



TECHNIQUES
DE L'INGÉNIEUR

LES FOCUS
TECHNIQUES DE L'INGÉNIEUR

LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

PROBLÈMES INVERSES EN TRAITEMENT DU SIGNAL ET DE L'IMAGE

Octobre / 2021



**TECHNIQUES
DE L'INGÉNIEUR**

Réf. : **TE5235 V1**

Date de publication :
10 novembre 2001

Date de dernière validation :
28 juin 2019

Problèmes inverses en traitement du signal et de l'image

Cet article est issu de : **Technologies de l'information | Le traitement du signal et ses applications**

par **Guy DEMOMENT, Jérôme IDIER,
Jean-François GIOVANNELLI,
Ali MOHAMMAD-DJAFARI**

Pour toute question :
Service Relation clientèle
Techniques de l'Ingénieur
Immeuble Pleyad 1
39, boulevard Ornano
93288 Saint-Denis Cedex

Par mail :
infos.clients@teching.com
Par téléphone :
00 33 (0)1 53 35 20 20

Document téléchargé le : **28/10/2021**

Pour le compte : **7200106152 - éditions ti // céline BLONBOU // 2.59.188.28**

© Techniques de l'Ingénieur | tous droits réservés

Problèmes inverses en traitement du signal et de l'image

par **Guy DEMOMENT**

*Professeur à l'université de Paris-Sud
Laboratoire des signaux et systèmes, Supélec*

Jérôme IDIER

*Chargé de recherche au CNRS
Laboratoire des signaux et systèmes, Supélec*

Jean-François GIOVANNELLI

*Maître de conférences à l'université de Paris-Sud
Laboratoire des signaux et systèmes, Supélec*

et **Ali MOHAMMAD-DJAFARI**

*Directeur de recherche au CNRS
Laboratoire des signaux et systèmes, Supélec*

1. Deux exemples d'inversion	TE 5 235 - 3
2. Aspects mathématiques	— 5
3. Régularisation	— 8
4. Approche bayésienne de l'inversion	— 12
5. Méthodes de minimisation de critère	— 20
6. Questions plus avancées	— 22
7. Conclusion	— 24
Références bibliographiques	— 24

Dans de nombreux domaines de la physique appliquée, tels que l'optique, le radar, la thermique, la spectroscopie, la géophysique, l'acoustique, la radioastronomie, le contrôle non destructif, le génie biomédical, l'instrumentation et l'imagerie en général, se pose le problème de la détermination de la distribution spatiale d'une grandeur scalaire ou vectorielle, souvent appelée **l'objet**, à partir de mesures. Selon les cas, ces mesures de l'objet sont directes – on parle alors d'**image** – ou indirectes – on parle alors de **projection** dans le cas de la tomographie, ou de **visibilité** en astronomie, par exemple. La résolution d'un tel problème d'imagerie peut être habituellement décomposée en trois étapes :

— un **problème direct** où, connaissant l'objet et le mécanisme d'observation, on établit une description mathématique des données observées. Ce modèle doit être assez précis pour fournir une description correcte du phénomène physique d'observation, et assez simple cependant pour se prêter à un traitement numérique ultérieur ;

— un **problème d'instrumentation** où l'on doit recueillir des données le plus informatives possible afin de résoudre le problème d'imagerie dans les meilleures conditions ;

— un **problème inverse** où l'on doit calculer une image acceptable de l'objet à partir du modèle et des données précédents.

Une bonne estimation de l'objet nécessite évidemment que ces trois sous-problèmes soient étudiés de manière coordonnée. Or, la caractéristique commune de ces problèmes de reconstruction ou de restauration d'image est qu'ils sont souvent mal-posés ou mal-conditionnés. Les problèmes de plus haut niveau que l'on rencontre en vision par ordinateur, tels que la segmentation d'image, le traitement du flot optique, la reconstruction de formes à partir d'ombrages, sont aussi des problèmes inverses et ils souffrent des mêmes difficultés.

Il existe, schématiquement, deux grandes communautés scientifiques qui s'intéressent à ces problèmes inverses, d'un point de vue méthodologique :

— celle de la **physique mathématique**, que l'on peut rattacher aux travaux fondateurs de Phillips, Twomey et Tikhonov dans les années 1960, dont P.C. Sabatier fut un des pionniers en France (avec son action thématique programmée du même nom), et dont une revue représentative est « *Inverse problems* » ;

— celle du **traitement statistique des données**, que l'on peut rattacher aux travaux de Franklin à la fin des années 1960, dont les frères Geman ont constitué les accélérateurs en traitement d'image, et dont une revue représentative est « *IEEE Transactions on Image Processing* ».

On peut dire, grossièrement, que les uns abordent le problème en dimension infinie, avec les questions d'existence, d'unicité et de stabilité qui deviennent très compliquées avec des problèmes directs non linéaires, et le résolvent numériquement en dimension finie ; alors que les autres partent d'un problème dont la discrétisation est déjà faite à la résolution souhaitée, et non remise en cause, et profitent du caractère fini du problème pour introduire une information a priori élaborée au travers de modèles probabilistes.

Nous nous proposons d'indiquer brièvement dans la suite quels sont, de notre point de vue, l'état de l'art et les questions ouvertes dans le domaine de la résolution des problèmes inverses, en accordant une place importante aux approches probabilistes.

Sigles et abréviations	
AIC	Akaike information criterion
AR	autorégressif
ARMA	auto-regressive moving average
ART	algebraic reconstruction technic
BIC	Bayesian information criterion
DSP	densité spectrale de puissance
EM	expectation maximization
EQM	erreur quadratique moyenne
EQMP	erreur quadratique moyenne de prédiction
MAP	maximum a posteriori
MUSIC	multiple signal classification
MV	maximum de vraisemblance
RIF	réponse impulsionnelle finie
SEM	stochastic expectation maximization
SVD	singular value decomposition
TF	transformée de Fourier
TFD	transformée de Fourier discrète
VCG	validation croisée généralisée

1. Deux exemples d'inversion

1.1 Restauration d'image

Le problème de la **restauration d'une image dégradée** par un processus d'observation étant fréquemment rencontré en ingénierie, nous le choisissons comme premier fil directeur afin d'illustrer les différents concepts qui seront introduits par la suite.

Une image est généralement définie comme une **fonction scalaire de deux variables** appartenant à une région donnée de l'espace. Bien que ce support soit souvent continu, il est habituellement échantillonné sur une grille rectangulaire. Cela définit un ensemble de *pixels*, et l'image est représentée par un signal bidimensionnel $x_{n,m}$ des valeurs des intensités des pixels, où les indices n et m sont ceux de leurs coordonnées.

Dans tous les problèmes de traitement d'image évoqués dans l'introduction [1] [2] [3] [4] [5], l'objet d'intérêt ne peut pas être observé directement : il l'est à travers un système de mesure en général imparfait, pour lequel, par exemple, l'image d'un point n'est pas un point mais une *tache de flou*, ou **réponse impulsionnelle**. Un modèle simple pour décrire un tel système est le modèle convolutif :

$$y(n, m) = \sum_{k, \ell} h(k, \ell) x(n-k, m-\ell) \quad (1)$$

qui est linéaire et invariant. En réalité, pour intégrer les erreurs de mesures et celles de modélisation, on écrit plutôt :

$$y(n, m) = \sum_{k, \ell} h(k, \ell) x(n-k, m-\ell) + b(n, m) \quad (2)$$

où b représente un terme d'erreur. L'opération à réaliser est donc une **déconvolution** en présence de bruit : il s'agit de retrouver l'image originale (ou plutôt d'en faire une estimation) à partir des données observées et de la réponse impulsionnelle.

En concaténant l'ensemble des valeurs des pixels de l'objet et des données dans deux vecteurs \mathbf{x} et \mathbf{y} respectivement, la relation (2) s'écrit sous la forme générique :

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b} \quad (3)$$

dans laquelle \mathbf{A} est la matrice de convolution constituée des coefficients $h(k, \ell)$ de la réponse impulsionnelle.

Pour notre exemple de déconvolution d'image, l'objet idéal est présenté figure 1. Il s'agit d'une image de 128×128 pixels dont les niveaux de gris sont codés entre 0 et 255. La figure 2 donne un exemple typique d'image dégradée. Elle a été obtenue par simulation numérique à partir du modèle de l'équation (3) : la réponse impulsionnelle est constante (égale à $1/49$) sur un carré de taille 7×7 et nulle en dehors. La convolution réalise donc une moyenne des 49 pixels du carré centré au pixel (n, m) de l'objet idéal \mathbf{x} , pour construire le pixel correspondant de l'image observée \mathbf{y} . Le bruit \mathbf{b} est une séquence pseudo-aléatoire gaussienne et spatialement blanche, de puissance $\sigma_b^2 = 1$.

Les grandes structures de l'image restent discernables, mais la résolution spatiale est insuffisante pour observer les petites structures. Les contours francs de l'objet initial, et notamment les détails de la caméra, ont disparu. Avec un modèle convolutif tel que (2), une interprétation fréquentielle du processus d'observation permet d'appréhender intuitivement ces dégradations. L'équation (2) est en effet équivalente à :

$$Y(v_x, v_y) = H(v_x, v_y) X(v_x, v_y) + B(v_x, v_y) \quad (4)$$

où v_x et v_y sont les fréquences spatiales, et X, Y et H les transformées de Fourier respectives de x, y et h . La figure 3 présente le module de H , fonction de transfert en fréquence du système d'observation. Il s'agit clairement d'un système passe-bas, les hautes fréquences de l'objet – inévitablement présentes lorsque celui-



Figure 1 – Objet original



Figure 2 – Données : image floue bruitée

ci contient des contours nets séparant des régions homogènes – sont très atténuées, ce qui explique la mauvaise résolution de l'image observée. Il existe en plus des zéros de transmission qui renforcent les difficultés du problème.

L'idée qui vient naturellement à l'esprit pour améliorer la résolution du dispositif est de résoudre numériquement l'équation linéaire (3), connaissant les caractéristiques du système d'observation (matrice \mathbf{A}) et les données mesurées (vecteur \mathbf{y}). Il s'agit donc d'*inverser* le problème. Comme \mathbf{x} et \mathbf{y} ne sont pas nécessairement de même dimension, il est naturel de minimiser un critère des moindres carrés, de la forme :

$$\mathcal{J}(\mathbf{x}) = \mathcal{G}(\mathbf{y} - \mathbf{A}\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \quad (5)$$

La solution des moindres carrés ou de norme minimale s'écrit :

$$\hat{\mathbf{x}}_{MC} = (\mathbf{A}^t \mathbf{A})^{-1} \mathbf{A}^t \mathbf{y} \quad \text{ou} \quad \hat{\mathbf{x}}_{IG} = \mathbf{A}^+ \mathbf{y} \quad (6)$$

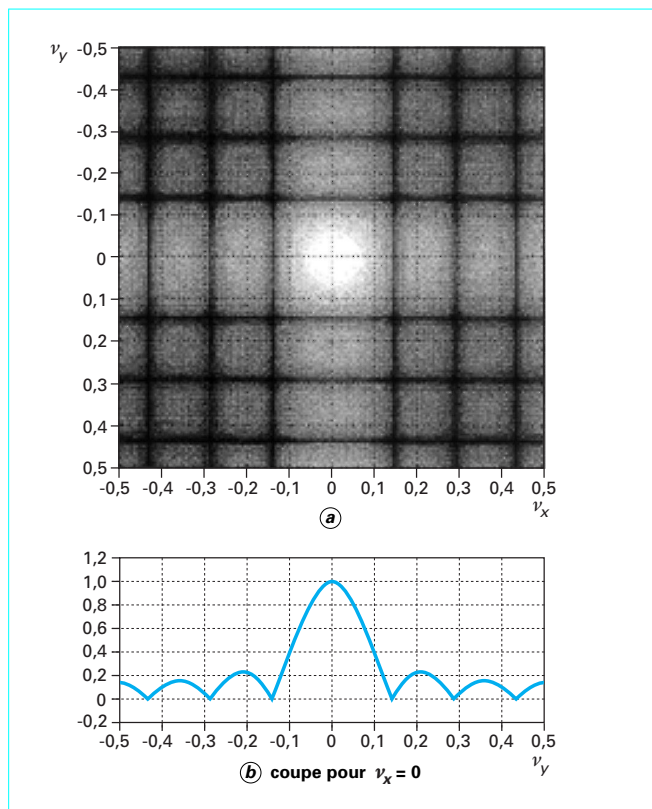


Figure 3 – Module de la fonction de transfert en fréquence du système d'observation H

selon que la matrice normale $A^t A$ est régulière ou qu'elle n'admet qu'une inverse généralisée A^\dagger . Notons que, si la matrice A est carrée et inversible, alors \hat{x}_{MC} s'obtient simplement en inversant la matrice A :

$$\hat{x}_{MC} = A^{-1}y \tag{7}$$

Ce choix des moindres carrés semble d'autant plus raisonnable que, sous hypothèse de bruit blanc centré, la solution est aussi un estimateur sans biais et à variance minimale, propriété souvent recherchée en statistique.

Le résultat dans notre exemple est représenté figure 4. Aucune amélioration n'est obtenue, bien au contraire : les structures spatiales qui apparaissent sont sans rapport avec celles de l'objet réel de la figure 1 et, de plus, la valeur des pixels dépasse en de nombreux endroits 10^4 alors que la valeur des pixels de l'objet d'origine reste inférieure à 255. Une interprétation fréquentielle permet ici encore d'analyser le résultat. La solution des moindres carrés (6) est obtenue par l'application d'un *filtre inverse* à l'image observée, c'est-à-dire, dans le domaine fréquentiel, à la multiplication par la fonction de transfert $1/H(v_x, v_y)$. Or la figure 3 montre clairement que certaines fréquences spatiales sont annulées (ou presque) par le processus d'observation au voisinage des zéros de transmission. Il s'agit des fréquences $v = 1/7 \approx 0,14$, $v = 2/7 \approx 0,29$ et $v = 3/7 \approx 0,43$. Cela signifie que ces composantes fréquentielles, qui seraient absentes de données parfaites, non bruitées, vont être considérablement amplifiées par le filtre inverse qui tend à corriger ce « trou » fréquentiel, et amplifie en fait exagérément les composantes correspondantes du bruit. On observe d'ailleurs bien, sur la figure 4, la présence dominante d'un motif régulier et

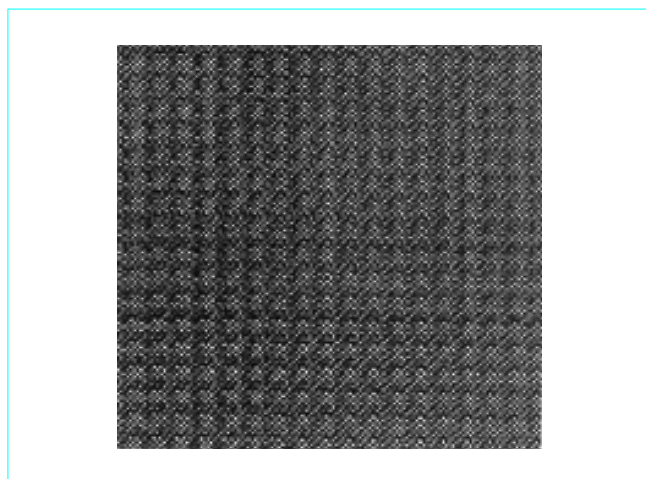


Figure 4 – Solution des moindres carrés

doublément périodique, de période 15 pixels environ, ce qui correspond exactement au premier zéro en $v \approx 0,14$. On comprend donc bien pourquoi la solution obtenue est inacceptable.

