad \text{la matrice Hessienne du } J, \quad \boldsymbol{H}^k = \nabla^2 J\left(\boldsymbol{x}^k\right)$$

sa matrice Hessienne, la majorité des méthodes d'optimisation locale calcule la solution d'une manière itérative suivante :

$$\begin{aligned} \boldsymbol{x}^{(k+1)} &= \boldsymbol{x}^{(k)} - \alpha^{(k)} \boldsymbol{D}^{(k)} \nabla J\left(\boldsymbol{x}^k\right) \\ &= \boldsymbol{x}^{(k)} - \alpha^{(k)} \boldsymbol{D}^{(k)} \boldsymbol{g}^k \\ &= \boldsymbol{x}^{(k)} + \alpha^{(k)} \boldsymbol{d}^{(k)} \end{aligned}$$

Suivant les différents choix pour $\alpha^{(k)}$ et $\boldsymbol{D}^{(k)}$ on peut distinguer les algorithmes suivants :

1. **Méthodes de premier ordre ou de gradient**
   - gradient à pas fixe : $\boldsymbol{D}^{(k)} = \boldsymbol{I}$, $\alpha^{(k)} = \alpha$ et $\boldsymbol{d}^{(k)} = -\boldsymbol{g}^{(k)}$
   - gradient à pas variable : $\boldsymbol{D}^{(k)} = \boldsymbol{I}$ et

$$\alpha^{(k)} = -\frac{\boldsymbol{g}^{(k)^t} \boldsymbol{g}^{(k)}}{\boldsymbol{g}^{(k)^t} \boldsymbol{H}^k \boldsymbol{g}^{(k)}} = \frac{\|\boldsymbol{g}^k\|^2}{\|\boldsymbol{g}^k\|_H^2}$$

   - gradient conjugué :

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha^{(k)} \boldsymbol{d}^{(k)} \quad \alpha^{(k)} = -\frac{\boldsymbol{d}^{(k)^t} \boldsymbol{g}^{(k)}}{\boldsymbol{d}^{(k)^t} \boldsymbol{H}^k \boldsymbol{d}^{(k)}}$$

$$\boldsymbol{d}^{(k+1)} = \boldsymbol{d}^{(k)} + \beta^{(k)} \boldsymbol{g}^{(k)} \quad \beta^{(k)} = -\frac{\boldsymbol{g}^{(k)^t} \boldsymbol{g}^{(k)}}{\boldsymbol{g}^{(k-1)^t} \boldsymbol{g}^{(k-1)}}$$

2. **Méthodes de second ordre**
   - Méthode de Newton : $\boldsymbol{D}^{(k)} = \left[\boldsymbol{H}^{(k)}\right]^{-1}$

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha^{(k)} \left[\boldsymbol{H}^{(k)}\right]^{-1} \boldsymbol{g}^{(k)}$$

   - Méthode de Newton approchée : $\boldsymbol{D}^{(k)} = \text{diag}\left\{d_1^{(k)}, \cdots, d_n^{(k)}\right\}$ avec

$$d_i^{(k)} = \boldsymbol{D}_{ii} = \left(\frac{\partial^2 J}{\partial {x_i}^2}\right)^{-1}$$

     les éléments diagonaux de la matrice Hessienne.

– Méthodes de Newton modifiées :

$$\boldsymbol{D}^{(k)} = \boldsymbol{D}^{(0)} = \left[\boldsymbol{H}^{(0)}\right]^{-1} = \boldsymbol{D}$$

ou

$$\boldsymbol{D}^{(k)} = \boldsymbol{D}^{(0)} = \mathrm{diag}\left\{d_1^{(0)}, \cdots, d_n^{(0)}\right\} \quad \text{avec} \quad d_i^{(0)} = \left(\frac{\partial^2 J}{\partial {x_i}^2}(\boldsymbol{x})^{(0)}\right)^{-1}$$

A.2 Cas des critères moindres carrés

Considérons maintenant le cas des critères de la forme

$$J(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{f}(\boldsymbol{x})\|^2 = \frac{1}{2}\sum_{i=1}^{M} |f_i(\boldsymbol{x})|^2$$

On a alors

$$\frac{\partial J(\boldsymbol{x})}{\partial x_n} = \sum_{i=1}^{M} \frac{\partial f_i}{\partial x_n} f_i$$

et

$$\frac{\partial^2 J(\boldsymbol{x})}{\partial x_n \partial x_m} = \sum_{i=1}^{M} \frac{\partial^2 f_i}{\partial x_n \partial x_m} f_i + \frac{\partial f_i}{\partial x_n}\frac{\partial f_i}{\partial x_m}$$

Si on note par

$$\boldsymbol{F} = \left[\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\right] = \left[\nabla f_1 \vdots \cdots \vdots \nabla f_M\right] = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \cdots & \frac{\partial f_M}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & & & \frac{\partial f_M}{\partial x_2} \\ \vdots & & & \\ & & & \frac{\partial f_M}{\partial x_{N-1}} \\ \frac{\partial f_1}{\partial x_N} & \frac{\partial f_2}{\partial x_N} & & \frac{\partial f_M}{\partial x_N} \end{pmatrix}$$

on a

$$\nabla^2 J(\boldsymbol{x}) = \boldsymbol{F}^t\boldsymbol{F} + \sum_{i=1}^{M} \frac{\partial^2 g_i}{\partial x_n \partial x_m} g_i$$

et on peut alors distinguer les algorithmes suivants :

– Méthode de Newton : $\boldsymbol{D}^{(k)} = \left[\nabla^2 J(\boldsymbol{x}^{(k)})\right]^{-1}$

– Méthode de Gauss-Newton : $\boldsymbol{D}^{(k)} = (\boldsymbol{F}^t\boldsymbol{F})^{-1}$

– Méthodes de Levenberg-Marquart :

$$\boldsymbol{D}^{(k)} = (\boldsymbol{F}^t\boldsymbol{F} + \lambda^{(k)}\boldsymbol{I})^{-1}$$

avec $\lambda^{(k)}$ une suite décroissante de nombres positifs.
– pour λ grand : gradient
– pour λ petit : Newton

– Méthodes de Quasi-Newton :

$$\boldsymbol{D}^{(k+1)} = \boldsymbol{D}^{(k)} + \frac{\boldsymbol{p}^{(k)}\boldsymbol{p}^{t^{(k)}}}{\boldsymbol{p}^{t^{(k)}}\boldsymbol{Q}^{(k)}} - \frac{\boldsymbol{D}^{(k)}\boldsymbol{Q}^{(k)}\boldsymbol{Q}^{t^{(k)}}\boldsymbol{D}^{(k)}}{\boldsymbol{Q}^{t^{(k)}}\boldsymbol{D}^{(k)}\boldsymbol{Q}^{(k)}} + \eta^{(k)}\tau^{(k)}\boldsymbol{v}^{(k)}\boldsymbol{v}^{t^{(k)}}$$

$$\begin{aligned}
\boldsymbol{p}^{(k)} &= \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} \\
\boldsymbol{Q}^{(k)} &= \nabla J(\boldsymbol{x}^{(k+1)}) - \nabla J(\boldsymbol{x}^{(k)}) \\
\boldsymbol{v}^{(k)} &= \frac{\boldsymbol{p}^{(k)}}{\boldsymbol{p}^{t^{(k)}}\boldsymbol{Q}^{(k)}} - \frac{\boldsymbol{D}^{(k)}\boldsymbol{Q}^{(k)}}{\tau^{(k)}} \\
\tau^{(k)} &= \boldsymbol{Q}^{t^{(k)}}\boldsymbol{D}^{(k)}\boldsymbol{Q}^{(k)}
\end{aligned}$$

  – Méthode de Davidon-Fletcher-Powel (DFP) : $\eta^{(k)} = 0$
  – Méthode de Broyden-Fletcher-Golfand-Shayno (BGFGS) : $\eta^{(k)} = 1$

A.3 Cas des problèmes inverses avec un bruit additif

Considérons maintenant le modèle :

$$y_i = h_i(\boldsymbol{x}) + b_i, \quad ,i = 1, \ldots ,M$$

ou sous sa forme vectorielle :

$$\boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}) + \boldsymbol{b},$$

où $\boldsymbol{y}$ sont les données et $\boldsymbol{x}$ les inconnues.

Le critère des moindres carrés s'écrit :

$$J(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{h}(\boldsymbol{x})\|^2 = [\boldsymbol{y} - \boldsymbol{h}(\boldsymbol{x})]^t[\boldsymbol{y} - \boldsymbol{h}(\boldsymbol{x})] = \sum_{i=1}^{M} (y_i - h_i(\boldsymbol{x}))^2$$

Lorsque le système est non linéaire, ce critère n'est pas quadratique et, en général, peut être multimodal. Cependant l'usage des méthodes du type gradient est courant pour calculer une solution correspondant à un minimum local de ce critère. L'objet de cette annexe est de fournir une description succincte de ces méthodes.

Notons le gradient du critère par :

$$\nabla J(\boldsymbol{x}) = \left[\frac{\partial J}{\partial x_j}\right] = \left[\frac{\partial J}{\partial x_1}, \cdots, \frac{\partial J}{\partial x_N}\right]^t$$

et la matrice Hessien du critère par :

$$\nabla^2 J(\boldsymbol{x}) = \left[\frac{\partial^2 J}{\partial x_k \partial x_l}\right] = \begin{pmatrix} \frac{\partial^2 J}{\partial {x_1}^2} & \frac{\partial^2 J}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 J}{\partial x_1 \partial x_N} \\ \frac{\partial^2 J}{\partial x_2 \partial x_1} & & & \\ \vdots & & & \\ & & & \frac{\partial^2 J}{\partial x_{N-1} \partial x_N} \\ \frac{\partial^2 J}{\partial x_N \partial x_1} & & \frac{\partial^2 J}{\partial x_N \partial x_{N-1}} & \frac{\partial^2 J}{\partial {x_N}^2} \end{pmatrix}$$

Notons aussi :

$$\nabla h_i(\boldsymbol{x}) = \left[\frac{\partial h_i}{\partial x_j}\right] = \left[\frac{\partial h_i}{\partial x_1}, \cdots, \frac{\partial h_i}{\partial x_N}\right]^t,$$

$$\nabla^2 h_i = \left[\frac{\partial^2 h_i}{\partial x_k \partial x_l}\right] = \begin{pmatrix} \frac{\partial^2 h_i}{\partial {x_1}^2} & \frac{\partial^2 h_i}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 h_i}{\partial x_1 \partial x_N} \\ \frac{\partial^2 h_i}{\partial x_2 \partial x_1} & & & \\ \vdots & & & \\ & & & \frac{\partial^2 h_i}{\partial x_{N-1} \partial x_N} \\ \frac{\partial^2 h_i}{\partial x_N \partial x_1} & & \frac{\partial^2 h_i}{\partial x_N \partial x_{N-1}} & \frac{\partial^2 h_i}{\partial {x_N}^2} \end{pmatrix}$$

et

$$\boldsymbol{H} = \left[\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}\right] = \left[\nabla h_1 \vdots \cdots \vdots \nabla h_M\right] = \begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_2}{\partial x_1} & \cdots & \frac{\partial h_M}{\partial x_1} \\ \frac{\partial h_1}{\partial x_2} & & & \frac{\partial h_M}{\partial x_2} \\ \vdots & & & \\ & & & \frac{\partial h_M}{\partial x_{N-1}} \\ \frac{\partial h_1}{\partial x_N} & \frac{\partial h_2}{\partial x_N} & & \frac{\partial h_M}{\partial x_N} \end{pmatrix}$$

On a alors

$$\begin{aligned}
\nabla J(\boldsymbol{x}) &= -2\left[\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}\right]^t [\boldsymbol{y}-\boldsymbol{h}(\boldsymbol{x})] \\
&= -2\left[\boldsymbol{H}(\boldsymbol{x})\right]^t [\boldsymbol{y}-\boldsymbol{h}(\boldsymbol{x})] \\
\nabla^2 J(\boldsymbol{x}) &= 2\left[\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}\right]^t \left[\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}\right] + 2\left[\sum_i [y_i - h_i(\boldsymbol{x})]\,\nabla^2 h_i(\boldsymbol{x})\right] \\
&= 2\left[\boldsymbol{H}(\boldsymbol{x})\right]^t [\boldsymbol{H}(\boldsymbol{x})] + 2\left[\sum_i [y_i - h_i(\boldsymbol{x})]\,\nabla^2 h_i(\boldsymbol{x})\right]
\end{aligned}$$

Notons que pour un système linéaire on a

$$y_i = h_i(\boldsymbol{x}) = \sum_{j=1}^{N} h_{ij} x_j \quad \Longrightarrow \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b},$$

avec

$$\begin{aligned}
\boldsymbol{A} &= \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1N} \\ h_{21} & \cdots & \cdots & \vdots \\ \vdots & & & \vdots \\ h_{M1} & \cdots & \cdots & h_{MN} \end{pmatrix} \\
J(\boldsymbol{x}) &= \|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|^2 = [\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}]^t[\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}] \\
\boldsymbol{H} &= \boldsymbol{A} \\
\nabla^2 h_i &= \left[\frac{\partial^2 h_i}{\partial \boldsymbol{x}^2}\right] = 0 \\
\nabla J &= -2\boldsymbol{A}^t[\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}] \\
\nabla^2 J &= 2\boldsymbol{A}^t\boldsymbol{A}
\end{aligned}$$

car $\frac{\partial^2 h_i}{\partial x_k \partial x_l} = 0, \forall k,l$.

Développement à l'ordre deux, en série de Taylor de $J(\boldsymbol{x})$ autour d'un point $\boldsymbol{x}^*$ donne

$$J(\boldsymbol{x}) = J(\boldsymbol{x}^*) + [\nabla J(\boldsymbol{x}^*)]^t [\boldsymbol{x}-\boldsymbol{x}^*] + \frac{1}{2}[\boldsymbol{x}-\boldsymbol{x}^*]^t \left[\nabla^2 J(\boldsymbol{x}^*)\right][\boldsymbol{x}-\boldsymbol{x}^*]$$

Conditions d'existence d'une solution :

$$\begin{cases} \nabla J(\boldsymbol{x}^*) = \boldsymbol{0} \\ \nabla^2 J(\boldsymbol{x}^*) \geq 0 \quad \text{(C.N. dnn)} \\ \nabla^2 J(\boldsymbol{x}^*) > 0 \quad \text{(C.S. dp)} \end{cases}$$

La première condition d'existence d'une solution donne lieu au système d'équations non linéaires :

$$\nabla J(\boldsymbol{x}^*) = \boldsymbol{0} \quad \Longrightarrow \quad [\boldsymbol{H}(\boldsymbol{x}^*)]^t [\boldsymbol{y}-\boldsymbol{h}(\boldsymbol{x}^*)] = \boldsymbol{0}$$

Méthode de Gauss :
Développement à l'ordre 1 de $\boldsymbol{y}$ donne :

$$\boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}^*) + \left[\frac{\partial \boldsymbol{h}(\boldsymbol{x}^*)}{\partial \boldsymbol{x}}\right]^t [\boldsymbol{x}-\boldsymbol{x}^*] + \boldsymbol{b} = \boldsymbol{h}(\boldsymbol{x}^*) + \boldsymbol{H}^t[\boldsymbol{x}-\boldsymbol{x}^*] + \boldsymbol{b}$$

$$\nabla J(\boldsymbol{x}) = \mathbf{0} \quad \Longrightarrow \quad \boldsymbol{x} = \boldsymbol{x}^* + [\boldsymbol{H}^t\boldsymbol{H}]^{-1}\boldsymbol{H}^t[\boldsymbol{y} - \boldsymbol{h}(\boldsymbol{x}^*)]$$

Algorithme :

$$\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(k-1)} + \alpha[\boldsymbol{H}^t\boldsymbol{H}]^{-1}\boldsymbol{H}^t[\boldsymbol{y} - \boldsymbol{h}(\boldsymbol{x}^{(k)})]$$

Méthode de Newton–Raphson :
Développement à l'ordre 2 de $\boldsymbol{y}$ donne :

$$\nabla J(\boldsymbol{x}) = \mathbf{0} \Longrightarrow \boldsymbol{x} = \boldsymbol{x}^* - \left[\nabla^2 J(\boldsymbol{x}^*)\right]^{-1}[\nabla J(\boldsymbol{x}^*)]$$

Algorithme :

$$\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(k-1)} + \alpha\left[\boldsymbol{H}^t\boldsymbol{H} + \sum_i [y_i - h_i(\boldsymbol{x}^{(k)})]\frac{\partial^2 h_i}{\partial \boldsymbol{x}^2}(\boldsymbol{x}^{(k)})\right]^{-1}\boldsymbol{H}^t[\boldsymbol{y} - \boldsymbol{h}(\boldsymbol{x}^{(k)})]$$

A.4 Comparaison entre les méthodes

Notons

$$\boldsymbol{H}_k = \left[\boldsymbol{H}(\boldsymbol{x}^{(k)})\right], \quad \text{et} \quad \boldsymbol{M}_k = \left[\boldsymbol{M}(\boldsymbol{x}^{(k)})\right].$$

On a alors dans les deux cas :

$$\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(k-1)} + \alpha\boldsymbol{M}_k^{-1}\boldsymbol{H}_k^t[\boldsymbol{y} - \boldsymbol{h}(\boldsymbol{x}^{(k)})]$$

– Méthode de Newton–Raphson :

$$\boldsymbol{M}_k = \left[\boldsymbol{H}_k^t\boldsymbol{H}_k + \sum_i [y_i - h_i(\boldsymbol{x}^{(k)})]\nabla^2 h_i(\boldsymbol{x}^{(k)})\right]$$

– Méthode de Gauss :

$$\boldsymbol{M}_k = [\boldsymbol{H}_k^t\boldsymbol{H}_k]$$

– Méthode du gradient conjugué :

$$\boldsymbol{M}_k = [\boldsymbol{H}_k^t\boldsymbol{H}_k], \quad \boldsymbol{H}_k = \left[\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}(\boldsymbol{x}^{(k)})\right] \quad \text{est supposée Toeplitz}$$

– Méthode du gradient :

$$\boldsymbol{M}_k = \boldsymbol{I}$$

Dans le cas d'un système linéaire on a

$$\boldsymbol{H}_k = \boldsymbol{H} = \boldsymbol{A}, \quad \boldsymbol{M}_k = \boldsymbol{M} = \boldsymbol{H}^t\boldsymbol{H}.$$

Ceci vaut dire qu'il y a l'équivalence entre la méthode de Newton–Raphson et la méthode de Gauss.

Si le système linéaire et invariant par translation on a

$$\boldsymbol{H}_k = \boldsymbol{H} = \boldsymbol{A} \quad \text{(Toeplitz)}, \quad \boldsymbol{M}_k = \boldsymbol{M} = \boldsymbol{H}^t\boldsymbol{H} \quad \text{(Toeplitz symetrique)}$$

Ceci vaut dire qu'il y a l'équivalence entre la méthode de Newton–Raphson, la méthode de Gauss et la méthode du gradient conjugué.

Annexe B

Algorithme de Gauss-Seidel pour optimisation d'un critère quadratique

Lorsqu'on cherche à optimiser un critère d'une manière itérative, il existe deux approches:

- les méthodes qui mettent à jour l'ensemble des variables à chaque itérations, comme par exemples des méthodes du type gradient ; et
- les méthodes qui mettent à jour seulement une variable à chaque itérations.

L'objet de cette annexe est étudier la méthode de Gausse-Seidel qui est une méthode de la deuxième approche pour le cas du critère MC ou celui d'un critère régularisé.

B.1 Introduction

Considérons le critère

$$J(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2$$

et son gradient :

$$\nabla J = -2\boldsymbol{A}^t(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})$$

Rappelons le principe d'un algorithme du gradient pour calculer la solution :

$$\begin{aligned}
\boldsymbol{x}^{(0)} &= \boldsymbol{0} \\
\boldsymbol{x}^{(k+1)} &= \boldsymbol{x}^{(k)} + \alpha^{(k)} \nabla J\left(\boldsymbol{x}^{(k)}\right) \\
&= \boldsymbol{x}^{(k)} + \alpha^{(k)} \boldsymbol{A}^t\left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{(k)}\right) = \boldsymbol{x}^{(k)} + \alpha^{(k)} \boldsymbol{A}^t\boldsymbol{y} - \alpha^{(k)} \boldsymbol{A}^t\boldsymbol{A}\boldsymbol{x}^{(k)}
\end{aligned}$$

Supposons maintenant, qu'à l'itération k, on voudrait remettre à jour seulement la variable x_k et écrivons alors le critère en fonction de cette seule variable :

$$\begin{aligned}
J(x_k) &= \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2 \\
&= \sum_{i=1}^{M} \left(y_i - \sum_{j=1}^{N} a_{ij} x_j \right)^2 \\
&= \sum_{i=1}^{M} \left(y_i - \sum_{j \neq k}^{N} a_{ij} x_j - a_{ik} x_k \right)^2 \\
&= \sum_{i=1}^{M} \left[\left(y_i - \sum_{j \neq k}^{N} a_{ij} x_j \right)^2 - 2 a_{ik} \left(y_i - \sum_{j \neq k}^{N} a_{ij} x_j \right) x_k + a_{ik}^2 x_k^2 \right] \\
\frac{\partial J}{\partial x_k} &= -2 \sum_{i=1}^{M} a_{ik} \left(y_i - \sum_{j=1}^{N} a_{ij} x_j \right) \\
&= -2 \sum_{i=1}^{M} a_{ik} \left(y_i - \sum_{j \neq k}^{N} a_{ij} x_j - a_{ik} x_k \right) \\
&= -2 \sum_{i=1}^{M} \left[a_{ik} \left(y_i - \sum_{j \neq k}^{N} a_{ij} x_j \right) - a_{ik}^2 x_k \right]
\end{aligned}$$

Le critère $J(x_k)$ est quadratique en x_k et sa dérivée est linéaire. On peut alors calculer explicitement la solution en annulant la dérivée :

$$\begin{aligned}
\frac{\partial J}{\partial x_k} = 0 \longrightarrow x_k &= \frac{1}{\sum_{i=1}^{M} a_{ik}^2} \sum_{i=1}^{M} a_{ik} \left(y_i - \sum_{j \neq k}^{N} a_{ij} x_j \right) \\
&= \frac{1}{\sum_{i=1}^{M} a_{ik}^2} \sum_{i=1}^{M} a_{ik} \left(y_i - \sum_{j=1}^{N} a_{ij} x_j + a_{ik} x_k \right) \\
&= x_k + \frac{1}{\sum_{i=1}^{M} a_{ik}^2} \sum_{i=1}^{M} a_{ik} \left(y_i - \sum_{j=1}^{N} a_{ij} x_j \right)
\end{aligned}$$

$$
\begin{aligned}
&= x_k + \frac{1}{< \boldsymbol{a}_{*k}, \boldsymbol{a}_{*k} >} \sum_{i=1}^{M} a_{ik} \left(y_i - [\boldsymbol{A}\boldsymbol{x}^{(k)}]_i \right) \\
&= x_k + \frac{1}{< \boldsymbol{a}_{*k}, \boldsymbol{a}_{*k} >} \left\langle \boldsymbol{a}_{*k}, (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{(k)}) \right\rangle
\end{aligned}
$$

Pour simplifier ces écritures, notons par

$$\boldsymbol{x} = \boldsymbol{x}_{(-k)} + \boldsymbol{x}_{(k)}$$

avec

$$\boldsymbol{x}_{(-k)} = [x_1, \cdots, x_{k-1}, 0, x_{k+1}, \cdots, x_N]^t, \text{ et } \boldsymbol{x}_{(k)} = [0, \cdots, 0, x_k, 0, \cdots, 0]^t$$

On a alors

$$J(x_k) = a x_k^2 + b x_k + c$$

avec

$$
\begin{aligned}
a &= \sum_i a_{ik}^2 = \|\boldsymbol{a}_{*k}\|^2 \\
b &= -2 \sum_i a_{ik} \left(y_i - \sum_{j \neq k}^{N} a_{ij} x_j \right) = -2 \left\langle \boldsymbol{a}_{*k}, (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_{(-k)}) \right\rangle \\
c &= \sum_i \left(y_i - \sum_{j \neq k}^{N} a_{ij} x_j \right)^2 = \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_{(-k)}\|^2
\end{aligned}
$$

L'algorithme de Gausse-Seidel est fondée sur cette relation et donné par :

$$
\begin{aligned}
\boldsymbol{x}^{(0)} &= \boldsymbol{0} \\
x_k^{(k+1)} &= x_k^{(k)} + \alpha^{(k)} \frac{\partial J}{\partial x_k} \left(\boldsymbol{x}^{(k)} \right) \\
&= x_k^{(k)} + \alpha^{(k)} \left[\boldsymbol{A}^t \left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{(k)} \right) \right]_k \\
&= x_k^{(k)} + \alpha^{(k)} \sum_{i=1}^{M} a_{ik} \left(y_i - [\boldsymbol{A}\boldsymbol{x}^{(k)}]_i \right) \quad \text{avec} \quad \alpha^{(k)} = \frac{1}{< \boldsymbol{a}_{*k}, \boldsymbol{a}_{*k} >}.
\end{aligned}
$$

Rappelons à titre de comparaison la Méthode de Kaczmarz que nous l'avons vue dans le chapitre concernant l'inversion généralisée :

$$
\begin{aligned}
\boldsymbol{x}^{(0)} &= \boldsymbol{0} \\
\boldsymbol{x}^{(k+1)} &= \boldsymbol{x}^{(k)} + \frac{\left(y_i - [\boldsymbol{A}\boldsymbol{x}^{(k)}]_i \right)}{< \boldsymbol{a}_{i*}, \boldsymbol{a}_{i*} >} \boldsymbol{a}_{i*} \\
&= \boldsymbol{x}^{(k)} + \frac{\left(y_i - \left\langle \boldsymbol{a}_{i*}, \boldsymbol{x}^{(k)} \right\rangle \right)}{< \boldsymbol{a}_{i*}, \boldsymbol{a}_{i*} >} \boldsymbol{a}_{i*}, \quad i = 1, \cdots, M, 1 \cdots, M, 1, \cdots
\end{aligned}
$$

B.2 Le cas d'un critère régularisé

Le critère et son gradient dans ce cas sont :

$$\begin{aligned} J(\boldsymbol{x}) &= \|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|^2+\lambda\|\boldsymbol{C}\boldsymbol{x}\|^2, \\ \nabla J &= -2\boldsymbol{A}^t(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})+2\lambda\boldsymbol{C}^t\boldsymbol{C}\boldsymbol{x}=-2\left[\boldsymbol{A}^t\boldsymbol{y}-(\boldsymbol{A}^t\boldsymbol{A}+2\lambda\boldsymbol{C}^t\boldsymbol{C})\boldsymbol{x}\right] \end{aligned}$$

La même démarche nous conduit :

$$\begin{aligned} J(x_k) &= \|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|^2+\lambda\|\boldsymbol{C}\boldsymbol{x}\|^2 \\ &= \sum_{i=1}^{M}\left(y_i-\sum_{j=1}^{N}a_{ij}x_j\right)^2+\sum_{j=1}^{N}\left(\sum_{n=1}^{N}c_{jn}x_n\right)^2 \\ &= \sum_{i=1}^{M}\left[\left(y_i-\sum_{j\neq k}^{N}a_{ij}x_j\right)^2-2a_{ik}\left(y_i-\sum_{j\neq k}^{N}a_{ij}x_j\right)x_k+a_{ik}^2x_k^2\right] \\ &\quad +\lambda\sum_{j\neq k}^{N}\left(\sum_{n\neq k}^{N}c_{jn}x_n+c_{jk}x_k\right)^2+\lambda\sum_{n\neq k}^{N}\left(c_{kn}x_n+c_{kk}x_k\right)^2 \\ \frac{\partial J}{\partial x_k} &= -2\sum_{i=1}^{M}a_{ik}\left(y_i-\sum_{j=1}^{N}a_{ij}x_j\right)+2\lambda\sum_{j=1}^{N}\sum_{n=1}^{N}c_{jk}c_{jn}x_n \\ &= -2\sum_{i=1}^{M}a_{ik}\left(y_i-\sum_{j\neq k}^{N}a_{ij}x_j-a_{ik}x_k\right)+2\lambda\sum_{n=1}^{N}x_n\sum_{j=1}^{N}c_{jk}c_{jn} \\ &= -2\sum_{i=1}^{M}\left[a_{ik}\left(y_i-\sum_{j\neq k}^{N}a_{ij}x_j\right)-a_{ik}^2x_k\right]+2\lambda\sum_{n\neq k}^{N}x_n\sum_{j=1}^{N}c_{jk}c_{jn}+2\lambda x_k\sum_{j=1}^{N}c_{jk}^2 \end{aligned}$$

Solution explicite:

$$\nabla J=\boldsymbol{0}\longrightarrow(\boldsymbol{A}^t\boldsymbol{A}+\lambda\boldsymbol{C}^t\boldsymbol{C})\boldsymbol{x}=\boldsymbol{A}^t\boldsymbol{y}\longrightarrow\boldsymbol{x}=(\boldsymbol{A}^t\boldsymbol{A}+\lambda\boldsymbol{C}^t\boldsymbol{C})^{-1}\boldsymbol{A}^t\boldsymbol{y}$$

Méthode du gradient :

$$\begin{aligned} \boldsymbol{x}^{(0)} &= \boldsymbol{0} \\ \boldsymbol{x}^{(k+1)} &= \boldsymbol{x}^{(k)}+\alpha^{(k)}\nabla J\left(\boldsymbol{x}^{(k)}\right) \\ &= \boldsymbol{x}^{(k)}+\alpha^{(k)}\left[\boldsymbol{A}^t\left(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}^{(k)}\right)-\lambda\boldsymbol{C}^t\boldsymbol{C}\boldsymbol{x}^{(k)}\right] \\ &= \boldsymbol{x}^{(k)}+\alpha^{(k)}\boldsymbol{A}^t\boldsymbol{y}-\alpha^{(k)}\left(\boldsymbol{A}^t\boldsymbol{A}+\frac{\lambda}{\alpha}\boldsymbol{C}^t\boldsymbol{C}\right)\boldsymbol{x}^{(k)} \end{aligned}$$

Algorithme du Gauss-Seidel :

$$\begin{aligned} \frac{\partial J}{\partial x_k} &= 0\longrightarrow \\ x_k &= \frac{1}{\sum_{i=1}^{M}a_{ik}^2+\lambda\sum_{j=1}^{N}c_{jk}^2}\sum_{i=1}^{M}a_{ik}\left(y_i-\sum_{j\neq k}^{N}a_{ij}x_j\right)-\lambda\sum_{n\neq k}^{N}x_n\sum_{j=1}^{N}c_{jk}c_{jn} \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sum_{i=1}^{M} a_{ik}^2 + \lambda \sum_{j=1}^{N} c_{jk}^2} \sum_{i=1}^{M} a_{ik} \left(y_i - \sum_{j=1}^{N} a_{ij} x_j + a_{ik} x_k \right) - \lambda \sum_{n=1}^{N} x_n \sum_{j=1}^{N} c_{jk} c_{jn} + \lambda x_k \sum_{j=1}^{N} c_{jk}^2 \\
&= x_k + \frac{1}{\|\boldsymbol{a}_{*k}\|^2 + \lambda \|\boldsymbol{c}_{*k}\|^2} \left\{ \sum_{i=1}^{M} a_{ik} \left(y_i - [\boldsymbol{A}\boldsymbol{x}^{(k)}]_i \right) + \lambda \sum_{j=1}^{N} c_{jk} \sum_{n=1}^{N} c_{jn} x_n \right\} \\
&= x_k + \frac{1}{\|\boldsymbol{a}_{*k}\|^2 + \lambda \|\boldsymbol{c}_{*k}\|^2} \left\{ \left[\boldsymbol{A}^t \left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{(k)} \right) \right]_k + \lambda \left[\boldsymbol{C}^t \boldsymbol{C} \boldsymbol{x}^{(k)} \right]_k \right\} \\
&= x_k + \frac{1}{\|\boldsymbol{a}_{*k}\|^2 + \lambda \|\boldsymbol{c}_{*k}\|^2} \left\{ [\boldsymbol{A}^t \boldsymbol{y}]_k - \left[\left(\boldsymbol{A}^t \boldsymbol{A} + \lambda \boldsymbol{C}^t \boldsymbol{C} \right) \boldsymbol{x}^{(k)} \right]_k \right\} \\
&= x_k + \frac{1}{\|\boldsymbol{a}_{*k}\|^2 + \lambda \|\boldsymbol{c}_{*k}\|^2} \left\{ \left\langle \boldsymbol{a}_{*k}, (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{(k)}) \right\rangle + \lambda \left[\boldsymbol{C}^t \boldsymbol{C} \boldsymbol{x}^{(k)} \right]_k \right\}
\end{aligned}
$$

Algorithme :

$$
\begin{aligned}
\boldsymbol{x}^{(0)} &= \boldsymbol{0} \\
x_k^{(k+1)} &= x_k^{(k)} + \alpha^{(k)} + \alpha^k \left\{ \left\langle \boldsymbol{a}_{*k}, (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{(k)}) \right\rangle + \lambda \left[\boldsymbol{C}^t \boldsymbol{C} \boldsymbol{x}^{(k)} \right]_k \right\} \\
&\qquad \text{avec} \quad \alpha^k = \frac{1}{\|\boldsymbol{a}_{*k}\|^2 + \lambda \|\boldsymbol{c}_{*k}\|^2}
\end{aligned}
$$

Annexe C

Expressions des différentes fonctions potentiels

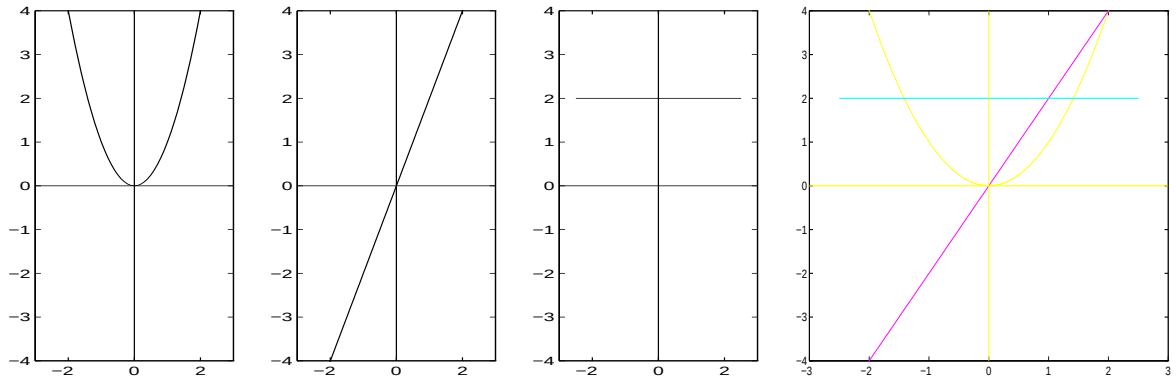
Dans cette annexe, nous faisons un inventaires des fonctions potentiels utilisées en modélisation markovienne pour la résolution des problèmes inverses.

Ces fonctions sont classées en deux familles : convexes et non convexes.

C.1 Potentiels convexes

Quadratique ou L_2:

$$\begin{aligned} \phi(t) &= t^2 \\ \phi'(t) &= 2t \\ \phi''(t) &= 2 \end{aligned}$$

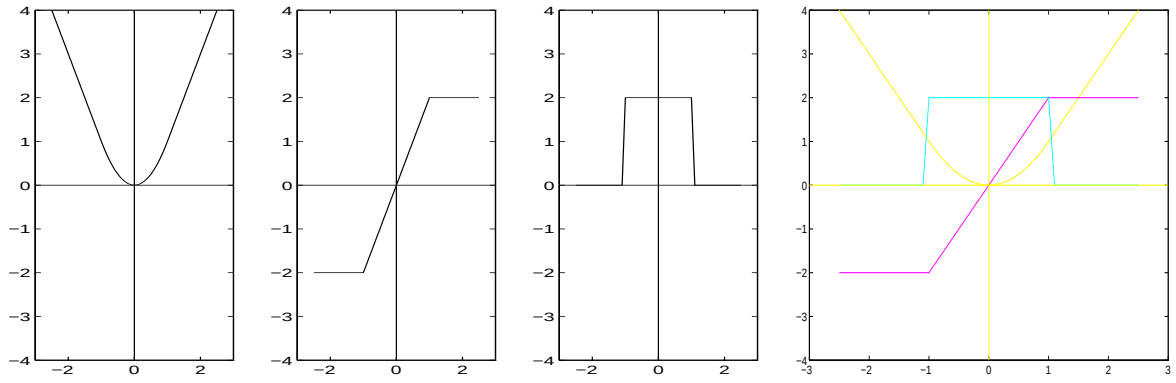


Quadratique prolongée en linéaire :

$$\phi(t) = \begin{cases} t^2 & |t| \leq 1 \\ 2t-1 & |t| > 1 \end{cases}$$

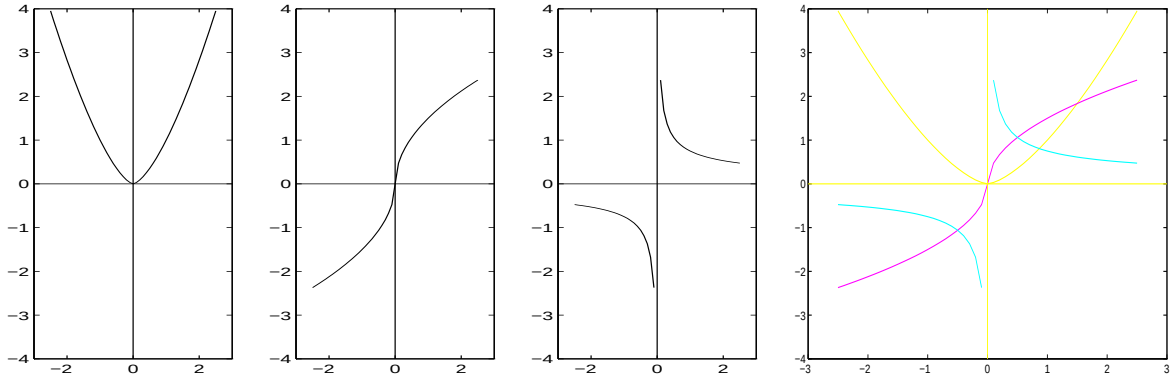
$$\phi'(t) = \begin{cases} 2t & |t| \leq 1 \\ 2 & |t| > 1 \end{cases}$$

$$\phi''(t) = \begin{cases} 2 & |t| \leq 1 \\ 0 & |t| > 1 \end{cases}$$



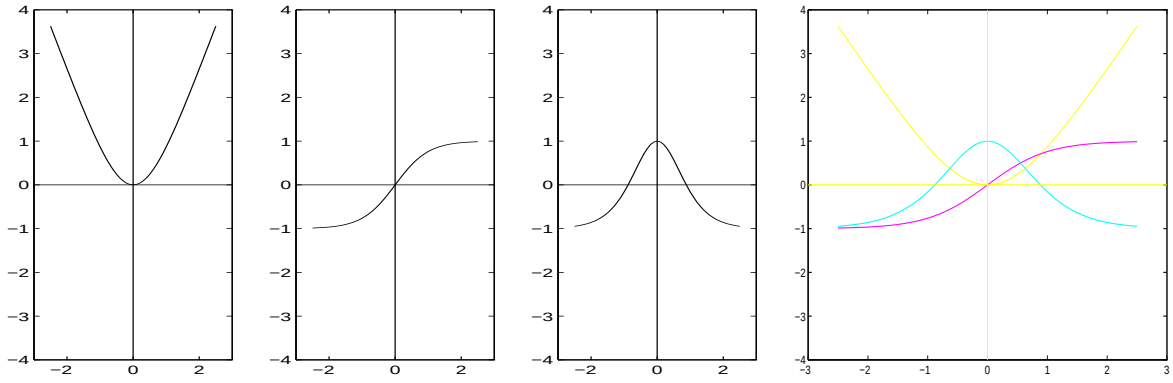
L_p :

$$\begin{aligned}\phi(t) &= |t|^p, \quad 1 < p \leq 2\\ \phi'(t) &= p\,\mathrm{sign}(t)|t|^{p-1}\\ \phi''(t) &= p(p-1)\,|t|^{p-2}\end{aligned}$$



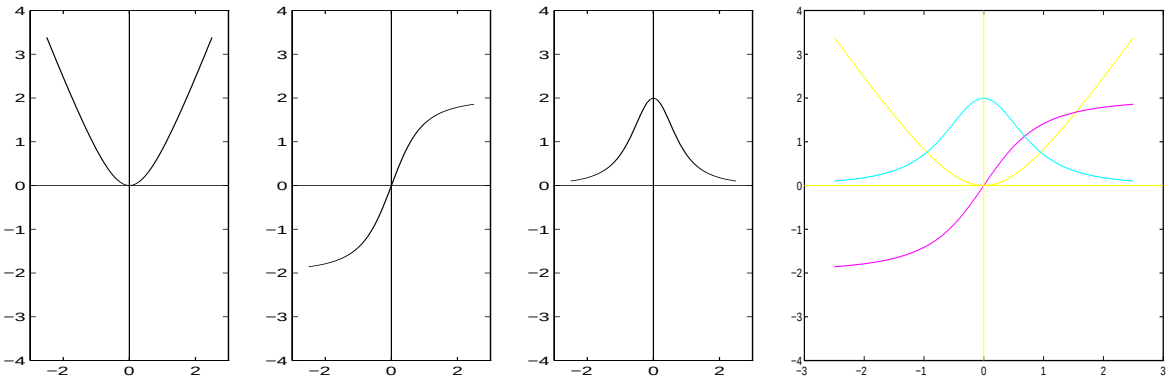
log cosh :

$$\begin{aligned}\phi(t) &= 2\log(\cosh(t))\\ \phi'(t) &= \frac{\sinh(t)}{\cosh(t)}\\ \phi''(t) &= \frac{1-\sinh^2(t)}{\cosh^2(t)}\end{aligned}$$



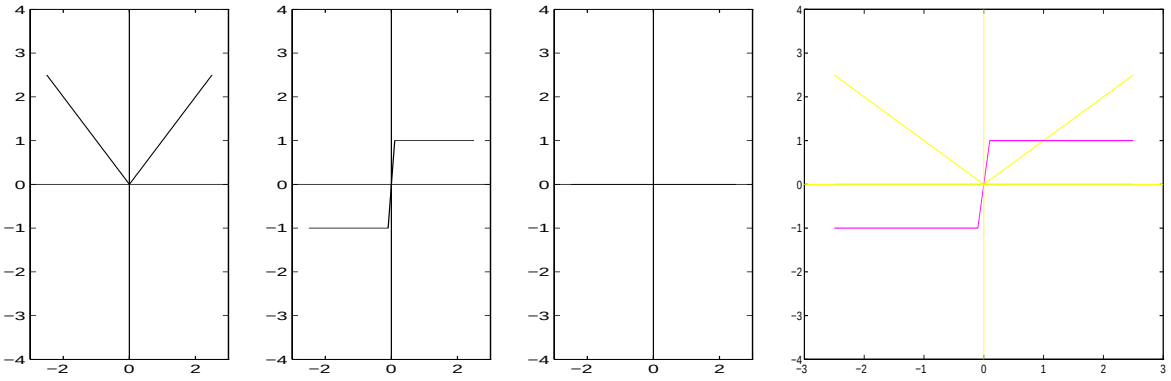
Racine carrée :

$$\begin{aligned}\phi(t) &= 2\sqrt{1+t^2}-2\\ \phi'(t) &= \frac{2t}{\sqrt{1+t^2}}\\ \phi''(t) &= \frac{2}{(1+t^2)^{3/2}}\end{aligned}$$



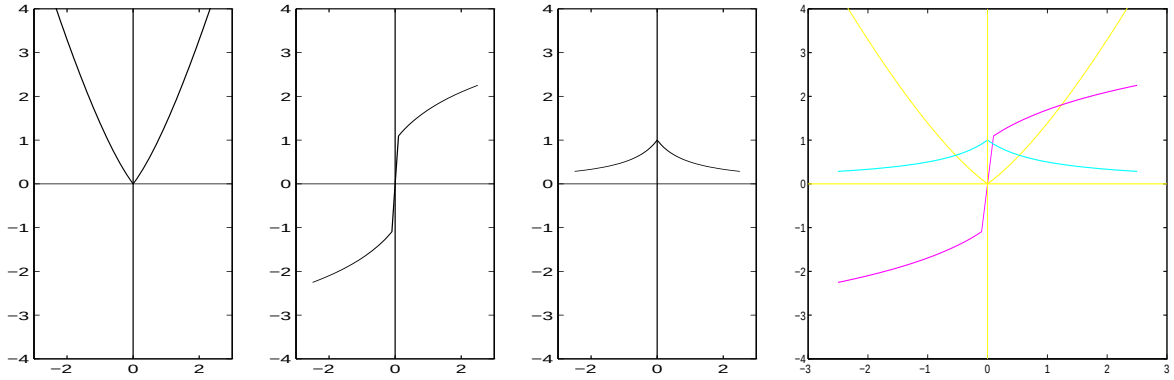
L_1**:**

$$\begin{aligned}\phi(t) &= |t|\\ \phi'(t) &= \text{sign}(t)\\ \phi''(t) &= 0\end{aligned}$$



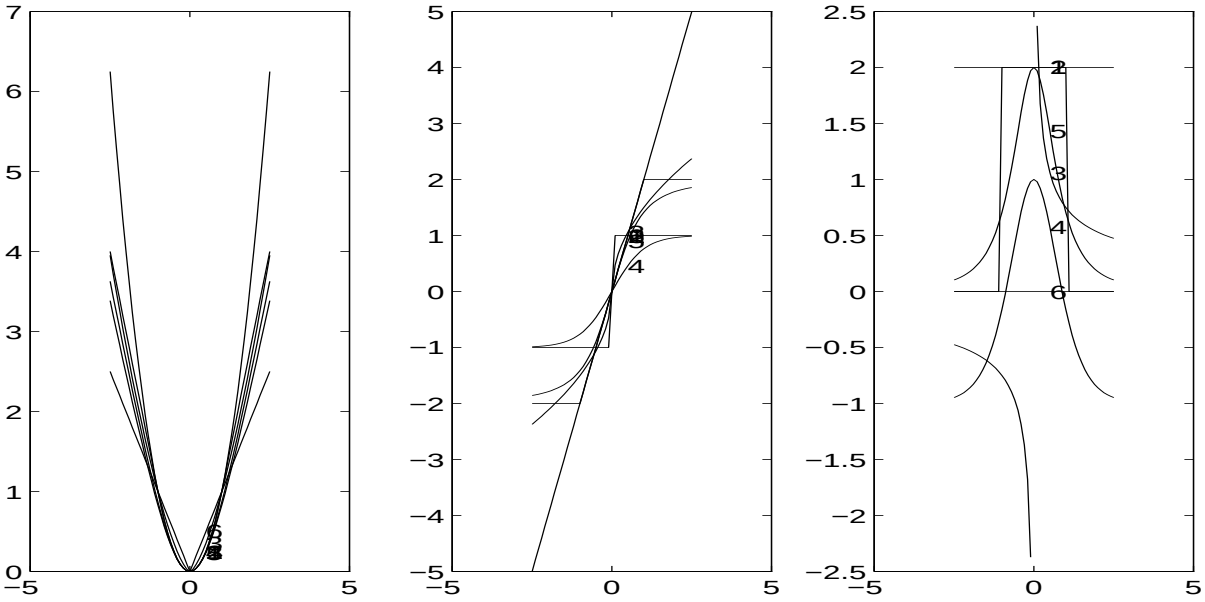
Logarithmique :

$$\begin{aligned}\phi(t) &= (|t|+1)\log(|t|+1)\\ \phi'(t) &= \mathrm{sign}(t)\,(1+\log(|t|+1))\\ \phi''(t) &= \frac{1}{|t|+1}\end{aligned}$$



Le tableau ci-dessous résume ces fonctions :

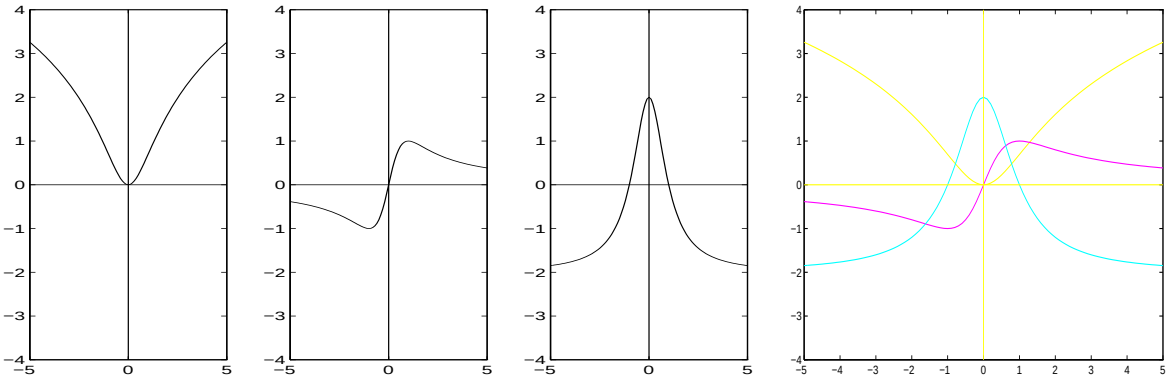
cas	$\phi(t)$	$\phi'(t)$	$\phi''(t)$	références
1	t^2	$2t$	2	Tikhonov
2	$\begin{cases} t^2 & \lvert t\rvert \le 1 \\ 2t-1 & \lvert t\rvert > 1 \end{cases}$	$\begin{cases} 2t & \lvert t\rvert \le 1 \\ 2 & \lvert t\rvert > 1 \end{cases}$	$\begin{cases} 2 & \lvert t\rvert \le 1 \\ 0 & \lvert t\rvert > 1 \end{cases}$	Hubert 81
3	$\lvert t\rvert^p, \quad 1 < p \le 2$	$p\,\mathrm{sign}(t)\lvert t\rvert^{p-1}$	$p(p-1)\,\lvert t\rvert^{p-2}$	Bouman & Sauer 93
4	$2\log(\cosh(t))$	$\frac{\sinh(t)}{\cosh(t)}$	$\frac{1-\sinh^2(t)}{\cosh^2(t)}$	Green 90
5	$2\sqrt{1+t^2}-2$	$\frac{2t}{\sqrt{1+t^2}}$	$\frac{2}{(1+t^2)^{3/2}}$	Aubert 94
6	$\lvert t\rvert$	$\mathrm{sign}(t)$	0	Rudin Osher 92, Besag 86
7	$(\lvert t\rvert+1)\log(\lvert t\rvert+1)$	$\mathrm{sign}(t)\,(1+\log(\lvert t\rvert+1))$	$\frac{1}{\lvert t\rvert+1}$	Zervakis 95
8	$(\lvert t-m\rvert+\lvert t\rvert+1)$ $\log\left(\frac{\lvert t-m\rvert+\lvert t\rvert+1}{\lvert t\rvert+1}+1\right)$	$\mathrm{sign}(t-m)$ $\log\left(\frac{\lvert t-m\rvert}{\lvert t\rvert+1}+1\right)$	$\frac{1}{\lvert t-m\rvert+\lvert t\rvert+1}$	Zervakis 95
9	$\begin{cases} t^2 & \lvert t\rvert \le 1 \\ 1-2\log\lvert t\rvert & \lvert t\rvert > 1 \end{cases}$	$\begin{cases} 2t & \lvert t\rvert \le 1 \\ \frac{2\lvert t\rvert}{t^2} & \lvert t\rvert > 1 \end{cases}$	$\begin{cases} 2 & \lvert t\rvert \le 1 \\ \frac{2}{t^2} & \lvert t\rvert > 1 \end{cases}$	AMD



C.2 Potentiels non convexes

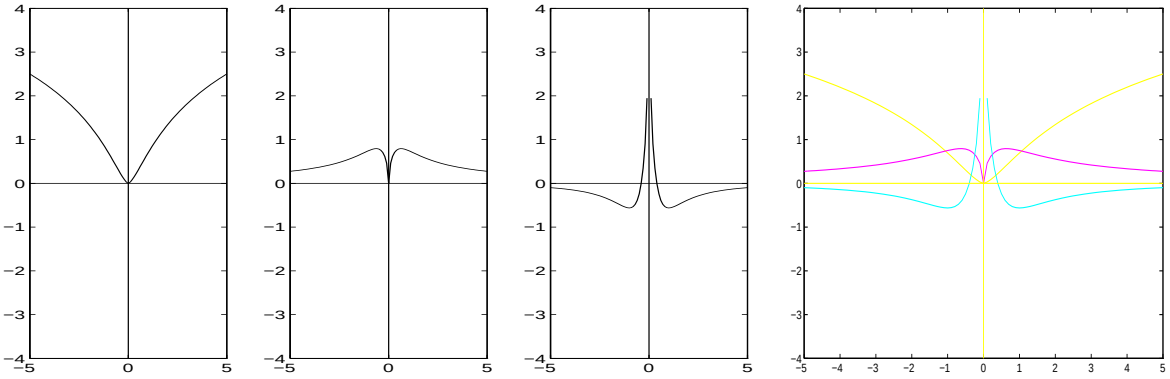
Hebert & Leahy, Perona & Malik :

$$\begin{aligned} \phi(t) &= \log(1+t^2) \\ \phi'(t) &= \frac{2t}{1+t^2} \\ \phi''(t) &= \frac{2(1-t^2)}{(1+t^2)^2} \end{aligned}$$



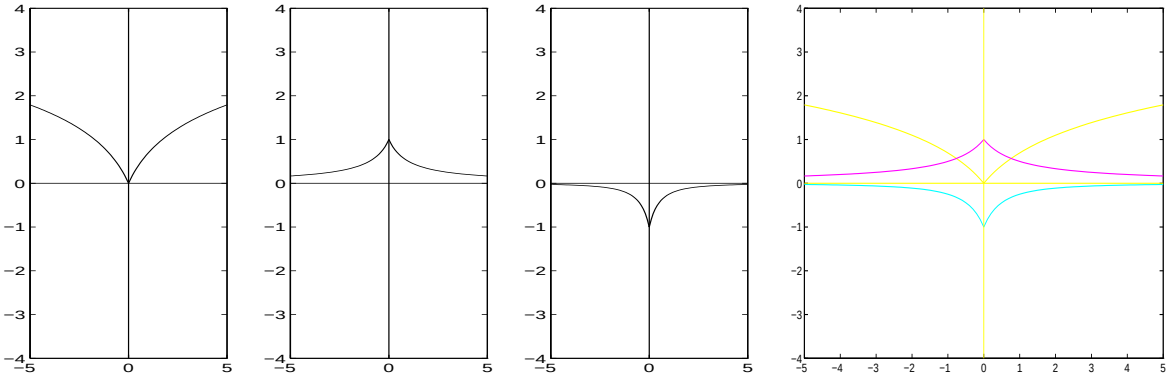
Extension :

$$\begin{aligned} \phi(t) &= \log(1+|t|^p) \\ \phi'(t) &= \frac{p\,|t|^{p-1}}{1+t^p} \\ \phi''(t) &= \frac{p(p-1)|t|^{p-2}(1-t^p)-p^2t^{2p-2}}{(1+t^p)^2} \end{aligned}$$



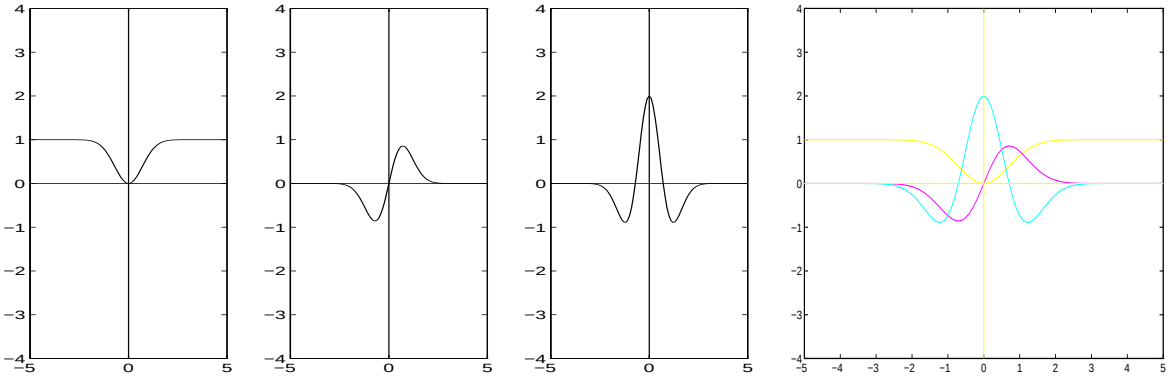
Cas particulier $p = 1$:

$$\begin{aligned}\phi(t) &= \log(1+|t|)\\ \phi'(t) &= \frac{\text{sign}(t)}{1+t}\\ \phi''(t) &= \frac{-1}{(1+t)^2}\end{aligned}$$



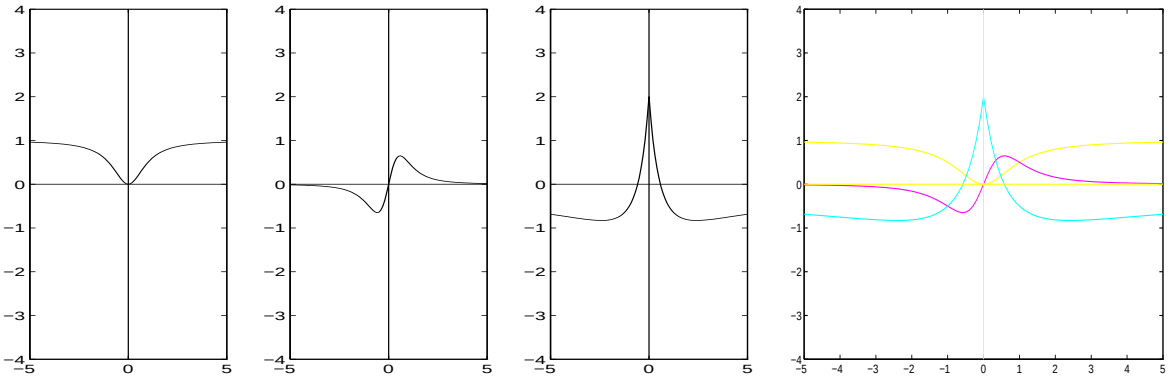
Gaussienne inversée :

$$\begin{aligned}\phi(t) &= 1-e^{-t^2}\\ \phi'(t) &= 2te^{-t^2}\\ \phi''(t) &= 2e^{-t^2}(1-2t^2)\end{aligned}$$



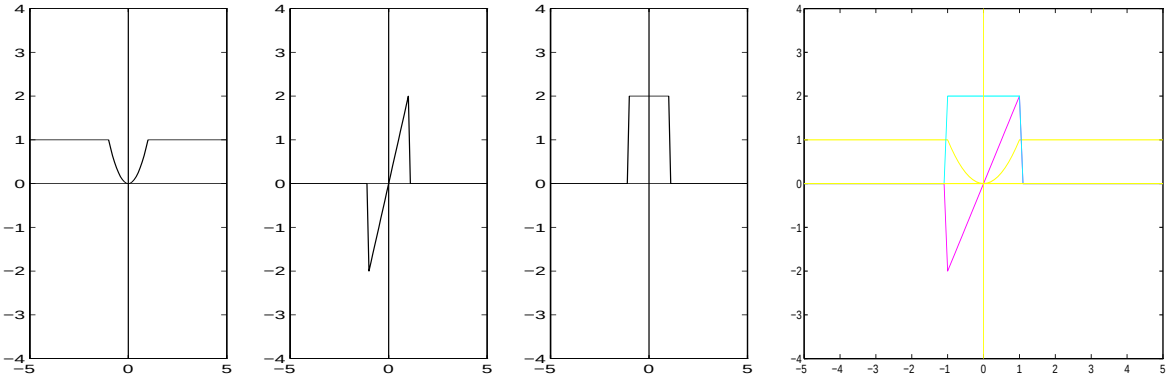
Lorentzienne inversée :

$$\begin{aligned}\phi(t) &= \frac{t^2}{1+t^2}\\ \phi'(t) &= \frac{2t}{(1+t^2)^2}\\ \phi''(t) &= \frac{2(1-3t^2)}{(1+t)^3}\end{aligned}$$



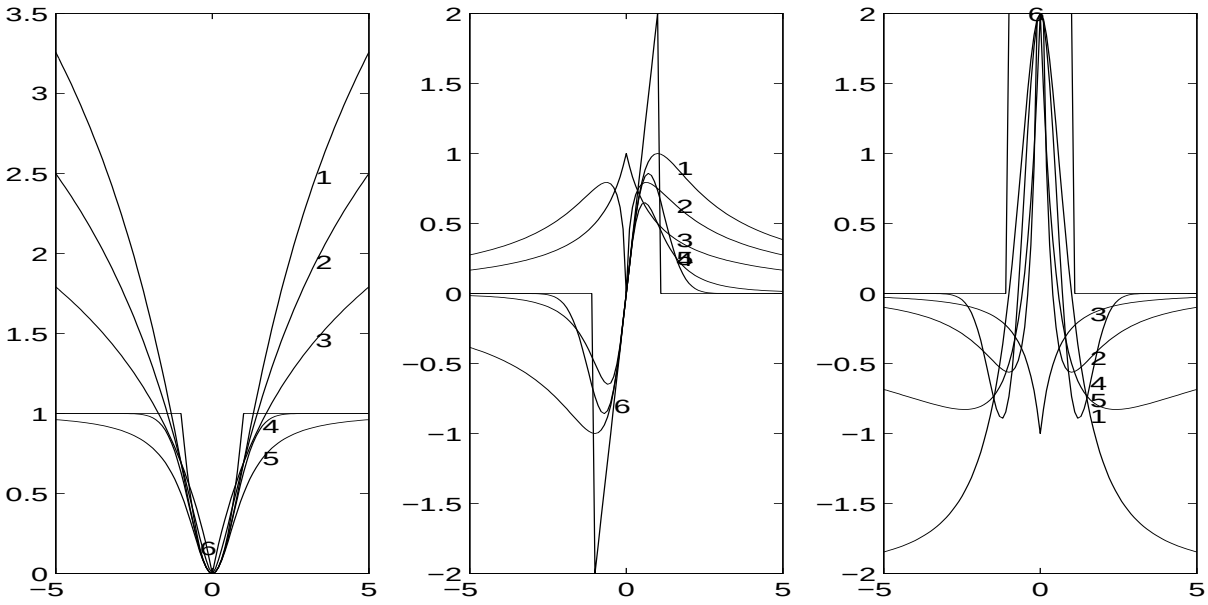
Quadratique tronquée :

$$\begin{aligned}\phi(t) &= \begin{cases} 2t & |t| \leq 1\\ 0 & |t| > 1\end{cases}\\ \phi'(t) &= \begin{cases} 2t & |t| \leq 1\\ 0 & |t| > 1\end{cases}\\ \phi''(t) &= \begin{cases} 2 & |t| \leq 1\\ 0 & |t| > 1\end{cases}\end{aligned}$$



Le tableau ci-dessous résume ces fonctions :

cas	$\phi(t)$	$\phi'(t)$	$\phi''(t)$	références
1	$\log(1+t^2)$	$\frac{2t}{1+t^2}$	$\frac{2(1-t^2)}{(1+t^2)^2}$	Hebert & Leahy 90
				Perona & Malik 90
2	$\log(1+\lvert t\rvert^p)$	$\frac{p\,\lvert t\rvert^{p-1}}{1+t^p}$	$\frac{p(p-1)\lvert t\rvert^{p-2}(1-t^p)-p^2t^{2p-2}}{(1+t^p)^2}$	AMD
3	$\log(1+\lvert t\rvert)$	$\mathrm{sign}(t)\frac{1}{1+t}$	$\frac{-1}{(1+t)^2}$	AMD
4	$-e^{-t^2}+1$	$2te^{-t^2}$	$2e^{-t^2}(1-2t^2)$	Perona & Malik 90
5	$\frac{t^2}{1+t^2}$	$\frac{2t}{(1+t^2)^2}$	$\frac{2(1-3t^2)}{(1+t)^3}$	Geman & McClure 85 90
6	$\min(1,t^2) = \begin{cases} t^2 & \lvert t\rvert \le 1 \\ 1 & \lvert t\rvert > 1 \end{cases}$	$\begin{cases} 2t & \lvert t\rvert \le 1 \\ 0 & \lvert t\rvert > 1 \end{cases}$	$\begin{cases} 2 & \lvert t\rvert \le 1 \\ 0 & \lvert t\rvert > 1 \end{cases}$	Blake & Zisserman 87
				Geman & Geman 84



Annexe D

Quelques exemples simples d'inférence bayésienne

Dans cette annexe, quelques exemples simples de l'inférence bayésienne sont présentés. L'objective étant plutôt pédagogique.