1.2 Synthèse de Fourier

Pour illustrer les différents concepts utilisés, nous utilisons un second fil directeur, constitué du problème de l'analyse spectrale. Ce problème, qui s'apparente aussi à celui de la **synthèse de Fourier** rencontré en radioastronomie par exemple, n'est pas habituellement abordé comme un problème inverse. Nous verrons cependant qu'il gagne à l'être, et que l'inversion se confond finalement avec l'inférence.

Le signal à traiter est typique de ceux que l'on peut recueillir en interférométrie optique visible, par exemple. Le problème est de déterminer le nombre de raies dans le spectre du signal analysé, leur position, leur amplitude, leur « largeur », bref d'effectuer une *estimation spectrale*, c'est-à-dire estimer :

$$S_x(v) = |TF\{x\}|^2$$

à partir d'une mesure imparfaite y (support de mesure limité, bruit de mesure, « fond » variable, etc.) de x .

Ce signal, dont le graphe est tracé figure 5, est bien connu de la communauté du traitement du signal puisqu'il fut utilisé dans un célèbre article de synthèse sur l'analyse spectrale au début des années 1980 [6]. Il est constitué de :

- deux sinusoïdes de même amplitude mais dont les fréquences sont séparées d'un intervalle de longueur inférieure à la résolution des méthodes classiques d'analyse spectrale ;
- d'une troisième sinusoïde bien séparée des deux premières, mais dont la puissance est inférieure de 20 dB ;
- d'un bruit coloré par passage d'un bruit blanc gaussien pseudo-aléatoire dans un filtre à moyenne glissante (ou filtre à réponse impulsionnelle finie, ou encore filtre MA), de manière à compliquer l'analyse.

L'étude du tracé de ce signal montre clairement un battement entre les fréquences des deux sinusoïdes les plus proches. Son spectre théorique est représenté figure 6.

On note :

- $y = [y_1, \dots, y_N]^t$ les données expérimentales ;
- $X(v)$ la transformée de Fourier (TF) cherchée (on s'intéresse en fait plutôt à $|X(v)|^2$) ;

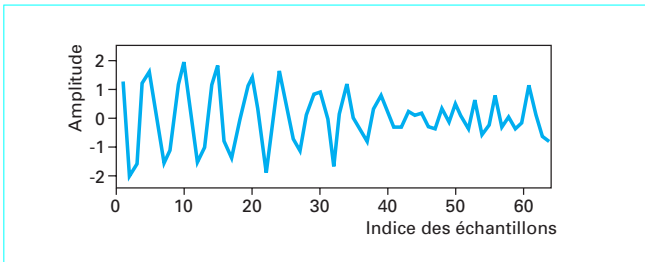


Figure 5 – Signal à analyser

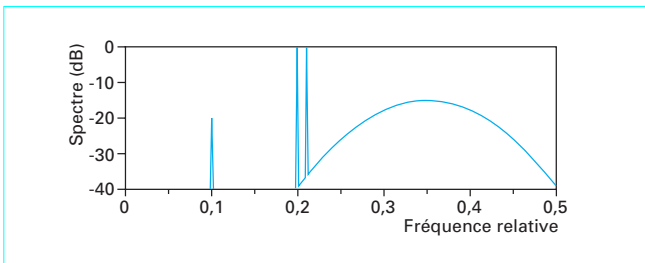


Figure 6 – Spectre théorique du signal à analyser

— $x_k = \int_0^1 X(v) \exp(2j\pi kv) dv$ les coefficients du développement en série de Fourier de $X(v)$ qui est périodique.

Le problème d'analyse spectrale peut être ainsi formulé [7] :

$$\text{Trouver } X(v) \quad v \in [0,1],$$

$$\text{tel que } x_k = y_k \quad k = 1, 2, \dots, N \tag{8}$$

On pourrait aussi : 1^o) rechercher le nombre de sinusoides et 2^o) estimer leur fréquence et leur amplitude, ce que nous ne ferons pas pour l'instant. La formulation (8) diffère de la formulation habituelle des méthodes dites « modernes » qui consiste à rechercher la densité spectrale de puissance d'un processus aléatoire à partir de y considéré comme un fragment d'une trajectoire. Elle s'apparente plutôt à celle des problèmes de synthèse de Fourier déjà évoqués. Remarquons aussi que cette formulation est *continue-discrète* puisque l'on recherche une fonction $X(v)$ à partir d'un nombre fini de données discrètes y .

Il existe évidemment une infinité de solutions à (8) puisque la série des x_k connus est tronquée. Le problème est donc *indéterminé*. Tous les prolongements de la suite des x_k connus qui assurent la convergence de la série de Fourier $X(v) = \sum_{k \in \mathbb{Z}} x_k \exp(-2j\pi kv)$ sont acceptables puisque compatibles avec les données expérimentales.

Ces deux exemples sont typiques des difficultés rencontrées dans la résolution de nombreux problèmes inverses [1] [2] [5] et il importe d'en comprendre la nature avant d'étudier les solutions permettant de les contourner.

2. Aspects mathématiques

Nous verrons plus loin (§ 2.3) que le caractère inacceptable de la solution des moindres carrés obtenue dans notre premier exemple (§ 1.1) est la conséquence du *mauvais conditionnement* de la matrice A qui a pour effet d'amplifier fortement le bruit affectant les données.

Mais il est important de comprendre que ce mauvais conditionnement est lui-même la conséquence d'une difficulté fondamentale de la résolution des problèmes inverses, difficulté déjà présente dans le problème continu initial dont l'équation (3) est une approximation discrète.

2.1 Problèmes mal-posés

Hadamard a défini les conditions pour qu'un problème mathématique soit **bien-posé** [8] :

(a) pour chaque donnée y dans une classe définie Y , il existe une solution x dans une classe prescrite X (*existence*) ;

(b) la solution est unique dans X (*unicité*) ;

(c) la dépendance de x par rapport à y est continue, c'est-à-dire que lorsque l'erreur δy sur la donnée y tend vers zéro, l'erreur δx induite sur la solution x tend aussi vers zéro (*continuité*).

L'exigence de continuité est reliée à celle de *stabilité* ou de *robustesse* de la solution. La continuité est cependant une condition nécessaire, mais pas suffisante, de stabilité. Un problème bien-posé peut être *mal-conditionné*, ainsi que nous le verrons au paragraphe 2.3.

Tous les problèmes classiques de la physique mathématique, comme celui de Dirichlet pour les équations elliptiques, ou celui de Cauchy pour les équations hyperboliques, sont bien-posés au sens de Hadamard. Mais les problèmes « inverses », obtenus à partir d'un problème « direct » en échangeant les rôles de la solution et des données, ne sont généralement pas bien-posés.

Prenons un exemple élémentaire emprunté à l'optique classique. Si la distribution d'énergie x est donnée dans le plan objet d'un instrument optique et si les caractéristiques de l'instrument sont connues, il est possible de calculer la distribution d'énergie y dans le plan image. Dans le cas de l'optique de Fourier, on trouve une relation linéaire entre x et y :

$$y(s, s') = \iint h(s, r; s', r') x(r, r') dr dr' \tag{9}$$

qui est un cas particulier d'équation de Fredholm de première espèce. Le noyau h de cette équation intégrale est la réponse impulsionnelle 2-D du système d'imagerie.

Supposons que $h(s, r; s', r') = h(s - r; s' - r')$ (l'équation (9) est alors une convolution) et que $h(r; r')$ soit une fonction à bande limitée. Il existe alors des fonctions x qui produisent une image nulle (songer à une fonction dont toutes les composantes de Fourier sont en dehors de la bande passante de l'instrument) et l'unicité n'est donc pas assurée. De plus, si y est le résultat d'une expérience, et donc inévitablement entachée de bruit, elle n'est pas nécessairement à bande limitée, ou peut être à bande plus large que celle de l'instrument. Dans ces conditions, il n'existe pas de solution à l'équation (9).

Comme les données sont incertaines ou bruitées, nous ne pouvons pas espérer résoudre exactement cette équation et la solution doit être approchée dans un certain sens. Le concept de distance entre signaux est alors un moyen naturel d'évaluer la qualité d'une approximation, ce qui explique pourquoi x et y sont souvent supposés appartenir à des espaces de Hilbert. Le problème (9) peut alors être réécrit comme :

$$y = Ax \quad \text{avec } x \in \mathcal{X} \quad \text{et } y \in \mathcal{Y} \tag{10}$$

où x et y sont maintenant des éléments d'espaces fonctionnels de dimension infinie \mathcal{X} et \mathcal{Y} , respectivement, et où $A : \mathcal{X} \rightarrow \mathcal{Y}$ est l'opérateur linéaire correspondant à l'équation intégrale (9).

Supposons que x et y appartiennent au même espace de Hilbert et que h soit de carré intégrable, condition remplie par beaucoup de systèmes d'imagerie. Le problème direct est alors bien-posé : une petite erreur δx sur les données entraîne une petite erreur δy sur la solution. Cette condition n'est cependant pas remplie dans le

problème inverse correspondant où c'est l'objet x qui doit être calculé à partir de la réponse $y: x = A^{-1}y$. En effet, les conditions nécessaires et suffisantes d'existence, d'unicité et de continuité de la solution s'écrivent alors respectivement :

$$\mathcal{Y} = \text{Im}A \quad \text{Ker}A = \{0\} \quad \text{Im}A = \overline{\text{Im}A}$$

où $\text{Im}A$ est l'image de A (c'est-à-dire l'ensemble des y qui sont images d'un $x \in \mathcal{X}$), $\text{Ker}A$ son noyau (c'est-à-dire l'ensemble des solutions à l'équation $Ax = 0$), et $\overline{\text{Im}A}$ la fermeture de $\text{Im}A$. Quand le noyau h est de carré intégrable, le théorème de Riesz-Fréchet indique que l'opérateur A est borné et compact [8]. Mais l'image d'un opérateur compact n'est pas fermée (sauf dans le cas dégénéré où sa dimension est finie). Cela signifie que l'opérateur inverse A^{-1} n'est pas borné, son image n'est pas fermée, et la troisième condition de Hadamard n'est pas remplie pour le problème inverse.

Pour mieux saisir ces notions abstraites, il est commode d'utiliser les propriétés spectrales des opérateurs compacts dans des espaces de Hilbert. La propriété la plus remarquable de ces opérateurs est en effet qu'ils peuvent être décomposés en valeurs singulières, comme les matrices (la fameuse *singular value decomposition*, ou SVD). Le système singulier d'un opérateur compact est défini comme l'ensemble des solutions des équations couplées :

$$A u_n = \sigma_n v_n \quad \text{et} \quad A^* v_n = \sigma_n u_n \quad (11)$$

où les valeurs singulières σ_n sont des nombres positifs, où les fonctions singulières u_n et v_n sont respectivement des éléments de \mathcal{X} et \mathcal{Y} , et où A^* est l'opérateur adjoint de A , qui existe puisque A est continu, et donc tel que : $\langle Ax, y \rangle_{\mathcal{Y}} = \langle x, A^*y \rangle_{\mathcal{X}}$ pour tout $x \in \mathcal{X}$ et $y \in \mathcal{Y}$.

Quand A est compact, il possède toujours un système singulier $\{u_n, v_n; \sigma_n\}$ avec les propriétés suivantes [8] :

- les σ_n étant ordonnées et comptées avec leur multiplicité (qui est finie) : $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq \dots \geq 0$, σ_n tend vers 0 quand $n \rightarrow \infty$ et, ou bien la limite est atteinte pour $n = n_0$ (auquel cas l'opérateur A est dégénéré), ou bien elle n'est atteinte pour aucune valeur finie de n ;

- les fonctions u_n forment une base orthonormale de $(\text{Ker}A)^\perp$, complément orthogonal de $\text{Ker}A$ dans la décomposition : $\mathcal{X} = \text{Ker}A \oplus (\text{Ker}A)^\perp$, et les fonctions v_n forment une base orthonormale de $(\text{Ker}A^*)^\perp = \overline{\text{Im}A}$, complément orthogonal de $\text{Ker}A^*$ dans la décomposition : $\mathcal{Y} = \text{Ker}A^* \oplus (\text{Ker}A^*)^\perp$.

Soit E l'ensemble des indices n tels que $\sigma_n \neq 0$ ($E \subseteq \mathbb{N}$). Le critère de Picard [8] permet d'établir qu'une fonction $y \in \mathcal{Y}$ est dans $\text{Im}A$ si, et seulement si :

$$y \in (\text{Ker}A^*)^\perp \quad \text{et} \quad \sum_{n \in E} \sigma_n^{-2} |\langle y, v_n \rangle|^2 < +\infty \quad (12)$$

Pour que la seconde condition (12) soit satisfaite, il faut, lorsque l'opérateur A n'est pas dégénéré ($E = \mathbb{N}$), que les composantes $\langle y, u_n \rangle$ du développement de l'image y sur l'ensemble des fonctions propres $\{v_n\}$ tendent vers zéro plus vite que les valeurs propres σ_n^2 lorsque $n \rightarrow \infty$. Cette condition sévère n'a aucune raison d'être satisfaite par une fonction arbitraire de $\overline{\text{Im}A}$. Remarquons cependant qu'elle l'est naturellement si $y = Ax$ est l'image parfaite résultant d'un objet x d'énergie finie. La solution s'écrit alors :

$$x = \sum_{n \in E} \sigma_n^{-1} \langle y, v_n \rangle u_n \quad (13)$$

mais cette solution, quand elle existe, est instable : une petite perturbation additive $\delta y = \varepsilon v_n$, par exemple, sur les données parfaites y entraîne une perturbation δx sur la solution calculée avec les données $y + \delta y$:

$$\delta x = \sigma_n^{-1} \langle \delta y, v_n \rangle u_n = \sigma_n^{-1} \varepsilon u_n \quad (14)$$

Le rapport $\|\delta x\|/\|\delta y\|$ vaut σ_n^{-1} qui peut être arbitrairement grand. L'opérateur linéaire inverse $A^{-1} : \mathcal{Y} \rightarrow \mathcal{X}$, défini par (13), n'est donc pas borné puisqu'il n'est pas possible de trouver une constante C telle que, pour tout $y \in \mathcal{Y}$, on ait $\|A^{-1}y\|_{\mathcal{X}} \leq C\|y\|_{\mathcal{Y}}$, ce qui est une condition nécessaire et suffisante pour que A^{-1} soit continu [9].

2.2 Inversion généralisée

La nécessité d'approfondir des problèmes qui ne sont pas mathématiquement bien-posés, mais cependant d'un grand intérêt pour les sciences de l'ingénieur, est à l'origine de deux branches récentes de l'analyse : la théorie de l'inversion généralisée et celle de la régularisation (§ 3).

Supposons que l'équation $Ax = 0$ ait des solutions non triviales. L'ensemble $\text{Ker}A \neq \{0\}$ de ces solutions est un sous-espace fermé de \mathcal{X} . C'est l'ensemble des « objets invisibles », puisqu'ils produisent une image y nulle. Supposons aussi que $\text{Im}A$ soit un sous-espace fermé de \mathcal{Y} . Un exemple est fourni par l'opérateur intégral correspondant à un filtrage passe-bas idéal :

$$(Ax)(r) = \int_{-\infty}^{+\infty} \frac{\sin \Omega(r-r')}{\pi(r-r')} x(r') dr' \quad (15)$$

Si nous choisissons $\mathcal{X} = \mathcal{Y} = L^2_{\mathbb{R}}$, le noyau est l'ensemble de toutes les fonctions x dont la transformée de Fourier s'annule dans la bande $[-\Omega, +\Omega]$, alors que l'image de A est l'ensemble des fonctions à bande limitée sur le même intervalle, qui est un sous-espace fermé de $L^2_{\mathbb{R}}$.

Un moyen de restaurer l'existence et l'unicité de la solution dans ces conditions est de redéfinir à la fois l'espace \mathcal{X} des solutions et l'espace \mathcal{Y} des données. Si nous choisissons un nouvel espace \mathcal{X}' qui est l'ensemble de toutes les fonctions orthogonales à $\text{Ker}A$ (dans le cas (15), \mathcal{X}' est l'ensemble des fonctions de carré sommable et à bande limitée à l'intervalle $[-\Omega, +\Omega]$), et si nous prenons pour nouvel espace \mathcal{Y}' des données $\text{Im}A$ (qui est à nouveau, dans le cas (15), l'ensemble des fonctions de carré sommable et à bande limitée à l'intervalle $[-\Omega, +\Omega]$), alors, pour tout $y \in \mathcal{Y}'$, il existe une unique $x \in \mathcal{X}'$ tel que $Ax = y$ (dans notre exemple (15), la solution est même triviale : $x = y$) et le nouveau problème est donc bien-posé.

Le fait qu'il soit en principe possible de choisir, dans la plupart des cas, les espaces \mathcal{X} et \mathcal{Y} pour que le problème devienne bien-posé, est de peu d'intérêt pratique puisque c'est l'application visée qui impose en général les espaces appropriés. Un autre moyen envisageable est de changer la notion même de solution.

■ Pseudo-solutions

Considérons tout d'abord le cas où A est injectif ($\text{Ker}A = \{0\}$) mais la condition d'existence n'est pas toujours satisfaite ($\text{Im}A \neq \mathcal{Y}$). L'ensemble des fonctions $x \in \mathcal{X}$ solutions du problème variationnel

$$x = \arg \min_{x \in \mathcal{X}} \|Ax - y\|_{\mathcal{Y}}^2 \quad (16)$$

où $\|\cdot\|_{\mathcal{Y}}$ désigne la norme dans \mathcal{Y} , sont appelées pseudo-solutions, ou solutions des moindres carrés du problème (10). En annulant la première variation de la fonctionnelle à minimiser dans (16), on obtient l'équation d'Euler :

$$A^*Ax = A^*y \quad (17)$$

qui fait intervenir l'opérateur auto-adjoint A^*A dont le système propre se déduit du système singulier de A . Quand $\text{Im}A$ est fermée, l'équation (17) possède toujours une solution, mais elle n'est pas unique si le noyau $\text{Ker}A$ n'est pas trivial. Lorsqu'il l'est, comme supposé ici, le caractère bien-posé a été restauré.

Solutions généralisées

Considérons maintenant le cas où la condition d'unicité n'est pas satisfaite ($\text{Ker}A \neq \{0\}$, le problème est indéterminé). L'ensemble des solutions de (17) étant un sous-ensemble fermé convexe de \mathfrak{X} , il contient donc un élément unique de norme minimale, qui est noté x^\dagger ou \hat{x}_{IG} , et appelé *solution généralisée* de (10). x^\dagger étant orthogonal à $\text{Ker}A$, cette manière de définir la solution est équivalente à choisir $\mathfrak{X}' = (\text{Ker}A)^\perp$.

Comme il existe un unique x^\dagger pour tout $y \in \mathfrak{Y}$, une application linéaire A^\dagger de \mathfrak{Y} dans \mathfrak{X} est définie par :

$$A^\dagger y = x^\dagger = \hat{x}_{IG} \tag{18}$$

L'opérateur A^\dagger est appelé *inverse généralisé* de A et il est continu [10].

Le problème du calcul de la solution généralisée de l'équation (10) est bien-posé si, et seulement si, $\text{Im}A$ est fermée. La raison essentielle est que c'est une condition nécessaire et suffisante pour que l'on puisse décomposer \mathfrak{Y} selon $\mathfrak{Y} = \text{Im}A \oplus (\text{Im}A)^\perp$, où \oplus désigne la somme directe et $(\text{Im}A)^\perp$ le complément orthogonal de $\text{Im}A$.

Cas des données discrètes

Lorsque les données sont discrètes, y est un vecteur de dimension N dans un espace euclidien. Ignorant les erreurs sur les données, un problème inverse linéaire avec des données discrètes peut être énoncé comme suit : étant donné un ensemble $\{F_i(x)\}_{i=1}^N$, de fonctionnelles linéaires définies sur \mathfrak{X} , et un ensemble $\{y_i\}_{i=1}^N$ de nombres, trouver une fonction $x \in \mathfrak{X}$ telle que :

$$y_i = F_i(x) \quad i = 1, \dots, N$$

En particulier, quand les fonctionnelles F_i sont continues sur \mathfrak{X} , le théorème de Riesz dit qu'il existe des fonctions ψ_1, \dots, ψ_N telles que :

$$F_i(x) = \langle x, \psi_i \rangle_{\mathfrak{X}}$$

L'exemple élémentaire de (9) prend cette forme quand $y(r)$ est mesurée sur un ensemble fini de points, disons r_1, \dots, r_N , et que \mathfrak{X} est un espace L^2 . Dans ce cas, nous avons :

$$\psi_i(s) = h(r_i, s)$$

Ce problème est un cas particulier de celui de l'équation (10) si nous définissons un opérateur A de \mathfrak{X} dans \mathfrak{Y} par la relation :

$$(Ax)_i = \langle x, \psi_i \rangle_{\mathfrak{X}} \quad i = 1, \dots, N$$

L'opérateur A n'est pas injectif : $\text{Ker}A$ est le sous-espace fermé de dimension infinie de toutes les fonctions x orthogonales au sous-espace engendré par les fonctions ψ_j . Inversement, l'image de A , $\text{Im}A$ est fermée : $\text{Im}A$ est simplement \mathfrak{Y} quand les fonctions ψ_j sont linéairement indépendantes ; sinon, c'est un sous-espace de dimension $N' < N$.

Exemple

Pour illustrer cette notion d'inversion généralisée, revenons à notre exemple de synthèse de Fourier du paragraphe 1.2 qui entre dans la catégorie définie dans le paragraphe précédent, avec les exponentielles de Fourier comme fonctions ψ_j . Pour imposer une solution unique dans la classe des solutions possibles effectuant un prolongement de la suite des x_k connus, nous pouvons choisir la solution inverse généralisée du problème initial (8) :

$$\begin{aligned} \hat{X}_{IG}(v) &= \arg \min_{X \in L^2_{\mathbb{C}}[0,1]} \int_0^1 |X(v)|^2 dv \\ \text{s.c. } x_k &= y_k \quad k = 1, \dots, N \end{aligned}$$

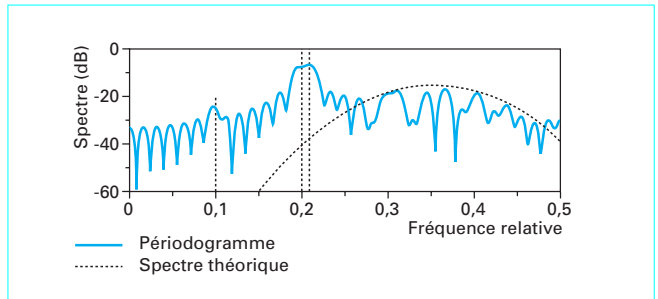


Figure 7 – Spectre du périodogramme et spectre théorique

Les relations de Plancherel et Parseval nous montrent que c'est équivalent à rechercher les coefficients :

$$\hat{x}_k = \int_0^1 \hat{X}(v) \exp(2j\pi kv) dv \quad k \in \mathbb{Z}$$

tels que :

$$\hat{x} = \arg \min_{x \in \ell^2_{\mathbb{C}}} \sum_{k \in \mathbb{Z}} |x_k|^2 \quad \text{s.c. } x_k = y_k \quad k = 1, \dots, N$$

La solution est triviale puisque le problème est séparable :

$$\hat{x}_k = \begin{cases} y_k & \text{si } k \in \{1, 2, \dots, N\} \\ 0 & \text{sinon} \end{cases}$$

On a donc :

$$\hat{X}_{IG}(v) = \sum_{k=1}^N y_k \exp(-2j\pi kv) \tag{19}$$

qui donne, après quadrature, le **périodogramme** de Schuster, à un coefficient près :

$$P_x(v) \stackrel{\text{def}}{=} \frac{1}{N} |X(v)|^2 = \frac{1}{N} \left| \sum_{n=1}^N x_n \exp(-2j\pi vn) \right|^2$$

Le périodogramme de notre exemple est tracé figure 7. Les deux raies au voisinage de la fréquence réduite 0,2 ne sont pas séparées, en raison d'un nombre insuffisant d'échantillons. La troisième raie en 0,1 se distingue mal du bruit dont le spectre est assez mal rendu. Nous avons résolu le problème de l'unicité mais la solution n'est pas pour autant satisfaisante.

2.3 Discrétisation

Une première description d'un problème direct fait souvent intervenir des fonctions de variables réelles (temps, fréquence, variables d'espace, ...), représentant les grandeurs physiques mises en jeu : grandeurs accessibles à la mesure et grandeurs d'intérêt inconnues. L'analyse du problème à ce niveau de description a permis de fournir une explication des difficultés rencontrées lors de l'inversion, en se plaçant dans des espaces fonctionnels de dimension infinie. Nous avons vu ainsi que dans le cas d'un problème inverse décrit par une équation intégrale de première espèce, l'inversion était un problème mal-posé au sens de Hadamard.

Cette analyse est cependant insuffisante. Les données expérimentales disponibles sont presque toujours constituées de la mesure de grandeurs physiques accessibles en un nombre nécessairement fini de points du domaine de définition de leurs variables. Elles sont donc

Parution : novembre 2001 - Dernière validation : juin 2019 - Ce document a été délivré pour le compte de 7200106152 - éditions tti // celine BLONDOU // 2.59.188.28

naturellement discrètes, et nous les regroupons dans le vecteur y comme nous l'avons fait dans le paragraphe 1.2. Mais l'objet inconnu est lui aussi discrétisé, soit d'emblée, soit en cours de résolution (comme dans l'exemple du périodogramme (§ 2.2)), par décomposition sur un nombre fini de fonctions. Si celles-ci sont des éléments d'une base de l'espace auquel l'objet appartient, la décomposition est nécessairement tronquée. En imagerie par exemple, on utilise comme fonctions de base dans la très grande majorité des cas les indicatrices des pixels, ou les sinus cardinaux, selon que l'objet est implicitement supposé à support limité ou à spectre limité. On commence aussi à utiliser des bases d'ondelettes [11]. Le point de départ est donc constitué d'un *modèle paramétré* par le vecteur x des coefficients de la décomposition, c'est-à-dire d'un ensemble d'hypothèses exclusives dont chacune est indexée par la valeur des coefficients. Cet espace d'hypothèses est donc l'ensemble des valeurs possibles de ces paramètres inconnus, $\mathcal{X} = \{x_j\}$. Le choix de ces fonctions de base fait évidemment partie du problème d'inversion, même s'il est assez peu abordé [12].

Dans le **cas discret** (ou plus exactement discret-discret), le problème change sensiblement car x et y appartiennent à des espaces de dimension finie et l'opérateur linéaire A devient une matrice A . L'équation (3) possède une solution unique de norme minimale $\hat{x}_{IG} = A^+y$ qui dépend continûment de y puisque l'inverse généralisée A^+ est alors toujours bornée [8]. Le problème est donc toujours bien-posé au sens de Hadamard. Mais, même dans ce cadre, le problème d'inversion a encore, d'un point de vue numérique cette fois, un caractère instable. La décomposition spectrale (14) est toujours valable, à la seule différence que le nombre de valeurs singulières de la matrice normale A^+A est maintenant fini et égal à M . Ces valeurs singulières peuvent être rarement calculées explicitement, mais dans le cas particulier de la restauration d'image avec un objet et un noyau à supports finis (ce qui est le cas de l'exemple proposé), la décomposition spectrale (14) est une décomposition de Fourier et les valeurs propres de A^+A sont les valeurs du spectre de Fourier du système d'imagerie, évaluées à des fréquences spatiales régulièrement espacées. Elles sont présentées sur la figure 3 et certaines sont clairement très proches de zéro. De manière plus générale, quand le système d'imagerie est du type passe-bas, ce qui est très fréquent, σ_n prend de très faibles valeurs lorsque $n \rightarrow M$. Même si nous excluons les valeurs singulières nulles, comme en utilisant A^+ , il y a toujours des valeurs singulières proches de zéro. La matrice A est donc *mal-conditionnée*. Les coefficients $\sigma_n^{-1}(\delta y, u_n)$ dans (14) deviennent très grands pour les σ_n proches de zéro, même si δy est petit.

De manière générale, que l'on soit dans le cas discret ou non, supposons que $\text{Im}A$ soit fermée pour que l'inverse généralisé A^+ soit continu. Désignons par δy une erreur sur les données y et par δx^+ l'erreur induite sur la solution inverse généralisée x^+ . La linéarité de (18) entraîne que $\delta x^+ = A^+ \delta y$, ce qui implique :

$$\|\delta x^+\|_x \leq \|A^+\| \|\delta y\|_y$$

où $\|A^+\|$ désigne la norme de l'opérateur continu A^+ . De manière analogue, l'équation (10) implique :

$$\|y\|_y \leq \|A\| \|x^+\|_x$$

En combinant ces deux relations, on obtient l'inégalité :

$$\frac{\|\delta x^+\|_x}{\|x^+\|_x} \leq \|A\| \|A^+\| \frac{\|\delta y\|_y}{\|y\|_y} \quad (20)$$

Il est important de noter que cette inégalité est précise dans un certain sens. Quand A est une matrice de dimensions $(N \times M)$ ou correspond à un problème inverse avec des données discrètes, alors l'inégalité peut devenir une égalité. Quand A est un opérateur sur des espaces de dimension infinie, alors on peut seulement

établir que le membre de gauche de l'inégalité (20) peut être arbitrairement proche du membre de droite. La quantité :

$$c = \|A\| \|A^+\| \geq 1 \quad (21)$$

est appelée **nombre de condition** du problème. Quand c est proche de l'unité, le problème est dit *bien-conditionné*, alors que quand il est grand devant l'unité, le problème est dit *mal-conditionné*.

Il est important en pratique d'avoir une idée du nombre de condition qui renseigne sur la stabilité numérique du problème. Quand A est une matrice de dimensions $(N \times M)$, $\|A\|$ est la racine carrée de la plus grande des valeurs propres de la matrice A^*A , de dimensions $(M \times M)$ (les valeurs propres positives de cette matrice coïncident avec celles de la matrice AA^*), et $\|A^+\|$ est l'inverse de la racine carrée de la plus petite de ces mêmes valeurs propres :

$$c = \sqrt{\lambda_{\max} / \lambda_{\min}}$$

2.4 Résumé

■ Quand nous avons affaire à un opérateur défini sur des espaces de Hilbert \mathcal{X} et \mathcal{Y} , et du type $A: \mathcal{X} \rightarrow \mathcal{Y}$, linéaire et continu, nous avons trois situations principales.

- Si A est injectif (l'équation $Ax = 0$ a comme seule solution la solution triviale $x = 0$, donc $\text{Ker}A = \{0\}$), et que son image est fermée et donnée par $\text{Im}A = \mathcal{Y}$, le problème inverse est bien-posé car l'opérateur inverse est continu.

- Si A n'est pas injectif, mais que $\text{Im}A$ est fermée, alors, en recherchant une pseudo-solution, le problème inverse devient bien-posé dans la mesure où l'inverse généralisé est continu.

- Si l'image $\text{Im}A$ n'est pas fermée, l'emploi d'une pseudo-solution ne peut, en soi, garantir l'existence et la continuité de la solution inverse.

■ Quand nous avons affaire à un opérateur linéaire défini sur des espaces de dimension finie \mathbb{R}^N et \mathbb{R}^M , et du type $A: \mathbb{R}^N \rightarrow \mathbb{R}^M$, nous avons à nouveau trois situations principales.

- Si p désigne le rang de la matrice associée à l'opérateur A et si $p = N = M$, alors A est injectif et $\text{Im}A = \mathbb{R}^M$; une solution existe toujours, qui est unique, et le problème inverse est bien défini.