D.1 Introduction et rappel des notations

Dans ces exemples nous utilisons les notations suivantes :

$$N(m,\lambda) = \sqrt{\frac{\lambda}{2\pi}} \exp\left[-\frac{\lambda}{2}(x-m)^2\right]$$

$$N(\boldsymbol{m},\boldsymbol{\Lambda}) = \frac{|\boldsymbol{\Lambda}|^{1/2}}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^t \boldsymbol{\Lambda}(\boldsymbol{x}-\boldsymbol{m})\right]$$

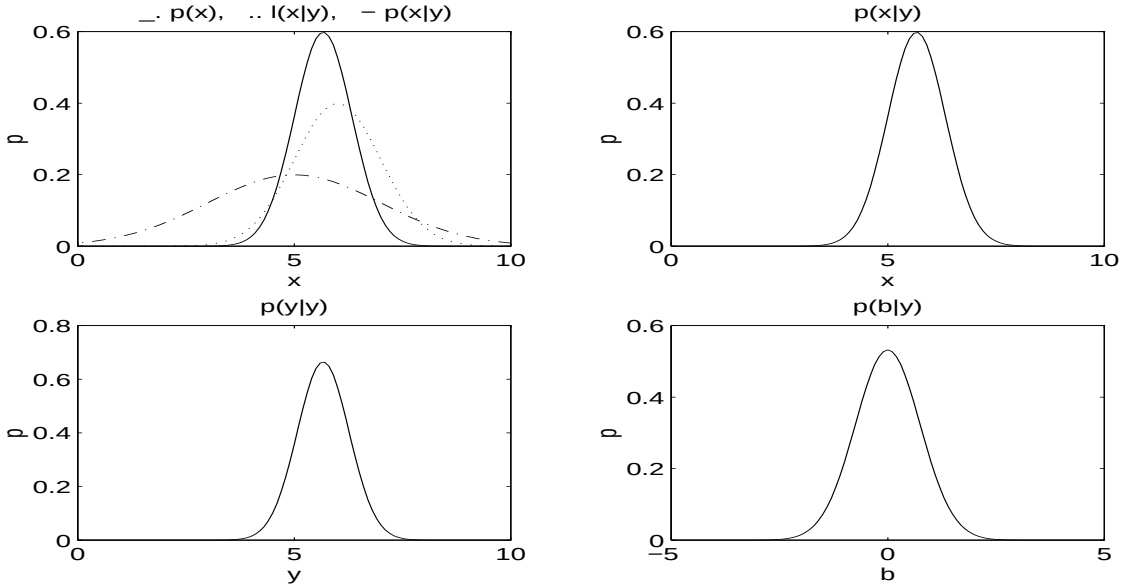
Problème 1 :

$$y = x + b$$

$$\begin{array}{ll}
b|\beta & \sim \mathbf{N}(0,\beta) \\
y|x,\beta & \sim \mathbf{N}(x,\beta) \\
\\
x|x_0,\theta & \sim \mathbf{N}(x_0,\theta) \\
\\
\begin{pmatrix} y \\ x \end{pmatrix} |\beta,x_0,\theta & \sim \mathbf{N}\left(\begin{pmatrix} x_0 \\ x_0 \end{pmatrix}, \begin{pmatrix} \theta+\beta & \theta \\ \theta & \theta \end{pmatrix}^{-1}\right) \\
\\
y|\beta,x_0,\theta & \sim \mathbf{N}\left(x_0, \frac{\theta\beta}{\theta+\beta}\right) \\
\\
x|y,\beta,x_0,\theta & \sim \mathbf{N}(x_m,\theta_m), \\
& \text{avec} \\
& x_m = \frac{1}{\theta_m}(\theta x_0 + \beta y) \\
& \theta_m = \theta + \beta \\
\\
y|y,\beta,x_0,\theta & \sim \mathbf{N}\left(x_m, \frac{\theta_m\beta}{\theta_m+\beta}\right) \\
b|y,\beta,x_0,\theta & \sim \mathbf{N}\left(0, \frac{2\theta_m\beta}{2\theta_m+\beta}\right)
\end{array}$$

$$\begin{array}{llll}
\beta = 0 & \longrightarrow & x_m = x_0, & \theta_m = \theta \\
\theta = 0 & \longrightarrow & x_m = y, & \theta_m = \beta \\
\beta \longrightarrow \infty & \longrightarrow & x_m = y, & \theta_m = \infty \\
\theta \longrightarrow \infty & \longrightarrow & x_m = x_0, & \theta_m = \infty \\
\\
\theta = r\beta & \longrightarrow & x_m = \frac{1}{r+1}(rx_0 + y), & \theta_m = (r+1)\beta
\end{array}$$

Exemple 1 : $y = 6, \quad \beta = 1, \quad x_0 = 5, \quad \theta = .5 \longrightarrow x_m = 5.6666, \quad \theta_m = 1.5$



Problème 2 :

$$y_i = x + b_i, \quad i = 1, \cdots, m \quad \longrightarrow \boldsymbol{y} = x\mathbf{1} + \boldsymbol{b}$$

$$\begin{aligned}
b_i|\beta &\sim \mathbf{N}(0,\beta) \\
\boldsymbol{b}|\beta &\sim \mathbf{N}(\mathbf{0},\beta\boldsymbol{I}) \\
y_i|x,\beta &\sim \mathbf{N}(x,\beta) \\
\boldsymbol{y}|x,\beta &\sim \mathbf{N}(x\mathbf{1},\beta\boldsymbol{I})
\end{aligned}$$

$$x|x_0,\theta \sim \mathbf{N}(x_0,\theta)$$

$$\begin{aligned}
\begin{pmatrix} y_i \\ x \end{pmatrix} |\beta,x_0,\theta &\sim \mathbf{N}\left(\begin{pmatrix} x_0 \\ x_0 \end{pmatrix}, \begin{pmatrix} \theta+\beta & \theta \\ \theta & \theta \end{pmatrix}^{-1} \right) \\
\begin{pmatrix} \boldsymbol{y} \\ x \end{pmatrix} |\beta,x_0,\theta &\sim \mathbf{N}\left(\begin{pmatrix} x_0\mathbf{1} \\ x_0 \end{pmatrix}, \begin{pmatrix} (\theta+\beta)\boldsymbol{I} & \theta\mathbf{1} \\ \theta\mathbf{1}^t & \theta \end{pmatrix}^{-1} \right)
\end{aligned}$$

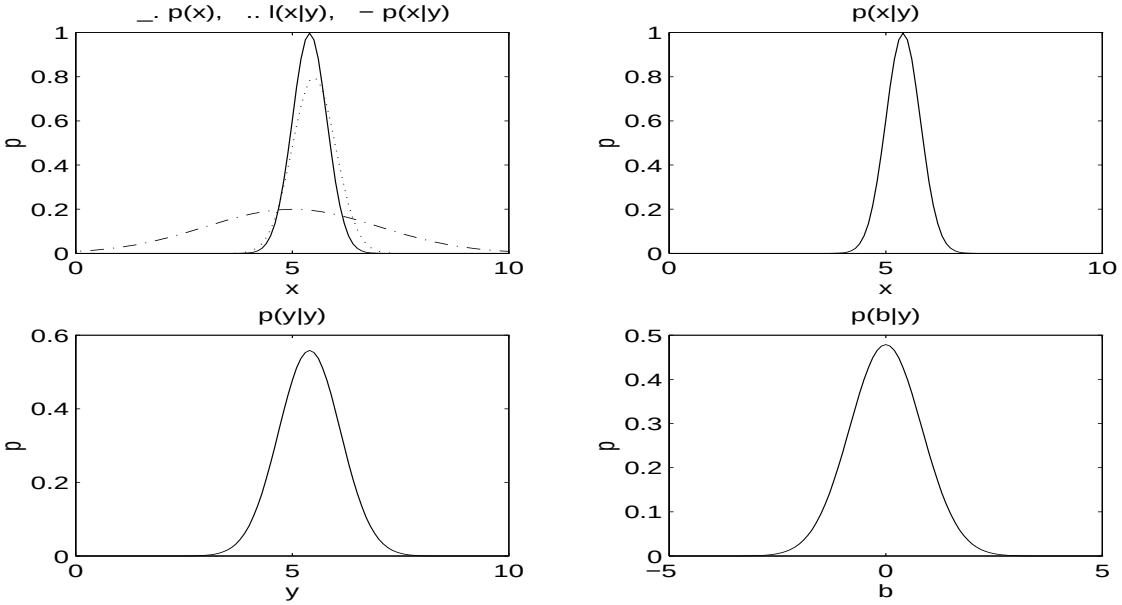
$$\begin{aligned}
y_i|\beta,x_0,\theta &\sim \mathbf{N}(x_0, \tfrac{\theta\beta}{\theta+\beta}) \\
\boldsymbol{y}|\beta,x_0,\theta &\sim \mathbf{N}(x_0\mathbf{1}, \tfrac{\theta\beta}{\theta+\beta}\boldsymbol{I})
\end{aligned}$$

$$\begin{aligned}
x|\boldsymbol{y},\beta,x_0,\theta &\sim \mathbf{N}(x_m,\theta_m), \\
&\text{avec} \\
&x_m = \tfrac{1}{\theta_m}(\theta x_0 + m\beta\bar{y}) \longrightarrow x_m = \tfrac{1}{\theta_m}(\theta_{m-1}x_{m-1} + \beta y_m) \\
&\theta_m = \theta + m\beta \longrightarrow \qquad \theta_m = \theta_{m-1} + \beta \\
&\bar{y} = \tfrac{1}{m}\textstyle\sum_i y_i
\end{aligned}$$

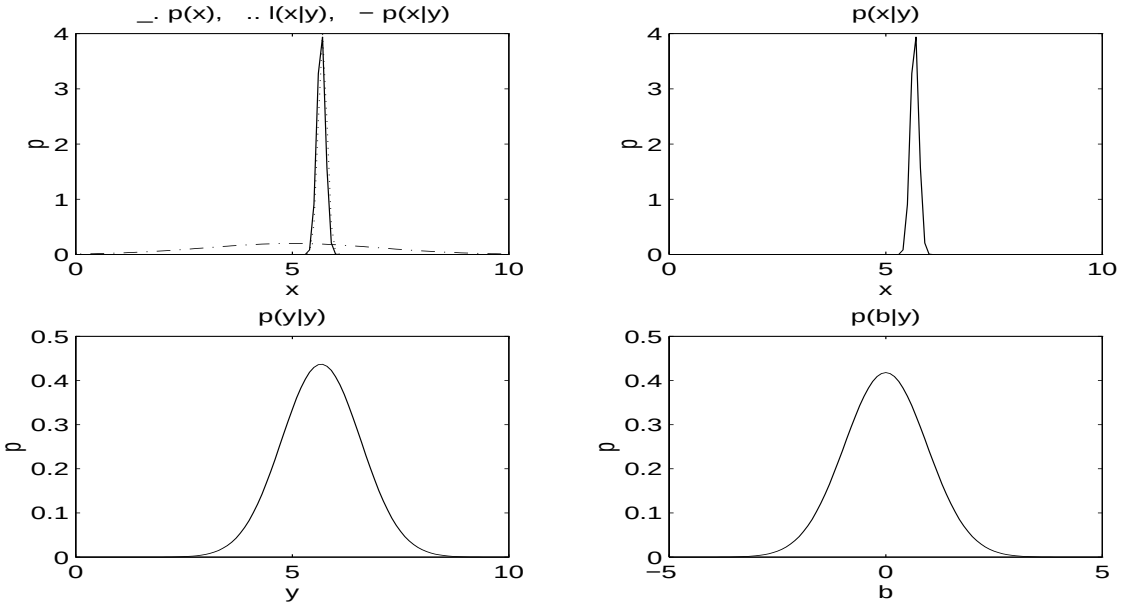
$$\begin{aligned}
y_k|\boldsymbol{y},\beta,x_0,\theta &\sim \mathbf{N}\left(x_m, \tfrac{\theta_m\beta}{\theta_m+\beta}\right) \\
b_k|\boldsymbol{y},\beta,x_0,\theta &\sim \mathbf{N}\left(0, \tfrac{2\theta_m\beta}{2\theta_m+\beta}\right)
\end{aligned}$$

$$\begin{aligned}
\beta = 0 \quad &\longrightarrow \quad x_m = x_0, && \theta_m = \theta \\
\theta = 0 \quad &\longrightarrow \quad x_m = \bar{y}, && \theta_m = m\beta \\
\beta \longrightarrow \infty \quad &\longrightarrow \quad x_m = \bar{y}, && \theta_m = \infty \\
\theta \longrightarrow \infty \quad &\longrightarrow \quad x_m = x_0, && \theta_m = \infty \\
m \longrightarrow \infty \quad &\longrightarrow \quad x_m = \bar{y}, && \theta_m = \infty \\
\\
\theta = r\beta \quad &\longrightarrow \quad x_m = \tfrac{1}{r+m}(rx_0 + m\bar{y}), && \theta_m = (r+m)\beta
\end{aligned}$$

Exemple 2.1 : $\boldsymbol{y} = [6{,}5]$, $m = 2$, $\bar{y} = 5.5$, $\beta = 1$, $x_0 = 5$, $\theta = .5$
$x_m = 5.4$, $\theta_m = 2.5$



Exemple 2.2 : $\boldsymbol{y} = [6{,}5{,}4{,}7{,}3{,}5{,}6{,}8{,}7{,}6]$, $m = 10$, $\bar{y} = 5.7$
$\beta = 1$, $x_0 = 5$, $\theta = .5$ $x_m = 5.667$, $\theta_m = 10.5$



Problème 3 :

$$y = ax + b$$

$$\begin{array}{ll}
b|\beta & \sim \mathbf{N}(0,\beta) \\
y|x,\beta & \sim \mathbf{N}(ax,\beta)
\end{array}$$

$$x|x_0,\theta \quad \sim \mathbf{N}(x_0,\theta)$$

$$\begin{pmatrix} y \\ x \end{pmatrix} |\beta,x_0,\theta \quad \sim \mathbf{N}\left(\begin{pmatrix} ax_0 \\ x_0 \end{pmatrix}, \begin{pmatrix} \theta + a^2\beta & a\theta \\ a\theta & \theta \end{pmatrix}^{-1} \right)$$

$$y|\beta,x_0,\theta \quad \sim \mathbf{N}\left(ax_0, \tfrac{\theta\beta}{\theta+a^2\beta}\right)$$

$$\begin{array}{ll}
x|y,\beta,x_0,\theta & \sim \mathbf{N}(x_m,\theta_m), \\
& \text{avec} \\
& x_m = \frac{1}{\theta_m}(\theta x_0 + a\beta y), \\
& \theta_m = \theta + a^2\beta
\end{array}$$

$$\begin{array}{ll}
y|y,\beta,x_0,\theta & \sim \mathbf{N}\left(ax_m, \frac{\theta_m\beta}{\theta_m+a^2\beta}\right) \\
b|y,\beta,x_0,\theta & \sim \mathbf{N}\left(0, \frac{2\theta_m\beta}{2\theta_m+a^2\beta}\right)
\end{array}$$

$$\begin{array}{lllll}
\beta = 0 & \longrightarrow & x_m = x_0, & & \theta_m = \theta \\
\theta = 0 & \longrightarrow & x_m = y/a, & & \theta_m = a^2\beta \\
\beta \longrightarrow \infty & \longrightarrow & x_m = y/a, & & \theta_m = \infty \\
\theta \longrightarrow \infty & \longrightarrow & x_m = x_0, & & \theta_m = \infty
\end{array}$$

$$\begin{array}{lllll}
\theta = r\beta & \longrightarrow & x_m = \frac{1}{r+a^2}(rx_0 + ay), & & \theta_m = (r + a^2)\beta \\
r \longrightarrow \infty & \longrightarrow & x_m = x_0, & & \theta_m = \infty \\
r = 0 & \longrightarrow & x_m = y/a, & & \theta_m = a^2\beta
\end{array}$$

Problème 4 :

$$y_i = ax + b_i, \quad i = 1, \cdots, m \quad \longrightarrow \boldsymbol{y} = ax\mathbf{1} + \boldsymbol{b}$$

$$\begin{array}{ll}
b_i|\beta & \sim \mathbf{N}(0,\beta) \\
\boldsymbol{b}|\beta & \sim \mathbf{N}(\mathbf{0},\beta\boldsymbol{I}) \\
y_i|x,\beta & \sim \mathbf{N}(ax,\beta) \\
\boldsymbol{y}|x,\beta & \sim \mathbf{N}(ax\mathbf{1},\beta\boldsymbol{I})
\end{array}$$

$$x|x_0,\theta \quad \sim \mathbf{N}(x_0,\theta)$$

$$\begin{array}{ll}
\begin{pmatrix} y_i \\ x \end{pmatrix} |\beta,x_0,\theta & \sim \mathbf{N}\left(\begin{pmatrix} ax_0 \\ x_0 \end{pmatrix}, \begin{pmatrix} a^2\theta+\beta & a\theta \\ a\theta & \theta \end{pmatrix}^{-1} \right) \\
\begin{pmatrix} \boldsymbol{y} \\ x \end{pmatrix} |\beta,x_0,\theta & \sim \mathbf{N}\left(\begin{pmatrix} ax_0\mathbf{1} \\ x_0 \end{pmatrix}, \begin{pmatrix} a^2\theta+\beta\boldsymbol{I} & a\theta\mathbf{1} \\ a\theta\mathbf{1}^t & \theta \end{pmatrix}^{-1} \right)
\end{array}$$

$$\begin{array}{ll}
y|\beta,x_0,\theta & \sim \mathbf{N}\left(ax_0, \frac{\theta\beta}{\theta+a^2\beta}\right) \\
\boldsymbol{y}|\beta,x_0,\theta & \sim \mathbf{N}(ax_0\mathbf{1}, \frac{\theta\beta}{\theta+a^2\beta}\boldsymbol{I})
\end{array}$$

$$\begin{array}{ll}
x|\boldsymbol{y},\beta,x_0,\theta & \sim \mathbf{N}(x_m,\theta_m), \\
& \text{avec} \\
& x_m = \frac{1}{\theta_m}(\theta x_0 + ma\beta\bar{y}) \longrightarrow x_m = \frac{1}{\theta_m}(\theta_{m-1}x_0 + a\beta y_m) \\
& \theta_m = (\theta + ma^2\beta) \longrightarrow \quad \theta_m = \theta_{m-1} + a^2\beta \\
& \bar{y} = \frac{1}{m}\sum_i y_i
\end{array}$$

$$\begin{array}{ll}
y_k|\boldsymbol{y},\beta,x_0,\theta & \sim \mathbf{N}\left(ax_m, \frac{\theta_m\beta}{\theta_m+ma^2\beta}\right) \\
b_k|\boldsymbol{y},\beta,x_0,\theta & \sim \mathbf{N}\left(0, \frac{2\theta_m\beta}{2\theta_m+ma^2\beta}\right)
\end{array}$$

$$\begin{array}{lll}
\beta = 0 & \longrightarrow \quad x_m = x_0, & \theta_m = \theta \\
\theta = 0 & \longrightarrow \quad x_m = \bar{y}/a, & \theta_m = ma^2\beta \\
\beta \longrightarrow \infty & \longrightarrow \quad x_m = \bar{y}/a, & \theta_m = \infty \\
\theta \longrightarrow \infty & \longrightarrow \quad x_m = x_0, & \theta_m = \infty \\
m \longrightarrow \infty & \longrightarrow \quad x_m = \bar{y}/a, & \theta_m = \infty
\end{array}$$

$$\begin{array}{lll}
\theta = r\beta & \longrightarrow \quad x_m = \frac{1}{r+ma^2}(ma\bar{y} + rx_0), & \theta_m = (r+ma^2)\beta \\
r \longrightarrow \infty & \longrightarrow \quad x_m = x_0, & \theta_m = \infty \\
r = 0 & \longrightarrow \quad x_m = \bar{y}/a, & \theta_m = ma^2\beta
\end{array}$$

Problème 5 :

$y_i = x + b_i, \quad i = 1, \cdots, m \quad \longrightarrow \boldsymbol{y} = ax\mathbf{1} + \boldsymbol{b}$

$b_i|\beta \quad \sim \mathbf{N}(0,\beta)$
$y_i|x,\beta \quad \sim \mathbf{N}(x,\beta)$

$x|x_0,\theta \quad \sim \mathbf{N}(x_0,\theta), \qquad \theta = r\beta, \quad r = \theta/\beta$
$\theta|a,b \quad \sim \mathbf{Gam}(a,rb), \quad a = r_0 b, \quad r_0 = a/b, \quad \mathrm{E}[\theta] = \frac{a}{rb} = \frac{r_0}{r}, \quad \mathrm{Var}[\theta] = \frac{a}{(rb)^2}$
$(x,\theta) \quad \sim \mathbf{NGam}(x_0,1,a,rb)$

$x|x_0,r,\beta \quad \sim \mathbf{N}(x_0,r\beta),$
$\beta|a,b \quad \sim \mathbf{Gam}(a,b), \quad a = r_0 b, \quad r_0 = a/b, \quad \mathrm{E}[\beta] = \frac{a}{b} = r_0, \quad \mathrm{Var}[\beta] = \frac{a}{b^2} = \frac{r_0}{b}$
$(x,\beta) \quad \sim \mathbf{NGam}(x_0,r,a,b)$

$x|x_0,r,r_0,a \quad \sim \mathbf{St}\left(x_0, rr_0, 2a\right), \quad \mathrm{E}[x] = x_0, \quad \mathrm{Var}[x] = \frac{2a}{2a-2}\frac{1}{rr_0}$
$y_i|x_0,r,r_0,a \quad \sim \mathbf{St}\left(x_0, \frac{r}{r+1}r_0, 2a\right), \quad \mathrm{E}[y] = x_0, \quad \mathrm{Var}[y] = \frac{2a}{2a-2}\frac{r+1}{rr_0}$

$x|\boldsymbol{y},x_0,r,r_0,a \quad \sim \mathbf{St}\left(x_m,(m+r)r_m,2a_m\right),$
avec
$x_m = \frac{1}{r+m}(rx_0 + m\bar{y}) \longrightarrow x_m = \frac{1}{r+m}[(r+m-1)x_{m-1} + y_m]$
$r_m = a_m/b_m$
$a_m = a + m/2, \quad b_m = b + ms^2/2 + \frac{1}{2}\frac{rm}{r+m}(x_0 - \bar{y})^2$
$\bar{y} = \frac{1}{m}\sum_{i=1}^m y_i, \quad s^2 = \sum_{i=1}^m (y_i - \bar{y})^2$

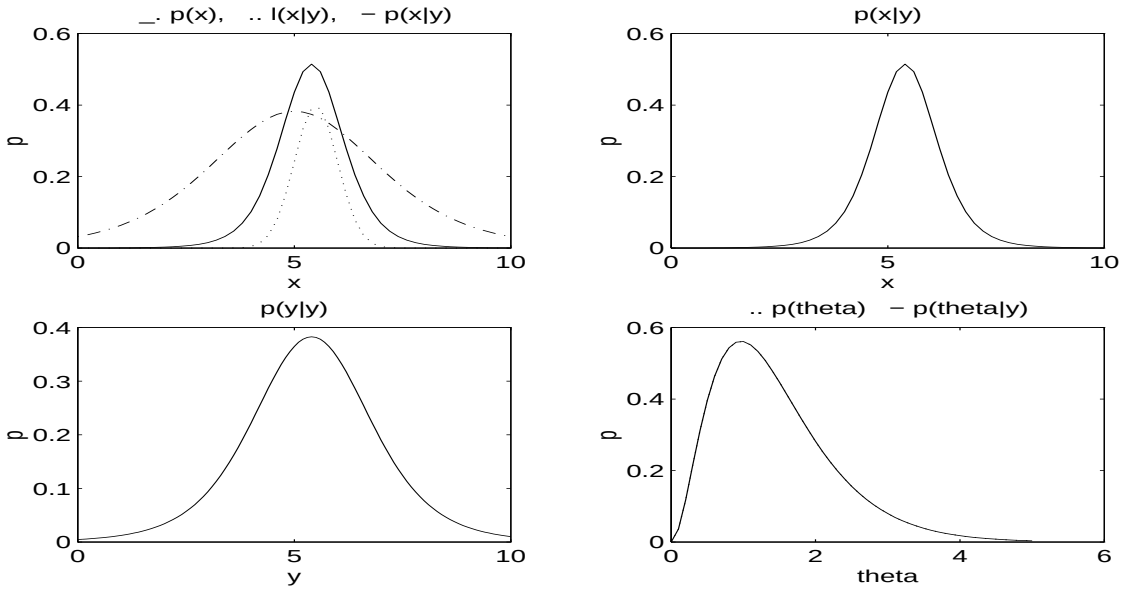
$\mathrm{E}[x|\boldsymbol{y}] = x_m$, si $2a_m > 1$
$\mathrm{Var}[x|\boldsymbol{y}] = \frac{1}{(m+r)r_m}\frac{2a_m}{2a_m-2}$, si $2a_m > 2$
$\mathrm{mode}(x|\boldsymbol{y}) = x_m$

$y_k|\boldsymbol{y},x_0,r,r_0,a \quad \sim \mathbf{St}\left(x_m, \frac{m+r}{m+r+1}r_m, 2a_m\right)$
$\beta|\boldsymbol{y},x_0,r,r_0,a \quad \sim \mathbf{Gam}(a_m,b_m)$
$\theta|\boldsymbol{y},x_0,r,r_0,a \quad \sim \mathbf{Gam}(a_m,rb_m)$
$\mathrm{E}[\beta|\boldsymbol{y}] = \frac{a_m}{b_m}, \quad \mathrm{Var}[\beta|\boldsymbol{y}] = \frac{a_m}{b_m^2}, \quad \mathrm{mode}(\beta|\boldsymbol{y}) = \frac{a_m - 1}{b_m}$
$\mathrm{E}[\theta|\boldsymbol{y}] = \frac{a_m}{rb_m}, \quad \mathrm{Var}[\theta|\boldsymbol{y}] = \frac{a_m}{r^2 b_m^2}, \quad \mathrm{mode}(\theta|\boldsymbol{y}) = \frac{a_m - 1}{rb_m}$

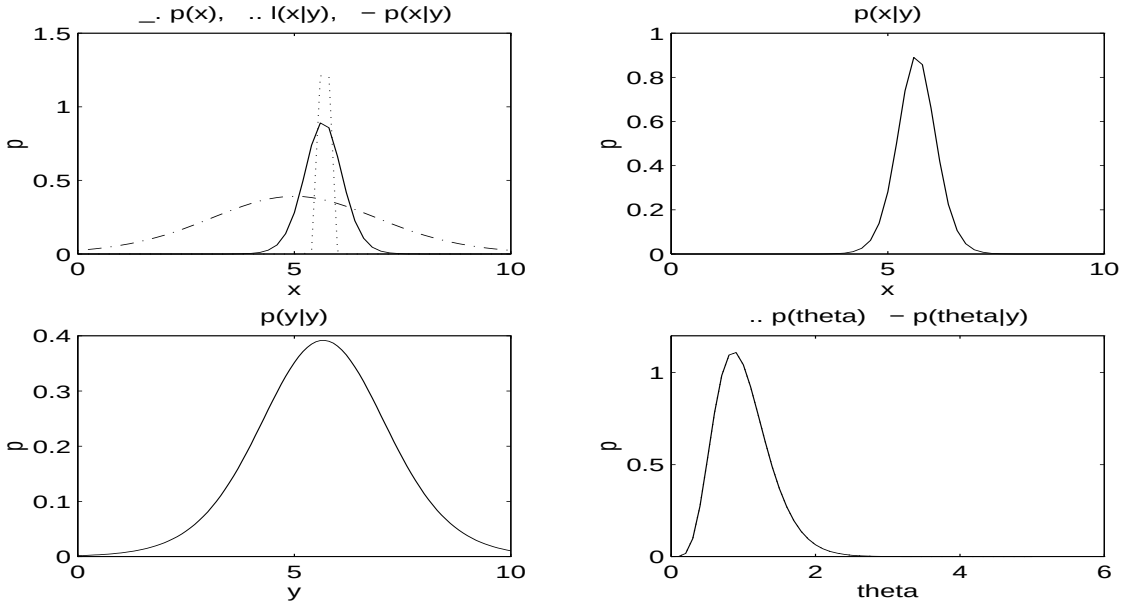
$\theta = 0 \longrightarrow r = 0 \quad \longrightarrow \quad x_m = \bar{y}, \quad b_m = b + ms^2/2$
$\theta \longrightarrow \infty \quad \longrightarrow \quad x_m = x_0, \quad b_m = b + ms^2/2 + \frac{m}{2}(x_0 - \bar{y})^2 = b + \frac{m}{2}[s^2 + (x_0 - \bar{y})^2]$
$m \longrightarrow \infty \quad \longrightarrow \quad x_m = \bar{y}, \quad x|\boldsymbol{y},x_0,r,r_0,a \sim \mathbf{N}(\bar{y},0)$

Cas particulier :
$x_0 = \bar{y} \quad \longrightarrow \quad x_m = \bar{y} \quad a_m = a + m/2, \quad b_m = b + (m/2)s^2,$
$x_0 = \bar{y}, a = \frac{1}{2}, b = \frac{s^2}{2} \quad \longrightarrow \quad x_m = \bar{y}, \quad a_m = \frac{m+1}{2}, b_m = \frac{m+1}{2}s^2,$
$r_m = \frac{1}{s^2},$
$\mathrm{Var}[x|\boldsymbol{y}] = \frac{m+1}{m-1}\frac{1}{m+r}s^2$

Exemple 5.1 : $\boldsymbol{y} = [6,5], \quad m = 2, \quad \bar{y} = 5.5, \quad s^2 = 0.25,$
$\beta = 1, x_0 = 5, \theta = .5, r = .5, \quad a = 2, b = 4, r_0 = .5, \quad \lambda_0 = 0.25, \mathrm{Var}[x] = 6$
$a_m = 3, b_m = 4.15, r_m = 0.7229, \quad x_m = 5.4, \lambda_m = 1.8072, \mathrm{Var}[x|\boldsymbol{y}] = .83$



Exemple 5.2 : $\boldsymbol{y} = [6,5,4,7,3,5,6,8,7,6], \quad m = 10, \quad \bar{y} = 5.7, \quad s^2 = 2.01,$
$\beta = 1, x_0 = 5, \theta = .5, r = .5, \quad a = 2, b = 4, r_0 = .5, \quad \lambda_0 = 0.25, \mathrm{Var}[x] = 4.667$
$a_m = 7, b_m = 13.88, r_m = 0.5042, \quad x_m = 5.667, \lambda_m = 5.294, \mathrm{Var}[x|\boldsymbol{y}] = .2204$



Annexe E

Quelques rappels sur les matrices

Dans cette annexe d'abord nous rappelons un certain nombres de définitions et de relations sur les matrices et le calcul matriciel. Ensuite, les propriétés d'un certain nombres de matrices que l'on trouve souvent en traitement du signal et des images sont étudiées. parmi ces matrices, une place particulière est donnée aux matrices de Toeplitz, de Hankel, circulantes, bloc-Toeplitz et bloc-circulantes.

E.1 Introduction

E.1.1 Définition

Soient $\mathcal{V}$ et $\mathcal{W}$ deux espaces vectoriels sur le même corps (le corps $\mathcal{R}$ des nombres réels ou le corps $\mathcal{C}$ des nombres complexes ou, d'une façon plus générale, le corps $\mathcal{K}$ des scalaires), munis de bases $\{\boldsymbol{e}_i, i = 1, \cdots, m\}$ et $\{\boldsymbol{f}_j, j = 1, \cdots, n\}$ respectivement. Relativement à ces bases, une application linéaire

$$\boldsymbol{A} : \mathcal{V} \longrightarrow \mathcal{W}$$

est représentée par la matrice $\boldsymbol{A}$ à m lignes et n colonnes :

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

les éléments a_{ij} de la matrice $\boldsymbol{A}$ étant définis de façon unique par les relations

$$\boldsymbol{A}\boldsymbol{e}_j = \sum_{j=1}^{N} a_{ij} \boldsymbol{f}_j, \quad i = 1, \cdots, m.$$

Autrement dit, le j-ème vecteur colonne

$$\boldsymbol{a}_{*j} = [\boldsymbol{A}]_{*j} = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$$

de la matrice $\boldsymbol{A}$ représente le vecteur $\boldsymbol{A}\boldsymbol{e}_i$ dans la base $\{\boldsymbol{f}_j, j = 1, \cdots, n\}$. On note

$$\boldsymbol{a}_{i*} = [\boldsymbol{A}]_{i*} = (\, a_{i1} \quad a_{i2} \quad \cdots \quad a_{in} \,)^t$$

le i-ème vecteur ligne de la matrice $\boldsymbol{A}$.

- Une matrice $\boldsymbol{A}$ d'éléments a_{ij} est notée $\boldsymbol{A} = [a_{ij}]$. Une matrice à m lignes et n colonnes est appelée matrice de type $[m,n]$, et on note $\mathcal{A}_{m,n}(\mathcal{K})$ l'espace vectoriel sur le corps $\mathcal{K}$ formé par les matrices de type $[m,n]$ à éléments dans $\mathcal{K}$. Un vecteur colonne est donc une matrice de type $[m,1]$ et un vecteur ligne une matrice de type $[1,n]$.
- étant donnée une matrice $\boldsymbol{A} \in \mathcal{A}_{m,n}(\mathcal{C})$, on note $\boldsymbol{A}^* \in \mathcal{A}_{n,m}(\mathcal{C})$ la matrice adjointe de la matrice $\boldsymbol{A}$, définie de façon unique par les relations

$$< \boldsymbol{Au},\boldsymbol{v} >_m = < \boldsymbol{u},\boldsymbol{A}^*\boldsymbol{v} >_n \,, \quad \forall\, \boldsymbol{u} \in \mathcal{C}^n, \boldsymbol{v} \in \mathcal{C}^m$$

qui entraînent $[\boldsymbol{A}^*]_{ij} = a^*_{ji}$. De la même façon, étant donnée une matrice $\boldsymbol{A} \in \mathcal{A}_{m,n}(\mathcal{R})$, on note $\boldsymbol{A}^t \in \boldsymbol{A}_{n,m}(\mathcal{R})$ la matrice transposée de la matrice $\boldsymbol{A}$, définie de façon unique par les relations

$$< \boldsymbol{Au},\boldsymbol{v} >_m = < \boldsymbol{u},\boldsymbol{A}^*\boldsymbol{v} >_n \; \forall\, \boldsymbol{u} \in \mathcal{R}^n, \boldsymbol{v} \in \mathcal{R}^m$$

qui entraînent $[\boldsymbol{A}^t]_{ij} = a_{ji}$.

E.1.2 Opérations élémentaires

- **Addition :**

$$\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}, \quad [\boldsymbol{C}]_{ij} = [\boldsymbol{A}]_{ij} + [\boldsymbol{B}]_{ij}$$

$$[\boldsymbol{A} + \boldsymbol{B} + \boldsymbol{C}] = [\boldsymbol{C} + \boldsymbol{B} + \boldsymbol{A}] = [\boldsymbol{B} + \boldsymbol{C} + \boldsymbol{A}]$$

- **Multiplication par un scalaire :**

$$\boldsymbol{C} = \lambda \boldsymbol{A}, \quad [\boldsymbol{C}]_{ij} = \lambda [\boldsymbol{A}]_{ij}$$

$$\lambda[\boldsymbol{A} + \boldsymbol{B}] = \lambda \boldsymbol{A} + \lambda \boldsymbol{B}$$

- **Multiplication de deux matrices :**
  À la composition des applications linéaires correspond la multiplication des matrices : si $\boldsymbol{A} = [a_{ij}]$ est une matrice de type $[I,J]$ et $B = [b_{jk}]$ est une matrice de type $[J,K]$, leur produit $\boldsymbol{AB}$ est la matrice $\boldsymbol{C} = [c_{ik}]$ de type $[I,K]$ définie par

$$\boldsymbol{C} = \boldsymbol{AB} \longrightarrow c_{ik} = \sum_{j=1}^{J} a_{ij} b_{jk}$$

  On rappelle que l'on a :

$$[\boldsymbol{A} + \boldsymbol{B}]^t = \boldsymbol{A}^t + \boldsymbol{B}^t, \quad [\boldsymbol{A} + \boldsymbol{B}]^* = \boldsymbol{A}^* + \boldsymbol{B}^*$$

$$[\boldsymbol{AB}]^t = \boldsymbol{B}^t\boldsymbol{A}^t, \quad [\boldsymbol{AB}]^* = \boldsymbol{B}^*\boldsymbol{A}^*$$

$$\lambda[\boldsymbol{AB}] = [\lambda\boldsymbol{A}]\boldsymbol{B} = \boldsymbol{A}[\lambda\boldsymbol{B}]$$

$$\boldsymbol{A}[\boldsymbol{B}+\boldsymbol{C}] = \boldsymbol{AB} + \boldsymbol{AC}$$

$$[\boldsymbol{A}+\boldsymbol{B}]\boldsymbol{C} = \boldsymbol{AC} + \boldsymbol{BC}$$

E.1.3 Matrices élémentaires

- Une matrice de type $[n,n]$ est dite matrice carrée d'ordre n. Si $\boldsymbol{A}$ est une matrice carrée, les éléments a_{ii} sont appelés *éléments diagonaux*, et les éléments $a_{ij}, i \neq j$, sont appelés *éléments hors-diagonaux*.
- Une matrice de type $[n,n]$ est dite *triangulaire supérieure* si $a_{ij} = 0$ pour $i > j$, et *triangulaire inférieure* si $a_{ij} = 0$ pour $i < j$.
- La *matrice unité* est la matrice $\boldsymbol{I} = [d_{ij}]$.
- Une matrice est *diagonale* si $a_{ij} = 0$, pour $i \neq j$ on la note

$$\boldsymbol{A} = \text{diag}\,[a_{ii}] = \text{diag}\,[a_{11}, a_{22}, \cdots, a_{nn}]$$

- Une matrice $\boldsymbol{A}$ est :

symétrique	si	$\boldsymbol{A}$ est réelle et $\boldsymbol{A} = \boldsymbol{A}^t$
hermitienne	si	$\boldsymbol{A} = \boldsymbol{A}^*$
orthogonale	si	$\boldsymbol{A}$ est réelle et $\boldsymbol{A}\boldsymbol{A}^t = \boldsymbol{A}^t\boldsymbol{A} = \boldsymbol{I}$
unitaire	si	$\boldsymbol{A}\boldsymbol{A}^* = \boldsymbol{A}^*\boldsymbol{A} = \boldsymbol{I}$
normale	si	$\boldsymbol{A}\boldsymbol{A}^* = \boldsymbol{A}^*\boldsymbol{A}$

E.1.4 Rang d'une matrice

- Le *rang* d'une matrice $\boldsymbol{A}$ de type $[M,N]$ est la dimension des sous espaces ligne et colonne engendrés respectivement par les vecteurs lignes $\{\boldsymbol{a}_{i*}, i = 1, \cdots, M\}$ et les vecteurs colonnes $\{\boldsymbol{a}_{*j}, j = 1, \cdots, N\}$ de $\boldsymbol{A}$.
- Une matrice est dite de *rang plein* lorsque $\text{rang}\,[\boldsymbol{A}] = \min(M,N)$.
- $\text{rang}\,[\boldsymbol{A}]$ est l'ordre de sa plus grande sous-matrice non singulière.

$$\text{rang}\,[\boldsymbol{A}] = \text{rang}\left[\boldsymbol{A}^t\right]$$

- Deux matrices sont dites équivalentes si elle ont le même rang.
- Si $\boldsymbol{B}$ est une matrice de rang plein régulière, alors les matrices $\boldsymbol{BA}$ ou $\boldsymbol{AB}$ ont le même rang que la matrice $\boldsymbol{A}$:

$$\text{rang}\,[\boldsymbol{AB}] = \text{rang}\,[\boldsymbol{BA}] = \text{rang}\,[\boldsymbol{A}]$$

E.1.5 Matrice inverse

- Une matrice $\boldsymbol{A}$ est inversible s'il existe une matrice, notée $\boldsymbol{A}^{-1}$ et appelée matrice inverse de la matrice $\boldsymbol{A}$, telle que $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$. Dans le cas contraire, on dit que la matrice est singulière.
- Si $\boldsymbol{A}$ et B sont des matrices inversibles on a :

$$[\boldsymbol{A}^t]^{-1} = [\boldsymbol{A}^{-1}]^t, \quad [\boldsymbol{A}^*]^{-1} = [\boldsymbol{A}^{-1}]^* \quad \text{et} \quad [\boldsymbol{A}\boldsymbol{B}]^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$$

- Une CNS pour qu'une matrice carrée de dimensions $[N,N]$ soit inversible (ou régulière) est qu'elle soit de rang plein, c'est à dire rang $[\boldsymbol{A}] = N$.

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, \qquad \boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{y}$$

$$(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C})^{-1} = \boldsymbol{C}^{-1}\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$$

E.1.6 Déterminant d'une matrice

$$\det[\boldsymbol{A}] = |\boldsymbol{A}| = \sum_{j=1}^{N} [\boldsymbol{A}]_{ij} \boldsymbol{A}_{ij}^{\ddagger}$$

$$\det[\boldsymbol{A}] = \det\left[\boldsymbol{A}^t\right]$$

$$\det[\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}] = \det[\boldsymbol{A}] \det[\boldsymbol{B}] \det[\boldsymbol{C}]$$

E.1.7 Matrice singulière

$\boldsymbol{A}$ est singulière si $\det[\boldsymbol{A}] = 0$.

$$\boldsymbol{A}\boldsymbol{x} = 0, \quad \boldsymbol{A}^t\boldsymbol{x} = \boldsymbol{0} \quad \forall \quad \boldsymbol{x} \neq \boldsymbol{0}.$$

E.1.8 Trace d'une matrice

La trace d'une matrice est définie par

$$\mathrm{tr}[\boldsymbol{A}] = \sum_{i=1}^{N} a_{ii}$$

On a les relations suivantes:

$$\mathrm{tr}[\boldsymbol{A}] = \sum_{i=1}^{N} a_{ii} = \sum_{i=1}^{N} \lambda_i$$

$$\mathrm{tr}[\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}] = \mathrm{tr}[\boldsymbol{C}\boldsymbol{A}\boldsymbol{B}] = \mathrm{tr}[\boldsymbol{B}\boldsymbol{C}\boldsymbol{A}]$$

$$\mathrm{tr}[\boldsymbol{A} + \boldsymbol{B} + \boldsymbol{C}] = \mathrm{tr}[\boldsymbol{A}] + \mathrm{tr}[\boldsymbol{B}] + \mathrm{tr}[\boldsymbol{C}]$$

La trace est invariante dans un changement de base.

E.1.9 Vecteurs propres et valeurs propre d'une matrice

- Un vecteur $\boldsymbol{u}$ est dit un *vecteur propre* de la matrice $\boldsymbol{A}$ si

$$\boldsymbol{A}\boldsymbol{u} = \lambda \boldsymbol{u};$$

  λ est alors une *valeurs propre* associée au vecteur propre $\boldsymbol{u}$.
- Pour une matrice carrée on définit :
  polynôme caractéristique: $f(\lambda) = \det[\boldsymbol{A} - \lambda \boldsymbol{I}]$
  Matrice caractéristique: $\boldsymbol{A} - \lambda \boldsymbol{I}$
  Équation caractéristique: $f(\lambda) = \det[\boldsymbol{A} - \lambda \boldsymbol{I} = 0]$
- Les *valeurs propres* $\{\lambda_i = \lambda_i(\boldsymbol{A}), i = 1, \cdots, n\}$ d'une matrice $\boldsymbol{A}$ d'ordre n sont les n racines, réelles ou complexes, distinctes ou confondues, du polynôme caractéristique

$$f(\lambda) = \det[\boldsymbol{A} - \lambda \boldsymbol{I}] = (-\lambda)^n + \sum_{i=1}^{N} a_{ij} (-\lambda)^{n-1} + \cdots + \det[\boldsymbol{A}] = 0$$

- Les v.p. d'une matrice ont une interprétation géométrique simple: Soit $\mathcal{E}^n$ une espace linéaire complexe à n dimensions et $\{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n\}$ une base dans cet espace. A la matrice $\boldsymbol{A}$ et à cette base on peut associer un opérateur linéaire sous la forme de

$$\boldsymbol{A}\,\boldsymbol{e}_j = \sum_{i=1}^{N} a_{ji}\,\boldsymbol{e}_i, \quad j = 1, \cdots, n$$

  Ainsi les λ_i sont les v.p. de l'opérateur $\boldsymbol{A}$, c'est à dire qu'il existe un vecteur

$$\boldsymbol{u} = u_1 e_1 + \cdots + u_i e_i + \cdots + u_n e_n \neq \boldsymbol{0},$$

  appartenant à l'espace $\mathcal{E}^n$ de façon à avoir $\boldsymbol{A}\boldsymbol{u} = \lambda \boldsymbol{u}$. Ce vecteur est appelé le vecteur propre de l'opérateur $\boldsymbol{A}$ correspondant à la v.p. λ.
- On montre facilement que

$$\det[\boldsymbol{A}] = \prod_{i=1}^{n} \lambda_i$$

  Une matrice est régulière ssi $\lambda_i \neq 0, i = 1, \cdots, n$.
- Le *spectre* de la matrice $\boldsymbol{A}$ est le sous-ensemble

$$\mathrm{sp}[\boldsymbol{A}] = \cap_{j=1}^{n} \{\lambda_i(\boldsymbol{A})\}$$

  du plan complexe.
- Le *rayon spectral* d'une matrice $\boldsymbol{A}$ est le nombre $\rho(\boldsymbol{A}) \geq 0$ défini par

$$\rho(\boldsymbol{A}) = \max_i \{|\lambda_i(\boldsymbol{A})|, \quad i = 1, \cdots, n\}$$

- A tout valeur propre λ d'une matrice $\boldsymbol{A}$ est associé (au moins) un vecteur propre $\boldsymbol{v} \neq 0$ tel que: $\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}$.
- Si $\lambda \in \mathrm{sp}[\boldsymbol{A}]$, le sous-espace vectoriel $\{\boldsymbol{v} \in \mathcal{V}; \boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}\}$ est appelé sous-espace propre correspondant à la valeur propre λ.

E.1.10 Valeurs singulières d'une matrice

Les valeurs singulières de la matrice $\boldsymbol{A}$ de dimensions $[M,N]$ sont données par

$$\mu_i^2 \boldsymbol{A}\boldsymbol{A}^* \boldsymbol{u}_i = \boldsymbol{u}_i, \quad i = 1,\cdots,r$$

$$\mu_i^2 \boldsymbol{A}^* \boldsymbol{A} \boldsymbol{v}_i = \boldsymbol{v}_i, \quad i = 1,\cdots,r$$

où $r = \text{rang}\,[\boldsymbol{A}]$. Si on définit $\lambda_i = \frac{1}{(\mu_i)^2}, \quad i = 1,\cdots,r$ en supposant $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$, et si on les complète par des zéros, c'est à dire $\{\lambda_i = 0, \quad i = r+1,\cdots,\max(M,N)\}$, alors on peut associer deux systèmes de v.p. $\{\boldsymbol{u}_i, \quad i = 1,\cdots,M\}$ et $\{\boldsymbol{v}_j, \quad j = 1,\cdots,N\}$ aux $\{\lambda_i\}$, tels que

$$\boldsymbol{A}\boldsymbol{u}_i = \lambda_i \boldsymbol{v}_i$$

$$\boldsymbol{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{u}_j$$

$$\boldsymbol{A}\boldsymbol{A}^* \boldsymbol{u}_i = \lambda_i^2 \boldsymbol{u}_i, \quad i = 1,\cdots,M$$

$$\boldsymbol{A}^*\boldsymbol{A} \boldsymbol{v}_i = \lambda_i^2 \boldsymbol{v}_i, \quad j = 1,\cdots,N.$$

λ_i^2,$\boldsymbol{u}_i$ et λ_i^2,$\boldsymbol{v}_i$ constituent les ystèmes caractéristiques de $\boldsymbol{A}\boldsymbol{A}^*$ et $\boldsymbol{A}^*\boldsymbol{A}$, respectivement. La matrice $\boldsymbol{A}$ peut se factoriser en le produit

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^t$$

où $\boldsymbol{U}$ est la matrice dont les colonnes sont les v.p. $\boldsymbol{u}_j$, $\boldsymbol{V}$ est la matrice dont les colonnes sont les v.p. $\boldsymbol{v}_i$ et $\boldsymbol{\Lambda}$ est la matrice diagonale formée de v.p. λ_i.

Le rang r de la matrice normale $\boldsymbol{A}^*\boldsymbol{A}$ est inférieur ou égale à $\inf(M,N)$. On complète les matrices $\boldsymbol{U}$ et $\boldsymbol{V}$ par les $(M-r)$ v.p. $\boldsymbol{u}_i$ et les $(N-r)$ v.p. $\boldsymbol{v}_i$ qui correspondent aux v.p. nulles, de façon à obtenir des bases complètes respectivement de Y et de X.

E.1.11 Matrices et les opérations linéaires

Si commençant par une base $\{\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n\}$, on introduit dans un espace euclidienne $\mathcal{E}^n$ un produit scalaire sous forme de

$$< \boldsymbol{x},\boldsymbol{y} >= x_1 y_1^* + \cdots + x_i y_i^* + \cdots + x_n y_n^*$$

alors une matrice hermitienne $\boldsymbol{A}$: $([\boldsymbol{A}]_{ji} = [\boldsymbol{A}]_{ij}^*)$ induit sur cette base un opérateur hermitienne, c'est à dire

$$< \boldsymbol{A}\boldsymbol{x},\boldsymbol{y} >=< \boldsymbol{x},\boldsymbol{A}\boldsymbol{y} > \quad \forall \boldsymbol{x},\boldsymbol{y} \in \mathcal{E}^n$$

– Le produit scalaire a les propriétés suivantes:

$$\forall \boldsymbol{x},\boldsymbol{x}_1,\boldsymbol{x}_1,\boldsymbol{x},\boldsymbol{y} \quad \text{et} \quad \boldsymbol{y}_2 \in \mathcal{E}^n$$

$$< \boldsymbol{x},\boldsymbol{x} > > 0 \quad \forall \boldsymbol{x} \neq 0$$

$$< (\boldsymbol{x}_1 + \boldsymbol{x}_2),\boldsymbol{x} >=< \boldsymbol{x}_1,\boldsymbol{y} > + < \boldsymbol{x}_2,\boldsymbol{y} >$$

$$< \alpha\boldsymbol{x},\boldsymbol{y} >= \alpha < \boldsymbol{x},\boldsymbol{y} >$$

$$< \boldsymbol{x},\boldsymbol{y} >=< \boldsymbol{y},\boldsymbol{x} >^*$$

- Si $\boldsymbol{A}$ est une matrice hermitienne on a

$$< \boldsymbol{A}\boldsymbol{x},\boldsymbol{x} >= \sum_i a_{ij} x_i x_j^*$$

Un telle forme est entièrement déterminée par une matrice hermitienne $\boldsymbol{A}$. L'ordre N et le rang de la matrice $\boldsymbol{A}$ sont appelés respectivement l'ordre et le rang de la forme $< \boldsymbol{A}\boldsymbol{x},\boldsymbol{x} >$.

- Une tache importante de la théorie des formes hermitiennes est la réduction de la forme hermitienne à la somme de carrées suivantes

$$< \boldsymbol{A}\boldsymbol{x},\boldsymbol{x} >= \sum_k \lambda_k |x_k|^2$$

où $x_k = c_{1k}x_1 + c_{2k}x_2 + \cdots$
Si r est le rang de $\boldsymbol{A}$ et x_i sont linéairement indépendantes, alors seul r coefficients λ_k dans cette équation son non-nulle.

- Soient $\mathcal{V}$ un espace vectoriel de dimension finie n, et $\mathcal{A} : \mathcal{V} \longrightarrow \mathcal{V}$ une application linéaire, représentée par une matrice carrée $\boldsymbol{A}$ relativement à une base $\{\boldsymbol{e}_i, i = 1,\cdots,m\}$. Cette même application, relativement à une autre base $\{\boldsymbol{f}_j, j = 1,\cdots,n\}$, est représentée par la matrice

$$\boldsymbol{B} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P},$$

où $\boldsymbol{P}$ est la matrice inversible dont le j-ème vecteur colonne est formé des composantes du vecteur $\boldsymbol{f}_j$ dans la base $\{\boldsymbol{e}_i, i = 1,\cdots,m\}$. La matrice $\boldsymbol{P}$ est appelée *matrice de passage de la base* $\{\boldsymbol{e}_j, j = 1,\cdots,n\}$ dans la base $\{\boldsymbol{f}_j, j = 1,\cdots,n\}$. Les deux matrices $\boldsymbol{A}$ et $\boldsymbol{B}$ sont dites *semblables* s'il existe une matrice inversible $\boldsymbol{P}$ telle que $\boldsymbol{B} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$.

- Une même application linéaire $\boldsymbol{A}$ étant représentée par différentes matrices selon la base choisie, le problème de la réduction des matrices consiste alors de trouver une base vis-à-vis de laquelle la matrice représentant l'application soit aussi simple que possible. Le cas le plus favorable est celui où il existe une matrice inversible $\boldsymbol{P}$ telle que la matrice $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ soit diagonale. Dans ce cas on dit que la matrice $\boldsymbol{A}$ est *diagonalisable*. Les éléments diagonaux de la matrice $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ sont alors les valeurs propres $\lambda_1,\lambda_2,\cdots,\lambda_n$ de la matrice $\boldsymbol{A}$, et le j-ème vecteur colonne $\boldsymbol{p}_j$ de la matrice $\boldsymbol{P}$ est formé des composantes d'un vecteur propre correspondant à λ_j, et on a :

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \mathrm{diag}\,[\lambda_i] \longrightarrow \boldsymbol{A}\boldsymbol{p}_j = \lambda_j \boldsymbol{p}_j, \quad j = 1,\cdots,n$$

Autrement dit, une matrice est diagonalisable ssi il existe une base de vecteur propres.

- Étant donnée une matrice carrée $\boldsymbol{A}$, il existe une matrice unitaire $\boldsymbol{U}$ telle que la matrice $\boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U}$ soit triangulaire.
- Étant donnée une matrice normale $\boldsymbol{A}$, il existe une matrice unitaire $\boldsymbol{U}$ telle que la matrice $\boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U}$ soit diagonale.
- Étant donnée une matrice symétrique $\boldsymbol{A}$, il existe une matrice orthogonale $\boldsymbol{O}$ telle que la matrice $\boldsymbol{O}^{-1}\boldsymbol{A}\boldsymbol{O}$ soit diagonale.
- On appelle *valeurs singulières* d'une matrice $\boldsymbol{A}$ carrée les racines carrées positives des valeurs propres de la matrice hermitienne $\boldsymbol{A}^*\boldsymbol{A}$ (ou $\boldsymbol{A}^t\boldsymbol{A}$ si la matrice $\boldsymbol{A}$ est réelle). Les valeurs singulières d'une matrice sont toujours ≥ 0. Elle sont toutes > 0 ssi la matrice $\boldsymbol{A}$ est inversible.
- Deux matrices $\boldsymbol{A}$ et $\boldsymbol{B}$ de type $[m,n]$ sont dites *équivalentes* s'il existe une matrice inversible $\boldsymbol{Q}$ d'ordre m et une matrice inversible $\boldsymbol{P}$ d'ordre n telle que $\boldsymbol{B} = \boldsymbol{Q}\boldsymbol{A}\boldsymbol{P}$.
- Pour une matrice réelle carrée $\boldsymbol{A}$, il existe deux matrices orthogonales $\boldsymbol{U}$ et $\boldsymbol{V}$ telle que

$$\boldsymbol{U}^t\boldsymbol{A}\boldsymbol{U} = \mathrm{diag}\,[\mu_i]\,.$$

- Pour une matrice complexe carrée $\boldsymbol{A}$, il existe deux matrices unitaires $\boldsymbol{U}$ et $\boldsymbol{V}$ telle que

$$\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U} = \mathrm{diag}\,[\mu_i]\,.$$

- Dans les deux cas, les nombres $\mu_i \geq 0$ sont les valeurs singulières de la matrice $\boldsymbol{A}$.
- Toutes les valeurs propres d'une matrice hermitienne sont réelles. Toute matrice hermitienne est diagonalisable et la matrice de passage sont unitaires.
- Soit $\boldsymbol{A}$ une matrice carrée représentant une application linéaire d'un espace $\mathcal{V}$ sur le corps $\mathcal{C}$, muni de son produit *canonique*.
- Le *quotient de Rayleigh* de la matrice $\boldsymbol{A}$ est l'application :

$$R_A : \{\boldsymbol{V} - \{0\}\} \longrightarrow \mathcal{C}$$

définie par :

$$R_A(\boldsymbol{v}) = \frac{<\boldsymbol{v},\boldsymbol{A}\boldsymbol{v}>}{<\boldsymbol{v},\boldsymbol{v}>} = \frac{\boldsymbol{v}^*\boldsymbol{A}\boldsymbol{v}}{\boldsymbol{v}^*\boldsymbol{v}}, \quad \boldsymbol{v} \neq 0$$

E.1.12 Décomposition tronquée en valeurs singulière

Si les λ_i décroissent rapidement alors on peut approcher la matrice $\boldsymbol{A}$ par une décomposition tronquée en valeurs singulières ($k < r = \mathrm{rang}\,[\boldsymbol{A}]$) :

$$\boldsymbol{A} = \sum_{i=1}^{k} \lambda_i \boldsymbol{u}_i \boldsymbol{v}_i^* + \boldsymbol{E}_k$$

où $\boldsymbol{E}_k$ est la matrice d'erreur de l'approximation. Si on définit J_k comme l'énergie de l'erreur par

$$J_k = \sum_{i=1}^{N}\sum_{j=1}^{N} |[\boldsymbol{E}_k]_{ij}|^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} [\boldsymbol{E}_k]_{ij}[\boldsymbol{E}_k]_{ij}^*$$

alors on montre facilement que l'on a

$$J_k = \sum_{i=k+1}^{r} \lambda_i$$

Si on définit α_k par

$$\alpha_k = \frac{J_k}{J_0} = 1 - \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{r} \lambda_i}$$

alors $0 \leq \alpha_k \leq 1$, $\alpha_0 = 1$, et $\alpha_r = 0$.

E.1.13 Inverse généralisée d'une matrice

Considérons l'équation

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$

où $\boldsymbol{A}$ est une matrice de dimension [m,n].

- **Définition de Moore :**
  Une matrice $\boldsymbol{G}$ de dimension [n,m] est dite inverse généralisée de la matrice $\boldsymbol{A}$ si

  $$\boldsymbol{A}\boldsymbol{G} = \boldsymbol{P}_A \quad \text{et} \quad \boldsymbol{G}\boldsymbol{A} = \boldsymbol{P}_G$$

  où $\boldsymbol{P}_A$ est l'opérateur (matrice) de projection des vecteurs de l'espace $\mathcal{E}^m$ sur $\mu(\boldsymbol{A})$ et $\boldsymbol{P}_G$ est l'opérateur (matrice) de projection des vecteurs de l'espace $\mathcal{E}^n$ sur $\mu(\boldsymbol{G})$.
- **Définition de Penrose :**
  Une matrice $\boldsymbol{G}$ de dimension [n,m] est dite inverse généralisée de la matrice $\boldsymbol{A}$ si

  $$\boldsymbol{A}\boldsymbol{G}\boldsymbol{A} = \boldsymbol{A}, \quad (\boldsymbol{A}\boldsymbol{G})^* = \boldsymbol{A}\boldsymbol{G}, \quad \boldsymbol{G}\boldsymbol{A}\boldsymbol{G} = \boldsymbol{G} \quad \text{et} \quad (\boldsymbol{G}\boldsymbol{A})^* = \boldsymbol{G}\boldsymbol{A}$$

- Ces définitions sont identiques si les normes associées aux espaces $\mathcal{E}^m$ et $\mathcal{E}^n$ sont du type euclidien: $\|\boldsymbol{x}\| = < \boldsymbol{x}^*, \boldsymbol{x} >^{1/2}$.

E.2 Quelques matrices célèbres

Matrices identité :

$$[\boldsymbol{I}]_{ij} = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & \text{ailleurs} \end{cases}, \quad j = 1, \cdots, N$$

$$\boldsymbol{I}^t\boldsymbol{I} = \boldsymbol{I}\boldsymbol{I}^t = \boldsymbol{I}, \quad \boldsymbol{I}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{I} = \boldsymbol{A}$$

Matrices diagonales :

$$\boldsymbol{D} = \text{diag}\,[d_1, \cdots, d_N] \longrightarrow [\boldsymbol{D}]_{ij} = d_i\,\delta_{ij}$$

$$\boldsymbol{D}^{-1} = \text{diag}\left[\frac{1}{d_1}, \cdots, \frac{1}{d_N}\right]$$

Matrices d'inversion d'ordre

$$\boldsymbol{J} = \begin{pmatrix} . & . & . & . & 1 \\ . & 0 & . & 1 & . \\ . & . & 1 & . & . \\ . & 1 & . & 0 & . \\ 1 & . & . & . & \end{pmatrix} \quad [\boldsymbol{J}]_{ij} = \begin{cases} 1 & i = N - j + 1, \quad j = 1, \cdots, N \\ 0 & \text{ailleurs} \end{cases}$$

$$[\boldsymbol{J}]^{2k} = \boldsymbol{I}, \quad [\boldsymbol{J}]^{2k+1} = \boldsymbol{J}, \quad \boldsymbol{J}^{-1} = \boldsymbol{J}$$

- $\boldsymbol{AJ}$ inverse l'ordre des éléments des lignes de la matrice $\boldsymbol{A}$.
- $\boldsymbol{JA}$ inverse l'ordre des éléments des colonnes de la matrice $\boldsymbol{A}$.
- $\boldsymbol{JAJ}$ inverse l'ordre des éléments des lignes et des colonnes de la matrice $\boldsymbol{A}$.
- $\boldsymbol{Jh}$ inverse l'ordre des éléments du vecteur $\boldsymbol{h}$.
- Si $\boldsymbol{h}$ est un vecteur symétrique on a $\boldsymbol{Jh} = \boldsymbol{h}$.
- Si $\boldsymbol{h}$ est un vecteur anti-symétrique on a $\boldsymbol{Jh} = -\boldsymbol{h}$.

Matrices symétriques :

$$[\boldsymbol{S}]_{ij} = [\boldsymbol{S}]_{ji}, \quad i,j = 1, \cdots, N, \quad \boldsymbol{S}^t = \boldsymbol{S}$$

- Si $\boldsymbol{S}$ est une matrice symétrique avec des éléments réels, alors ses v.p. sont réelles et il existe au moins une matrice $\boldsymbol{R}$ permettant d'avoir: $\boldsymbol{R}^t\boldsymbol{SR} = \boldsymbol{D}$, $\boldsymbol{S} = \boldsymbol{RDR}^t$, où $\boldsymbol{D}$ est une matrice diagonale.
- Étant donnée une matrice symétrique $\boldsymbol{A}$, il existe une matrice orthogonale $\boldsymbol{O}$ telle que la matrice $\boldsymbol{O}^{-1}\boldsymbol{AO}$ soit diagonale.

Matrices anti-symétriques :

$$[\boldsymbol{R}]_{ij} = -[\boldsymbol{R}]_{ji}, \quad i,j = 1, \cdots, N, \quad i \neq j \quad \boldsymbol{R}^t = -\boldsymbol{R}$$

Matrices symétriques par rapport à la deuxième diagonale :

$$[\boldsymbol{S}]_{ij} = [\boldsymbol{S}]_{(N-j+1)(N-i+1)}, \quad i,j = 1, \cdots, N, \quad \boldsymbol{JSJ} = \boldsymbol{S}^t, \quad \boldsymbol{JS}^t\boldsymbol{J} = \boldsymbol{S}$$

Matrices anti-symétriques par rapport à la deuxième diagonale :

$$[\boldsymbol{R}]_{ij} = -[\boldsymbol{R}]_{(N-i+1)(N-j+1)}, \quad i,j = 1, \cdots, N, \quad i \neq j, \quad \boldsymbol{JSJ} = -\boldsymbol{S}^t$$

Matrices hermitienne :

$$[\boldsymbol{H}]_{ij} = [\boldsymbol{H}^*]_{ji}, \quad i,j = 1, \cdots, N, \quad \boldsymbol{H}^* = \boldsymbol{H}^*$$

$[\boldsymbol{H}]_{ii}$ sont réels. Si $\boldsymbol{H}$ est une matrice hermitienne $\boldsymbol{H}^{-1}$ (à condition qu'elle existe) est aussi hermitienne, et on a $\boldsymbol{x}^t\boldsymbol{Hx} = \alpha\|\boldsymbol{x}\|^2$ est réel. $\boldsymbol{A}$ toute matrice hermitienne on peut toujours associer un système de vecteurs propres orthonormés complet. Les v.p. de $\boldsymbol{H}$ sont réelles et ses vecteurs propres sont orthogonaux.

Matrices anti-hermitiennes :

$$[\boldsymbol{K}]_{ij} = -[\boldsymbol{K}^*]_{ji}, \quad i,j = 1,\cdots,N, \quad \boldsymbol{K}^* = -\boldsymbol{K}^*$$

Si $\boldsymbol{K}$ est une matrice anti-hermitienne $\boldsymbol{K}^{-1}$ (à condition qu'elle existe) est aussi anti-hermitienne, et on a $\boldsymbol{x}^t\boldsymbol{K}\boldsymbol{x} = -j\alpha\|\boldsymbol{x}\|^2$ est imaginaire Les v.p. de $\boldsymbol{K}$ sont imaginaires et ses vecteurs propres sont orthogonaux. Une matrice quelconque $\boldsymbol{A}$ peut toujours être décomposée de la façon suivante :

$$\boldsymbol{A} = \frac{1}{2}[\boldsymbol{A} + \boldsymbol{A}^t] + \frac{1}{2}[\boldsymbol{A} - \boldsymbol{A}^t] = \frac{1}{2}[\boldsymbol{H} + \boldsymbol{K}].$$

Matrices idempotentes :

$$\boldsymbol{Q}^t\boldsymbol{Q} = \boldsymbol{Q}$$

Si $\boldsymbol{A}$ est une matrice idempotente et symétrique on a:

$$\boldsymbol{A}^2 = \boldsymbol{A}^t\boldsymbol{A} = \boldsymbol{A}$$

$$\boldsymbol{A}^2\boldsymbol{x}_i = \boldsymbol{A}\boldsymbol{x}_i = \lambda_i\boldsymbol{x}_i = \boldsymbol{A}(\lambda_i\boldsymbol{x}_i) = \lambda_i^2\boldsymbol{x}_i \longrightarrow \lambda_i = 0 \text{ ou } 1.$$

Matrices orthogonales :

$$\boldsymbol{O}^t\boldsymbol{O} = \boldsymbol{O}\boldsymbol{O}^t = \boldsymbol{O}\boldsymbol{O}^{-1} = \boldsymbol{O}^{-1}\boldsymbol{O} = \boldsymbol{I}$$

Deux matrices $\boldsymbol{A}$ et $\boldsymbol{B}$ sont *similaires* s'il existe une matrice $\boldsymbol{C}$ telle que

$$\boldsymbol{B} = \boldsymbol{C}^{-1}\boldsymbol{A}\boldsymbol{C}$$

On a alors : $\boldsymbol{A} = \boldsymbol{C}^{-1}\boldsymbol{B}\boldsymbol{C}$. Deux matrices $\boldsymbol{A}$ et $\boldsymbol{B}$ sont *orthogonalement similaires* ou *congruentes* s'il existe une matrice orthogonale $\boldsymbol{O}$ telle que $\boldsymbol{B} = \boldsymbol{O}^{-1}\boldsymbol{A}\boldsymbol{O}$. On a alors $\boldsymbol{A} = \boldsymbol{O}^{-1}\boldsymbol{B}\boldsymbol{O}$.