- Si $p < N$, alors l'unicité n'est pas assurée, mais elle peut être rétablie en considérant une inversion généralisée.

- Si $p < M$, alors l'existence ne peut être assurée pour des données quelconques, mais elle peut l'être en considérant à nouveau une inversion généralisée.

Pour des opérateurs dans des espaces de dimension finie, l'inverse ou l'inverse généralisé est toujours continu. En conséquence, l'utilisation d'une inversion généralisée suffit à garantir le caractère bien-posé. Mais il faut se souvenir qu'un problème bien-posé peut être mal-conditionné, et que dans ce cas il est nécessaire d'employer les méthodes de régularisation que nous allons présenter au paragraphe 3, comme pour les problèmes mal-posés.

3. Régularisation

Quand l'image $\text{Im}A$ d'un opérateur linéaire à inverser n'est pas fermée, alors l'inverse A^{-1} , ou l'inverse généralisé A^+ , n'est pas défini partout sur \mathcal{Y} et n'est pas continu. C'est le cas par exemple des opérateurs compacts non dégénérés (ou de rang non fini), et il est facile de voir que le nombre de condition du problème est infini. Des techniques de résolution appropriées sont alors requises, mais il faut

voir aussi qu'un problème bien-posé, mais sévèrement mal-conditionné, se comporte en pratique comme un problème mal-posé et doit être traité par les mêmes techniques.

Que l'on soit en dimension finie ou infinie, un **régulariseur** de l'équation (10) $y = Ax$ est une famille d'opérateurs $\{R_\alpha; \alpha \in \Lambda\}$ tels que :

- 1) $\forall \alpha \in \Lambda, R_\alpha$ est un opérateur continu de \mathcal{Y} dans \mathcal{X} ;
- $\forall y \in \text{Im}A, \lim_{\alpha \rightarrow 0} R_\alpha y = A^\dagger y$.

Autrement dit, puisque l'opérateur inverse A^{-1} ne présente pas les propriétés de continuité ou de stabilité requises, on construit une famille d'opérateurs continus, indexés par un paramètre de réglage α (appelé *coefficient de régularisation*), et incluant comme cas limite A^\dagger . Appliqué à des données parfaites y , R_α fournit une approximation de x^\dagger d'autant meilleure que $\alpha \rightarrow 0$. Cependant, quand R_α est appliqué à des données $y_b = Ax + b$ inévitablement entachées d'un bruit b , on obtient une solution *approchée* $x_b = R_\alpha y_b$, et l'on a :

$$R_\alpha y_b = R_\alpha y + R_\alpha b \tag{22}$$

Le second terme diverge quand $\alpha \rightarrow 0$. Il s'ensuit qu'un compromis doit être trouvé entre deux termes antagonistes, l'erreur d'approximation (premier terme) et l'erreur due au bruit (second terme). Cela se fait, au sein d'une famille donnée d'opérateurs R_α , en ajustant la valeur du coefficient de régularisation α .

La plupart des méthodes qui ont été proposées depuis les années 1970 pour résoudre et stabiliser les problèmes mal-posés entrent, d'une façon ou d'une autre, dans ce schéma général. On peut les classer en deux grandes familles : celles qui procèdent par contrôle d'une dimension $-\Lambda$ est alors un ensemble discret – et celles qui opèrent par minimisation d'un critère composite ou par optimisation sous contrainte $-\Lambda$ est alors \mathbb{R}_+ . Nous nous intéresserons ici principalement à la seconde famille mais nous consacrons le paragraphe suivant à la première.

3.1 Contrôle de dimension

3.1.1 Décomposition tronquée en valeurs singulières

Un exemple typique de méthodes de la première famille est fourni par l'examen de (14) : pour supprimer le mauvais conditionnement du problème, il suffit de tronquer le développement en retenant les composantes correspondant à des valeurs singulières suffisamment grandes pour que les termes d'erreur de la forme $\sigma_n^{-1} \langle \delta y, u_n \rangle$ restent faibles. C'est la *décomposition tronquée en valeurs singulières* [8]. Elle est très efficace pour assurer la stabilité numérique. Se pose tout de même le problème du choix de l'ordre de la troncature, qui joue ici le rôle de l'inverse d'un coefficient de régularisation. Mais le principal défaut de cette approche est que l'on renonce à rétablir les composantes spectrales qui ont été trop dégradées par le dispositif d'imagerie. Comme pour la définition du critère de résolution de Rayleigh en optique, on n'utilise aucune autre information sur l'objet recherché que le fait que son énergie soit finie, alors que, la plupart du temps, on sait qu'il est par exemple positif, ou bien qu'il comporte des régions à variations spatiales « douces » séparées par des contours nets, ou bien qu'il est à valeurs bornées, ou bien encore à support borné. Si l'on veut aller au-delà de cette résolution de Rayleigh, il est indispensable de pouvoir prendre en compte ce genre d'information *a priori*.

3.1.2 Changement de discrétisation

Dans la décomposition tronquée en valeurs singulières, c'est le dispositif d'imagerie qui impose, en quelque sorte, la discrétisation

au travers des fonctions singulières de l'opérateur correspondant. Mais on peut aussi contourner les difficultés soulevées par le mauvais conditionnement de la matrice **A**, conséquence d'une discrétisation de l'objet sur une grille cartésienne par exemple, en choisissant une *paramétrisation parcimonieuse* de l'objet mieux adaptée aux propriétés *a priori* de celui-ci. C'est ce que font les méthodes de décomposition sur une base d'ondelettes [11] [13], par exemple. Le principe reste le même : on procède par *seuillage* parmi les coefficients de la décomposition, en éliminant le sous-espace dominé par les composantes du bruit.

Ce mode de discrétisation résout, par principe, le problème de stabilité ou de mauvais conditionnement analysé plus haut, mais ne fournit pas pour autant une solution satisfaisante. Tout dépend de la décomposition choisie. Pour l'illustrer, nous allons étudier deux des paramétrisations les plus utilisées dans les problèmes d'analyse spectrale.

Dans la première, le signal à analyser est supposé être une *réalisation d'un processus aléatoire* à temps discret, stationnaire (pour pouvoir parler de spectre ou, plus précisément dans ce cas, de *densité spectrale de puissance*), et *autorégressif* (c'est-à-dire engendré par le passage d'un processus aléatoire $u(n)$ blanc et stationnaire dans un filtre linéaire, invariant, et de fonction de transfert $G(z)$ « tout pôles »). Le filtre générateur est décrit par l'équation récurrente :

$$x(n) = a_1 x(n-1) + \dots + a_d x(n-d) + u(n)$$

et possède une fonction de transfert en z :

$$G(z) = \frac{1}{1 - a_1 z^{-1} - \dots - a_d z^{-d}}$$

Les coefficients a_k sont appelés coefficients de régression ou coefficients autorégressifs (AR). Si la variance de $u(n)$ est notée σ_u^2 , on a le résultat consécutif :

$$g_x(v) = \sigma_u^2 |G(e^{2j\pi v})|^2 \tag{23}$$

Le problème d'analyse spectrale se réduit ainsi à celui de l'estimation de ces $d + 1$ paramètres (d paramètres AR et la variance du bruit). De nombreuses variantes existent selon la nature du cadre statistique adopté et les approximations faites [6].

La figure 8 donne un exemple de spectre obtenu avec une méthode du maximum de vraisemblance, ou des moindres carrés (elles sont en effet équivalentes avec les hypothèses habituellement faites), et un ordre d déterminé par le critère classique de Rissanen-Schwarz (BIC : *Bayesian information criterion*). Le spectre est beaucoup plus régulier que celui du périodogramme (figure 7), mais la résolution de la méthode reste médiocre.

Un autre exemple de paramétrisation parcimonieuse est fourni par la figure 9. Le signal est cette fois modélisé comme une somme de sinusoides pures et d'un bruit blanc, ce qui est manifestement mieux adapté à notre exemple. Sous ces hypothèses, on peut mettre en évidence des propriétés géométriques intéressantes des « sous-espace signal » (entendons par là les sinusoides) et « sous-espace bruit » dont les dimensions respectives apparaissent dans le spectre des valeurs propres de la matrice de corrélation du signal analysé. La détermination de ces valeurs propres, et la résolution d'une équation polynomiale associée, permettent de déterminer le nombre et les valeurs des fréquences pures contenues dans le signal [6]. Mais, dans la pratique, la corrélation du signal n'est pas connue et doit être estimée à partir de ses échantillons. De plus, tout ceci repose sur l'hypothèse d'un bruit blanc, non vérifiée dans notre exemple. Les trois fréquences pures réellement présentes dans le signal sont correctement estimées, mais sept autres, correspondant au bruit coloré, ont été aussi extraites.

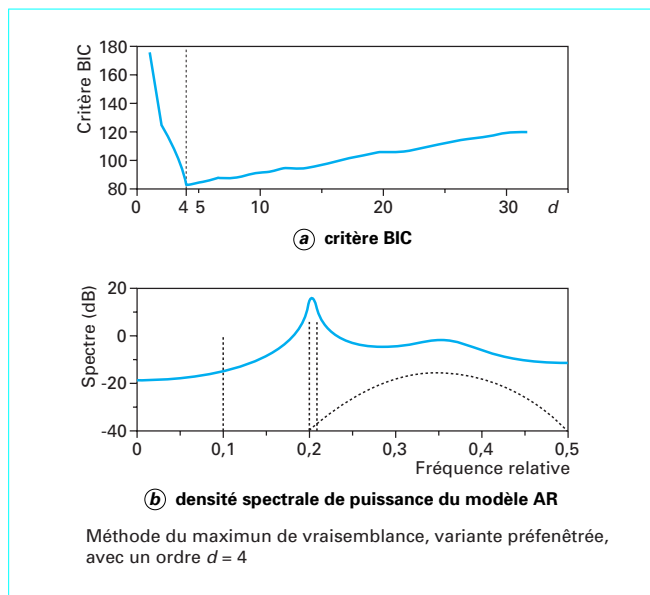


Figure 8 – Spectre obtenu par une méthode du maximum de vraisemblance ou des moindres carrés

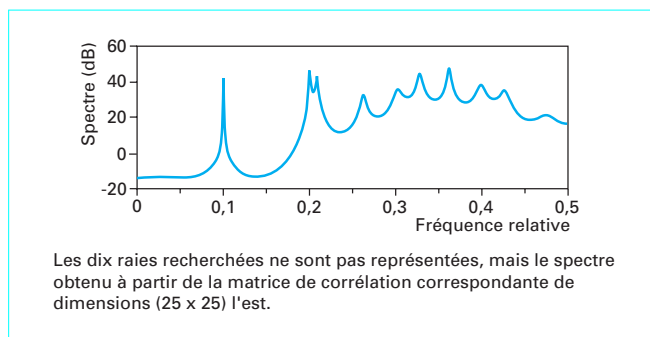


Figure 9 – Spectre obtenu par la méthode MUSIC

3.1.3 Méthodes itératives

Une famille de méthodes très populaires est constituée des méthodes *itératives*, de la forme :

$$x^{(n+1)} = x^{(n)} + \alpha (y - Ax^{(n)}) \quad n = 0, 1, \dots \quad (24)$$

où $0 < \alpha < 2/\|A\|$ (méthode de Bialy [14]). Si A est un opérateur linéaire borné non négatif (c'est-à-dire tel que $\langle Ax, x \rangle \geq 0, \forall x \in \mathcal{X}$), et si $y = Ax$ possède au moins une solution, alors la suite des $x^{(n)}$ converge, et :

$$\lim_{n \rightarrow \infty} x^{(n)} = Px^{(0)} + x_{IG}$$

où P est l'opérateur de projection orthogonale sur $\text{Ker } A$ et x_{IG} la solution inverse généralisée. Cette méthode recherche en fait le point fixe de l'opérateur $G : Gx = \alpha y + (I - \alpha A)x$, mais si A est compact et \mathcal{X} de dimension infinie, alors $(I - \alpha A)$ n'est pas une contraction et la méthode diverge. De plus, nous avons vu aussi que, même en dimension finie, la solution inverse généralisée était souvent dominée par le bruit.

La condition de non négativité exclut beaucoup d'opérateurs, mais la méthode peut être appliquée à la résolution de l'équation normale $A^*y = A^*Ax$ puisque A^*A est un opérateur non négatif. On obtient ainsi la méthode de Landweber [14] :

$$x^{(n+1)} = x^{(n)} + \alpha A^* (y - Ax^{(n)}) \quad n = 0, 1, \dots$$

avec $0 < \alpha < 2/\|A^*A\|$. La méthode bien connue de Gerchberg-Papoulis-Van-Cittert pour l'extrapolation d'un signal à spectre limité est un cas particulier de la méthode de Bialy-Landweber. C'est dans la même catégorie des méthodes itératives que l'on peut ranger la méthode de Lucy [15], très populaire en astronomie.

Toutes ces méthodes ne peuvent fournir de solution acceptable qu'à la condition de limiter le nombre d'itérations (qui joue alors le rôle de l'inverse d'un coefficient de régularisation). Cela se fait souvent de manière empirique car le cadre théorique initial ne prenant pas en compte le bruit d'observation, une théorie de réglage du nombre d'itérations pour limiter l'amplification du bruit ne peut être qu'hétéronome [16]. Cela explique pourquoi nous nous focalisons sur les méthodes de régularisation de la seconde famille qui sont, de ce point de vue, plus autonomes.

3.2 Minimisation d'un critère composite

Le cadre adopté dans la suite relève de la seconde famille des méthodes de régularisation dont la caractéristique principale est de demander à la solution de réaliser un compromis entre une fidélité aux données mesurées et une fidélité à une information *a priori* [17]. Ce compromis est établi à l'aide d'un unique critère d'optimalité. Cette démarche peut s'interpréter de la manière suivante.

La solution des moindres carrés \hat{x}_{MC} minimise l'énergie de l'erreur entre le modèle Ax et les données y . En ce sens, elle procure la plus grande fidélité aux données. Mais lorsque le bruit d'observation est à large bande, l'équation (14) montre que les composantes de l'objet restauré ou reconstruit à des fréquences spatiales élevées ont des amplitudes élevées qui résultent principalement du bruit. La solution

\hat{x}_{MC} , bien que statistiquement sans biais, est alors rejetée car on s'attend à ce que le vrai objet ait des variations spatiales sensiblement plus douces. Il faut donc introduire un peu d'infidélité aux données pour obtenir une solution moins rugueuse que celle des moindres carrés et plus conforme à l'idée que l'on s'en fait *a priori*. Un moyen désormais classique d'y parvenir est fourni par la minimisation d'un critère composite [8] [18] [19]. L'idée de base est de renoncer à l'espoir d'accéder à la solution exacte à partir de données imparfaites, de considérer que l'équation (3) définit une

classe de solutions admissibles $\{\hat{x} : \|y - Ax\| \leq \|b\|\}$, et de rechercher dans cette classe une solution qui peut être considérée comme étant physiquement raisonnable, c'est-à-dire compatible avec une certaine information *a priori*. Cela s'obtient usuellement

en recherchant une solution $\hat{x}(\alpha, y)$ qui minimise un critère de la forme :

$$\mathcal{J}(x) = \mathcal{G}(y - Ax) + \alpha \mathcal{F}(x), \quad \alpha \in \mathbb{R}_+^* \quad (25)$$

spécifiquement conçu pour :

- que la solution soit, jusqu'à un certain point, fidèle aux données (premier terme du critère) ;
- renforcer certaines propriétés souhaitables qui résument notre connaissance *a priori* sur la solution (second terme).

Cette manière de procéder s'oppose ainsi aux méthodes de régularisation par contrôle de dimension qui, dans le cas d'un problème mal-posé ou mal-conditionné, contournent la difficulté en cherchant le minimum de (5) dans un sous-espace de dimension réduite, après éventuellement un changement de base approprié.