Matrices de décalages :

$\boldsymbol{Z}$ est une matrices de décalage vers le bas (ou vers la droite) si :

$$\boldsymbol{Z} = \begin{pmatrix} 0 & . & . & . & 0 \\ 1 & 0 & . & . & 0 \\ 0 & 1 & 0 & . & 0 \\ . & 0 & 1 & 0 & 0 \\ . & . & 0 & 1 & 0 \end{pmatrix}, \quad [\boldsymbol{Z}]_{ij} = \begin{cases} 1 & i = j+1 \\ 0 & \text{ailleurs} \end{cases}, \quad j = 1,\cdots N-1$$

$\boldsymbol{Z}^t$ est une matrices de décalage vers le haut.

Si $\boldsymbol{Z}$ est une matrice de décalage vers le haut on a:

$$\begin{array}{ll} \boldsymbol{B} = \boldsymbol{Z}^\alpha\boldsymbol{A}\boldsymbol{Z}^\beta & [\boldsymbol{B}]_{mn} = [\boldsymbol{A}]_{m+\alpha n-\beta} \\ \boldsymbol{B} = \boldsymbol{Z}^\alpha\boldsymbol{A}\boldsymbol{Z}^{t\beta} & [\boldsymbol{B}]_{mn} = [\boldsymbol{A}]_{m+\alpha n+\beta} \\ \boldsymbol{B} = \boldsymbol{Z}^{t\alpha}\boldsymbol{A}\boldsymbol{Z}^\beta & [\boldsymbol{B}]_{mn} = [\boldsymbol{A}]_{m-\alpha n-\beta} \\ \boldsymbol{B} = \boldsymbol{Z}^{t\alpha}\boldsymbol{A}\boldsymbol{Z}^{t\beta} & [\boldsymbol{B}]_{mn} = [\boldsymbol{A}]_{m-\alpha n+\beta} \end{array}$$

Matrices de décalage circulaire :

$\boldsymbol{P}_0$ est une matrice de décalage circulaire vers le haut (ou vers la gauche) si :

$$\boldsymbol{P}_0 = \begin{pmatrix} 0 & 1 & 0 & . & 0 \\ 0 & 0 & 1 & 0 & . \\ 0 & . & 0 & 1 & 0 \\ . & . & . & 0 & 1 \\ 1 & . & . & . & 0 \end{pmatrix}, \quad [\boldsymbol{P}_0]_{ij} = \begin{cases} 1 & i = 1, \cdots, N-1 \\ 1 & j = 1, i = N \\ 0 & \text{ailleurs} \end{cases}$$

$\boldsymbol{P}_0$ est une matrice circulante. Elle peut donc être factorisée de façon suivante:

$$\boldsymbol{P}_0 = \boldsymbol{W} D \boldsymbol{W}^{-1}$$

où D est une matrice diagonale et les éléments des matrices $\boldsymbol{W}$ et $\boldsymbol{W}^{-1}$ sont données par:

$$[\boldsymbol{W}]_{pq} = \exp\left[\frac{2\pi}{N}(p-1)(q-1)\right] p,q = 1, \cdots, N$$

$$[\boldsymbol{W}^{-1}]_{pq} = \frac{1}{N}\exp\left[\frac{-2\pi}{N}(p-1)(q-1)\right], \quad p,q = 1, \cdots N$$

$\boldsymbol{P}_1 = \boldsymbol{P}_0^t$ est une matrices de décalage circulaire vers le bas (ou vers la droite).

$$\boldsymbol{P}_1^{-1} = \boldsymbol{P}_1^t = \boldsymbol{P}_0, \quad \boldsymbol{P}_0^{-1} = \boldsymbol{P}_0^t = \boldsymbol{P}_1$$

Matrices triangulaires supérieure :

$$\boldsymbol{U} = \begin{pmatrix} u_{11} & u_{12} & u_{13} & u_{14} & u_{15} \\ 0 & u_{22} & u_{23} & u_{24} & u_{25} \\ 0 & 0 & u_{33} & u_{34} & u_{35} \\ 0 & 0 & 0 & u_{44} & u_{45} \\ 0 & 0 & 0 & 0 & u_{55} \end{pmatrix}$$

$$\boldsymbol{U}_1 = \begin{pmatrix} 1 & u_{12} & u_{13} & u_{14} & u_{15} \\ 0 & 1 & u_{23} & u_{24} & u_{25} \\ 0 & 0 & 1 & u_{34} & u_{35} \\ 0 & 0 & 0 & 1 & u_{45} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{U}_0 = \begin{pmatrix} 0 & u_{12} & u_{13} & u_{14} & u_{15} \\ 0 & 0 & u_{23} & u_{24} & u_{25} \\ 0 & 0 & 0 & u_{34} & u_{35} \\ 0 & 0 & 0 & 0 & u_{45} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

L'inverse d'une matrice du type $\boldsymbol{U}_1$ est aussi une matrice du type $\boldsymbol{U}_1$.

$$\det[\boldsymbol{U}] = \prod_{i=1}^{n} u_{ii}, \quad \det[\boldsymbol{U}_1] = 1, \quad \det[\boldsymbol{U}_0] = 0$$

Matrices triangulaires inférieure :

$$\boldsymbol{L} = \begin{pmatrix} l_{11} & 0 & 0 & 0 & 0 \\ l_{21} & l_{22} & 0 & 0 & 0 \\ l_{31} & l_{32} & l_{33} & 0 & 0 \\ l_{41} & l_{42} & l_{43} & l_{44} & 0 \\ l_{51} & l_{52} & l_{53} & l_{54} & l_{55} \end{pmatrix}$$

$$\boldsymbol{L}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ l_{21} & 1 & 0 & 0 & 0 \\ l_{31} & l_{32} & 1 & 0 & 0 \\ l_{41} & l_{42} & l_{43} & 1 & 0 \\ l_{51} & l_{52} & l_{53} & l_{54} & 1 \end{pmatrix} \boldsymbol{L}_0 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ l_{21} & 0 & 0 & 0 & 0 \\ l_{31} & l_{32} & 0 & 0 & 0 \\ l_{41} & l_{42} & l_{43} & 0 & 0 \\ l_{51} & l_{52} & l_{53} & l_{54} & 0 \end{pmatrix}$$

L'inverse d'une matrice du type $\boldsymbol{L}_1$ est aussi une matrice du type $\boldsymbol{L}_1$.

$$\det[\boldsymbol{L}] = \prod_{i=1}^{n} l_{ii}, \quad \det[\boldsymbol{L}_1] = 1, \quad \det[\boldsymbol{L}_0] = 0$$

Matrices de Hessenberg supérieure :

$$\boldsymbol{H} = \begin{pmatrix} h_{11} & h_{12} & h_{13} & h_{14} & h_{15} \\ h_{12} & h_{22} & h_{23} & h_{24} & h_{25} \\ 0 & h_{23} & h_{33} & h_{34} & h_{35} \\ 0 & 0 & h_{34} & h_{44} & h_{45} \\ 0 & 0 & 0 & h_{45} & h_{55} \end{pmatrix}$$

Matrices de Hessenberg inférieure :

$$\boldsymbol{H} = \begin{pmatrix} h_{11} & h_{12} & 0 & 0 & 0 \\ h_{21} & h_{22} & h_{23} & 0 & 0 \\ h_{31} & h_{32} & h_{33} & h_{34} & 0 \\ h_{41} & h_{42} & h_{43} & h_{44} & h_{45} \\ h_{51} & h_{52} & h_{53} & h_{54} & h_{55} \end{pmatrix}$$

Matrices tridiagonale :

$$\boldsymbol{B} = \begin{pmatrix} b_{11} & b_{12} & 0 & 0 & 0 \\ b_{21} & b_{22} & b_{23} & 0 & 0 \\ 0 & b_{32} & b_{33} & b_{34} & 0 \\ 0 & 0 & b_{43} & b_{44} & b_{45} \\ 0 & 0 & 0 & b_{54} & b_{55} \end{pmatrix}$$

Matrice de transformation de Householder

Définition
Soit $\boldsymbol{v}$ un vecteur complexe de dimension $[N]$. Alors $\boldsymbol{v}^{*t}\boldsymbol{v}$ est un scalaire réel et $\boldsymbol{v}\boldsymbol{v}^{*t}$ est une matrice carrée de dimensions $[N,N]$. Soit $\boldsymbol{I}$ une matrice identité de mêmes dimensions. Alors la matrice

$$\boldsymbol{H}_{\boldsymbol{v}} = \boldsymbol{I} - \frac{2}{\boldsymbol{v}^*\boldsymbol{v}}[\boldsymbol{v}\boldsymbol{v}^*]$$

est une matrice de réflexion ou matrice de Householder.

Propriétés :

- $\boldsymbol{H}_{\boldsymbol{v}}$ est hermitienne et orthonormale.

- La transformation $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x}$ avec $\boldsymbol{H}$ une matrice de Householder est appelée la transformation de Householder. Une telle transformation conserve l'énergie du vecteur original $\boldsymbol{x}$. (Ceci est vrai pour n'importe quelle transformation orthonormale.)
- Soit $\boldsymbol{e}_j$ un vecteur base ($[\boldsymbol{e}_j]i = \delta_{ij}$), alors pour un vecteur quelconque $\boldsymbol{u}$ il existe une matrice de Householder $\boldsymbol{H}$ telle que

$$\boldsymbol{H}\boldsymbol{u} = \sigma \boldsymbol{e}_j$$

ssi σ est choisi telle que

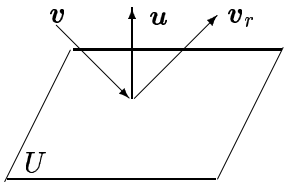
$$\boldsymbol{u}^*\boldsymbol{u} = |\sigma^2| = \sum_{i=1}^{N} u_i u_i^* \quad \text{ou bien} \quad \sigma = \pm \frac{u_j}{|u_j|}\sqrt{\boldsymbol{u}^*\boldsymbol{u}}$$

On a aussi $\boldsymbol{u}^*\boldsymbol{H} = \sigma^* \boldsymbol{e}_j^t$

- Soit $\boldsymbol{v}$ un vecteur, et $\boldsymbol{v}_r$ son image à travers un hyperplan U sous-espace orthogonale au vecteur $\boldsymbol{u}$ (symétrie plane). La matrice de transformation de Householder $\boldsymbol{H}_{\boldsymbol{u}}$ est une matrice qui nous permet de faire correspondre $\boldsymbol{v}_r$ à $\boldsymbol{v}$ par

$$\boldsymbol{v}_r = \boldsymbol{H}_{\boldsymbol{u}}\, \boldsymbol{v}$$

où $\boldsymbol{H}_{\boldsymbol{u}} = \boldsymbol{I} - \beta \boldsymbol{u}\boldsymbol{u}^t$ avec $\beta = \frac{2}{(\boldsymbol{u}^t\boldsymbol{u})}$



Matrice de transformation de Householder hyperbolique

Soit $\boldsymbol{\Phi}$ une matrice diagonale avec des éléments diagonaux de $+1$ et 1. Soit $\boldsymbol{v}$ est un vecteur complexe. Alors

$$\boldsymbol{v}^*\Phi\boldsymbol{v} = \sum_{i=1}^{N} |v_i|^2 [\Phi]_{ii}$$

est la norme hyperbolique du vecteur $\boldsymbol{v}$. Une matrice $\boldsymbol{Q}$ qui satisfait

$$\boldsymbol{Q}\boldsymbol{\Phi}\boldsymbol{Q}^* = \boldsymbol{\Phi}$$

est une matrice hypernormale. Une telle matrice préserve la norme hyperbolique des vecteurs, c'est à dire: si $\boldsymbol{y}^t = \boldsymbol{x}^t\boldsymbol{Q}$ alors $\boldsymbol{y}^t\boldsymbol{Q}\boldsymbol{y} = \boldsymbol{x}^t\boldsymbol{Q}\boldsymbol{x}$. Une matrice définie par

$$\boldsymbol{H} = \Phi - 2\frac{\boldsymbol{Q}\boldsymbol{Q}^t}{\boldsymbol{Q}^t\Phi\boldsymbol{Q}}$$

est appelée une matrice de transformation hyperbolique de Householder. $\boldsymbol{H}$ est hermitienne et hypernormale par rapport à la matrice $\boldsymbol{\Phi}$, c'est à dire $\boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{H}^t = \boldsymbol{\Phi}$.

Propriétés :

- $\det[\boldsymbol{H}] = 1$; $\boldsymbol{H}$ est une matrice non-singulière.

- $\boldsymbol{H}$ n'est, en général, pas symétrique, mais elle a une symétrie hyperbolique, c'est à dire $\boldsymbol{H}^t\boldsymbol{A}\boldsymbol{H} = \boldsymbol{H}\boldsymbol{A}\boldsymbol{H}^t$.
- Les valeurs propres d'une matrice hypernormale forment des paires réciproquement conjuguées, c'est à dire que si λ est une valeur propre de cette matrice $\frac{1}{\lambda^*}$ l'est aussi. De plus si λ est multiple d'ordre n alors $\frac{1}{\lambda^*}$ est aussi de même ordre n.

Décomposition orthogonale par transformation de Householder

Soit $\boldsymbol{A}$ une matrice de dimensions $[N,M]$ et de rang $N \leq M$. Il existe une matrice orthogonale $\boldsymbol{H_u}$ de dimensions $[N,M]$ telle que $S = \boldsymbol{H}\boldsymbol{u}\boldsymbol{A}$ soit triangulaire supérieure. La matrice $\boldsymbol{A}$ peut être décomposée en le produit d'une matrice orthogonale $\boldsymbol{H}\boldsymbol{u}$ et d'une matrice triangulaire supérieure $\boldsymbol{S}$ de rang plein— $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{u}\boldsymbol{S}$.

Matrices centro-symétriques

$$[\boldsymbol{G}]_{ij} = [\boldsymbol{G}]_{ji} = [\boldsymbol{G}]_{N+1-i,N+1-j}, \quad i,j = 1,\cdots,N$$

Exemple :

$$\boldsymbol{G} = \begin{pmatrix} g_{11} & g_{12} & g_{13} & g_{14} & g_{15} \\ g_{12} & g_{22} & g_{23} & g_{24} & g_{14} \\ g_{13} & g_{23} & g_{33} & g_{23} & g_{13} \\ g_{14} & g_{24} & g_{23} & g_{22} & g_{12} \\ g_{15} & g_{14} & g_{13} & g_{12} & g_{11} \end{pmatrix}$$

- Si $\boldsymbol{G}$ est une matrice centro-symétrique on a:

$$\boldsymbol{JG} = \boldsymbol{G}, \quad \boldsymbol{GJ} = \boldsymbol{G}, \quad \boldsymbol{JGJ} = \boldsymbol{G}, \quad \text{et} \quad \boldsymbol{JG}^{-1}\boldsymbol{J} = \boldsymbol{G}^{-1}$$

- Cette dernière relation montre que l'inverse d'une matrice centro-symétrique est aussi une matrice centro-symétrique.
- Si $\boldsymbol{G}$ est une matrice centro-symétrique de dimension $[N,N]$, alors elle peut être partitionnée de la manière suivante :
  - si N est pair ($N = 2r$) on a :

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{JB}^*\boldsymbol{J} & \boldsymbol{JA}^*\boldsymbol{J} \end{pmatrix}$$

  - si N est impair ($N = 2r+1$) on a :

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{x} & \boldsymbol{B} \\ \boldsymbol{x}^* & \alpha & \boldsymbol{x}^*\boldsymbol{J} \\ \boldsymbol{JB}^*\boldsymbol{J} & \boldsymbol{Jx} & \boldsymbol{JA}^*\boldsymbol{J} \end{pmatrix}$$

    où $\boldsymbol{A}$ et $\boldsymbol{B}$ sont deux matrices de dimensions $[r,r]$, $\boldsymbol{J}$ est une matrice d'inversion d'ordre de dimensions $[r,r]$, $\boldsymbol{x}$ est un vecteur de dimension r et α est un scalaire.
- Si $\boldsymbol{G}$ est une matrice centro-symétrique de dimension $[N,N]$, alors
  - si N est pair on a

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{C}^t \\ \boldsymbol{C} & \boldsymbol{JAJ} \end{pmatrix}$$

    où $\boldsymbol{A}$,$\boldsymbol{C}$ et $\boldsymbol{J}$ sont des matrices de dimensions $[N/2,N/2]$ et

$$\boldsymbol{A}^t = \boldsymbol{A} \quad \text{et} \quad \boldsymbol{C}^t = \boldsymbol{JCJ}$$

    On montre aussi que $\boldsymbol{G}$ est orthogonalement similaire à et on montre que si $\boldsymbol{G}$ a des valeurs propres (v.p.) distinctes, elle a $[N/2]$ vecteurs propres symétriques ($\boldsymbol{h}_i, i = 1,\cdots,N/2$) et $[N/2)]$ vecteurs propres v.p. anti-symétriques ($\boldsymbol{g}_i, i = 1,\cdots,N/2$).

$$\begin{pmatrix} \boldsymbol{A} - \boldsymbol{JC} & 0 \\ 0 & \boldsymbol{A} + \boldsymbol{JC} \end{pmatrix}$$

    Dans ce cas les v.p. symétriques $\boldsymbol{h}_i$ de $\boldsymbol{G}$ s'obtiennent par

$$\boldsymbol{h}_i = \frac{1}{\sqrt{2}}[\boldsymbol{u}_i, -[\boldsymbol{J}\boldsymbol{u}_i]^t]^t$$

De même les v.p. anti-symétriques $\boldsymbol{g}_i$ de $\boldsymbol{G}$ s'obtiennent par

$$\boldsymbol{g}_i = \frac{1}{\sqrt{2}}[\boldsymbol{v}_i, +[\boldsymbol{J}\boldsymbol{v}_i]^t]^t$$

où $\boldsymbol{u}_i$ et $\boldsymbol{v}_i$ sont respectivement les v.p. orthonormés des matrices $\boldsymbol{A} - \boldsymbol{J}\boldsymbol{C}$ et $\boldsymbol{A} + \boldsymbol{J}\boldsymbol{C}$.

– si N est impair on a

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{x} & \boldsymbol{C}^t \\ \boldsymbol{x}^t & q & \boldsymbol{x}^t\boldsymbol{J} \\ \boldsymbol{C} & \boldsymbol{J}\boldsymbol{x} & \boldsymbol{J}\boldsymbol{A}\boldsymbol{J} \end{pmatrix}$$

où $\boldsymbol{A}$, $\boldsymbol{C}$ et $\boldsymbol{J}$ sont des matrices de dimensions $[N/2-1, N/2-1]$ et

$$\boldsymbol{A}^t = \boldsymbol{A} \quad \text{et} \quad \boldsymbol{C}^t = \boldsymbol{J}\boldsymbol{C}\boldsymbol{J}$$

et on montre que si $\boldsymbol{G}$ a des valeurs propres distinctes elle a $[(N+1)/2)]$ v.p. symétriques $(\boldsymbol{h}_i, \quad i = 1, \cdots, (N+1)/2))$ et $[(N-1)/2)]$ v.p. anti-symétriques $(\boldsymbol{g}_i, \quad i = 1, \cdots, (N-1)/2))$
On montre aussi que $\boldsymbol{G}$ est orthogonalement similaire à

$$\begin{pmatrix} \boldsymbol{A} - \boldsymbol{J}\boldsymbol{C} & 0 & 0 \\ 0 & q & \sqrt{2}\boldsymbol{x}^t \\ 0 & \sqrt{2}\boldsymbol{x} & \boldsymbol{A} + \boldsymbol{J}\boldsymbol{C} \end{pmatrix}$$

Dans ce cas les v.p. symétriques $\boldsymbol{h}_i$ de $\boldsymbol{G}$ s'obtiennent par

$$\boldsymbol{h}_i = \frac{1}{\sqrt{2}}[\boldsymbol{u}_i^t, 0, -[\boldsymbol{J}\boldsymbol{u}_i]^t]^t$$

De même les v.p. anti-symétriques $\boldsymbol{g}_i$ de $\boldsymbol{G}$ s'obtiennent par

$$\boldsymbol{g}_i = \frac{1}{\sqrt{2}}[\boldsymbol{v}_i^t, \sqrt{2}\alpha_i, [\boldsymbol{J}\boldsymbol{v}_i]^t]^t$$

où $\boldsymbol{u}_i$ sont les v.p. orthonormés de la matrices $\boldsymbol{A} - \boldsymbol{J}\boldsymbol{C}$, et $[\boldsymbol{v}_i, \alpha_i]^t$ sont les v.p. orthonormés de la matrice

$$\begin{pmatrix} \boldsymbol{A} + \boldsymbol{J}\boldsymbol{C} & \sqrt{2}\boldsymbol{x} \\ \sqrt{2}\boldsymbol{x}^t & q \end{pmatrix}$$

Dans tous les cas ces v.p. sont uniques (à une constante multiplicative près) et mutuellement orthogonaux entre eux.
Si $\boldsymbol{a} = [a_0, a_1, \cdots, a_{(N-1)}]^t$ est un v.p. de $\boldsymbol{G}$, c'est à dire $\boldsymbol{G}\boldsymbol{a} = \lambda\boldsymbol{a}$, le polynôme

$$\boldsymbol{A}(z) = \sum_{k=0}^{N-1} a_k z^{-k}$$

est un polynôme d'ordre $(N-1)$, appelé polynôme propre de la matrice $\boldsymbol{G}$. On montre que si $z = z_i$ est une racine de $\boldsymbol{A}(z)$, alors $z = \frac{1}{z_i}$ en est une aussi.

Matrices d'innovation diagonales (MID)
(En Anglais : Diagonal Innovation matrices DIM)

Considérons les trois vecteurs $\boldsymbol{x}_{m-1}$ de dimension $[m-1]$, $\boldsymbol{z}_{m-1}$ de dimension $[m-1]$, et de $\boldsymbol{y}_m$ de dimension $[m]$. Une matrice MID est une matrice qui est générée par les trois vecteurs génératrices $\boldsymbol{x}_{m-1}$, $\boldsymbol{z}_{m-1}$, et $\boldsymbol{y}_m$ de façon suivante:

$$\boldsymbol{R} = \begin{pmatrix} y_1 & \vdots & x_1 & \vdots & x_1 & \vdots & x_1 & \vdots & x_1 \\ \cdots & & & & & & & & \\ z_1 & & y_2 & \vdots & x_2 & \vdots & x_2 & \vdots & x_2 \\ \cdots & \cdots & \cdots & & & & & & \\ z_1 & & z_2 & & y_3 & \vdots & x_3 & \vdots & x_3 \\ \cdots & \cdots & \cdots & \cdots & \cdots & & & & \\ z_1 & & z_2 & & z_3 & & y_4 & \vdots & x_4 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & & & \\ z_1 & & z_2 & & z_3 & & z_4 & & y_5 \end{pmatrix}, \quad \begin{cases} \boldsymbol{x}_{m+1}^t = [\boldsymbol{x}_m^t, x_{m+1}] \\ \boldsymbol{z}_{m+1}^t = [\boldsymbol{z}_m^t, z_{m+1}] \\ \boldsymbol{y}_{m+1}^t = [\boldsymbol{y}_m^t, y_{m+1}] \end{cases}$$

Matrices d'innovation périphériques (MIP)
(En Anglais : Peripheral Innovation matrices PIM)

Une matrice MIP est une matrice qui générée par les trois vecteurs génératrices $\boldsymbol{x}_{m-1}$, $\boldsymbol{z}_{m-1}$, et $\boldsymbol{y}_m$ de façon suivante:

$$\boldsymbol{R} = \begin{pmatrix} y_1 & \vdots & x_1 & \vdots & x_2 & \vdots & x_3 & \vdots & x_4 \\ \cdots & & & & & & & & \\ z_1 & & y_2 & \vdots & x_1 & \vdots & x_2 & \vdots & x_3 \\ \cdots & \cdots & \cdots & & & & & & \\ z_2 & & z_1 & & y_3 & \vdots & x_1 & \vdots & x_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & & & & \\ z_3 & & z_2 & & z_1 & & y_4 & \vdots & x_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & & \\ z_4 & & z_3 & & z_2 & & z_1 & & y_5 \end{pmatrix}, \begin{cases} \boldsymbol{x}_{m+1}^t = [x_{m+1}, \boldsymbol{x}_m^t] \\ \boldsymbol{z}_{m+1}^t = [z_{m+1}, \boldsymbol{z}_m^t] \\ \boldsymbol{y}_{m+1}^t = [y_{m+1}, \boldsymbol{y}_m^t] \end{cases}$$

Les matrices du types $\boldsymbol{T}$ (Toeplitz), $\boldsymbol{C}$ (circulante), et $\boldsymbol{H}$ (Hankel) sont des cas particuliers des matrices MIP. L'inversion des matrices du type MID et MIP demande un coût de calcul de l'ordre de $O(N_2)$.

Matrices de Vandermonde

Matrice de Vandermonde est une matrice dans laquelle les éléments de chaque colonne sont constitués par la puissance n-ième d'un scalaire. Cette matrice est entièrement définie par le vecteur $\boldsymbol{x} = [x_1, \cdots, x_n]$.

$$\boldsymbol{V} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \cdots & \cdots & \cdots & \cdots \\ x_1^{m-1} & x_2^{m-1} & \cdots & x_n^{m-1} \end{pmatrix}$$

$$\det[\boldsymbol{V}] = \prod_{i \neq j} (x_j - x_i)$$

E.3 Matrices blocs

Une matrice quelconque peut toujours être décomposée en blocs. Si ces blocs ont tous les mêmes dimensions on appellera la matrice *matrice-bloc*.
Soit $\boldsymbol{R} = [\boldsymbol{R}_{IJ}]$ une matrice-bloc dont chaque bloc est une matrice de dimensions [P,P]. Alors $\boldsymbol{R}$ est:

bloc-symétrique	si	$\boldsymbol{R}_{IJ} = \boldsymbol{R}_{JI}$
bloc-antisymétrique	si	$\boldsymbol{R}_{IJ} = -\boldsymbol{R}_{JI}$
bloc-Hankel	si	$\boldsymbol{R}_{IJ} = \boldsymbol{R}_{I+J}$
bloc-Toeplitz	si	$\boldsymbol{R}_{IJ} = \boldsymbol{R}_{I-J}$

Tous ce que l'on peut dire sur les matrices sont valables sur les matrices-bloc si tous les blocs sont inversibles. Par exemple:
Soit $\boldsymbol{R} = [\boldsymbol{R}_{IJ}]$, $\boldsymbol{I}$, $j = 1, \cdots, N$ une matrice-bloc avec $\boldsymbol{R}_{IJ}$ des matrices de dimensions [P,P]. Si toutes les $\boldsymbol{R}_{IJ}$ sont inversibles, alors $\boldsymbol{R}$ peut être factorisée soit par $\boldsymbol{R} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^t$, soit par $\boldsymbol{U}\boldsymbol{R}\boldsymbol{U}^t = \boldsymbol{D}$ où $\boldsymbol{L}$ est une matrice bloc-triangulaire inférieure avec des matrices d'identités sur son diagonale, $\boldsymbol{U}$ est une matrice bloc-triangulaire supérieure, $\boldsymbol{D}$ est une matrice bloc-diagonale, et $\boldsymbol{U} = \boldsymbol{L}^{-1}$.

Inversion d'une matrice décomposée en blocs

Soit $\boldsymbol{A}$ une matrice quelconque telle que

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}$$

Si $\boldsymbol{A}_{22}^{-1}$ existe, alors $\boldsymbol{A}$ peut être factorisée de la façon suivante:

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \\ 0 & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21} & 0 \\ 0 & \boldsymbol{A}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{I} & 0 \\ \boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21} & \boldsymbol{I} \end{pmatrix}$$

On a aussi

$$\text{rang}\,[\boldsymbol{A}] = \text{rang}\left[\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\right] + \text{rang}\,[\boldsymbol{A}_{22}]$$

$\boldsymbol{A}^{-1}$ existe ssi l'inverse de la matrice $\boldsymbol{T} = \boldsymbol{A}11 - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}$ existe. Dans ce cas si on note par $\boldsymbol{B} = \boldsymbol{A}^{-1}$ on a:

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{pmatrix}$$

$$\boldsymbol{B}_{11} = \boldsymbol{L}^{-1} = (\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1} = \boldsymbol{A}_{11}^{-1} + \boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{B}_{22}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}$$

$$\boldsymbol{B}_{22} = \boldsymbol{D}^{-1} = (\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1} = \boldsymbol{A}_{22}^{-1} + \boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{B}_{11}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}$$

$$\boldsymbol{B}_{12} = -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{B}_{22} = -\boldsymbol{B}_{11}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}$$

$$\boldsymbol{B}_{21} = -\boldsymbol{A}_{22}\boldsymbol{A}_{21}\boldsymbol{B}_{11} = -\boldsymbol{B}_{22}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}$$

Les matrices $\boldsymbol{L}$ et $\boldsymbol{D}$ sont appelées *le complément Shur* de la matrice $\boldsymbol{A}$.

Cas particulier 1 :
Si $\boldsymbol{A}$ est une matrice bloc triangulaire inférieure, c'est à dire si $\boldsymbol{A}_{21} = 0$, $\boldsymbol{B}$ est aussi triangulaire inférieure, c'est à dire $\boldsymbol{B}_{21} = 0$ et on a:

$$\boldsymbol{B}_{11} = \boldsymbol{A}_{11}^{-1}, \quad \boldsymbol{B}_{22} = \boldsymbol{A}_{22}^{-1}, \quad \boldsymbol{B}_{12} = -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}$$

Cas particulier 2 :
Si $\boldsymbol{A}$ est une matrice bloc triangulaire supérieure, c'est à dire si $\boldsymbol{A}_{12} = 0$, $\boldsymbol{B}$ est aussi triangulaire supérieure, c'est à dire $\boldsymbol{B}_{12} = 0$ et on a:

$$\boldsymbol{B}_{11} = \boldsymbol{A}_{11}^{-1}, \quad \boldsymbol{B}_{22} = \boldsymbol{A}_{22}^{-1}, \quad \boldsymbol{B}_{21} = -\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}$$

Cas particulier 3 :
Si $\boldsymbol{A}$ est une matrice bloc-diagonale, c'est à dire si $\boldsymbol{A}_{12} = \boldsymbol{A}_{21} = 0$, $\boldsymbol{B}$ est aussi bloc-diagonale, c'est à dire $\boldsymbol{B}_{12} = \boldsymbol{B}_{21} = 0$ et on a:

$$\boldsymbol{B}_{11} = \boldsymbol{A}_{11}^{-1}, \quad \boldsymbol{B}_{22} = \boldsymbol{A}_{22}^{-1}$$

Cas particulier 4 :
Si $\boldsymbol{A}_{22}$ est un scalaire et $\boldsymbol{A}_{21}$ et $\boldsymbol{A}_{12}$ sont des vecteurs, on a:

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{x} \\ \boldsymbol{z}^t & y \end{pmatrix}, \quad \boldsymbol{B} = \boldsymbol{A}^{-1} = \frac{1}{\alpha}\begin{pmatrix} \alpha\boldsymbol{A}_{11}^{-1} + \boldsymbol{w}\boldsymbol{v}^t & \boldsymbol{w} \\ \boldsymbol{v}^t & 1 \end{pmatrix}$$

où α, $\boldsymbol{w}$, et $\boldsymbol{v}$ s'obtient par:

$$\alpha = \frac{1}{(y - \boldsymbol{z}^t\boldsymbol{A}_{11}^{-1}\boldsymbol{x})} = \frac{|\boldsymbol{A}|}{|\boldsymbol{A}_{11}|}, \quad \boldsymbol{w} = -\boldsymbol{A}_{11}^{-1}\boldsymbol{x}, \quad \boldsymbol{v} = -\boldsymbol{A}_{11}^{-t}\boldsymbol{z}$$

E.4 Matrices de Toeplitz et de Hankel

E.4.1 Matrices de Toeplitz

Une matrice de dimensions [M,N] est dite de Toeplitz si

$$[\boldsymbol{T}]_{ij} = t_{(i-j)}, \quad i = 1,\cdots,M, \quad j = 1,\cdots,N$$

$$\boldsymbol{T} = \begin{pmatrix} t_0 & t_{-1} & . & . & t_{-(N-2)} & t_{(N-1)} \\ t_1 & t_0 & . & . & . & t_{-(N-2)} \\ . & . & . & . & . & . \\ . & . & t_0 & . & . & . \\ t_{(M-2)} & . & . & . & t_0 & t_{-1} \\ t_{(M-1)} & t_{(M-2)} & . & . & t_1 & t_0 \end{pmatrix}$$

Dans le cas où $\boldsymbol{T}$ est carrée on a $[\boldsymbol{T}]_{ij} = t_{(i-j)}, \quad i,j = 1,\cdots,N$. Si de plus $\boldsymbol{T}$ est symétrique on a $[\boldsymbol{T}]_{ij} = t_{|i-j|}, \quad i,j = 1,\cdots,N$.

Par la suite nous considérons seulement les matrices de Toeplitz carrées de dimensions [N,N].

$$\boldsymbol{T} = \begin{pmatrix} t_0 & t_{-1} & . & . & t_{-(N-2)} & t_{-(N-1)} \\ t_1 & t_0 & . & . & . & t_{-(N-2)} \\ . & . & . & . & . & . \\ . & . & t_0 & . & . & . \\ t_{(N-2)} & . & . & . & t_0 & t_{-1} \\ t_{(N-1)} & t_{(N-2)} & . & . & t_1 & t_0 \end{pmatrix}$$

Une matrice de Toeplitz est entièrement déterminée par sa première ligne

$$[\boldsymbol{T}]_{1*} = [t_0,t_{-1},\cdots,t_{-(N-2)},t_{-(N-1)}]^t$$

et sa première colonne

$$[\boldsymbol{T}]_{*1} = [t_0,t_1,\cdots,t_{(N-2)},t_N - 1]^t$$

c'est à dire par $(2N-1)$ éléments $\{t_{-(N-1)},\cdots,t_0,\cdots,t_{(N-1)}\}$.

Si $\boldsymbol{T}_N$ est une matrice de Toeplitz de dimensions [N,N] et si on note par $\{\boldsymbol{T}_n, \quad n = 1,\cdots,N\}$ les sous-matrices de $\boldsymbol{T}_N$ on a les relations suivantes:

$$\boldsymbol{J}_n\boldsymbol{T}_n\boldsymbol{J}_n = \boldsymbol{T}_n^t, \quad n = 1,\cdots,N$$

où $\boldsymbol{J}_n$ est une matrice d'inversion d'ordre de dimensions [n,n]. Une matrice de Toeplitz peut être factorisée de façon suivante:

$$\boldsymbol{T} = \boldsymbol{L}\boldsymbol{U}$$

De plus si elle est hermitienne elle peut être factorisée de façon suivante:

$$\boldsymbol{T} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^*$$

Les lignes successives de $\boldsymbol{L}$ sont les coefficients des prédicateurs linéaires d'ordre 0,1, $\cdots$,$N-1$, et les éléments diagonaux de la matrice $\boldsymbol{D}$ sont les variances d'erreurs de prédictions correspondantes.

Si $\boldsymbol{T}$ est une matrice de Toeplitz, on a:

$$\boldsymbol{T}^{-1} = \frac{1}{\lambda_{n-1}} \begin{pmatrix} 1 & \boldsymbol{b}_{n-1}^t \\ \boldsymbol{a}_{n-1} & \boldsymbol{S}_{n-1} \end{pmatrix}$$

Les éléments de $\boldsymbol{S}_{n-1}$ sont déterminés complètement par λ_{n-1}, $\boldsymbol{a}_{n-1}$, et $\boldsymbol{b}_{n-1}$.

Une décomposition moins connue est:

$$\boldsymbol{T}^{-1} = \frac{1}{\lambda_{n-1}} [\boldsymbol{A}!\, \boldsymbol{B}! - \boldsymbol{B}\boldsymbol{A}]$$

où $\boldsymbol{A}!$ et $\boldsymbol{B}$ sont des matrices Toeplitz triangulaire supérieur type $\boldsymbol{L}_1$, et $\boldsymbol{A}$ et $\boldsymbol{B}!$ sont des matrices matrices Toeplitz triangulaire inférieur type $\boldsymbol{U}_1$. Les éléments de ces matrices sont données par:

$$[\boldsymbol{A}]_{jk} = [\widetilde{\boldsymbol{a}}_{n-1}]_{k-j}, \quad [\boldsymbol{B}]_{jk} = [\widetilde{\boldsymbol{b}}_{n-1}]_{k-j}$$

$$[\boldsymbol{A}!]_{jk} = [\boldsymbol{a}_{n-1}]_{j-k} + \delta_{j-k}, \quad [\boldsymbol{B}!]_{jk} = [\boldsymbol{b}_n - 1]_{kj} + \delta_{j-k}, \quad 1 \leq j,k \leq n$$

et où pour tout vecteur $\boldsymbol{g}_m$, $\quad \boldsymbol{g}_m(k) = 0, \forall k \leq 0$ ou $k > m$.

E.4.2 Matrices de Hankel

$\boldsymbol{H}$ est une matrice de Hankel si tous ses éléments se trouvant sur des lignes parallèle à sa deuxième diagonale sont identiques. Par exemple:

$$\boldsymbol{H} = \begin{pmatrix} h_N & h_{-N+1} & . & . & h_{-1} & h_0 \\ h_{-N+1} & . & . & . & h_0 & h_1 \\ \vdots & . & . & . & . & \vdots \\ h_{-1} & h_0 & . & . & . & h_{M-1} \\ h_0 & h_1 & . & . & & h_M \end{pmatrix}$$

Si $\boldsymbol{H}$ est une matrice carrée on la note par $\boldsymbol{H}_N$.

$\boldsymbol{H}_N$ est une matrice de Hankel d'ordre N si

$$[\boldsymbol{H}]_{ij} = h_{i+j-2}, \quad i,j = 1, \cdots, N$$

$$\boldsymbol{T} = \begin{pmatrix} h_0 & h_1 & . & . & h_{N-2} & h_{N-1} \\ h_1 & h_2 & . & . & h_{N-1} & h_N \\ . & . & . & . & . & . \\ . & . & h_{N-1} & . & . & h_{2N-3} \\ . & . & . & . & h_{2N-2} & h_{2N-2} \\ h_N & h_{N+1} & & . & h_{2N-2} & h_{2N-1} \end{pmatrix}$$

Une matrice de Hankel est toujours symétrique: $\boldsymbol{H}^t = \boldsymbol{H}$. Elle est hermitienne ssi tous ses éléments sont réels. Si T est une matrice de Toeplitz, alors les matrices

$$\boldsymbol{H}_1 = \boldsymbol{T}\boldsymbol{J}, \quad \text{et} \quad \boldsymbol{H}_2 = \boldsymbol{J}\boldsymbol{T}^t$$

sont des matrices de Hankel.

Les trois systèmes d'équations suivantes sont équivalentes:

$$\boldsymbol{H}\boldsymbol{a} = -\boldsymbol{d}, \quad \boldsymbol{T}[\boldsymbol{J}\boldsymbol{a}] = -\boldsymbol{d}, \quad \boldsymbol{T}\boldsymbol{a} = -\boldsymbol{J}\boldsymbol{d}$$

E.5 Matrices circulantes

Une matrice circulante est une matrice qui est définie par seule sa première ligne (ou seule par sa première colonne). $\boldsymbol{C}$ est une matrice *circulante à droite (circulante vers le haut)* si :

$$[\boldsymbol{C}]_{ij} = \begin{cases} c_{j-i} & \text{si } \; j-i \geq 0 \\ c_{n-j+1} & \text{si } \; j-i < 0 \end{cases}$$

Exemple :

$$\boldsymbol{C} = \begin{pmatrix} c_0 & c_1 & c_2 & \cdots & \cdots & c_{n-1} \\ c_{n-1} & c_0 & c_1 & \cdots & \cdots & c_{n-2} \\ c_{n-2} & c_{n-1} & c_0 & \cdots & \cdots & . \\ \vdots & c_{n-2} & c_{n-1} & c_0 & \cdots & \vdots \\ . & \cdots & \cdots & \cdots & \cdots & c_1 \\ c_1 & c_2 & \cdots & \cdots & c_{n-1} & c_0 \end{pmatrix}$$

Une matrice circulante à droite est une matrice de Toeplitz qui est symétrique par rapport à sa deuxième diagonale. $\boldsymbol{C}$ est une matrice *circulante à gauche (ou circulante vers le bas)* si :

$$[\boldsymbol{C}]_{ij} = \begin{cases} c_{n+1-i-j} & \text{si } \; j+i \leq n+1 \\ c_{2n+1-i-j} & \text{si } \; j+i > n+1 \end{cases}$$

Exemple :

$$\boldsymbol{C} = \begin{pmatrix} c_{n-1} & \cdots & \cdots & c_2 & c_1 & c_0 \\ c_{n-2} & c_{n-1} & \cdots & \cdots & c_0 & c_{n-1} \\ c_{n-3} & \cdots & \cdots & c_0 & c_{n-1} & c_{n-2} \\ \vdots & \cdots & c_0 & c_{n-1} & c_{n-2} & \vdots \\ c_1 & \cdots & \cdots & \cdots & \cdots & c_2 \\ c_0 & c_{n-1} & c_{n-2} & \cdots & c_2 & c_1 \end{pmatrix}$$

Une matrice circulante à gauche est une matrice de Hankel qui est symétrique par rapport à sa diagonale. Si $\boldsymbol{C}$ est une matrice circulante de dimensions $[N,N]$ avec des éléments réels $[\boldsymbol{C}]_{ij}$ on a:

$$[\boldsymbol{C}]_{ij} = c_{(i-j) \bmod N}, \quad \boldsymbol{C}^t = \boldsymbol{C}^{-1}$$

$$\boldsymbol{C} = \begin{pmatrix} c_0 & c_{N-1} & . & . & c_2 & c_1 \\ c_1 & c_0 & . & . & . & c_2 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ c_{N-2} & . & . & . & c_0 & c_{N-1} \\ c_{N-1} & c_{N-2} & . & . & c_1 & c_0 \end{pmatrix}$$

Les valeurs propres d'une matrice circulante sont les coefficients de la TFD de sa première ligne et ses vecteurs propres sont les vecteurs de base de la TFD.

Si $\boldsymbol{C}$ est une matrice circulante, alors

$$\boldsymbol{C} = \boldsymbol{F}\boldsymbol{\Lambda}\boldsymbol{F}^* = \boldsymbol{F}^*\boldsymbol{\Lambda}_c^*\boldsymbol{F}$$

où

- $\boldsymbol{F}$ est la matrice de la TFD :

$$\boldsymbol{F} = \frac{1}{\sqrt{N}}[\boldsymbol{w}_0,\boldsymbol{w}_1,\cdots,\boldsymbol{w}_p,\cdots,\boldsymbol{w}_{n-1}]$$

$$[\boldsymbol{w}_p]_q = \exp\left[-j\frac{2\pi pq}{N}\right], \quad p = 0,\cdots,N-1 \quad q = 1,\cdots,N$$

- $\boldsymbol{\Lambda}_c$ est une matrice diagonale contenant les valeurs propres de la matrice $\boldsymbol{C}$, c'est à dire les coefficients de la TFD de sa première ligne.
- $\boldsymbol{F}$ est une matrice qui multipliée par un vecteur $\boldsymbol{x}$ donne un vecteur $\bar{\boldsymbol{x}}$ qui contient les valeurs de sa TFD.

$$\boldsymbol{F}\boldsymbol{x} = \bar{\boldsymbol{x}} \quad \text{et} \quad \boldsymbol{x} = \boldsymbol{F}^*\bar{\boldsymbol{x}}$$

E.6 Matrices bloc-Toeplitz et bloc-circulantes

E.6.1 Matrices bloc-Toeplitz

La matrice $\boldsymbol{R}$ est dite bloc-Toeplitz si $\boldsymbol{R}$ est une matrice-bloc dont la IJ-ème bloc $\boldsymbol{R}_{IJ}$ est fonction de $I-J$):

$$\boldsymbol{R}_{IJ} = \boldsymbol{R}_{I-J}.$$

Si chaque bloc est aussi une matrice de Toeplitz alors $\boldsymbol{R}$ est une matrice Toeplitz-bloc-Toeplitz.

Matrices bloc-Toeplitz avec des blocs Toeplitz

Soit $\boldsymbol{R}_{P,M}$ est une matrice bloc d'ordre M avec les blocs $\boldsymbol{R}_{IJ}$, $I,J = 1,\cdots,M$, qui aux même sont des matrices de Toeplitz de dimensions $[P,P]$, c'est à dire:

$$\boldsymbol{R}_{PM} = \begin{pmatrix} R_0 & R_{-1} & . & . & R_{-(M-2)} & R_{-(M-1)} \\ R_1 & R_0 & . & . & . & R_{-(M-2)} \\ . & . & . & . & . & . \\ . & . & R_0 & . & . & . \\ R_{(M-2)} & . & . & . & R_0 & R_{-1} \\ R_{(M-1)} & R_{(M-2)} & . & . & R_1 & R_0 \end{pmatrix}$$

avec

$$\boldsymbol{R}_J = \begin{pmatrix} r_0 & r_{-1} & . & . & r_{-(P-2)} & r_{-(P-1)} \\ r_1 & r_0 & . & . & . & r_{-(P-2)} \\ . & . & . & . & . & . \\ . & . & r_0 & . & . & . \\ r_{(P-2)} & . & . & . & r_0 & r_{-1} \\ r_{(P-1)} & r_{(P-2)} & . & . & r_1 & r_0 \end{pmatrix}$$

Soit $\boldsymbol{J}_{P,M}$ une matrice de dimensions $[PM,PM]$ telle que

$$\boldsymbol{J}_{PM} = \begin{pmatrix} . & . & . & . & \boldsymbol{J}_P \\ . & 0 & . & \boldsymbol{J}_P & . \\ . & . & \boldsymbol{J}_P & . & . \\ . & \boldsymbol{J}_P & . & 0 & . \\ \boldsymbol{J}_P & . & . & . & . \end{pmatrix} \quad \text{avec} \quad \boldsymbol{J}_P = \begin{pmatrix} . & . & . & . & 1 \\ . & 0 & . & 1 & . \\ . & . & 1 & . & . \\ . & 1 & . & 0 & . \\ 1 & . & . & . & . \end{pmatrix}$$

une matrice de dimensions $[P,P]$. Alors on a les propriétés suivantes:

- $\boldsymbol{R}_{P,M}$ et son inverse sont toutes les deux des matrices centrosymétriques et on a

$$\boldsymbol{J}_{P,M}\boldsymbol{R}_{P,M}\boldsymbol{J}_{P,M} = \boldsymbol{R}_{P,M}^t \quad \text{et} \quad \boldsymbol{J}_{P,M}\boldsymbol{R}_{P,M}^{-1}\boldsymbol{J}_{P,M} = \boldsymbol{R}_{P,M}^{-t}$$

- $\boldsymbol{R}_{P,M}$ qui est une matrice $\boldsymbol{TT}$ de $[M,M]$ blocs de matrice de Toeplitz de dimensions $[P,P]$ peut être transformée à une matrice $\boldsymbol{TT}$ de $[P,P]$ blocs de matrices de Toeplitz de dimensions $[M,M]$ avec de simples permutations de ses éléments. Ceci peut être démontré en définissant une matrice de permutation $\boldsymbol{E}$ de dimensions $[PM,PM]$ telle que

$$\boldsymbol{E} = (\, \boldsymbol{E}_0 \quad . \quad . \quad \boldsymbol{E}_{P-1} \,)$$

avec $\boldsymbol{E}_j$ une matrice de dimensions [M,MP] telle que

$$\boldsymbol{E}_j = \begin{pmatrix} \boldsymbol{e}_i & . & . & . & . \\ . & \boldsymbol{e}_i & . & 0 & . \\ . & . & . & . & . \\ . & 0 & . & \boldsymbol{e}_i & . \\ . & . & . & . & \boldsymbol{e}_i \end{pmatrix}$$

avec $\boldsymbol{e}_i$ un vecteur ligne dont toutes ses composantes sont nulles sauf la j-ème qui est égale à 1. On montre alors que

$$\boldsymbol{E}\boldsymbol{R}_{P,M}\boldsymbol{E}^t = \widetilde{\boldsymbol{R}}_{M,P}$$

où $\widetilde{\boldsymbol{R}}_{M,P}$ est une matrice $\boldsymbol{TT}$ de [PxP] blocs de matrices de Toeplitz de dimensions [M,M]. La ij-ème bloc de $\widetilde{\boldsymbol{R}}_{M,P}$ s'obtient par $\boldsymbol{E}_j\boldsymbol{R}_{P,M}\ \boldsymbol{E}_j^t$ et le kl-ème élément de ce bloc est $\boldsymbol{e}_j\boldsymbol{R}_{l-k}\boldsymbol{e}_j^t = r_{j-i,l-k}$.

$$\boldsymbol{E}\boldsymbol{R}_{P,M}\boldsymbol{E}^t = \begin{pmatrix} \boldsymbol{E}_0 \\ . \\ . \\ \boldsymbol{E}_{P-1} \end{pmatrix} \boldsymbol{R}_{P,M}[\boldsymbol{E}^t, \cdots, \boldsymbol{E}_{p-1}^t]$$

E.6.2 Matrices bloc-circulantes

Si $\boldsymbol{A}$ est une matrice bloc-circulante, composée de [P,P] blocs de matrices circulante de dimensions [K,K], $\boldsymbol{A}$ peut être factorisée de façon suivante:

$$\boldsymbol{A} = \boldsymbol{F}^*\boldsymbol{D}\boldsymbol{F}$$

$$\boldsymbol{F} = \begin{pmatrix} F & . & . & . & . \\ . & F & . & 0 & . \\ . & . & . & . & . \\ 0 & . & F & . & . \\ . & . & . & . & F \end{pmatrix} D = \begin{pmatrix} D_{11} & . & . & . & D_{1p} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ D_{P1} & . & . & . & D_{PP} \end{pmatrix}$$

Chaque matrice D_{ij} est une matrice diagonale contenant les valeurs propres de la ij-ème bloc (matrice circulante) de $\boldsymbol{A}$.

E.7 Inversion récursive des matrices

Dans cette section nous allons présenter une méthode d'inversion récursive des matrices qui est basée sur le partitionnement de matrices en blocs.

Cas général

Si $\boldsymbol{A}_n$ est une matrice définie positive d'ordre n, alors on peut la partionner de façon suivante:

$$\boldsymbol{A}_n = \begin{pmatrix} \boldsymbol{A}_{n-1} & \boldsymbol{a}_n \\ \boldsymbol{b}_n^t & a_{nn} \end{pmatrix}$$

où $\boldsymbol{a}_n$ et $\boldsymbol{b}_n$ sont définie par

$$\boldsymbol{a}_n = [a_{1n}, a_{2n}, \cdots, a_{n-1,n}]^t \quad \text{et} \quad \boldsymbol{b}_n = [b_{1n}, n_{2n}, \cdots, b_{n-1,n}]^t$$

Si $\boldsymbol{Q}_n = \boldsymbol{A}_n^{-1}$ est aussi partitionnée par

$$\boldsymbol{Q}_n = \boldsymbol{A}_n^{-1} = \begin{pmatrix} \boldsymbol{Q}_{n-1} & \boldsymbol{Q}_n \\ \boldsymbol{v}_n^t & q_{nn} \end{pmatrix}$$

on a

$$\boldsymbol{Q}_{n-1} = \boldsymbol{A}_{n-1}^{-1} + \alpha_n \boldsymbol{Q}_n \boldsymbol{v}_n^t$$

où

$$\alpha_n = a_{nn} - \boldsymbol{b}_n \boldsymbol{A}_{n-1}^{-1} \boldsymbol{a}_n$$

$$q_{nn} = \frac{1}{\alpha_n}, \quad \boldsymbol{Q}_n = -\frac{1}{\alpha_n}[\boldsymbol{A}_{n-1}^{-1}\boldsymbol{a}_n], \quad \boldsymbol{v}_n = \frac{1}{\alpha_n}[\boldsymbol{A}_{n-1}^{-1}\boldsymbol{b}_n].$$

Cas des matrices symétriques

Si $\boldsymbol{A}$ est une matrice réelle, symétrique et définie positive, alors son inverse $\boldsymbol{A}^{-1}$ ainsi que toutes ses sous-matrices principales sont réelles, symétriques et définie positives. Si on note alors par $\boldsymbol{A}_n^{-1} = \boldsymbol{Q}_n$ et par $[\boldsymbol{Q}_n]_{ij} = q_{ij}^{(n)}$ on a

$$q_{nn}^{(n)} = \frac{1}{\left(a_{nn} - \sum_{i=1}^{n-1} a_{in}\beta_i^{(n)}\right)}$$

$$q_{ij}^{(n)} = q_{ij}^{(n-1)} - q_{in}^{(n)}\beta_i^{(n)} = q_{ij}^{(n-1)} - \beta_i^{(n)} q_{jn}^{(n)}$$

$$q_{in}^{(n)} = -q_{in}^{(n)}\beta_i^{(n)}, \quad i = 1, \cdots, n-1, \quad \beta_i^{(n)} = \sum_{k=1}^{n-1} q_{ik}^{(n-1)} a_{kn}$$

avec initialisation $\beta_1^{(1)} = -1 \quad$ et $\quad q_1^{(1)} = \frac{1}{a_{11}}$.
On a aussi la propriétés suivante

$$|\boldsymbol{A}_n| = \prod_{i=1}^{N} \frac{1}{q_{ii}^{(i)}} = \frac{1}{q_{11}^{(1)} q_{22}^{(2)} \cdots q_{nn}^{(n)}}$$

Cas des matrices de Toeplitz symétriques

$$\beta_n = \sum_{k=1}^{n-1} q_{1k}^{(n-1)} a_{1,n+1-k}, \quad q_{11}^{(n)} = \frac{q_{11}^{(n-1)}}{1-\beta_n^2}, \quad q_{1n}^{(n)} = -\beta_n q_{11}^{(n)}$$

Si $\boldsymbol{A}$ est une matrice définie positive, alors on montre que $0 \leq \beta_n < 1$, et que $q_{11}^{(n)}$ croit d'une façon monotone en fonction de n, c'est à dire que $(q_{11}^{(n)}/q_{11}^{(n-1)}) \geq 1$. D'autre part si β_n devient nul à l'ordre n, alors $q_{11}^{(n)}$ devient aussi nul, et les éléments de $n+1$ème inverse ne changent plus.

Cas des matrices de corrélation d'un signal stationnaire

La matrice de corrélation d'un signal stationnaire exponentiellement est une matrice de Toeplitz symétrique avec la propriété suivante

$$\frac{a_{12}}{a_{11}} = \frac{a_{13}}{a_{12}} = \cdots = \frac{a_{1N}}{a_{N-1}}$$

Dans ce cas on montre que l'on peut écrire

$$\beta_2 = \frac{a_{12}}{a_{11}}, \quad \beta_k = 0 \quad \text{pour} \quad k > 2,$$

$$q_{11}^{(n)} = q_{nn}^{(n)} = \frac{1}{a_{11}(1-\beta_2^2)}, \quad q_{12}^{(n)} = q_{21}^{(n)} = \frac{-\beta_2}{a_{11}(1-\beta_2^2)}$$

$$q_{jj}^{(n)} = \frac{1+\beta_2^2}{a_{11}(1-\beta_2^2)}, \quad j = 2, \cdots, n-1, \quad q_{ij}(n) = 0 \quad |i-j| > 2$$

D'une façon plus schématique $\boldsymbol{A}_n^{-1}$ est sous la forme de

$$\boldsymbol{A}_n^{-1} = \frac{1}{a_{11}(1-\beta_2^2)} \begin{pmatrix} 1 & -\beta_2 & . & . & . \\ -\beta_2 & (1+\beta_2^2) & . & 0 & . \\ . & . & . & . & . \\ . & 0 & . & (1+\beta_2^2) & -\beta_2 \\ . & . & . & -\beta_2 & 1 \end{pmatrix}$$

Cas des matrices de Toeplitz symétriques et tridiagonale

Un cas particulier des matrices de Toeplitz symétrique est une matrice dont les éléments du m-ième diagonal sont tous égale à zéro pour m supérieure à k. Pour $k = 2$ on a une matrice tridiagonale. Ceci est par exemple le cas de la matrice de covariance d'un processus AR d'ordre 1.

$$\boldsymbol{A}_n = \begin{pmatrix} a_{11} & a_{12} & 0 & . & 0 \\ a_{12} & a_{11} & . & 0 & . \\ 0 & . & . & . & 0 \\ . & 0 & . & a_{11} & \\ & 0 & . & a_{12} & a_{11} \end{pmatrix}$$

Pour cette matrice la valeur de β_n devient

$$\beta_n = \frac{(-1)^n (a_{12}/a_{11})^{n-1}}{\prod_{j=2}^{n-1}(1-\beta_j^2)^{n-j}} = (-1)^n a_{12}^{n-1} \det\left[\boldsymbol{A}_{n-1}^{-1}\right], \quad n = 2,3,\cdots$$

Une propriétés intéressante de ces matrices est le faite que même si la matrice est définie positive à l'ordre n, elle peut ne pas l'être à l'ordre n+1. Ceci peut être le cas si par exemple $|\beta_{n+1}| \geq 1$. On peut montrer que si $|\beta_2| \leq 1/2$, alors $|\beta_n| \longrightarrow 0$ quand $n \longrightarrow \infty$. Ainsi la matrice reste définie positive pour toute valeur de n, et si $|\beta_2| > 1/2$, la matrice deviendra non définie positive.

Cas des matrices triangulaires

Si $\boldsymbol{A}$ est une matrice réelle, triangulaire et définie positive d'ordre n, alors son inverse $\boldsymbol{A}_n^{-1}$ ainsi que toutes ses sous-matrices principales sont réelles, triangulaires et définie positives. Ceci est un cas particulier de (1.1) et (1.2) avec $b_n = 0$ et $\boldsymbol{v}_n = 0$. Ainsi les éléments q_{ij} de la matrice $\boldsymbol{Q}_n$ sont données par

$$q_{ij} = \begin{cases} -\dfrac{1}{a_{ii}} \displaystyle\sum_{k=1}^{j-1} q_{ik} a_{kj} & \text{pour } j > i \\ \dfrac{1}{a_{ii}} & \text{pour } j = i \\ 0 & \text{pour } j < i \end{cases}$$

Cas de Toeplitz et triangulaires

Si $\boldsymbol{A}$ est une matrice de Toeplitz, réelle, triangulaire et définie positive d'ordre n, alors son inverse $\boldsymbol{A}_n^{-1}$ ainsi que toutes ses sous- matrices principales sont des matrices de Toeplitz, réelles, triangulaires et définie positives.

$$\boldsymbol{A}_n = \begin{pmatrix} a_0 & a_1 & . & . & a_{n-2} & a_{n-1} \\ 0 & a_0 & a_1 & . & & a_{n-2} \\ . & 0 & . & . & & \\ . & . & . & . & & . \\ . & . & . & . & & . \\ . & 0 & & & a_0 & a_1 \\ . & . & . & 0 & & a_0 \end{pmatrix}$$

Dans ce cas on a $a_{ij} = a_{i-j}$, $\quad q_{ij} = q_{i-j} = 0$ pour $i > j$, et on a $q_0 = 1/a_0$, et

$$q_{ij} = -\frac{1}{a_0} \sum_{k=1}^{j-1} q_{k-i} a_{j-k}$$

Si on note par $m = j - i$ et $n = k - i$, alors on a

$$q_m = -\frac{1}{a_0} \sum_{n=1}^{m-1} q_n a_{m-n}$$

$$\boldsymbol{Q}_n = \boldsymbol{A}_n^{-1} = \begin{pmatrix} q_0 & q_1 & . & . & q_{n-2} & q_{n-1} \\ 0 & q_0 & q_1 & . & . & q_{n-2} \\ . & 0 & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & 0 & q_0 & q_1 \\ 0 & . & . & . & 0 & q_0 \end{pmatrix}$$

E.8 Résolution des systèmes d'équations linéaires

Dans cette section nous allons présenter les deux problèmes essentiels du calcul matriciel qui sont la résolution d'un système linéaire et le calcul des valeurs propres et des vecteurs propres d'une matrice. L'objective étant de mettre en évidence les difficultés de ces deux problèmes lors que la matrice est mal conditionnée.

Les deux problèmes fondamentaux de l'analyse numérique matricielles sont:

1. Résolution d'un système linéaire $\boldsymbol{A}\boldsymbol{u} = \boldsymbol{b}$.
2. Calcul des valeurs propres et des vecteurs propres d'une matrice $\boldsymbol{A}$.

Pour résoudre un problème du type (il) ou (ii), on commence naturellement par choisir une méthode déterminée (selon des critères qui seront précisés au cours des chapitres suivants). Or, quelle qui soit la méthode, ce résultat n'est pas en général le résultat exacte, car certaines erreurs sont inévitables au cours du calcul ce sont les erreurs d'arrondi.

Alors que les erreurs d'arrondi sont toujours présentes dans un calcul, le deuxième type d'erreurs n'apparaît que dans certaines méthodes. Une méthode directe conduit à la solution exacte du problème en un nombre fini d'opérations élémentaires s'il n'y avait pas d'erreurs d'arrondi.

Dans une méthode récursive la solution $\boldsymbol{u}$ du problème est la limite d'une suite $\boldsymbol{u}_k$ de solutions exactes des sous-matrices de $\boldsymbol{A}$. Dans une méthode itérative la solution $\boldsymbol{u}$ du problème est la limite d'une suite $\boldsymbol{u}_k$ de solutions approchées. Comme on est bien obligé d'arrêter les calcul à l'itération k on commet ainsi une erreur de troncature mesurée par le nombre $\|\boldsymbol{u} - \boldsymbol{u}_k\|$ (pour une norme vectorielle donnée). Le choix d'une méthode dépends des propriétés de la matrice : matrice pleine quelconque matrice creuse; matrice symétrique; matrice-bandes; etc.

Une des premières caractéristiques de la matrice à considérer est le nombre et, éventuellement la répartition des éléments nuls de la matrice. C'est ainsi que l'on distingue les matrices pleines et le matrices creuses. Les problème de résolution de système linéaire les plus faciles à traiter numériquement sont ceux dont les matrices sont symétriques définies positives.

Les problème de calcul des valeurs propres et des vecteurs propres d'une matrice les plus faciles à traiter numériquement sont ceux des matrices sont symétriques. Autant il existe les trois types de méthodes directe, récursive et itérative pour le problème (1), les méthodes de calcul des valeurs propres ne sauraient qu'être itératives. En effet, les valeurs propres d'une matrice sont les racines d'un polynôme de degré n. Or l'existence d'une méthode directe pour le calcul des valeurs propres équivaudrait à affirmer qu'on peut calculer les racines d'un polynôme arbitraire en un nombre fini d'opérations élémentaires.

E.8.1 Conditionnement d'un problème de résolution des systèmes linéaires

Considérons la matrice inversible

$$\boldsymbol{A} = \begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix}$$

et son inverse

$$\boldsymbol{A}^{-1} = \begin{pmatrix} 25 & -41 & 10 & -6 \\ -41 & 68 & -17 & 10 \\ 10 & -17 & 5 & -3 \\ -6 & 10 & -3 & 2 \end{pmatrix}$$

et les deux vecteurs $\boldsymbol{x}^t = [x_1, \cdots, x_4]$ et $\boldsymbol{y}^t = [32,23,33,31]$. La solution exacte de l'équation $\boldsymbol{Ax} = \boldsymbol{y}$ est $\boldsymbol{x}^t = [1,1,1,1]$. Considérons maintenant le vecteur perturbé $[\boldsymbol{y} + \delta\boldsymbol{y}]^t = [32.1,22.9,33.1,30.9]$ très légèrement différent à $\boldsymbol{y}$. La solution de l'équation $\boldsymbol{Ax} = [\boldsymbol{y} + \delta\boldsymbol{y}]$ est

$$\boldsymbol{x}^t = [9.2, -12.6, 4.5, -1.1]$$

Autrement dit une erreur relative de 1/200 sur les données $\boldsymbol{y}$ entraîne une erreur relative de 10/1 sur le résultat (soit un rapport d'amplification des erreurs relatives de l'ordre de 2000) de la solution du système linéaire. Considérons également une matrice $\boldsymbol{A}$ légèrement modifiée :

$$\boldsymbol{A} = \begin{pmatrix} 10 & 7 & 8.1 & 7.2 \\ 7.08 & 5.04 & 6 & 5 \\ 8 & 5.98 & 9.89 & 9 \\ 6.99 & 4.99 & 9 & 9.98 \end{pmatrix}$$

cette fois ci encore la solution de l'équation $\boldsymbol{Ax} = \boldsymbol{y}$ est $\boldsymbol{x}^t = [-81{,}137, -34{,}22]$. Là encore, de très petites variations des éléments de la matrice modifient complètement le résultat. Pourtant, la matrice $\boldsymbol{A}$ du système a un bon aspect: elle est symétrique, son déterminant vaut 1, et la matrice inverse $\boldsymbol{A}^{-1}$ est tout aussi sympathique! Ces deux exemples montrent qu'il est utile d'étudier le conditionnement d'une matrice quand on a à résoudre un système d'équations linéaires même si la solution est calculée par une méthode directe. Analyse des erreurs permet d'établir facilement les relations suivantes: soient $\boldsymbol{Ax} = \boldsymbol{y}$, $\boldsymbol{A}(\boldsymbol{x} + \delta\boldsymbol{x}) = \boldsymbol{y} + \delta\boldsymbol{y}$. Alors on a:

$$\frac{\|\delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leq \|\boldsymbol{A}\| \, \|\boldsymbol{A}^{-1}\| \frac{\|\delta\boldsymbol{y}\|}{\|\boldsymbol{y}\|}$$

soient

$$\boldsymbol{Ax} = \boldsymbol{y}, \quad (\boldsymbol{A} + \Delta\boldsymbol{A})(\boldsymbol{x} + \delta\boldsymbol{x}) = \boldsymbol{y}$$

Alors

$$\frac{\|\delta\boldsymbol{x}\|}{\|\boldsymbol{x} + \delta\boldsymbol{x}\|} \leq \|\boldsymbol{A}\| \, \|\boldsymbol{A}^{-1}\| \frac{\|\Delta\boldsymbol{A}\|}{\|\boldsymbol{A}\|}$$

On définit ainsi le nombre de conditionnement de la matrice $\boldsymbol{A}$ par

$$\text{cond}\,\{\boldsymbol{A}\} = \|\boldsymbol{A}\| \, \|\boldsymbol{A}^{-1}\|$$

Ces deux inégalités montre que le nombre cond $\{\boldsymbol{A}\}$ mesure la sensibilité de la solution $\boldsymbol{x}$ du système linéaire $\boldsymbol{Ax} = \boldsymbol{y}$ vis- à-vis des variations sur les données $\boldsymbol{A}$ et $\boldsymbol{y}$, qualité que l'on appelle le conditionnement du système linéaire considéré. Un système linéaire est bien-ou-mal-conditionné selon que cond $\{\boldsymbol{A}\}$ est petit ou grand. Par la suite nous présentons un certains nombre de théorèmes et résultats (sans démonstrations) concernant cond $\{\boldsymbol{A}\}$.