Le choix des fonctionnelles \mathcal{F} et \mathcal{G} est un choix qualitatif qui détermine la manière selon laquelle la régularisation est effectuée. Inversement, le choix du **coefficient de régularisation** α est quantitatif et permet d'ajuster le compromis entre les deux sources d'information. Une fidélité parfaite aux données est obtenue avec $\alpha = 0$, alors qu'une fidélité parfaite à l'a priori est obtenue si $\alpha = \infty$.

La littérature sur le sujet est dominée par quelques fonctionnelles dont voici les plus fréquentes [17].

■ **Distances quadratiques**

La distance euclidienne entre deux objets \mathbf{x}_1 et \mathbf{x}_2 est $\|\mathbf{x}_1 - \mathbf{x}_2\|_{\mathbf{P}}$ où \mathbf{P} est une matrice symétrique définie non négative qui est choisie pour traduire certaines caractéristiques souhaitées de la mesure de proximité. Une distance de ce type constitue le choix habituel pour \mathcal{G} dans le cas où le bruit \mathbf{b} est supposé gaussien, centré et indépendant de \mathbf{x} . Elle est aussi très souvent utilisée pour \mathcal{F} afin de pénaliser les objets \mathbf{x} de grande amplitude.

Appliqué à notre exemple d'analyse spectrale (qui est, rappelons-le, du type continu-discret), ce mode de régularisation nous conduit à rechercher la transformée de Fourier du signal vérifiant :

$$\hat{X}(v) = \arg \min_{X \in L^2_{\mathbb{C}}[0,1]} \left[\|\mathbf{x}_N - \mathbf{y}\|^2 + \alpha \int_0^1 |X(v)|^2 dv \right]$$

où $\mathbf{x}_N = [x_1, \dots, x_N]^t$ et $x_k = \int_0^1 X(v) \exp(2j\pi kv) dv$

La solution s'écrit :

$$\hat{X}^{(\alpha)}(v) = \frac{1}{\alpha + 1} \hat{X}_{IG}(v) \tag{26}$$

Le spectre ainsi régularisé est donc simplement proportionnel au périodogramme (19) que l'on obtient comme cas limite ($\alpha \rightarrow \infty$). Ce type de régularisation ne convient donc pas dans cet exemple.

■ **Mesures de rugosité**

Une manière très simple de mesurer la rugosité d'une image consiste à lui appliquer un opérateur de différentiation approprié, puis à calculer la norme euclidienne de l'image résultante. Comme l'opération de différentiation est linéaire, la mesure de rugosité résultante est quadratique :

$$\mathcal{F}(\mathbf{x}) = \|\nabla^k(\mathbf{x})\|^2 = \|\mathbf{D}_k \mathbf{x}\|^2 \tag{27}$$

L'ordre k de l'opérateur des différences $\nabla(\cdot)$ vaut habituellement 1 ou 2. La mesure (27) est minimale lorsque toutes les composantes de \mathbf{x} sont égales (si $k = 1$).

Appliqué à notre exemple de déconvolution d'image, ce mode de régularisation conduit au résultat de la figure 10. L'amélioration est cette fois sensible par rapport à la solution des moindres carrés de la figure 4, mais de nombreux détails de l'objet original de la figure 1 ne sont pas visibles, notamment ceux de la caméra. Nous y reviendrons au paragraphe 5.

■ **Distance de Kullback**

Dans de nombreux problèmes de traitement d'image, une caractéristique essentielle à préserver est la positivité des intensités des pixels. Il existe plusieurs manières d'y parvenir, l'une d'entre elles étant de considérer que l'objet positif peut être identifié, après normalisation, à une distribution de probabilité, puis d'utiliser des mesures de distance entre lois de probabilité. La pseudo-distance de Kullback :

$$\mathcal{F}(\mathbf{x}) = \mathcal{D}_{KL}(\mathbf{x}, m) = \sum_{j=1}^M x_j \ln \frac{x_j}{m_j} \tag{28}$$

où l'objet de référence m est un vecteur positif, est souvent utilisée.



Figure 10 – Déconvolution linéaire-quadratique

Conclusion

Le critère (25) résume une vision de la régularisation que l'on peut qualifier de *déterministe* puisque la seule distribution de probabilité utilisée est, au moins implicitement au travers du choix de la fonctionnelle \mathcal{G} , celle du bruit. Elle a donné lieu à des développements théoriques importants, essentiellement en physique mathématique. Cependant, les questions du choix de la fonctionnelle régularisante $\mathcal{F}(\mathbf{x})$ et du coefficient de régularisation α y sont encore largement ouvertes.

3.3 Choix du coefficient de régularisation

Dans de nombreux problèmes inverses, un réglage précis du coefficient de régularisation n'est pas nécessaire, la solution n'étant sensible qu'à des variations d'environ un ordre de grandeur de α . Il est alors possible, avec un peu d'expérience, de choisir empiriquement α , dans un mode *supervisé*. Dans le cas contraire il existe, dans le cadre de ce paragraphe, quelques **méthodes de réglage automatique**, à partir des données, c'est-à-dire en mode *non supervisé* [20].

■ **Contrôle de l'énergie de l'erreur résiduelle**

Une des idées les plus intuitives et les plus anciennes pour régler la valeur de α qui intervient dans le critère régularisé, ou pénalisé (25), est de considérer α comme un multiplicateur de Lagrange dans le problème équivalent :

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \mathcal{F}(\mathbf{x}) \quad \text{s.c.} \quad \mathcal{G}(\mathbf{y} - \mathbf{A}\mathbf{x}) = c \tag{29}$$

Le degré de régularisation est fixé par la valeur de c . Lorsque \mathcal{G} est quadratique, on préconise souvent de choisir $c = N$ si \mathcal{G} est pondérée par l'inverse de la matrice de covariance du bruit, ou $N\sigma^2$ si \mathcal{G} est le critère des moindres carrés ordinaires et si le bruit est blanc, de variance σ^2 . Mais un tel choix ne conduit pas toujours à une régularisation acceptable. Une explication en est que, dans beaucoup de problèmes, le graphe de la fonction $\mathcal{G}(\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}) = \mathcal{G}(\alpha)$ est presque horizontal sur une large plage de valeurs de α : toute erreur dans l'estimation de σ^2 entraîne donc de grandes variations dans la valeur de α qui satisfait la contrainte (29).

■ **Méthode de la « courbe en L »**

On peut aussi utiliser une méthode alternative qui a fait ses preuves dans des problèmes inverses linéaires de la forme (25) et dans

Parution : novembre 2001 - Dernière validation : juin 2019 - Ce document a été délivré pour le compte de 7200106152 - éditions t / 2.59.188.28

le cas où la fonctionnelle de régularisation $\mathcal{F}(\mathbf{x})$ est quadratique, la méthode de la « courbe en L » [21]. Cette dernière consiste à tracer, en échelle log-log, la fonctionnelle de régularisation $\mathcal{F}(\hat{\mathbf{x}}(\alpha))$ en fonction du critère des moindres carrés $\|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}(\alpha)\|_2^2$ en faisant varier le coefficient de régularisation α . Cette courbe a en général l'allure caractéristique d'un L (d'où son nom) et la valeur de α correspondant à l'angle de ce L fournit un bon compromis entre les exigences contradictoires de fidélité aux données et de fidélité à l'a priori.

Pour comprendre pourquoi il en va ainsi, on sait que si \mathbf{x}_0 désigne la solution exacte, alors l'erreur $\hat{\mathbf{x}}(\alpha) - \mathbf{x}_0$ peut être décomposée en deux : une erreur de *perturbation* due à la présence de l'erreur de mesure \mathbf{b} et une erreur de *régularisation* due à l'emploi d'un opérateur régularisant au lieu d'un opérateur inverse (voir (22)). La partie verticale de la courbe en L, décrite pour de faibles valeurs de α , correspond à des solutions pour lesquelles $\mathcal{F}(\hat{\mathbf{x}}(\alpha))$ est très sensible aux variations de α car l'erreur de mesure \mathbf{b} domine $\hat{\mathbf{x}}(\alpha)$ et ne satisfait pas la condition de Picard discrète [21]. La partie horizontale de la courbe, décrite pour des valeurs élevées de α , correspond à des solutions pour lesquelles c'est la norme au carré des résidus $\|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}(\alpha)\|_2^2$ qui est la plus sensible aux variations de α car $\hat{\mathbf{x}}(\alpha)$ est dominée par l'erreur de régularisation, pour autant que $\mathbf{y} - \mathbf{b}$ satisfasse la condition de Picard discrète.

■ Validation croisée

On veut évidemment trouver une valeur du coefficient de régularisation α telle que la solution régularisée $\hat{\mathbf{x}}(\alpha, \mathbf{y})$ soit aussi proche que possible du véritable objet \mathbf{x} . Avec le choix de distances quadratiques pour \mathcal{F} et \mathcal{G} , il est naturel de choisir aussi une distance quadratique Δ_x pour mesurer l'écart entre $\hat{\mathbf{x}}(\alpha, \mathbf{y})$ et \mathbf{x} . Une méthode raisonnable de choix de α serait donc de choisir la valeur qui minimise ce risque en moyenne, c'est-à-dire l'erreur quadratique moyenne d'estimation :

$$EQM(\alpha, \mathbf{x}) = \int \Delta_x(\alpha, \mathbf{x}, \mathbf{y}) p(\mathbf{y}|\mathbf{x}) d\mathbf{y} \tag{30}$$

Malheureusement, la solution de ce problème dépend de l'objet réel qui est évidemment inconnu. Comme la solution régularisée $\hat{\mathbf{x}}(\alpha, \mathbf{y})$ peut être vue aussi comme un prédicateur des observations au travers de $\hat{\mathbf{y}}(\alpha, \mathbf{y}) = \mathbf{A}\hat{\mathbf{x}}(\alpha, \mathbf{y})$, on pourrait aussi mesurer l'écart entre les observations réelles et prédites avec une fonction de perte quadratique $\Delta_y(\alpha, \mathbf{x}, \mathbf{y})$ et rechercher la valeur de α qui minimise le risque moyen correspondant, qui est dans ce cas l'erreur quadratique moyenne de prédiction :

$$EQMP(\alpha, \mathbf{x}) = \int \Delta_y(\alpha, \mathbf{x}, \mathbf{y}) p(\mathbf{y}|\mathbf{x}) d\mathbf{y} \tag{31}$$

mais, là encore, la solution dépend de l'objet réel. La difficulté peut être cependant surmontée car le critère EQMP(α, \mathbf{x}) peut être estimé par **validation croisée généralisée** (VCG). Le principe de base en est le suivant [22]. Soit $\hat{\mathbf{x}}(\alpha, \mathbf{y}^{[-k]})$ le minimiseur du critère :

$$\mathcal{J}^{[-k]}(\mathbf{x}) = \sum_{n \neq k, n=1}^N [y_n - (\mathbf{A}\mathbf{x})_n]^2 + \alpha \|\mathbf{x}\|_2^2 \tag{32}$$

c'est-à-dire l'objet restauré en utilisant toutes les données *sauf* l'échantillon y_k . Il est possible de l'utiliser ensuite pour prédire la donnée manquante :

$$\hat{\mathbf{y}}_k^{[-k]}(\alpha) = \left[\mathbf{A}\hat{\mathbf{x}}(\alpha, \mathbf{y}^{[-k]}) \right]_k \tag{33}$$

La méthode consiste à rechercher la valeur de α qui minimise une énergie pondérée de l'erreur de prédiction :

$\alpha_{VCG} = \arg \min_{\alpha} V(\alpha)$, avec :

$$V(\alpha) = \frac{1}{N} \sum_{k=1}^N w_k^2(\alpha) \left[y_k - \hat{y}_k^{[-k]}(\alpha) \right]^2 \tag{34}$$

où les coefficients $w_k^2(\alpha)$ sont introduits pour éviter que le critère (34) n'ait des propriétés indésirables, telles que le manque d'invariance dans des rotations arbitraires de l'espace des observations ou l'absence de minimum. Ils sont donnés par :

$$w_k(\alpha) = \frac{1 - [\mathbf{B}(\alpha)]_{k,k}}{1 - \text{trace}[\mathbf{B}(\alpha)]/M}$$

où $[\mathbf{B}(\alpha)]_{k,k}$ est le k^e élément diagonal de la matrice :

$$\mathbf{B}(\alpha) = \mathbf{A}(\mathbf{A}\mathbf{A}^t + \alpha\mathbf{Q})^{-1}\mathbf{A}^t$$

Le calcul du minimum s'appuie sur le caractère linéaire-quadratique du problème qui permet d'établir la relation plus simple :

$$V(\alpha) = \frac{N \|\mathbf{I} - \mathbf{B}(\alpha)\|_F^2}{[\text{trace}\{\mathbf{I} - \mathbf{B}(\alpha)\}]^2} \tag{35}$$

qui montre clairement que la fonction de validation croisée généralisée $V(\alpha)$ est en fait la somme des carrés des erreurs résiduelles pondérée par un coefficient dépendant de α . De nombreux résultats pratiques expliquent que cette méthode ait été souvent utilisée dans des problèmes 1-D. Son emploi en traitement d'image est plus récent [23] [24].

Conclusion

Ces méthodes de choix du coefficient de régularisation n'ont de justification claire que dans le cas des critères régularisés quadratiques. Par ailleurs, la question du choix de la fonctionnelle régularisante \mathcal{F} n'a pas reçu, dans ce cadre déterministe, de réponse autre que des justifications empiriques. Le but du paragraphe suivant est de montrer qu'il est possible d'en trouver une dans le cadre de la théorie de l'information.

4. Approche bayésienne de l'inversion

Il existe au moins deux raisons qui poussent à inscrire la résolution des problèmes inverses dans un cadre bayésien. Tout d'abord, c'est dans ce cadre qu'ont été introduites les fonctions d'énergie locale et les modélisations markoviennes qui ont marqué durablement le traitement d'image à bas niveau. Mais c'est aussi lui qui offre les réponses les plus cohérentes et les plus complètes à des problèmes laissés ouverts dans les autres approches, comme le choix des hyperparamètres ou l'optimisation d'un critère multimodal.

4.1 Inversion et inférence statistique

Pour expliciter le lien qui unit inversion et inférence statistique, il est utile, à ce stade, de résumer l'analyse faite au paragraphe 2. Après discrétisation, le problème direct prend la forme générale $\mathbf{A}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$, où \mathbf{A} est un opérateur reliant l'objet inconnu \mathbf{x} aux données expérimentales \mathbf{y} . Il prend même souvent la forme explicite $\mathbf{y} = \mathbf{A}(\mathbf{x})$, voire linéaire $\mathbf{y} = \mathbf{A}\mathbf{x}$. L'inversion, c'est-à-dire le calcul de \mathbf{x} connaissant \mathbf{A} et \mathbf{y} , est très souvent un problème mal-posé, en deux sens.

Parution : novembre 2001 - Dernière validation : juin 2019 - Ce document a été délivré pour le compte de 7200106152 - éditions ti // celine BLONBOU // 2.59.188.28

■ Premièrement, l'opérateur \mathbf{A} est souvent singulier dans le sens où il existe une classe \mathcal{X} d'objets $\mathbf{x} \in \mathcal{X}$ invisibles, c'est-à-dire tels que $\mathbf{Ax} = \mathbf{0}$ (le noyau $\text{Ker } \mathbf{A} = \mathcal{X}$ n'est donc pas vide). Tout élément de \mathcal{X} peut être ajouté à toute solution pour donner une autre solution aussi compatible avec les données. Nous ne pouvons donc inverser la relation directe pour déterminer uniquement \mathbf{x} à partir de \mathbf{y} . Ce manque d'unicité fait que le problème inverse discret est mal-posé au sens d'Hadamard. Cette situation survient à chaque fois que la réponse instrumentale détruit une partie de l'information au sujet de l'objet.

■ Deuxièmement, aucun dispositif expérimental n'est complètement affranchi d'une incertitude, dont l'origine la plus simple est la précision finie des mesures. Il est donc plus réaliste de considérer que l'objet recherché et les mesures sont reliés par une équation de la forme $\mathbf{y} = \mathbf{A}(\mathbf{x}) \diamond \mathbf{b}$ dans laquelle \mathbf{A} est un opérateur décrivant la partie essentielle de l'expérience, et $\diamond \mathbf{b}$ prend en compte la dégradation de cette représentation idéale par différentes sources d'erreur (de discrétisation, de mesure) regroupées sous le terme de *bruit*. Quand le mécanisme d'observation peut être approché par une distorsion linéaire et l'ajout d'un bruit, alors cette équation se réduit à $\mathbf{y} = \mathbf{Ax} + \mathbf{b}$ (3). La présence de ce bruit a pour effet d'« élargir » l'ensemble \mathcal{X} puisque tout élément x tel que $\mathbf{Ax} = \boldsymbol{\epsilon}$, où $\boldsymbol{\epsilon}$ est « petit » par rapport au niveau supposé de bruit, peut être ajouté à toute solution possible pour obtenir une autre solution acceptable. En pratique, cela se traduit par l'instabilité des tentatives d'inversion directe de l'équation (3), des petits changements dans les données entraînant de grandes variations de la solution calculée. Ce n'est plus, cette fois, nécessairement un problème mal-posé au sens d'Hadamard puisque la solution peut très bien – lorsque le noyau \mathcal{X} est vide – être unique et dépendre continûment des données. L'instabilité provient du *mauvais conditionnement* de \mathbf{A} [25].

On voit donc que, dans ces problèmes mal-posés, l'obtention d'une solution n'est pas du tout un problème de déduction mathématique, c'est un problème d'inférence, c'est-à-dire de traitement de l'information : comment tirer les meilleures conclusions possibles de l'information incomplète qui est disponible.

Toute méthode d'inférence scientifique devrait, pour être acceptable, 1° prendre en compte toute l'information pertinente disponible et, 2° éviter soigneusement de supposer une information disponible alors qu'elle ne l'est pas. La modélisation probabiliste est un moyen commode et cohérent de décrire une situation d'information incomplète. Nous allons voir comment elle conduit à une approche statistique bayésienne.

4.2 Distribution directe

Il faut préciser d'emblée que tout problème auquel on s'attaque par une approche bayésienne doit être *bien-posé* dans le sens où une information suffisante doit être apportée pour permettre d'attribuer sans ambiguïté les distributions de probabilité nécessaires au calcul. Cela signifie, au minimum, qu'un ensemble exhaustif de possibilités doit être spécifié au début de chaque problème. Nous l'appellerons *espace des données* (ou des *épreuves*) s'il s'agit des résultats possibles de l'expérience, ou *espace des hypothèses* s'il spécifie les hypothèses que nous voulons vérifier. Il est utile aussi de distinguer deux classes de problèmes, appelées **estimation** et **choix de modèle**. L'estimation étudie les conséquences du choix d'un modèle particulier, supposé « vrai », alors que le choix de modèle a pour but d'en retenir un par comparaison avec un ou plusieurs autres.

Dans un problème d'estimation, on suppose que le modèle est vrai pour une valeur (inconnue) de ses paramètres, et on explore les contraintes imposées aux paramètres par les données. L'espace des hypothèses est alors l'ensemble de toutes les valeurs possibles des paramètres, $\mathcal{X} = \{x_i\}$. Les données consistent en un ou

plusieurs échantillons ; pour rendre le problème bien-posé, l'espace de tous les échantillons possibles, $\mathcal{S} = \{y_j\}$, doit être aussi précisé. Les espaces \mathcal{X} et \mathcal{S} peuvent être tous deux discrets ou continus.

Désignant par x_0 la vraie valeur du paramètre, nous pouvons nous attaquer au problème d'estimation en calculant la probabilité que chacune des valeurs possibles du paramètre soit la vraie valeur. Désignons par D la proposition affirmant les valeurs des données expérimentales réellement observées, par H la proposition $\mathbf{x}_0 = \mathbf{x}$ affirmant que l'une des valeurs possibles du paramètre, \mathbf{x} , est la vraie valeur \mathbf{x}_0 , et par I l'environnement logique du problème. C'est cet environnement logique qui définit notre problème en spécifiant l'espace des hypothèses, l'espace des données, comment les hypothèses (valeurs des paramètres) et les données sont reliées, et toute information supplémentaire que nous pourrions avoir sur les hypothèses ou les données. Nous pourrions définir I comme la proposition affirmant :

- 1° que la vraie valeur du paramètre est dans \mathcal{X} ;
- 2° que les données observées consistent en N échantillons de l'espace \mathcal{S}^N ;
- 3° la manière dont les paramètres sont reliés aux données (le modèle direct \mathbf{A}) ;
- 4° toute information supplémentaire. Bien entendu, la nature physique des paramètres et des données est spécifiée implicitement dans \mathcal{X} , \mathcal{S} et \mathbf{A} .

La première étape dans toute méthode d'inférence statistique destinée à résoudre un problème tel que (3) consiste à choisir une distribution de probabilité décrivant notre information – ou notre incertitude – sur les erreurs $\mathbf{b} : q(\mathbf{b} | I)$. C'est une étape essentielle puisqu'elle permet d'en déduire la distribution *directe*, ou d'*échantillonnage* :

$$p(\mathbf{y} | \mathbf{x}, I) = q(\mathbf{y} - \mathbf{Ax} | I) \tag{36}$$

Dans la très grande majorité des cas, on choisit une distribution gaussienne centrée pour les erreurs, indépendante de x , ce qui donne :

$$p(\mathbf{y} | \mathbf{x}, I) = [(2\pi)^N |\mathbf{R}|]^{-1/2} \exp\left\{-\frac{1}{2} \|\mathbf{y} - \mathbf{Ax}\|_{\mathbf{R}^{-1}}^2\right\}$$

où \mathbf{R} désigne la matrice de covariances de la distribution $q(\mathbf{b} | I)$. Elle est souvent diagonale, voire même proportionnelle à l'identité. Une question se pose immédiatement : quel sens doit-on donner à un tel choix et dans quelles situations un tel modèle est-il approprié ?

Avec une **interprétation fréquentiste** d'une probabilité, la distribution pour le bruit devrait être la distribution de fréquence de ses valeurs dans un très grand nombre de répétitions des mesures. Elle est alors justifiée par référence au théorème central limite qui dit, sous des conditions assez larges, que si le bruit dans un échantillon des données résulte d'un grand nombre d'effets élémentaires cumulés, « aléatoires » et indépendants, la distribution gaussienne est une bonne approximation de sa véritable distribution de fréquence. Mais, excepté les fluctuations d'origine électronique dans une chaîne de mesure, le bruit n'est pas en général le résultat d'un grand nombre d'effets indépendants (songeons par exemple aux erreurs de discrétisation qui dépendent de la solution x_0). De plus, pour pouvoir effectuer une inférence avec cette interprétation, il serait nécessaire que nous disposions de nombreux résultats d'autres mesures pour pouvoir déterminer ces fréquences, ce qui est une situation expérimentale extrêmement rare.

Cette « hypothèse » gaussienne n'est donc pas une hypothèse sur le caractère « aléatoire » du bruit. Nous ne prétendons nullement que le phénomène donnant naissance au bruit soit véritablement aléatoire et suive une distribution gaussienne. Ce n'est même pas une hypothèse à proprement parler ; c'est plutôt le choix le moins compromettant – ou le plus conservateur – que nous puissions faire pour la distribution du bruit dans une situation d'incertitude. Nous supposons ici deux choses, 1° que le bruit peut prendre toute valeur réelle mais qu'il est de moyenne nulle, c'est-à-dire qu'il n'y a pas

d'erreur de mesure systématique (ou que s'il y en a, nous avons été capables de la détecter et de la corriger), et 2° que nous nous attendons à ce qu'il y ait une « échelle typique » du bruit, c'est-à-dire que les grandes contributions du bruit ne sont pas aussi probables que les petites.

Autrement dit, nous pensons que la distribution pour le bruit doit avoir une moyenne nulle et un écart-type fini, même si nous n'avons pas d'idée précise de la valeur de ce dernier. Par contre, nous n'avons aucune idée de l'existence ou non de cumulants d'ordre supérieur à deux. Dans ces conditions, le choix le moins compromettant vis-à-vis des caractéristiques que nous ignorons – que l'on peut justifier formellement par des principes informationnels [26] – est celui d'une distribution gaussienne. Et si l'on suspecte que les composantes du bruit affectant les N échantillons ont des échelles différentes ou sont corrélées, on peut prendre en compte ces informations par l'intermédiaire de la matrice de covariance de la distribution. Il n'est pas nécessaire d'en spécifier la valeur, mais si elle est inconnue, ses éléments, regroupés dans un vecteur d'*hyperparamètres* θ , viendront en général compliquer le problème. On les appelle **paramètres de nuisance** pour cette raison.

Ce choix pour la distribution du bruit est approprié à chaque fois que cette information est tout ce que nous savons du bruit. Comme c'est une situation fréquente, il est souvent fait. Si nous avons des informations supplémentaires sur le bruit, qui nous conduiraient à choisir une distribution non gaussienne, nous pouvons les inclure de la même manière, mais cela ne conduira à des résultats nettement améliorés que si la distribution diffère nettement d'une gaussienne. Il existe ainsi des situations – comme l'imagerie à faible taux de comptage de particules – où les données sont entières et de faible valeur. Le choix d'une distribution poissonnienne peut alors améliorer les résultats.

4.3 Maximisation de vraisemblance

Munis de cette seule distribution directe $p(\mathbf{y} | \mathbf{x}, \theta, I)$, nous pourrions définir la solution du problème inverse comme étant celle du **maximum de vraisemblance**, la vraisemblance étant la distribution directe dans laquelle la variable \mathbf{y} prend sa valeur observée et le paramètre \mathbf{x} devient la variable :

$$\hat{\mathbf{x}}_{MV} = \arg \max_{\mathbf{x} \in \mathcal{X}} p(\mathbf{y} | \mathbf{x}, \theta, I)$$

La **solution des moindres carrés** est un cas particulier de celle du maximum de vraisemblance, lorsque la distribution directe est gaussienne :

$$\hat{\mathbf{x}}_{MC} = \arg \min_{\mathbf{x} \in \mathcal{X}} [\mathbf{y} - \mathbf{Ax}]^t \mathbf{R}^{-1} [\mathbf{y} - \mathbf{Ax}]$$

Introduite de cette manière, il s'agit toujours d'une méthode des moindres carrés *pondérés* (par la matrice \mathbf{R}^{-1}) qui possède la propriété indispensable d'invariance par changement d'unités dans \mathcal{X} et \mathcal{Y} . Dans beaucoup de situations simples, cette méthode d'inférence fournit toute l'information recherchée, mais dans les problèmes inverses dont la paramétrisation n'est pas parcimonieuse, la distribution directe ne contient pas toute l'information nécessaire pour rendre le problème bien-posé et elle ne fournit pas tout l'appareil technique nécessaire au calcul.

■ Dans le cas particulier d'un problème indéterminé $\mathbf{y} = \mathbf{Ax}$ où \mathbf{A} est singulier (problème dit d'*inversion généralisée*), il n'y a pas de « bruit », et donc pas de distribution directe, sauf dans le sens rudimentaire où $p(\mathbf{y} | \mathbf{x}, I)$ est constant si \mathbf{x} est dans la classe \mathcal{C} des antécédents possibles de \mathbf{y} , et nul sinon. La vraisemblance étant constante dans la classe \mathcal{C} , sa maximisation ne nous est d'aucune aide pour choisir dans cette classe. La difficulté du problème ne réside pas dans la présence d'un bruit « aléatoire » per-

turbant les données, mais plutôt dans le fait que notre information soit incomplète, bien qu'essentiellement non bruitée.

■ Dans le cas linéaire (3), la matrice \mathbf{A} du problème direct est souvent *mal-conditionnée*. L'opérateur de résolution $\mathbf{A}^+ = (\mathbf{A}^t \mathbf{R}^{-1} \mathbf{A}) \mathbf{R}^{-1} \mathbf{A}^t$ est instable et la solution $\hat{\mathbf{x}}_{MV} = \mathbf{A}^+ \mathbf{y}$, est inacceptable : l'amplification du bruit est excessive.

■ Le problème peut avoir des paramètres de nuisance sans intérêt pour nous, et en grand nombre. Ainsi, lorsque la matrice \mathbf{R} est pleine, il vient s'ajouter $N(N-1)/2$ hyperparamètres qui, lorsqu'ils sont inconnus, doivent en général être estimés par maximum de vraisemblance comme les paramètres d'intérêt \mathbf{x} , et le maximum global peut alors n'être plus un point mais toute une région.

■ Nous pouvons avoir une information hautement pertinente sur la solution recherchée. Par exemple, nous pouvons savoir qu'elle doit être positive, ou qu'elle doit satisfaire certaines contraintes (comme en imagerie astronomique où l'intégrale de l'objet peut être connue par ailleurs), ou qu'elle est constituée de régions homogènes séparées par des contours nets. De telles informations ne sont pas contenues dans la distribution directe et il serait dommage de ne pas en tenir compte.

■ Dans beaucoup de problèmes, il est nécessaire de disposer non seulement d'une solution, mais aussi d'une mesure de la confiance à lui accorder. Munis de la seule distribution directe (36), les *intervalles de confiance* fournis par l'approche fréquentiste ne nous renseignent que sur le comportement à *long terme* de la solution, c'est-à-dire sur son comportement moyen dans un très grand nombre de répétitions de l'expérience. Or nous ne disposons que des résultats d'une seule expérience, d'ailleurs très souvent non reproductible.

■ Enfin, l'estimation des paramètres d'un modèle supposé valable n'est souvent qu'une étape, et l'on peut avoir besoin de juger les mérites relatifs de différents modèles.

Il est donc nécessaire d'aller au-delà de l'inférence par maximum de vraisemblance. Toutes les extensions évoquées précédemment sont « automatiquement » fournies par l'approche bayésienne.

4.4 Inférence bayésienne

L'inférence bayésienne est ainsi nommée parce qu'elle fait un large usage de la règle de Bayes, conséquence elle-même d'une règle fondamentale du calcul des probabilités, la **règle du produit** [27]. Soit H une hypothèse dont nous voulons évaluer la véracité, D un ensemble de données en rapport avec cette hypothèse, et I une proposition définissant l'environnement logique du problème. La règle du produit stipule que :

$$P(HD | I) = P(H | D, I) P(D | I) = P(D | H, I) P(H | I)$$

où, par exemple, $P(H | D, I)$ désigne de manière habituelle la probabilité que H soit vraie sachant D et I . On en tire la règle de Bayes :

$$P(H | D, I) = \frac{P(H | I) P(D | H, I)}{P(D | I)}$$

qui n'est rien d'autre qu'une règle d'*apprentissage* : elle nous dit comment nous devons ajuster la probabilité attribuée à la véracité d'une hypothèse lorsque notre état de connaissance change avec l'acquisition de données. La probabilité *a posteriori* pour H , $P(H | D, I)$, est obtenue en multipliant sa probabilité *a priori*, $P(H | I)$, par la probabilité d'avoir observé les données D en supposant l'hypothèse vraie, $P(D | H, I)$, et en divisant le tout par la probabilité d'avoir observé les données indépendamment du fait que l'hypothèse soit

vraie ou non, $P(D|I)$. Ce dernier terme, parfois appelé *vraisemblance globale*, joue le rôle d'une constante de normalisation.