Théorème 3 *Soit $\boldsymbol{A}$ une matrice inversible, et soit $\boldsymbol{x}$ et $(\boldsymbol{x}+\delta\boldsymbol{x})$ les solutions des systèmes linéaires $\boldsymbol{Ax} = \boldsymbol{y}$ et $\boldsymbol{A}(\boldsymbol{x} + \delta\boldsymbol{x}) = \boldsymbol{y} + \delta\boldsymbol{y}$. Alors l'inégalité*

$$\frac{\|\delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leq cond\{\boldsymbol{A}\}\,\frac{\|\delta\boldsymbol{y}\|}{\|\boldsymbol{y}\|}$$

est satisfaite, et c'est la meilleure possible— pour une matrice $\boldsymbol{A}$ donnée, on peut trouver des vecteurs $\boldsymbol{y} \neq 0$ et $\delta\boldsymbol{y} \neq 0$ tels qu'elle devienne une égalité.

Théorème 4 *Soit $\boldsymbol{A}$ une matrice inversible, et soit $\boldsymbol{x}$ et $(\boldsymbol{x}+\delta\boldsymbol{x})$ les solutions des systèmes linéaires $\boldsymbol{Ax} = \boldsymbol{y}$ et $(\boldsymbol{A} + \Delta\boldsymbol{A})(\boldsymbol{x} + \delta\boldsymbol{x}) = \boldsymbol{y}$. Alors l'inégalité*

$$\frac{\|\delta\boldsymbol{x}\|}{\|\boldsymbol{x} + \delta\boldsymbol{x}\|} \leq cond\{\boldsymbol{A}\}\,\frac{\|\Delta\boldsymbol{A}\|}{\|\boldsymbol{A}\|}$$

est satisfaite, et c'est la meilleure possible— pour une matrice $\boldsymbol{A}$ donnée, on peut trouver des vecteurs $\boldsymbol{y} \neq 0$ et une matrice $\Delta\boldsymbol{A} \neq 0$ tels qu'elle devienne une égalité. On a par ailleurs l'inégalité

$$\frac{\|\delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leq cond\{\boldsymbol{A}\}\,\frac{\|\Delta\boldsymbol{A}\|}{\|\boldsymbol{A}\|}\,\frac{1}{1 - \|\boldsymbol{A}^{-1}\|\,\|\Delta\boldsymbol{A}\|}$$

Théorème 5 *Pour toute matrice $\boldsymbol{A}$*

1. *cond $\{\boldsymbol{A}\} \geq 1$, cond $\{\boldsymbol{A}\} =$ cond $\left\{\boldsymbol{A}^{-1}\right\}$, cond $\{\alpha\boldsymbol{A}\} =$ cond $\{\boldsymbol{A}\}$ si $\alpha \neq 0$.*
2. *$cond_2(\boldsymbol{A}) = \frac{\mu_n(\boldsymbol{A})}{\mu_1(\boldsymbol{A})}$ où $\mu_1(\boldsymbol{A})$ et $\mu_n(\boldsymbol{A})$ sont respectivement la plus petite et la plus grande valeur singulière de la matrice $\boldsymbol{A}$.*
3. *Si $\boldsymbol{A}$ est une matrice normale; alors*

$$cond_2(\boldsymbol{A}) = \frac{\max_i |\lambda_i(\boldsymbol{A})|}{\min_i |\lambda_i(\boldsymbol{A})|}$$

   où $\lambda_i(\boldsymbol{A})$ sont les valeurs propres de $\boldsymbol{A}$.
4. *Si $\boldsymbol{A}$ est une matrice unitaire ou orthogonale $cond_2(\boldsymbol{A}) = 1$.*
5. *Le conditionnement $cond_2(\boldsymbol{A})$ est invariante par transformation unitaire:*

$$\boldsymbol{UU}^* = I \longrightarrow cond_2(\boldsymbol{AU}) = cond_2(\boldsymbol{UA}) = cond_2(\boldsymbol{U}^*\boldsymbol{AU}) = cond_2(\boldsymbol{A})$$

Dans l'exemple (1) on a

$$\lambda_1 = 0.01015 < \lambda 2 = 0.8431 < \lambda 3 = 3.858 < \lambda 4 = 30.2887$$

$$\text{cond}_2(\boldsymbol{A}) = \frac{\lambda_4}{\lambda_1} = 2984$$

On note que le calcul du $\text{cond}_2(\boldsymbol{A})$ repose sur la connaissance des valeurs propres extrêmes de la matrice $\boldsymbol{A}$. En l'absence de cette information on peut utiliser la relation suivante

$$\|\boldsymbol{A}\|_2 \leq \|\boldsymbol{A}\|_E \leq \sqrt{n}\|\boldsymbol{A}\|_2$$

à condition toutefois qu'on connaisse la matrice $\boldsymbol{A}^{-1}$. Un système linéaire est autant mieux conditionné que le nombre $\text{cond}\{\boldsymbol{A}\}$ est voisin de 1 (sous-entendu relativement à une norme matricielle subordonnée particulière). Un système linéaire $\boldsymbol{Ax} = \boldsymbol{b}$ à matrice orthogonale (ou unitaire) est très bien conditionné puisque $\text{cond}_2(\boldsymbol{A}) = 1$. Les transformations orthogonales ou unitaires préservent le conditionnement $\text{cond}_2(\boldsymbol{A})$.

E.8.2 Conditionnement d'un problème de calcul des valeurs singulières d'une matrice

Considérons la matrice carrée d'ordre n

$$\boldsymbol{A}(\epsilon) = \begin{pmatrix} 0 & & & \epsilon \\ 1 & 0 & & \\ & 1 & & \\ & & \cdot & \cdot \\ & & 1 & 0 \end{pmatrix}$$

Pour $n = 40$ les valeurs propres de $\boldsymbol{A}(0)$ sont égales à 0 : par contre les valeurs propres de $\boldsymbol{A}(10^{-4})$ ont toutes les mêmes module 10^{-1} (les racine n-ème de ϵ). On constate ainsi que l'on a besoin dans ce problème de calcul des valeurs propres; de mesurer le conditionnement de la matrice.

Théorème 6 (de Bauer-Fike) *Soit $\boldsymbol{A}$ une matrice diagonalisable; $\boldsymbol{P}$ une matrice telle que*

$$\boldsymbol{P}^{-1}\boldsymbol{AP} = diag\,[\lambda_i]$$

et $\|.\|$ une norme matricielle telle que

$$\|\,diag\,[d_i]\,\| = \max_i \|d_i\|$$

pour toute matrice diagonale. Alors pour toute matrice $\Delta\boldsymbol{A}$

$$sp\,[\boldsymbol{A} + \Delta\boldsymbol{A}] \in \prod_{i=1}^{N} x_i \quad où \quad x_i = \{z \in C || z - \lambda_i| \leq cond\,\{\boldsymbol{P}\}\,\|\Delta\boldsymbol{A}\|\}$$

E.9 Méthodes récursives de résolution des systèmes d'équations linéaires

Principe

On considère le système d'équations linéaires

$$\boldsymbol{R}\boldsymbol{a} = -\boldsymbol{d}$$

où $\boldsymbol{R}$ est une matrice de dimensions $[N,N]$; $\boldsymbol{a}$ un vecteur de dimension $[N]$ contenant les paramètres inconnus; et $\boldsymbol{d}$ un vecteur de données de dimension $[N]$. Supposons que $\boldsymbol{R}$ soit partitionnée de façon suivante:

$$\boldsymbol{R}_{m+1} = \begin{pmatrix} \boldsymbol{R}_m & \boldsymbol{x}_m \\ \boldsymbol{z}_m^t & y_m \end{pmatrix}, \quad 1 \leq m \leq N-1$$

$$\boldsymbol{R}_m\boldsymbol{a}_m = -\boldsymbol{d}_m, \quad \boldsymbol{R}_{m+1}\boldsymbol{a}_{m+1} = -\boldsymbol{d}_{m+1}, \quad \boldsymbol{d}_{m+1}^t = [\boldsymbol{d}_m^t, d_{m+1}]$$

Appliquant le lemme d'inversion de matrices en bloc à la matrice ainsi partionnée; on obtient:

$$\alpha_m \boldsymbol{R}_{m+1}^{-1} = \begin{pmatrix} \alpha_m \boldsymbol{R}_m^{-1} + \boldsymbol{w}_m \boldsymbol{v}_m^t & \boldsymbol{w}_m \\ \boldsymbol{v}_m^t & 1 \end{pmatrix}, \quad \leq m \leq N-1$$

où

$$\alpha_m = y_m + \boldsymbol{z}_m^t \boldsymbol{w}_m = y_m + \boldsymbol{v}_m^t \boldsymbol{x}_m = \frac{\det[\boldsymbol{R}_{m+1}]}{\det[\boldsymbol{R}_m]}$$

$$\boldsymbol{w}_m = -\boldsymbol{R}_m^{-1}\boldsymbol{x}_m, \quad \boldsymbol{v}_m = -\boldsymbol{R}_m^{-t}\boldsymbol{z}_m$$

Si on s'intéresse à la résolution de l'équation (1.1); plutôt qu'à l'inversion de la matrice (1.2); ces équations peuvent être écrites sous une autre forme:

$$\boldsymbol{a}_{m+1} = \begin{pmatrix} \boldsymbol{a}_m \\ 0 \end{pmatrix} + k_{m+1} \begin{pmatrix} \boldsymbol{w}_m \\ 1 \end{pmatrix}$$

$$k_{m+1} = -\frac{\beta_m}{\alpha_m}, \quad \beta_m = \boldsymbol{v}_m^t \boldsymbol{d}_m + d_{m+1} = \boldsymbol{z}_m^t \boldsymbol{a}_m + d_{m+1}$$

Le coût de calcul de cet algorithme est $O(N^3)$. Son intérêt particulier est dans le fait qu'on dispose de la valeur du déterminant de la matrice à chaque itération; ce qui est utile dans le contrôle de stabilité de l'algorithme pour l'inversion des matrices mal-conditionnées. D'autre part; parmi tous les paramètres utilisés dans cet algorithme, c'est à dire $\boldsymbol{w}_m$,$\boldsymbol{v}_m$,α_m et β_m, seul ce dernier dépend des données.

Cas des matrices Symétriques

Dans ce cas on a

$$\boldsymbol{x}_m = \boldsymbol{z}_m = \boldsymbol{r}_m, \quad \text{et} \quad \boldsymbol{w}_m = \boldsymbol{v}_m = \boldsymbol{R}_m^{-1}\boldsymbol{r}_m$$

$$\alpha_m = y_m + \boldsymbol{r}_m^t \boldsymbol{w}_m$$

$$\alpha_m \boldsymbol{R}_{m+1}^{-1} = \begin{pmatrix} \alpha_m \boldsymbol{R}_m^{-1} + \boldsymbol{v}_m \boldsymbol{v}_m^t & \boldsymbol{v}_m \\ \boldsymbol{v}_m^t & 1 \end{pmatrix}, \quad \leq m \leq N-1$$

Cas des matrices MID

$$\boldsymbol{w}_m = -\boldsymbol{R}_m^{-1}\boldsymbol{x}_m, \quad \boldsymbol{w}_{m+1} = -\boldsymbol{R}_{m+1}^{-1}\boldsymbol{x}_{m+1}$$

En utilisant les Matrices MID; c'est à dire:

$$\boldsymbol{x}_{m+1}^t = [\boldsymbol{x}_m^t; x_{m+1}], \quad \boldsymbol{z}_{m+1}^t = [\boldsymbol{z}_m^t; z_{m+1}]; \quad \boldsymbol{y}_{m+1}^t = [\boldsymbol{y}_m^t; y_{m+1}]$$

on obtient une équation de récurrence pour $\boldsymbol{w}_{m+1}$ et $\boldsymbol{v}_{m+1}$ en fonction de $\boldsymbol{w}_m$ et $\boldsymbol{v}_m$.

$$\boldsymbol{w}_{m+1} = \begin{pmatrix} \boldsymbol{w}_m \\ 0 \end{pmatrix} + k_{m+1}^+ \begin{pmatrix} \boldsymbol{w}_m \\ 1 \end{pmatrix}, \quad \boldsymbol{v}_{m+1} = \begin{pmatrix} \boldsymbol{v}_m \\ 0 \end{pmatrix} + k_{m+1}^- \begin{pmatrix} \boldsymbol{v}_m \\ 1 \end{pmatrix}$$

$$k_{m+1}^+ = -\frac{\beta_m^+}{\alpha_m} = \frac{\boldsymbol{v}_m^t\boldsymbol{x}_m + x_{m+1}}{\alpha_m}, \quad k_{m+1}^- = -\frac{\beta_m^-}{\alpha_m} = \frac{\boldsymbol{z}_m^t\boldsymbol{w}_m + z_{m+1}}{\alpha_m}$$

$$\alpha_m = \boldsymbol{z}_m^t\boldsymbol{w}_m + y_{m+1} = \boldsymbol{v}_m^t\boldsymbol{x}_m + y_{m+1}$$

On peut encore simplifier ces équations en trouvant une relation de récurrence pour α_m,β_m^+,β_m^- et β_m :

$$\alpha_m = \alpha_{m-1} + k_m^+\beta_{m-1}^- + (y_{m+1} - y_m)$$

$$\beta_m^+ = (x_{m+1} - y_{m+1}) + \alpha_m, \quad \beta_m^- = (z_{m+1} - y_{m+1}) + \alpha_m$$

$$\beta_m = \beta_{m-1}(1 - k_m^-) + (d_{m+1} - d_m)$$

Ces équations simplifient plus si $y_m = x_m$ et / ou $y_m = z_m$. L'inversion de ces matrice nécessitent $O(p^2)$ opérations élémentaires.

Cas des matrices MIP

Il n'existe pas encore des méthodes rapides directes pour inversion des matrices MIP dans le cas général car ces matrices peuvent avoir un rang de déplacement entre $0,\cdots,N$; $[N;N]$ étant les dimensions de la matrice. Par contre un cas particulier de ces matrices est la matrice de Toeplitz ($y_i = y_0$,$i = 1$,..,N) non-symétrique ($x_i \neq z_i$) ou symétrique ($x_i = z_i$).

Cas des matrices de Toeplitz

Une matrice de Toeplitz satisfait la propriété suivant:

$$\boldsymbol{J}\boldsymbol{R}_m\boldsymbol{J} = \boldsymbol{R}_m^t, \quad m = 1,\cdots,N$$

où $\boldsymbol{J}$ est une matrice d'inversion d'ordre de dimensions $[M,M]$; ce qui vaut dire qu'une matrice de Toeplitz est une matrice symétrique par rapport à sa deuxième diagonale. Si on note par

$$\widetilde{\boldsymbol{x}}_m = [x_1; x_2, \cdots, x_m]^t, \quad \widetilde{\boldsymbol{z}}_m = [x_1; x_2, \cdots, x_m]^t$$

$$\boldsymbol{w}_{m+1} = -\boldsymbol{R}_{m+1}^{-1}\boldsymbol{J}\widetilde{\boldsymbol{x}}_{m+1}, \quad \boldsymbol{w}_{m+1} = -\boldsymbol{J}\boldsymbol{R}_{m+1}^{-t}\widetilde{\boldsymbol{x}}_{m+1}$$

$$\boldsymbol{v}_{m+1}^t = -\widetilde{\boldsymbol{z}}_{m+1}^t\boldsymbol{J}\boldsymbol{R}_{m+1}^{-1} = -\widetilde{\boldsymbol{z}}_{m+1}^t\boldsymbol{R}_{m+1}^{-t}\boldsymbol{J}$$

$$\boldsymbol{J}\boldsymbol{w}_{m+1} = \begin{pmatrix} \boldsymbol{J}\boldsymbol{w}_m \\ 0 \end{pmatrix} + k_{m+1}^+ \begin{pmatrix} \boldsymbol{v}_m \\ 1 \end{pmatrix}, \quad \boldsymbol{J}\boldsymbol{v}_{m+1} = \begin{pmatrix} \boldsymbol{J}\boldsymbol{v}_m \\ 0 \end{pmatrix} + k_{m+1}^- \begin{pmatrix} \boldsymbol{w}_m \\ 1 \end{pmatrix}$$

$$k_{m+1}^+ = -\frac{\beta_m^+}{\alpha_m} = -\frac{\boldsymbol{w}_m^t\widetilde{\boldsymbol{x}}_m + x_{m+1}}{\alpha_m}, \quad k_{m+1}^- = -\frac{\beta_m^-}{\alpha_m} = -\frac{\widetilde{\boldsymbol{z}}_m^t\boldsymbol{w}_m + z_{m+1}}{\alpha_m}$$

$$\alpha_m = \boldsymbol{z}_m^t\boldsymbol{w}_m + y_0 = \boldsymbol{x}_m^t\boldsymbol{v}_m + y_0$$

Notons que les récusions pour $\boldsymbol{w}$ et $\boldsymbol{v}$ sont couplées (ce qui n'était pas le cas pour les matrice MID). Le terme α_m peut être calculé d'une façon plus efficace par

$$\alpha_m = \alpha_{m-1}(1 - k_m^+k_m^-)$$

On peut obtenir ces mêmes relations pour les matrices de Hankel en remarquant qu'une matrice de Hankel $\boldsymbol{H}$ et une matrice de Toeplitz $\boldsymbol{T}$ sont reliées par $\boldsymbol{H} = \boldsymbol{T}\boldsymbol{J}$ ou $\boldsymbol{H} = \boldsymbol{J}\boldsymbol{T}$. Ainsi les trois relations suivantes décrivent un même système d'équations linéaires:

$$\boldsymbol{H}\boldsymbol{a} = -\boldsymbol{d}, \quad \boldsymbol{T}(\boldsymbol{J}\boldsymbol{a}) = -\boldsymbol{d}, \quad \boldsymbol{T}\boldsymbol{a} = -\boldsymbol{J}\boldsymbol{d}$$

il suffit alors de remplacer $\boldsymbol{d}$ par $\boldsymbol{J}\boldsymbol{d}$ pour obtenir les relations nécessaires pour des matrices de Hankel.

Cas des matrices de Toeplitz Symétriques

Algorithme de Levinson :
Dans ce cas on a la relation $\boldsymbol{w} = \boldsymbol{v}$ ce qui va permettre de simplifier encore plus les relations précédentes. On a

$$k^+ = k^- = k$$

$$\boldsymbol{J}\boldsymbol{w}_{m+1} = \begin{pmatrix} \boldsymbol{J}\boldsymbol{w}_m \\ 0 \end{pmatrix} + k_{m+1} \begin{pmatrix} \boldsymbol{v}_m \\ 1 \end{pmatrix}$$

$$\beta = \boldsymbol{v}_{m+1}^t\widetilde{\boldsymbol{x}}_m + x_{m+1}, \quad \alpha_m = \alpha_{m-1}(1 - k_m^2)$$

Algorithme de Durbin :
Dans le cas particulier où le le vecteur $\boldsymbol{d}$peut être écrit sous la forme

$$\boldsymbol{d}_N^t = [\boldsymbol{d}_{N-1}^t + d_N]$$

comme c'est le cas du problème de la modélisation AR; on a la relation

$$\boldsymbol{w}_m = \boldsymbol{J}\boldsymbol{a}_m, \quad m = 1,2,\cdots,N-1$$

Cas des matrices circulantes

Dans le cas des matrices circulantes on a

$$\boldsymbol{z}_{N-1} = \boldsymbol{J}\boldsymbol{x}_{N-1}, \quad \boldsymbol{v}_{N-1} = \boldsymbol{J}\boldsymbol{w}_{N-1}$$

Il faut noter que l'on peut diagonaliser les matrices circulantes par TFD et obtenir des algorithmes non récursifs.

Cas des matrices de Toeplitz en bande

E.10 Diverses propriétés

Lemme de factorisation d'une matrice

Soit $\boldsymbol{A}$ une matrice de dimension [N,P]. On peut factoriser la matrice $[\boldsymbol{I}_N \pm \boldsymbol{A}\boldsymbol{A}^t]$ de façon suivante:

$$\boldsymbol{I}_N \pm \boldsymbol{A}\boldsymbol{A}^t = [\boldsymbol{I}_N \pm \boldsymbol{A}\boldsymbol{R}^{-1}\boldsymbol{A}^t].[\boldsymbol{I}_N \pm \boldsymbol{A}\boldsymbol{R}^{-1}\boldsymbol{A}^t]^t$$

où $\boldsymbol{R} = \boldsymbol{I}_{\boldsymbol{P}} + [\boldsymbol{I}_{\boldsymbol{P}} \pm \boldsymbol{A}^t\boldsymbol{A}]^{1/2}$.

Racines carrées d'une matrice définie positive

Toute matrice définie positive admet une racine carrée. On appelle racine carrée d'une matrice $\boldsymbol{A}$ toute matrice carrée $\boldsymbol{S}$, telle que

$$\boldsymbol{A} = \boldsymbol{S}\boldsymbol{S}^t, \quad \boldsymbol{S} = \boldsymbol{A}^{1/2}, \quad \boldsymbol{S}^t = \boldsymbol{A}^{t/2}.$$

Une telle factorisation n'est pas unique. Si $\boldsymbol{T}$ est orthogonale ($\boldsymbol{T}\boldsymbol{T}^t = \boldsymbol{I}$), alors $\boldsymbol{S}\boldsymbol{T}$ est aussi racine carrée de $\boldsymbol{A}$ puisque $\boldsymbol{A} = \boldsymbol{S}\boldsymbol{T}\boldsymbol{T}^t\boldsymbol{S}^t = \boldsymbol{S}\boldsymbol{S}^t$.

Décomposition de Cholesky d'une matrice définie positive

Soit $\boldsymbol{A}$ une matrice symétrique, réelle, et définie positive d'ordre N. Il existe alors une matrice $\boldsymbol{O}$ orthogonale réelle d'ordre N telle que $\boldsymbol{D} = \boldsymbol{O}^{-1}\boldsymbol{A}\boldsymbol{O} = \boldsymbol{O}^t\boldsymbol{A}\boldsymbol{O}$ soit diagonale. De plus, si $\boldsymbol{A}$ est de rang $K < N$, elle admet une décomposition de la forme

$$\boldsymbol{A} = \boldsymbol{S}_1\boldsymbol{S}_1^t - \boldsymbol{S}_2\boldsymbol{S}_2^t$$

où $\boldsymbol{S}_1$ et $\boldsymbol{S}_2$ sont de dimensions respectives [N,α] et [N,$K-\alpha$], α étant le nombre de valeurs propres positives non nulles de $\boldsymbol{A}$. La décomposition de Cholesky consiste à imposer aux racines carrées d'être des matrices triangulaires. La décomposition est alors unique.

Le rang de déplacement d'une matrice

Le rang de déplacement d'une matrice est le rang de la différence entre la matrice initiale et la matrice décalée d'un cran le long de sa diagonale principale. On définit deux rangs de déplacement, inférieur δ_- et supérieur δ_+, à partir des deux déplacements possibles (vers le haut-gauche, et vers le bas-droite):

$$\delta_+(\boldsymbol{R}) = \text{rang}\left[\boldsymbol{R} - \boldsymbol{Z}\boldsymbol{R}\boldsymbol{Z}^t\right], \quad \delta_-(\boldsymbol{R}) = \text{rang}\left[\boldsymbol{R} - \boldsymbol{Z}^t\boldsymbol{R}\boldsymbol{Z}\right]$$

où $\boldsymbol{Z}$ est une matrice de décalage vers le bas donnée par:

$$\boldsymbol{Z} = \begin{pmatrix} 0 & . & . & . & 0 \\ 1 & 0 & . & . & 0 \\ 0 & 1 & 0 & . & 0 \\ . & 0 & 1 & 0 & 0 \\ . & . & 0 & 1 & 0 \end{pmatrix}$$

Pour une matrice quelconque $\boldsymbol{R}$ on peut toujours construire deux partitions:

$$\boldsymbol{R} = \begin{pmatrix} r_0 & \boldsymbol{x}^t \\ \boldsymbol{y} & \boldsymbol{R}_i \end{pmatrix} = \begin{pmatrix} \boldsymbol{R}_s & \boldsymbol{u} \\ \boldsymbol{v}^t & s_0 \end{pmatrix}$$

Ainsi, pour cette matrice on a:

$$\boldsymbol{R} - \boldsymbol{Z}\boldsymbol{R}\boldsymbol{Z}^t = \begin{pmatrix} \boldsymbol{R}_i - \boldsymbol{R}_s & \boldsymbol{u} \\ \boldsymbol{v}^t & s_0 \end{pmatrix}, \quad \text{et} \quad \boldsymbol{R} - \boldsymbol{Z}^t\boldsymbol{R}\boldsymbol{Z} = \begin{pmatrix} r_0 & \boldsymbol{x}^t \\ \boldsymbol{y} & \boldsymbol{R}_i - \boldsymbol{R}_s \end{pmatrix}$$

Dans le cas des matrices symétriques on a $\boldsymbol{u} = \boldsymbol{v} = \boldsymbol{s}$, $\boldsymbol{x} = \boldsymbol{y} = \boldsymbol{r}$. On peut alors construire deux partitions de la matrice initiale $\boldsymbol{R}$:

$$\boldsymbol{R} = \begin{pmatrix} r_0 & \boldsymbol{r}^t \\ \boldsymbol{r} & \boldsymbol{R}_i \end{pmatrix} = \begin{pmatrix} \boldsymbol{R}_s & \boldsymbol{s} \\ \boldsymbol{s}^t & s_0 \end{pmatrix}$$

Dans ce cas on a

$$\boldsymbol{R} - \boldsymbol{Z}\boldsymbol{R}\boldsymbol{Z}^t = \begin{pmatrix} \boldsymbol{R}_i - \boldsymbol{R}_s & \boldsymbol{s} \\ \boldsymbol{s}^t & s_0 \end{pmatrix} \quad \text{et} \quad \boldsymbol{R} - \boldsymbol{Z}^t\boldsymbol{R}\boldsymbol{Z} = \begin{pmatrix} r_0 & \boldsymbol{r}^t \\ \boldsymbol{r} & \boldsymbol{R}_i - \boldsymbol{R}_s \end{pmatrix}$$

Pour une matrice de Toeplitz, on vérifie facilement, qu'on a: $\delta_+(\boldsymbol{T}) = \delta_-(\boldsymbol{T}) = 2$ On démontre aussi facilement [] que:

$$\delta_+(\boldsymbol{R}) = \delta_-\boldsymbol{R}^{-1} \quad \text{et} \quad \delta_-(\boldsymbol{R}) = \delta_+\boldsymbol{R}^{-1}\det\left[\delta_+(\boldsymbol{R}) - \delta_-(\boldsymbol{R})\right] \leq 2$$

Distance à Toeplitz

Le rang de déplacement $\delta_-(\boldsymbol{R})$ ou $\delta_+(\boldsymbol{R})$ ne constitue pas à proprement parler une mesure de la distance à Toeplitz puisque $\delta_-(\boldsymbol{T}) = \delta_-(\boldsymbol{T}) = 2$. C'est la raison pour laquelle on utilise aussi une définition différente du rang de déplacement qui fourni directement une *distance à Toeplitz*.

Soit $\boldsymbol{R}$ une matrice-bloc quelconque de dimensions $[NP,NP]$ $\boldsymbol{R} = [\boldsymbol{R}_{ji}]$, $i,j = 1,\cdots,N$ où $\boldsymbol{R}_{ij}$ sont des matrices de dimensions $[P,P]$. Le déplacement est alors défini par

$$\delta\boldsymbol{R} = [\boldsymbol{R}_{(i+1),(j+1)} - \boldsymbol{R}_{ij}], \quad i,j = 1,\cdots,N-1$$

et le rang de déplacement devient $\alpha =$ partie entière de $[\frac{\text{rang}[\delta\boldsymbol{R}]}{P}]$.

- Quand les éléments de $\boldsymbol{R}$ sont scalaires, $P = 1$ et $\alpha = \text{rang}\,[\delta\boldsymbol{R}]$.
- Si $\boldsymbol{R}$ est une matrice quelconque, $\delta\boldsymbol{R}$ peut être de rang plein et alors $\alpha = N - 1$.
- Si $\boldsymbol{R}$ est Toeplitz, alors $\alpha = 0$. On a en général $0 \leq \alpha \leq N - 1$, et α peut servir directement de mesure de la distance de $\boldsymbol{R}$ à Toeplitz.

Annexe F

Quelques transformations utiles

L'objet de cette annexe est de rappeler un certain nombre de transformation linéaires utiles en 1-D, 2-D et n-D. Nous donnons aussi l'expression de la transformée inverse de ces transformations lorsqu'elle existe.