Une part importante de l'inférence statistique est fondée sur l'emploi d'une information *a priori* sur les grandeurs à estimer, qui vient s'ajouter à l'information apportée par les données. Il n'est donc pas surprenant, si l'on se réfère à la nature profonde du principe de la régularisation exposé au paragraphe 3, qu'elle présente un lien étroit avec l'inférence bayésienne.

Dans le cas d'un problème inverse tel que (3), et en supposant que les distributions de probabilité concernées sont à densité, l'information *a priori* sur l'objet x s'exprime, dans un contexte bayésien, sous la forme d'une densité de probabilité *a priori* $p(\mathbf{x} | \theta)$. La règle de Bayes nous permet de la combiner avec l'information contenue dans les données pour obtenir la loi *a posteriori* :

$$p(\mathbf{x} | \mathbf{y}, \mathbf{A}, \theta) = \frac{p(\mathbf{x} | \theta) p(\mathbf{y} | \mathbf{x}, \mathbf{A}, \theta)}{p(\mathbf{y} | \mathbf{A}, \theta)} \quad (37)$$

Dans cette équation, θ est un vecteur d'*hyperparamètres* constitué des paramètres des distributions *a priori* des erreurs et de l'objet, et $p(\mathbf{y} | \mathbf{x}, \mathbf{A}, \theta)$ désigne la loi des données conditionnée par la vraie solution \mathbf{x} . Elle est complètement déterminée par la connaissance du modèle direct (3) et de la loi de probabilité du bruit. Le dernier terme assure la normalisation de la loi *a posteriori* :

$$p(\mathbf{y} | \mathbf{A}, \theta) = \int p(\mathbf{y} | \mathbf{x}, \mathbf{A}, \theta) p(\mathbf{x} | \theta) d\mathbf{x} \quad (38)$$

L'approche bayésienne fournit ainsi une réponse claire et sans ambiguïté au problème de l'inversion des données expérimentales : après avoir observé *les seules données* \mathbf{y} , l'incertitude sur l'objet \mathbf{x} est entièrement décrite par la distribution de probabilité (37). Mais étant donné une distribution de probabilité pour un paramètre \mathbf{x} , continu ou discret, quelle « meilleure » estimation peut-on en faire, et avec quelle précision ? Il n'y a pas de réponse unique à cette question ; le problème relève de la théorie de la décision qui répond à la question « *Que devons-nous faire ?* » Cela implique des jugements de valeur et va par conséquent au-delà des principes de l'inférence qui répond seulement à la question « *Que savons-nous ?* » On peut ainsi déduire de (37) aussi bien un estimateur ponctuel qu'une région d'incertitude [4] [19]. Un choix d'estimateur très fréquent consiste à attribuer à \mathbf{x} la valeur qui maximise la distribution *a posteriori* :

$$\hat{\mathbf{x}}_{\text{MAP}} = \arg \max_{\mathbf{x}} p(\mathbf{x} | \mathbf{y}, \mathbf{A}, \theta) \quad (39)$$

Mais ce n'est qu'une des solutions possibles. Cette estimation MAP correspond à la minimisation d'un coût de décision moyen avec une fonction de coût tout-ou-rien, limite (lorsque $\varepsilon \rightarrow 0$) du coût moyen $P(\|\hat{\mathbf{x}} - \mathbf{x}_0\| > \varepsilon)$. D'autres fonctions de coût ont soulevé l'intérêt dans le cadre de la modélisation des images par champs markoviens. Elles conduisent à maximiser des probabilités marginales [4] [28].

4.5 Liens avec les méthodes déterministes

Dans le cas qui nous intéresse ici, c'est-à-dire un problème inverse en dimension finie, il est clair que régulariser selon le principe général indiqué au paragraphe 3.2, et donc minimiser un critère tel que (25) est équivalent à choisir l'image qui maximise la loi *a posteriori* suivante :

$$p(\mathbf{x} | \mathbf{y}, \mathbf{A}, \theta) \propto \exp \left\{ - \frac{\mathcal{G}(\mathbf{y} - \mathbf{A}\mathbf{x}) + \alpha \mathcal{F}(\mathbf{x})}{2\sigma^2} \right\} \quad (40)$$

où σ^2 désigne la variance du bruit. Cette loi de probabilité n'est qu'un des choix possibles puisque toute fonction strictement monotone autre que l'exponentielle aurait convenu. Mais ce choix convient

particulièrement bien ici puisque avec le modèle linéaire (3), sous les hypothèses usuelles de normalité et d'indépendance du bruit, dès lors que \mathcal{G} est une norme euclidienne, la loi conditionnelle $p(\mathbf{y} | \mathbf{x}, \mathbf{A}, \theta)$ est réellement :

$$p(\mathbf{y} | \mathbf{x}, \mathbf{A}, \theta) \propto \exp \left\{ - (1/2 \sigma^2) \mathcal{G}(\mathbf{y} - \mathbf{A}\mathbf{x}) \right\} \quad (41)$$

Pour que l'analogie soit complète, la loi *a priori* doit aussi prendre la forme suivante :

$$p(\mathbf{x} | \theta) \propto \exp \left\{ - (\alpha/2 \sigma^2) \mathcal{F}(\mathbf{x}) \right\} \quad (42)$$

et il faut, pour qu'elle soit rigoureuse, que la loi *a posteriori* (40) soit propre, une condition suffisante étant que (41) et (42) le soient elles aussi.

Beaucoup de fonctions d'énergie locale utilisées en traitement d'image ont été introduites dans un cadre bayésien : elles définissent \mathbf{x} comme un champ de Markov [28] [29] [30]. Mais bien que le point de vue *énergétique* soit aussi représenté dans la communauté des traiteurs d'image, tous les critères de la forme (25) sont formellement réinterprétables dans un cadre bayésien. La question n'est pas de savoir si l'interprétation bayésienne représente une justification des autres approches, mais plutôt de voir ce qu'elle apporte en réponse aux problèmes évoqués précédemment. En plus de sa grande cohérence, elle fournit des outils originaux pour le choix des hyperparamètres (*marginalisation*) et pour la minimisation du critère composé (*algorithmes pseudo-aléatoires*), et la variété des coûts qu'elle permet d'utiliser conduit à des solutions qui n'ont pas d'équivalent dans le cadre énergétique (*régression*).

4.6 Cas linéaire et gaussien

Les lois gaussiennes, associées à des modèles directs linéaires, fournissent une structure linéaire d'estimation, donc un cadre algorithmique extrêmement commode. En revanche, elles ne permettent d'incorporer que des informations frustes, fondamentalement limitées à des caractéristiques d'ordre 2. Ainsi, dans la théorie standard de la régularisation [17], le choix d'un terme quadratique de fidélité aux données : $\mathcal{G}(\mathbf{y} - \mathbf{A}\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\mathbf{P}}^2$ équivaut à celui d'une distribution normale pour le bruit : $q(\mathbf{b} | \mathbf{R}_b, I) \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_b)$, avec $\mathbf{R}_b \propto \mathbf{P}^{-1}$. De même, le choix d'une pénalisation quadratique : $\mathcal{F}(\mathbf{x}) = \|\mathbf{D}_k \mathbf{x}\|^2$ équivaut lui aussi à celui d'une distribution *a priori* normale pour l'objet : $p(\mathbf{x} | \mathbf{R}_x, I) \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_x)$, avec $\mathbf{R}_x \propto (\mathbf{D}_k^t \mathbf{D}_k)^{-1}$, en supposant bien sûr que la matrice $\mathbf{D}_k^t \mathbf{D}_k$ soit définie positive. La régularisation déterministe linéaire quadratique est donc rigoureusement équivalente à l'estimation linéaire gaussienne. La solution est explicite :

$$\hat{\mathbf{x}} = (\mathbf{A}^t \mathbf{R}_b^{-1} \mathbf{A} + \mathbf{R}_x^{-1})^{-1} \mathbf{R}_b^{-1} \mathbf{A}^t \mathbf{y} \quad (43)$$

et présente la caractéristique remarquable d'être une fonction linéaire des données \mathbf{y} . Il est donc intéressant de l'étudier dans le domaine de Fourier dans le cas particulier où \mathbf{A} désigne une convolution discrète (circulante) et où les signaux sont supposés stationnaires (à corrélations circulantes) [1]. Si l'on désigne respectivement par $Y(v)$, $H(v)$, $\hat{X}(v)$, $g_x(v)$ et $g_b(v)$ les transformées de Fourier discrètes des données, de la réponse impulsionnelle du système d'observation, de la solution régularisée, de sa fonction de corrélation et de celle du bruit, on voit que :

$$\hat{X}(v) = W(v) Y(v)$$

où :

$$W(v) = \frac{|H(v)|^2 g_x(v)}{|H(v)|^2 g_x(v) + g_b(v)} \frac{1}{H(v)} \quad (44)$$

est analogue à la fonction de transfert en fréquence d'un filtre de Wiener. Le filtre de restauration est donc constitué de la mise en cascade d'un *filtre inverse* (fournissant une solution naïve et inacceptable) et d'un *filtre de lissage* (empêchant l'amplification incontrôlée du bruit dans la bande de fréquence où $|H(v)^2 g_x(v) \ll g_b(v)$). Cette inversion linéaire quadratique ou linéaire gaussienne occupant une place prépondérante dans les problèmes d'inversion, une réaction courante est de dire : « *Les problèmes inverses, ce n'est pas compliqué, il suffit de lisser les données avant d'inverser.* » Cette vision n'est pas fautive et suffit pour beaucoup de problèmes, mais elle est réductrice, ne permet pas d'aller plus loin, et induit un schéma cascadié qui n'est justifié que dans le cadre linéaire quadratique. Ainsi, dans notre exemple de déconvolution d'image, la fonction de transfert $H(v)$ est la réponse fréquentielle du système d'observation présentée sur la figure 3. Cette réponse étant du type passe-bas, le filtre de Wiener qui fournit la solution régularisée n'effectue qu'une *égalisation spectrale* dans une étroite bande de fréquence, et ne peut donc restaurer le contenu spectral de l'objet pour les fréquences élevées, ce qui est pourtant indispensable pour restaurer les contours francs présents dans l'image originale. Nous verrons plus bas comment on peut y parvenir, mais il faudra sortir du cadre gaussien et la solution ne sera plus un filtrage linéaire des données.

4.7 Choix des hyperparamètres

L'interprétation bayésienne permet d'étendre de manière appréciable la gamme des méthodes de détermination des hyperparamètres. Toutes les méthodes décrites dans les paragraphes précédents nécessitent en effet, pour leur mise en œuvre effective, de choisir la valeur du coefficient de régularisation α , et plus généralement, de tous les hyperparamètres θ définissant les mesures de distance \mathcal{F} et \mathcal{G} : la variance du bruit, les paramètres de corrélation de l'objet, et les paramètres des fonctions d'énergie locale. La détermination de θ est l'étape la plus délicate des méthodes de restauration et de reconstruction d'image. Bien que ce problème soit encore ouvert, l'approche bayésienne fournit des outils cohérents pour l'aborder.

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 x . 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 x , pour une valeur donnée de θ , au travers de la distribution *a posteriori* de l'équation (37). Le second infère sur θ grâce à une relation analogue :

$$p(\theta | \mathbf{y}, \mathbf{A}) = \frac{p(\theta | \mathbf{A}) p(\mathbf{y} | \theta, \mathbf{A})}{p(\mathbf{y} | \mathbf{A})} \tag{45}$$

On retrouve d'ailleurs là une caractéristique de l'utilisation de la règle de Bayes : la vraisemblance $p(\mathbf{y} | \theta, \mathbf{A})$ 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 », c'est-à-dire si les données y contiennent suffisamment d'information, l'influence de la distribution *a priori* $p(\theta | \mathbf{A})$ 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 :

$$p(\mathbf{y} | \theta, \mathbf{A}) = \int p(\mathbf{x}, \mathbf{y} | \theta, \mathbf{A}) d\mathbf{x} \tag{46}$$

$$= \int p(\mathbf{y} | \mathbf{x}, \theta, \mathbf{A}) p(\mathbf{x} | \theta) d\mathbf{x}$$

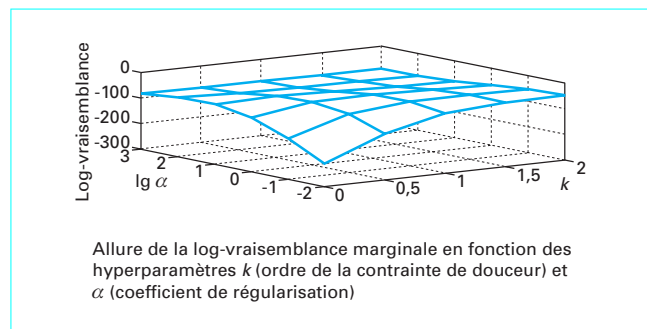


Figure 11 – Modèle AR-long

Une telle intégrale conduit très rarement à un résultat explicite. Une exception remarquable est fournie par une distribution conjointe $p(\mathbf{x}, \mathbf{y} | \theta, \mathbf{A})$ gaussienne, ce qui est justement le cas de l'analyse spectrale avec un **modèle autorégressif long** [31].

On constate en effet, avec les méthodes usuelles du paragraphe 3.1, que l'augmentation de l'ordre d du modèle AR permet de mieux faire ressortir les pics correspondant aux trois sinusoïdes effectivement présentes dans le signal de notre second exemple, mais aux prix de l'apparition de fluctuations importantes dans le spectre. Pour contourner le problème de la détermination de l'ordre, mal résolu avec les critères usuels, on retient un modèle d'ordre maximal $d = N - 1$ (modèle « AR-long »), et on contrôle les fluctuations erratiques du spectre en lui imposant une contrainte de *douceur* en minimisant la mesure de *rugosité* d'ordre k :

$$\mathcal{F}(a) = \mathcal{R}_k[B(v)] \stackrel{\text{def}}{=} \int_{-1/2}^{+1/2} \left| \frac{\partial^k B(v)}{\partial v^k} \right|^2 dv \tag{47}$$

où $B(v)$ est la réponse en fréquence du filtre blanchisseur du signal (23) :

$$B(v) \stackrel{\text{def}}{=} 1/H(v) = 1 - \sum_{m=1}^{N-1} a_m \exp(-2jmv) \tag{48}$$

On aboutit ainsi à un problème des moindres carrés sous contrainte, ou de minimisation d'un critère composé (25) qui possède une interprétation bayésienne simple (40). Les paramètres de réglage de la méthode ne sont plus l'ordre d du modèle, mais le degré k de douceur de la solution et le coefficient de régularisation α . La détermination de ces deux *hyperparamètres* par maximum de vraisemblance marginale permet d'obtenir une méthode *non supervisée*.

Les résultats obtenus sont présentés figure 11. On observe que la nappe de log-vraisemblance ne présente pas de maximum bien net, et qu'il existe plutôt une ligne de crête sur laquelle la vraisemblance varie peu. Autrement dit, sur cette ligne, une augmentation de l'ordre k de la contrainte de douceur, accompagnée d'une diminution conjointe du coefficient de régularisation α , ne change pas notablement la valeur de la vraisemblance.

On observe donc conséquemment qu'il est possible d'obtenir, pour des valeurs différentes de l'ordre k , des spectres voisins, tels que celui qui est présenté figure 12. Ce phénomène peut s'expliquer qualitativement par le fait que la méthode doit réaliser un compromis entre la douceur nécessaire pour bien rendre l'aspect régulier du spectre du bruit, et l'absence de douceur du spectre des trois composantes harmoniques du signal. Avec ce type de signal et ce mode de régularisation, l'amélioration attendue n'est donc pas vraiment convaincante, même si le terme de régularisation permet effectivement de mieux contrôler les fluctuations du spectre dues à l'augmentation de l'ordre du modèle.