F.1 Transformations 1-D

$$F(u) = \int f(x)h(u,x)\,\mathrm{d}x$$

Identité :	$h(u,x) = \delta(u-x)$	$F(u) = \int f(x)\delta(u-x)\,\mathrm{d}x = f(x)$
Convolution :	$h(u,x) = h(u-x)$	$F(u) = \int f(x)h(u-x)\,\mathrm{d}x = f(x) * h(x)$
Corrélation :	$h(u,x) = h(u+x)$	$F(u) = \int f(x)h(u+x)\,\mathrm{d}x = f(x) \star h(x)$
Fourier :	$h(u,x) = \exp[-jux]$	$F(u) = \int f(x)\exp[-jux]\,\mathrm{d}x$
Laplace :	$h(u,x) = \exp[-ux]$	$F(u) = \int f(x)\exp[-ux]\,\mathrm{d}x$

Transformations unitaires :

$$F(u) = \int f(x) h(u,x)\,\mathrm{d}x$$
$$f(x) = \int F(u) h^*(u,x)\,\mathrm{d}u$$

Fourier :	$h(u,x) = \exp[-jux]$	$F(u) = \frac{1}{\sqrt{2\pi}} \int f(x) \exp[-jux]\,\mathrm{d}x$ $f(x) = \frac{1}{\sqrt{2\pi}} \int F(u) \exp[+jux]\,\mathrm{d}u$
Cosinus :	$h(u,x) = \cos(ux)$	$F(u) = \frac{2}{\sqrt{\pi}} \int_0^\infty f(x) \cos(ux)\,\mathrm{d}x$ $f(x) = \frac{2}{\sqrt{\pi}} \int_0^\infty F(u) \cos(ux)\,\mathrm{d}u$
Sinus :	$h(u,x) = \sin(ux)$	$F(u) = \frac{2}{\sqrt{\pi}} \int_0^\infty f(x) \sin(ux)\,\mathrm{d}x$ $f(x) = \frac{2}{\sqrt{\pi}} \int_0^\infty F(u) \sin(ux)\,\mathrm{d}u$
Hartley :	$h(u,x) = \frac{1}{\sqrt{2\pi}}[\cos(ux) + \sin(ux)]$	$F(u) = \frac{1}{\sqrt{2\pi}} \int_0^\infty f(x)[\cos(ux) + \sin(ux)]\,\mathrm{d}x$ $f(x) = \frac{1}{\sqrt{2\pi}} \int_0^\infty F(u)[\cos(ux) + \sin(ux)]\,\mathrm{d}u$
Hilbert :	$h(u,x) = \frac{1}{\pi}(u-x)^{-1}$	$F(u) = \frac{1}{\pi} \int \frac{f(x)}{(x-u)}\,\mathrm{d}x$ $f(x) = \frac{1}{\pi} \int \frac{F(u)}{(x-u)}\,\mathrm{d}u$
Hankel :	$h(u,x) = J_\nu(ux)$	$F(u) = \int_0^\infty f(x) J_\nu(ux)\,\mathrm{d}x$ $f(x) = \int_0^\infty F(u) J_\nu(ux)\,\mathrm{d}u$
Mellin :	$h(u,x) = x^{u-1}$	$F(u) = \int f(x) x^{u-1}\,\mathrm{d}x$
Abel :	$h(u,x) = \frac{1}{\Gamma(\alpha)}(x-u)^{\alpha-1}$	$F(u) = \frac{1}{\Gamma(\alpha)} \int f(x)(x-u)^{\alpha-1}\,\mathrm{d}x$
Weyl :	$h(u,x) = \frac{1}{\Gamma(\alpha)}(u-x)^{\alpha-1}$	$F(u) = \frac{1}{\Gamma(\alpha)} \int f(x)(u-x)^{\alpha-1}\,\mathrm{d}x$

F.2 Transformations 2-D

$$F(u,v) = \iint f(x,y)h(u,v;x,y)\,\mathrm{d}x\,\mathrm{d}y$$

Identité :
$$h(u,v;x,y) = \delta(u-x;v-y)$$
$$F(u,v) = \iint f(x,y)\delta(u-x;v-y)\,\mathrm{d}x\,\mathrm{d}y = f(x,y)$$

Séparable :
$$h(u,v;x,y) = h_1(u,x)h_2(v,y)$$
$$F(u,v) = \iint f(x,y)h_1(u,x)h_2(v,y)\,\mathrm{d}x\,\mathrm{d}y$$

Séparable & Symétrique :
$$h(u,v;x,y) = h_1(u,x)h_1(v,y)$$
$$F(u,v) = \iint f(x,y)h_1(u,x)h_1(v,y)\,\mathrm{d}x\,\mathrm{d}y$$

Convolution :
$$h(u,v;x,y) = h(u-x;v-y)$$
$$F(u,v) = \iint f(x,y)h(u-x;v-y)\,\mathrm{d}x\,\mathrm{d}y = f(x,y) * h(x,y)$$

Corrélation :
$$h(u,v;x,y) = h(u+x;v+y)$$
$$F(u,v) = \iint f(x,y)h(u+x;v+y)\,\mathrm{d}x\,\mathrm{d}y = f(x,y) \star h(x,y)$$

Laplace :
$$h(u,v;x,y) = \exp\left[-(ux+vy)\right]$$
$$F(u,v) = \iint f(x,y)\exp\left[-(ux+vy)\right]\,\mathrm{d}x\,\mathrm{d}y$$

Transformations 2-D unitaires

Fourier : $h(u,v;x,y) = \exp\left[-j(ux+vy)\right]$
$F(u,v) = \iint f(x,y)\exp\left[-j(ux+vy)\right]\,\mathrm{d}x\,\mathrm{d}y$

Cosinus : $h(u,v;x,y) = \cos(ux)\cos(vy)$
$F(u,v) = \iint f(x,y)\cos(ux)\cos(vy)\,\mathrm{d}x\,\mathrm{d}y$

Sinus : $h(u,v;x,y) = \sin(ux)\sin(vy)$
$F(u,v) = \iint f(x,y)\sin(ux)\sin(vy)\,\mathrm{d}x\,\mathrm{d}y$

Abel : $h(u,v;x,y) = [(x-u)^2+(y-v)^2]^{(\alpha-1)/2}$
$F(u,v) = \iint f(x,y)[(u-x)^2+(v-y)^2]^{-(\alpha-1)/2}\,\mathrm{d}x\,\mathrm{d}y$

Weyl : $h(u,v;x,y) = [(u-x)^2+(v-y)^2]^{-(\alpha-1)/2}$
$F(u,v) = \iint f(x,y)[(u-x)^2+(v-y)^2]^{-(\alpha-1)/2}\,\mathrm{d}x\,\mathrm{d}y$

Hilbert : $h(u,v;x,y) = [(u-x)^2+(v-y)^2]^{-1/2}$
$F(u,v) = \iint \dfrac{f(x,y)}{\sqrt{(u-x)^2+(v-y)^2}}\,\mathrm{d}x\,\mathrm{d}y$

Radon : $h(r,\phi;x,y) = \delta(r - x\cos\phi - y\sin\phi)$
$F(r,\phi) = \iint f(x,y)\delta(r - x\cos\phi - y\sin\phi)\,\mathrm{d}x\,\mathrm{d}y$

F.3 Transformations n-D

$$F(\boldsymbol{u}) = \int f(\boldsymbol{x}) h(\boldsymbol{u},\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}$$

Identité : $h(\boldsymbol{u},\boldsymbol{x}) = \delta(\boldsymbol{u}-\boldsymbol{x})$ $\quad F(\boldsymbol{u}) = \int f(\boldsymbol{x})\delta(\boldsymbol{u}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = f(\boldsymbol{x})$

Convolution : $h(\boldsymbol{u},\boldsymbol{x}) = h(\boldsymbol{u}-\boldsymbol{x})$ $\quad F(\boldsymbol{u}) = \int f(\boldsymbol{x}) h(\boldsymbol{u}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = f(\boldsymbol{x}) * h(\boldsymbol{x})$

Corrélation : $h(\boldsymbol{u},\boldsymbol{x}) = h(\boldsymbol{u}+\boldsymbol{x})$ $\quad F(\boldsymbol{u}) = \int f(\boldsymbol{x}) h(\boldsymbol{u}+\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = f(\boldsymbol{x}) \star h(\boldsymbol{x})$

Laplace : $h(\boldsymbol{u},\boldsymbol{x}) = \exp\left[-\boldsymbol{u}^t\boldsymbol{x}\right]$ $\quad F(\boldsymbol{u}) = \int f(\boldsymbol{x}) \exp\left[-\boldsymbol{u}^t\boldsymbol{x}\right]\,\mathrm{d}\boldsymbol{x}$

Transformations unitaires

$$F(\boldsymbol{u}) = \int f(\boldsymbol{x}) h(\boldsymbol{u},\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$
$$f(\boldsymbol{x}) = \int F(\boldsymbol{u}) h^*(\boldsymbol{u},\boldsymbol{x})\,\mathrm{d}\boldsymbol{u}$$

Fourier : $h(\boldsymbol{u},\boldsymbol{x}) = \exp\left[-j\boldsymbol{u}^t\boldsymbol{x}\right]$

$$F(\boldsymbol{u}) = \int f(\boldsymbol{x}) \exp\left[-j\boldsymbol{u}^t\boldsymbol{x}\right]\,\mathrm{d}\boldsymbol{x}$$
$$f(\boldsymbol{x}) = \int F(\boldsymbol{u}) \exp\left[+j\boldsymbol{u}^t\boldsymbol{x}\right]\,\mathrm{d}\boldsymbol{u}$$

Abel : $h(\boldsymbol{u},\boldsymbol{x}) = ||\boldsymbol{x}-\boldsymbol{u}||^{(\alpha-1)/2}$ $\quad F(\boldsymbol{u}) = \int f(\boldsymbol{x}) ||\boldsymbol{x}-\boldsymbol{u}||^{(\alpha-1)/2}\,\mathrm{d}\boldsymbol{x}$

Weyl : $h(\boldsymbol{u},\boldsymbol{x}) = ||\boldsymbol{u}-\boldsymbol{x}||^{(\alpha-1)/2}$ $\quad F(\boldsymbol{u}) = \int f(\boldsymbol{x}) ||\boldsymbol{u}-\boldsymbol{x}||^{(\alpha-1)/2}\,\mathrm{d}\boldsymbol{x}$

Hilbert : $h(\boldsymbol{u},\boldsymbol{x}) = ||\boldsymbol{u}-\boldsymbol{x}||^{-1/2}$ $\quad F(\boldsymbol{u}) = \int f(\boldsymbol{x}) ||\boldsymbol{u}-\boldsymbol{x}||^{-1/2}\,\mathrm{d}\boldsymbol{x}$

Radon : $h(r,\boldsymbol{\xi};\boldsymbol{x}) = \delta(r-\boldsymbol{\xi}^t\boldsymbol{x})$ $\quad F(r,\boldsymbol{\xi}) = \int f(\boldsymbol{x})\delta(r-\boldsymbol{\xi}^t\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$

F.4 Transformations linéaires dans le cas discret

$$g(m) = \sum_{n=0}^{N-1} f(n)h(m,n), \quad m = 0,\cdots,N-1 \quad \longrightarrow \quad \boldsymbol{g} = \boldsymbol{H}\boldsymbol{f}$$

Identité : $h(m,n) = \delta(m-n), \quad m,n = 0,\cdots,N-1$

$\boldsymbol{H}$ matrice Identité

$$\boldsymbol{H} = \begin{pmatrix} 1 & 0 & \cdots & & & 0 \\ 0 & 1 & & & & \vdots \\ \vdots & & & & & \\ & & & & & \\ & & & & & \vdots \\ \vdots & & & & & 0 \\ 0 & \cdots & & & 0 & 1 \end{pmatrix}$$

Convolution : $h(m,n) = h(m-n), \quad m,n = 0,\cdots,N-1$

$\boldsymbol{H}$ matrice de Toeplitz

$$\boldsymbol{H} = \begin{pmatrix} h_0 & h_{-1} & h_{-2} & \cdots & & \cdots & h_{-N+1} \\ h_1 & h_0 & h_{-1} & & & & \vdots \\ h_2 & h_1 & & & & & \vdots \\ \vdots & & & & & & \\ & & & & & & \\ \vdots & & & & & & h_{-2} \\ \vdots & & & & & & h_{-1} \\ h_{N-1} & \cdots & & & h_2 & h_1 & h_0 \end{pmatrix}$$

Corrélation : $h(m,n) = h(m+n), \quad m,n = 0,\cdots,N-1$

$\boldsymbol{H}$ matrice de Hankel

$$\begin{pmatrix} h_0 & h_1 & h_2 & & h_3 & \cdots & h_{N-1} \\ h_1 & h_2 & h_3 & & & h_{N-1} & \vdots \\ h_2 & h_3 & & & & & \\ h_3 & & & & & & \\ \vdots & & & & & & \vdots \\ & & & & & & h_{2N-4} \\ \vdots & h_{N-1} & & & & h_{2N-4} & h_{2N-3} \\ h_{N-1} & \cdots & & & h_{2N-4} & h_{2N-3} & h_{2N-2} \end{pmatrix}$$

Transformations unitaires dans le cas discret

Fourier : $h(m,n) = \frac{1}{N} \exp\left[-j2\pi mn/N\right], \quad m,n = 0,\cdots,N-1$

$\boldsymbol{H}$ matrice circulante de TFD

$$\boldsymbol{H} = \frac{1}{N} \begin{pmatrix} 1 & 1 & 1 & \cdots & \cdots & 1 \\ 1 & w^1 & w^2 & & & w^{N-1} \\ 1 & w^2 & w^4 & & & w^{2(N-1)} \\ \vdots & & & & & \\ & & & & & w^{(N-3)(N-1)} \\ \vdots & & & & & w^{(N-2)(N-1)} \\ 1 & w^{N-1} & & & w^{(N-2)(N-1)} & w^{(N-1)(N-1)} \end{pmatrix}$$

avec $w = \exp\left[-j2\pi/N\right]$

Cosinus : $h(m,n) = \begin{cases} \frac{1}{\sqrt{N}} & \text{si } m = 0 \\ \frac{2}{\sqrt{N}} \cos(\pi m(n+1/2)/N) & \text{si } m \neq 0 \end{cases}, \quad m,n = 0,\cdots,N-1$

$\boldsymbol{H}$ matrice circulante de TCD

$$\boldsymbol{H} = \begin{pmatrix} 1 & 1 & 1 & \cdots & \cdots & 1 \\ 1 & \cos(w) & \cos(2w) & & & \cos[(N-1)w] \\ 1 & \cos(2w) & & & & \vdots \\ \vdots & & & & & \\ & & & & & \cos[(N-3)(N-1)w] \\ \vdots & & & & & \cos[(N-2)(N-1)w] \\ 1 & \cos[(N-1)w] & & & & \cos[(N-1)(N-1)w] \end{pmatrix}$$

avec $w = -j2\pi/N$

Sinus : $h(m,n) = \frac{2}{\sqrt{N}} \sin(\pi m(n+1/2)/N), \quad m,n = 0,\cdots,N-1$

$\boldsymbol{H}$ matrice circulante de TSD

$$\boldsymbol{H} = \begin{pmatrix} 1 & 1 & 1 & \cdots & \cdots & 1 \\ 1 & \sin(w) & \sin(2w) & & & \sin[(N-1)w] \\ 1 & \sin(2w) & & & & \vdots \\ \vdots & & & & & \\ & & & & & \sin[(N-3)(N-1)w] \\ \vdots & & & & & \sin[(N-2)(N-1)w] \\ 1 & \sin[(N-1)w] & & & & \sin[(N-1)(N-1)w] \end{pmatrix}$$

avec $w = -j2\pi/N$

Walsh : $h(m,n) = \prod_{i=0}^{k} (-1)^{b_i(m) b_{n-1-i}(n)}, \quad N = 2^k$

$\boldsymbol{H}$ matrice symétrique avec des éléments -1 ou +1

$b_k(z)$ est le k ème bit de la représentation binaire de z.

$$\boldsymbol{H} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{pmatrix}$$

Hadamard : $h(m,n) = (-1)^{\sum_{i=0}^{k} b_i(m) b_i(n)}, \quad N = 2^k$

$\boldsymbol{H}$ matrice symétrique avec des éléments -1 ou +1

$b_k(z)$ est le k ème bit de la représentation binaire de z.

Exemple: $z = 6 = 110 \longrightarrow b_0(z) = 0, b_1(z) = 1, b_2(z) = 2$.

$$\boldsymbol{H} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix}$$

Table des figures

Annexe G

Classement bibliographique

Problèmes inverses, régularisation :
[TA77, Tit85b, BDMP88, BPT88, Cul71, KV90], [Dem89, DL91, Dem91]
Inverse généralisée :
[Lan51, Bia59, NW74, Nas81, SF87]
Approche bayésienne :
[BT72, KW93], [Cul79, GK92, GHW79, TBKT91, Tit85a, HT87], [FDG93, GMG92]
Modèles Markoviens :
[Din88, Din90, GY95, HL92, JWHC84, You88], [Nik95, NMD96b]
Préservation de contour :
[GL93, LP88, Ter86, Ter88], [Nik95, NMD96b]
Calcul matriciel :
[Cia82, Cia88, GVL89]
Déconvolution :
[Dem85, DR85]
Déconvolution, algorithmes rapides :
[Com84, DR91]
Déconvolution aveugle :
[Hog74, BGR80, Gou92, Hol93, Sch93, LC95]
Déconvolution impulsionnelle :
[CGM85, Lav93, CID93, Cha93, Gou92] [IG90, Idi91, IG93b, IG93a]
Tomographie d'impédance :
[HWWT91, KM90, Web90]
Tomographie ultrasonore :
[Gre83, HH84, Kak79, **?**, Lef85, Mos86, SLW84]
Radio astronomie :
[Bra56, Luc74, Hog74, Fie82, PR84, Hol93, Lan89, Lan91] [TMS84]
Cristallographie :
[Nav85, Nav86, Nav90]
Imagerie Radar :
[AA82]
Choix de l'*a priori* :
[O'S94, O'S95, Zel77, MDD88a, MDD88c]
Sismique :
[Sai90, AH78]

POCS :
[Bre67a, Bre67b, SS82, Com93]
Maximum d'entropie :
[SW48, Jay68, Jay82, Jay85, Bal82, JS84], [BS80, BBF+83], [MDD86, MDD87a, MDD87b, MDD88b, MDD88c, MD94], [Ber95, LB93]

Bibliographie

[AA82] M.F. Adams and A.P. Anderson. Synthetic aperture tomographic (sat) imaging for microwave diagnostics. *Proceedings of the IEE*, 129:83–88, 1982.

[AH78] V.K. Arya and H.D. Holden. Deconvolution of seismic data-an overview. *IEEE Transactions on Geoscience and Remote Sensing*, GE-16:95–98, 1978.

[Bal82] R. Balian. *Du microscopique au macroscopique, cours de physique statistique de l'École Polytechnique*, volume 1. Ellipses, Paris, 1982.

[BBF+83] R.K. Bryan, M. Bansal, W. Folkhard, C. Nave, and D.A. Marvin. Maximum-entropy calculation of the electron density at 4 a resolution of pf1 filamentous bacteriophage. *Proc. Natl. Acad. Sci. USA*, 80:4728–4731, 1983.

[BDMP88] M. Bertero, C. De Mol, and E.R. Pike. Linear inverse problems with discrete data: II. stability and regularization. *Inverse Problems*, 4:3, 1988.

[Ber95] Jean-François Bercher. *Développement de critères de nature entropique pour la résolution des problèmes inverses linéaires.* PhD thesis, Université de Paris-Sud, Orsay, février 1995.

[Bes74] Julian E. Besag. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society B*, 36(2):192–236, 1974.

[Bes86] Julian E. Besag. On the statistical analysis of dirty pictures (with discussion). *Journal of the Royal Statistical Society B*, 48(3):259–302, 1986.

[Bes89] Julian E. Besag. Digital image processing : Towards Bayesian image analysis. *Journal of Applied Statistics*, 16(3):395–407, 1989.

[BG93] Julian E. Besag and Peter Green. Spatial statistics and Bayesian computation. *J. R. Statist. Soc. B*, 55:25–37, 1993.

[BGR80] A. Benveniste, M. Goursat, and G. Ruget. Robust identification of nonmnimum phase system : Blind adjustment of a linear equalizer in data communications. *IEEE Transactions on Automatic and Control*, AC-25, 1980.

[Bia59] H. Bialy. Iterative behandlung linearen funktionalgleichungen. *Arch. Ration. Mech. Anal.*, 4:166–176, 1959.

[BPT88] M. Bertero, T.A. Poggio, and V. Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, 76:869–889, August 1988.

[Bra56] R.N. Bracewell. Strip integration in radio astronomy. *Aust. J. Phys.*, 9:198–201, 1956.

[Bre67a] L. M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *Zh. vȳchisl. Mat. mat. Fiz.*, 7:147–156, 1967.

[Bre67b] L. M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming.

U.S.S.R. Computational Mathematics and Mathematical Physics, 7:200–217, 1967.

[BS80] R.K. Bryan and J. Skilling. Deconvolution by maximum entropy as illustrated by application to the jet of m87. *Monthly Notices of the Royal Astronomical Society*, 191:69–79, 1980.

[BT72] G.E.P Box and G.C. Tiao. *Bayesian inference in statistical analysis.* Addison-Wesley publishing, 1972.

[BZ87] A. Blake and A. Zisserman. *Visual reconstruction.* The MIT Press, Cambridge, 1987.

[CGM85] C.Y. Chi, J. Goustias, and Jerry M. Mendel. A fast maximum-likelihood estimation and detection algorithm for Bernoulli-Gaussian processes. In *Proceedings of IEEE ICASSP*, pages 1297–1300, 1985.

[Cha93] Frédéric Champagnat. *Déconvolution impulsionnelle et extensions pour la caractérisation des milieux inhomogènes en échographie.* PhD thesis, Université de Paris-Sud, Orsay, décembre 1993.

[Cia82] Philippe G. Ciarlet. *Introduction à l'analyse numérique et à l'optimisation.* Masson, Paris, 1982.

[Cia88] Philippe G. Ciarlet. *Introduction à l'analyse numérique matricielle et à l'optimisation.* Collection mathématiques appliquées pour la maîtrise. Masson, Paris, 1988. JFG.

[CID93] Frédéric Champagnat, Jérôme Idier, and Guy Demoment. Deconvolution of sparse spike trains accounting for wavelet phase shifts and colored noise. In *Proceedings of IEEE ICASSP*, volume III, pages 452–455, Minneapolis, U.S.A., 1993.

[Com84] D. Commenges. The deconvolution problem : Fast algorithms including the preconditioned conjugate-gradient to compute a MAP estimator. *IEEE Transactions on Automatic and Control*, AC-29:229–243, 1984.

[Com93] P. L. Combettes. The foundation of set theoretic estimation. *Proceedings of the IEEE*, 81(2):182–208, February 1993.

[Cul71] J. Cullum. Numerical differentiation and regularization. *SIAM Journal of Numerical Analysis*, 8:254–265, 1971.

[Cul79] J. Cullum. The effective choice of the smoothing norm in regularization. *Math. Comp.*, 33:149–170, 1979.

[Dem85] Guy Demoment. Déconvolution des signaux. Cours de l'ESE 3086, GPI–LSS, 1985.

[Dem89] Guy Demoment. Image reconstruction and restoration : Overview of common estimation structure and problems. *IEEE Transactions on Acoustics Speech and Signal Processing*, ASSP-37(12):2024–2036, December 1989.

[Dem91] Guy Demoment. Problèmes inverses. *Flux*, (141):42–43, novembre 1991.

[Din88] Jean-Marc Dinten. Tomographic reconstruction with a limited number of projections : regularization using a Markov model. Technical report, Université de Paris-Sud, 1988.

[Din90] Jean-Marc Dinten. *Tomographie à partir d'un nombre limité de projections : régularisation par champs markoviens.* PhD thesis, Université de Paris-Sud, janvier 1990.

[DL91] Guy Demoment and André Lannes. Les problèmes inverses. *Le Courrier du CNRS*, 77:–, juin 1991. numéro spécial sur le traitement du signal et de l'image.

[DR85] Guy Demoment and Roger Reynaud. Fast minimum-variance deconvolution. *IEEE Transactions on Acoustics Speech and Signal Processing*, ASSP-33:1324–1326, 1985.

[DR91] Guy Demoment and Roger Reynaud. Fast RLS algorithms and chandrasekhar equations. In S. Haykin, editor, *Adaptive Signal Processing*, pages 357–367, San Diego, July 1991. SPIE Conf.

[FDG93] Natalie Fortier, Guy Demoment, and Yves Goussard. GCV and ML methods of determining parameters in image restoration by regularization: Fast computation in the spatial domain and experimental comparison. *Journal of Visual Communication and Image Representation*, 4(2):157–170, June 1993.

[Fie82] J.R. Fienup. Phase retrieval algorithms: a comparison. *Applied Optics*, 21(15):2758–2769, August 1982.

[Gem90] Donald Geman. *École d'Été de Probabilités de Saint-Flour XVIII - 1988*, volume 1427, chapter Random fields and inverse problems in imaging, pages 117–193. Springer-Verlag, lecture notes in mathematics edition, 1990.

[GG84] Stuart Geman and Donald Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, November 1984.

[GHW79] G. H. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223, mai 1979. JFG.

[GK92] N.P. Galatsanos and A.K. Katsaggelos. Methods for choosing the regularization parameter and estimating the noise variance in image restoration and thier telation. *IEEE Transactions on Image Processing*, IP-1(3):322–336, July 1992. JFG.

[GL93] Muhittin Gökmen and Ching-Chung Li. Edge detection and surface reconstruction using refined regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-15(5):492–499, May 1993.

[GMG92] Élisabeth Gassiat, Fabrice Monfront, and Yves Goussard. On simultaneous signal estimation and parameter identification using a generalized likelihood approach. *IEEE Transactions on Information Theory*, IT-38:157–162, January 1992.

[Gou92] Yves Goussard. Blind deconvolution of sparse spike trains using stochastic optimization. In *Proceedings of IEEE ICASSP*, volume IV, pages 593–596, 1992.

[GR92] Stuart Geman and Georges Reynolds. Constrained restoration and recovery of discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-14(3):367–383, March 1992.

[Gre83] J.K. Greenleaf. Computerized tomography with ultrasound. *Proceedings of the IEEE*, 71:330, 1983.

[GVL89] G.H. Golub and C.F. Van Loan. *Matrix computations (2nd Edition)*. xxx, xxx, 1989.

[GY95] Donald Geman and Chengda Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE Transactions on Image Processing*, IP-4(7):932–946, July 1995.

[HH84] D. Hiller and H. Hermert. System analysis of ultrasound reflection mode computerized tomography. *IEEE Transactions on Sonics and Ultrasonics*, SU-31:240–250, 1984.

[HL92] Thomas J. Hebert and Richard Leahy. Statistic-based MAP image reconstruction from poisson data using Gibbs prior. *IEEE Transactions on Signal Processing*, SP-40(9):2290–2303, September 1992.

[Hog74] Hogbom. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astron. Astrophys. Suppl.*, 15:417–426, 1974.

[Hol93] Holmes. Blind deconvolution of quantum-limited incoherent imagery : Maximum-likelihood approach. *Journal of the Optical Society of America*, 9(7), July 1993.

[HT87] P. Hall and D. M. Titterington. Common structure of techniques for choosing smoothing parameter in regression problems. *Journal of the Royal Statistical Society B*, 49(2):184–198, 1987. JFG.

[HWWT91] P. Hua, E.J. Woo, J.G. Webster, and W.J. Tompkins. Iterative reconstruction methods using regularization and optimal current patterns in electrical impedance tomography. *IEEE Transactions on Medical Imaging*, MI-10:621–628, December 1991.

[Idi91] Jérôme Idier. *Imagerie des milieux stratifiés : modélisation markovienne, application à la déconvolution sismique.* PhD thesis, Université de Paris-Sud, Orsay, septembre 1991.

[IG90] Jérôme Idier and Yves Goussard. Stack algorithm for recursive deconvolution of Bernoulli-Gaussian processes. *IEEE Transactions on Geoscience and Remote Sensing*, GE-28(5):975–978, September 1990.

[IG93a] Jérôme Idier and Yves Goussard. Multichannel seismic deconvolution. *IEEE Transactions on Geoscience and Remote Sensing*, GE-31(5):961–979, September 1993.

[IG93b] Jérôme Idier and Yves Goussard. Markov modeling for Bayesian restoration of two-dimensional layered structures. *IEEE Transactions on Information Theory*, IT-39(4):1356–1373, July 1993.

[Jay68] E. T. Jaynes. Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics*, SSC-4(3):227–241, September 1968.

[Jay82] E. T. Jaynes. On the rationale of maximum-entropy methods. *Proceedings of the IEEE*, 70(9):939–952, September 1982.

[Jay85] E. T. Jaynes. Where do we go from here? In Jr. C.R. Smith & T. Grandy, editor, *Maximum-Entropy and Bayesian Methods in Inverse Problems*, pages 21–58. 1985.

[JS84] R. Johnson and J. Shore. Which is better entropy expression for speech processing : -slogs or logs? *IEEE Transactions on Acoustics Speech and Signal Processing*, ASSP-32:129–137, 1984.

[JWHC84] Valen Johnson, Wing Wong, Xiaoping Hu, and Chin-Tu Chen. Image restoration using Gibbs priors: Boundary modeling, treatement of blurring, and selection of hyperparameter. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-13(5):413–425, 1984.

[Kak79] Avinash C. Kak. Computerized tomography with X-ray, emission, and ultrasound sources. *Proceedings of the IEEE*, 67(9):1245–1272, September 1979.

[KM90] R. V. Kohn and A. McKenney. Numerical implementation of a variational method for electrical impedance tomography. *Inverse Problems*, pages 389–414, 1990.

[KV90] Nicolaos Karayiannis and Anastasios Venetsanopoulos. Regularization theory in image restoration – the stabilizing functional approach. *IEEE Transactions on Acoustics Speech and Signal Processing*, ASSP-38(7):1155–1179, July 1990.

[KW93] Daniel Keren and Michael Werman. Probabilistic analysis of regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-15(10):982–995, 1993.

[Lan51] L. Landweber. An iteration formula for fredholm integral equations of the first kind. *Amer. J. Math.*, 73:615–624, 1951.

[Lan89] Lannes. Backprojection mechanisms in phase-closure imaging. bispectral analysis of the phase-restoration process. *Experimental Astronomy*, 1:47–76, 1989.

[Lan91] A. Lannes. Phase and amplitude calibration in aperture synthesis. algebraic structures. *Inverse Problems*, 7:261–298, 1991.

[Lav93] M. Lavielle. Bayesian deconvolution of Bernoulli-Gaussian processes. *Signal Processing*, 33:67–79, 1993.

[LB93] Guy Le Besnerais. *Méthode du maximum d'entropie sur la moyenne, critères de reconstruction d'image et synthèse d'ouverture en radio-astronomie.* PhD thesis, Université de Paris-Sud, Orsay, décembre 1993.

[LC95] J.S. Liu and R. Chen. Blind deconvolution via sequential imputations. *Biometrika*, 90(430):567–576, 1995. MF.

[Lef85] J.P. Lefebvre. La tomographie d'impédance acoustique. *Traitement du Signal*, 2:103–110, 1985.

[LP88] David Lee and Theo Pavlidis. One-dimensional regularization with discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-10(6):417–438, November 1988.

[Luc74] L. B. Lucy. An iterative technique for the rectification of observed distributions. *The Astronomical Journal*, 79(6):745–754, 1974.

[MD94] Ali Mohammad-Djafari. Maximum d'entropie et problèmes inverses en imagerie. *Traitement du Signal*, pages 87–116, 1994.

[MDD86] Ali Mohammad-Djafari and Guy Demoment. Maximum entropy diffraction tomography. In *Proceedings of IEEE ICASSP*, pages 1–34, Tokyo, Japan, 1986.

[MDD87a] Ali Mohammad-Djafari and Guy Demoment. Maximum entropy Fourier synthesis with application to diffraction tomography. *Applied Optics*, 26(10):1745–1754, 1987.

[MDD87b] Ali Mohammad-Djafari and Guy Demoment. Tomographie de diffraction et synthèse de Fourier à maximum d'entropie. *Revue de Physique Appliqué*, 22:153–167, 1987.

[MDD88a] Ali Mohammad-Djafari and Guy Demoment. *Image restoration and reconstruction using entropy as a regularization functional*, volume 2, pages 341–355. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1988.

[MDD88b] Ali Mohammad-Djafari and Guy Demoment. Maximum entropy reconstruction in X ray and diffraction tomography. *IEEE Transactions on Medical Imaging*, MI-7(4):345–354, 1988.

[MDD88c] Ali Mohammad-Djafari and Guy Demoment. Utilisation de l'entropie dans les problèmes de restauration et de reconstruction d'images. *Traitement du Signal*, 5(4):235–248, 1988.

[Mos86] M. Moshfeghi. Ultrasound reflection-mode tomography using fan-shaped-beam insonification. *IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control*, UFFC-33:299–314, 1986.

[Nas81] M.Z. Nashed. Operator-theoretic and computational approaches to ill-posed problems with applications to antenna theory. *IEEE Transactions on Antennas and Propagation*, 29:220–231, 1981.

[Nav85] Jorge Navaza. On the maximum–entropy estimate of the electron density function. *Acta Crystallographica*, A–41:232–244, 1985.

[Nav86] Jorge Navaza. The use of non-local constraints in maximum-entropy electron density reconstruction. *Acta Crystallographica*, A–42:212–223, 1986.

[Nav90] Jorge Navaza. Accurate solutions of the maximum entropy equations. their impact on the foundations of direct methods. In *Int. School Comp.*, Bishenberg, Germany, 1990.

[Nik95] Mila Nikolova. *Inversion markovienne de problèmes linéaires mal posés. application à l'imagerie tomographique.* PhD thesis, Université de Paris-Sud, Orsay, février 1995.

[NMD96a] Mila Nikolova and Ali Mohammad-Djafari. Eddy current tomography using a binary Markov model. *Signal Processing*, 49:119–132, 1996. Mila Signal Processing.

[NMD96b] Mila Nikolova and Ali Mohammad-Djafari. Eddy current tomography using a binary Markov model. *Signal Processing*, 49(5):119–132, May 1996.

[NW74] M.Z. Nashed and G. Wahba. Generalized inverses in reproducing kernel spaces: An approach to regularization of linear operators equations. *SIAM Journal of Mathematical Analysis*, 5:974–987, 1974.

[O'S94] Joseph A. O'Sullivan. Divergence penalty for image regularization. In *Proceedings of IEEE ICASSP*, volume V, pages 541–544, Adelaide, Australia, April 1994.

[O'S95] Joseph A. O'Sullivan. Roughness penalties on finite domains. *IEEE Transactions on Image Processing*, IP-4(9):1258–1268, September 1995.

[PR84] T. J. Pearson and A. C. S. Readhead. Image formation by self-calibration in radio astronomy. *Ann. Rev. Astron. Astrophys.*, 22:97–130, 1984.

[Sai90] N. Saito. Superresolution of noisy band-limited data by data adaptive regularization and its application to seismic trace inversion. In *Proceedings of IEEE ICASSP*, pages 1237–1240, Albuquerque, U.S.A., April 1990.

[Sch93] Schultz. Multiframe blind deconvolution of astronomical images. *Journal of the Optical Society of America*, 10(5), May 1993.

[SF87] Didier Saint-Félix. *Restauration d'image : régularisation d'un problème mal-posé et algorithmique associée.* PhD thesis, Thèse de doctorat d'État, Université de Paris-Sud, septembre 1987.

[SLW84] C.F. Schueler, H. Lee, and G. Wade. Fundamentals of digital ultrasonic imaging. *IEEE Transactions on Sonics and Ultrasonics*, SU-31:195–217, 1984.

[SS82] M. I. Sezan and H. Stark. Image restoration by the method of convex projections: Part 2 - applications and numerical results. *IEEE Transactions on Medical Imaging*, MI-1(2):95–101, October 1982.

[SW48] C.E. Shannon and W. Weaver. The mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[TA77] A. Tikhonov and V. Arsenin. *Solutions of Ill-Posed Problems.* Winston, Washington DC, 1977.

[TBKT91] A. Thompson, J. C. Brown, J. W. Kay, and D. M. Titterington. A study of methods of choosing the smoothing parameter in image restoration by regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-13(4):326–339, April 1991.

[Ter86] D. Terzopoulos. Regularization of inverse visual problems involving discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(4):413–424, July 1986.

[Ter88] D. Terzopoulos. The computation of visual surface representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-10(4):417–438, 1988.

[Tit85a] D. M. Titterington. Common structure of smoothing techniques in statistics. *International Statistical Review*, 53(2):141–170, 1985. JFG,Mila.

[Tit85b] D. M. Titterington. General structure of regularization procedures in image reconstruction. *Å*, 144:381–387, 1985.

[TMS84] Thompson, Moran, and Swenson. *Interferometry and Synthesis in Radioastronomy.* Wiley Interscience, New-York, 1984.

[Web90] J.G. Webster. *Electrical Impedance Tomography.* Adam Hilger, IOP Publishing Ltd, 1990.

[You88] Laurent Younès. Estimation and annealing for Gibbsian fields. *Annales de l'institut Henri Poincaré*, 24(2):269–294, février 1988.

[Zel77] Zellner. Maximal data information prior distributions. In A. Aykac and C. Brumat, editors, *New Developments in the Applications of Bayesian Methods*, pages 211–232. North-Holland, Amsterdam, 1977.