Parution : novembre 2001 - Dernière validation : juin 2019 - Ce document a été délivré pour le compte de 7200106152 - éditions ti // celine BLONDOU // 2.59.188.28

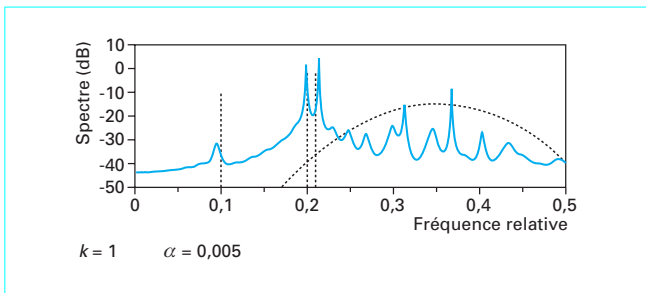


Figure 12 – Spectre obtenu avec un modèle AR-long

Pour contourner la difficulté posée par le calcul explicite d'une vraisemblance marginale, on peut introduire des « variables cachées » \mathbf{z} qui viennent compléter les observations \mathbf{y} de manière que la nouvelle vraisemblance $p(\mathbf{y}, \mathbf{z} | \theta, \mathbf{A})$ 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) [32]. La nécessité de telles approches stochastiques est apparue à la suite de l'impossibilité de mettre en œuvre des méthodes convergentes de maximisation de vraisemblance par les techniques d'optimisation classiques, la vraisemblance n'étant pas calculable.

On peut aussi remarquer que la distribution conjointe, ou **vraisemblance généralisée** :

$$\begin{aligned} p(\mathbf{y}, \mathbf{x} | \theta, \mathbf{A}) &= p(\mathbf{x} | \mathbf{y}, \theta, \mathbf{A}) p(\mathbf{y} | \theta, \mathbf{A}) \\ &= p(\mathbf{y} | \mathbf{x}, \theta, \mathbf{A}) p(\mathbf{x} | \theta) \end{aligned} \quad (49)$$

résume toute l'information propre au premier niveau d'inférence, et vouloir en faire la maximisation conjointe par rapport à \mathbf{x} et θ . Le problème d'intégration soulevé par (46) est évidemment évacué. Par contre, la vraisemblance généralisée n'a pas, en général, d'interprétation statistique claire, et il peut même arriver que l'estimateur correspondant ne soit pas défini, la vraisemblance pouvant ne comporter aucun maximum, même local [33].

4.8 Modèle *a priori*

Il est souvent reproché à l'estimation bayésienne de dépendre de la connaissance d'un hypothétique « vrai modèle » aléatoire ayant engendré l'objet à reconstruire. Implicitement, pour formuler ce reproche, il faut admettre que la réalité peut être « enfermée » dans un modèle mathématique. Vaste débat... Dans le cas de l'approche probabiliste de l'inversion telle que nous la concevons, l'interprétation fréquentiste des probabilités entretient une confusion fâcheuse : il nous semble important de préciser que nos hypothèses probabilistes ne sont pas des hypothèses sur le caractère « aléatoire » de l'objet, ou du bruit affectant les données, mais des choix d'un mode de représentation d'une information *a priori* incomplète – ou d'une connaissance incertaine – compatible avec l'outil d'inférence choisi. Cette situation n'a d'ailleurs rien de singulier puisqu'il est rare que dans un problème réel, l'information *a priori* disponible le soit sous une forme directement adaptée au cadre théorique choisi pour le traitement.

Les avantages de l'approche bayésienne ne résident pas dans l'information supplémentaire introduite par l'*a priori* : les interprétations énergétiques, déterministes, de la fonctionnelle de régularisation $\mathcal{F}(\mathbf{x})$ du paragraphe 4.5 montrent qu'elle ne lui est pas propre, et l'information introduite sur les paramètres de nuisance est *diffuse* la plupart du temps. En revanche, l'approche bayésienne donne accès à une couche d'outils qui n'existe pas dans les autres approches :

- la **marginalisation** (on intègre simplement ce qui ne nous intéresse pas *hors du problème*, au lieu d'être obligé d'imaginer une répétition virtuelle d'expériences « aléatoires ») ;
- la **régression** (l'espérance conditionnelle n'a pas d'équivalent dans les autres approches) ;
- l'**échantillonnage stochastique** (méthodes de Monte Carlo, algorithmes de recuit simulé, algorithmes génétiques) qui ne peut se concevoir sans cela.

Cela dit, la conversion de cette information *a priori* en une distribution de probabilité est un problème difficile, encore largement ouvert en statistique. Pour les paramètres de l'objet \mathbf{x} , l'*a priori* est souvent choisi *de manière pragmatique* (§ 4.8.1). Il existe cependant des *règles formelles* qui conduisent à des choix raisonnables [34] [35] [36], et sont surtout utilisées pour les paramètres de nuisance. Elles mènent souvent à des *a priori impropres*, qui ne posent pas de difficulté particulière quand ils sont correctement manipulés [37]. En voici quelques exemples.

■ Certaines méthodes reposent sur la **théorie des groupes de transformation** pour déterminer la mesure de référence « naturelle » pour le problème et satisfaire à certains principes d'invariance. Mais en pratique, cette approche n'a guère permis que 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* [37].

■ D'autres méthodes reposent sur des **principes informationnels**. Il s'agit principalement des méthodes dites à *maximum d'entropie* dans lesquelles on recherche une distribution qui soit la plus proche d'une distribution de référence (au sens d'une distance de Kullback) tout en vérifiant une information incomplète connue *a priori* [26] [38]. 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 (moments). On travaille alors dans la famille des *distributions exponentielles*.

■ Un autre principe formel consiste à utiliser des ***a priori* conjugués**, famille de lois telles que les distributions *a posteriori* correspondantes soient aussi dans la famille [34]. Cela ne présente d'intérêt que si ladite famille est aussi petite que possible, et paramétrée. Dans ce cas, le passage de l'*a priori* à l'*a posteriori*, par application de la règle de Bayes, se réduit à la mise à jour des paramètres. L'intérêt de cette méthode est essentiellement d'ordre technique puisque l'*a posteriori* est toujours calculable, au moins jusqu'à un certain point. On peut aussi lui trouver une justification partielle par un raisonnement d'invariance : si les données \mathbf{y} modifient $p(\mathbf{x} | I)$ en $p(\mathbf{x} | \mathbf{y}, I)$, l'information apportée par \mathbf{y} sur \mathbf{x} est évidemment limitée ; elle ne devrait pas conduire à une modification de toute la structure de $p(\mathbf{x} | I)$, mais seulement de ses paramètres. Mais il est clair que la principale motivation est sa commodité d'emploi. Les *a priori* conjugués sont habituellement associés à des distributions directes particulières qui permettent toujours de les établir, ce sont les *familles exponentielles* [39]. Le caractère « automatique » de cette méthode de choix est cependant trompeur, des *hyperparamètres* supplémentaires apparaissent inévitablement, qu'il faut aussi régler.

■ Nous allons plutôt nous intéresser ici à une dernière classe très importante, celle des **constructions faites « à la main »**, c'est-à-dire qui ne s'appuient pas sur des principes de portée générale comme les précédentes, mais introduisent de manière pragmatique des modèles probabilistes traduisant au mieux les propriétés attendues des solutions, au cas par cas. C'est dans cette catégorie qu'entrent les *modèles markoviens* qui ont connu un développement spectaculaire en imagerie depuis 1984 [29], 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. Aussi, avant d'illustrer ces derniers avec notre premier exemple (§ 4.8.2), nous commencerons par reprendre le

second, celui de l'analyse spectrale (§ 4.8.1), afin d'illustrer simplement quelle peut être l'influence d'un modèle *a priori* sur le résultat d'une inversion.

4.8.1 Analyse spectrale

Le principal obstacle à l'obtention d'une résolution suffisante avec les méthodes d'analyse classiques du type *périodogramme* étant le nombre limité d'échantillons disponibles, une approche très différente de l'inversion généralisée du paragraphe 2.2, ou de la modélisation autorégressive du paragraphe 3.1, consiste à aborder le problème comme celui d'une *extrapolation* de signal [40], c'est-à-dire à rechercher un vecteur à valeurs complexes $\mathbf{X} = [X_1, X_2, \dots, X_M]^T$, de dimension $M \gg N$, et tel que l'on ait :

$$\mathbf{y} = \mathbf{F}\mathbf{X} \tag{50}$$

où \mathbf{F} est une matrice de Fourier *rectangulaire* d'éléments $F_{n,m} = \frac{1}{M} \exp(2j\pi nm/M)$ et de dimensions $(N \times M)$. Ce problème est donc toujours mal-posé puisque indéterminé, mais cette fois de dimension finie. Comme on peut, par TFD inverse, associer au vecteur complexe recherché \mathbf{X} un vecteur réel \mathbf{x} de même longueur M , très supérieure à celle du vecteur des données \mathbf{y} , le problème peut être effectivement vu aussi comme un problème d'extrapolation de signal. On choisit de la régulariser de la manière désormais habituelle (40) :

$$\hat{\mathbf{X}} = \arg \min \{ \|\mathbf{y} - \mathbf{F}\mathbf{X}\|_2^2 + \alpha \mathcal{F}(\mathbf{X}) \} \tag{51}$$

où $\mathcal{F}(\mathbf{X})$ est la fonctionnelle de régularisation à choisir. Cela est, nous l'avons vu (§ 4.5), équivalent à rechercher le *mode* (MAP) d'une distribution *a posteriori* :

$$p(\mathbf{X} | \mathbf{y}, \theta) \propto p(\mathbf{X} | \theta) p(\mathbf{y} | \mathbf{X}, \theta)$$

où la vraisemblance $p(\mathbf{y} | \mathbf{X}, \theta)$ serait de la forme normale habituelle :

$$p(\mathbf{y} | \mathbf{X}, \theta) = (2\pi\sigma_d^2)^{-N/2} \exp \left[-\frac{1}{2\sigma_d^2} \|\mathbf{y} - \mathbf{F}\mathbf{X}\|_2^2 \right]$$

$\mathbf{X} \in \mathbb{C}^M$, $\mathbf{y} \in \mathbb{R}^N$, mais il faut bien voir cependant qu'une telle vraisemblance ne peut pas s'interpréter comme une version décentrée d'une distribution de probabilité pour des erreurs :

$$p(\mathbf{y} | \mathbf{X}, \theta) = q(\mathbf{y} - \mathbf{F}\mathbf{X} | \theta)$$

puisqu'il n'y a aucun « bruit » dans notre problème – celui-ci s'écrit en effet $\mathbf{y} = \mathbf{F}\mathbf{X}$ et non pas $\mathbf{y} = \mathbf{F}\mathbf{X} + \mathbf{b}$ – et, partant, aucune distribution directe, sauf dans le sens rudimentaire où $P(y(n) | \mathbf{X}) = 1$ ou 0 selon que \mathbf{X} est, ou pas, compatible avec les données $y(n)$. Les données observées \mathbf{y} nous disent seulement que le « vrai » \mathbf{X}_0 doit être dans la classe \mathcal{C} pour laquelle $P(y(n) | \mathbf{X}) = 1$. La vraisemblance présente donc un plateau horizontal, pour tous les $\mathbf{X} \in \mathcal{C}$, ce qui explique que la recherche d'une solution au maximum de vraisemblance soit sans issue. C'est le choix d'une distribution *a priori* pour \mathbf{X} qui permettra, éventuellement, d'obtenir une distribution *a posteriori* $p(\mathbf{X} | \mathbf{y}, \theta)$ suffisamment piquée pour pouvoir prendre une décision.

On a montré par (26) que le choix d'un *a priori* gaussien pour \mathbf{X} nous ramène à la solution du *périodogramme* – c'est-à-dire à une extrapolation par des zéros – et qu'il faut choisir une distribution « à longue queue » pour aller au-delà. Plusieurs choix sont possibles [41], et [40] propose une loi de Cauchy, parfaitement utilisable malgré la non existence de ses moments, caractéristique bien connue et à l'origine de la malédiction dont elle semble frappée en statistique fréquentiste. On choisit un coefficient de régularisation nul, et le seul hyperparamètre à régler est le paramètre d'échelle de la loi de Cauchy.

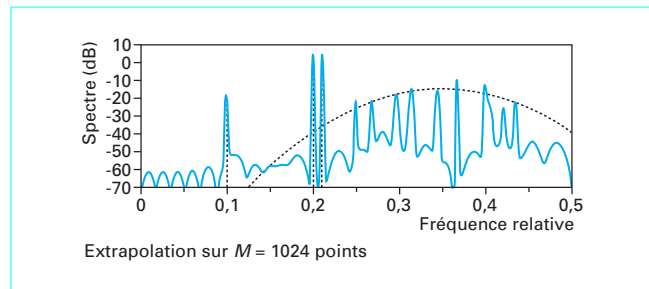


Figure 13 – Spectre obtenu avec un *a priori* de Cauchy

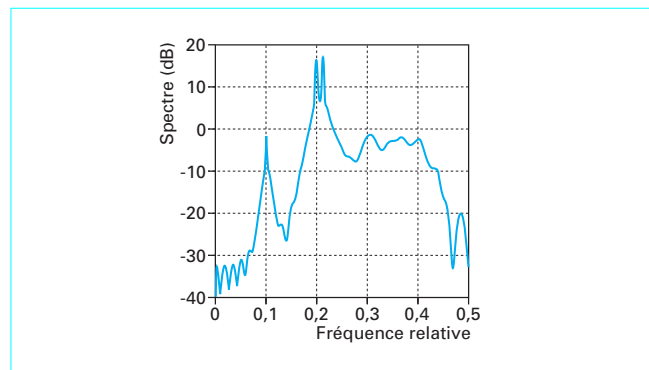


Figure 14 – Spectre obtenu avec un mélange dans l'*a priori*

Un exemple de résultat obtenu est présenté (figure 13). Les trois fréquences pures du signal sont très bien localisées et leurs amplitudes respectives bien restituées. On remarquera aussi la dynamique de l'amplitude du spectre estimé. Par contre, la composante du spectre due au bruit n'est pas très bien rendue, même si l'enveloppe des pics correspondants en donne une idée correcte. Cela provient bien sûr du choix de la loi *a priori* pour \mathbf{X} qui est séparable, ce qui convient très bien pour des composantes harmoniques sans rapport entre elles, mais ne permet pas de traduire les propriétés de continuité fréquentielle et de variation douce de la composante spectrale due au bruit. Il faudrait réaliser un « mélange » dans la fonctionnelle de régularisation. Cette voie a été explorée [42], avec les résultats présentés (figure 14) et on se reportera à [42] pour la description de la méthode.

4.8.2 Restauration d'image

Dans le cas de la déconvolution (premier exemple), nous avons vu au paragraphe 3.2 que la régularisation linéaire-quadratique ne permettait pas de restaurer correctement les contours francs présents dans l'objet original. L'interprétation bayésienne du paragraphe 4.6 nous en a fourni une explication : le choix d'un modèle *a priori* implicitement gaussien et stationnaire pour l'image inconnue conduit à effectuer, lors de l'inversion, un filtrage linéaire des données, et celui-ci est finalement un compromis entre le filtrage inverse nécessaire pour restituer des contours francs et le lissage accentué nécessaire pour limiter l'influence du bruit. Le résultat aboutit, dans ce contexte linéaire, à n'effectuer qu'une égalisation spectrale dans une étroite bande de fréquences, en raison du caractère passe-bas du système d'observation. Il faut donc, pour aller au-delà, *quitter ce cadre linéaire* en adoptant un modèle *a priori* plus approprié.

Un modèle gaussien stationnaire n'est effectivement pas bien adapté à la description d'une image telle que celle de la figure 1, constituée essentiellement de zones plutôt homogènes séparées par des contours plutôt francs. De façon générale, la traduction quantitative de propriétés locales d'une image (régions homogènes séparées par des contours, par exemple) peut être envisagée dans le cadre très large des fonctions d'énergie qui ont été introduites en traitement d'image en considérant que l'objet \mathbf{x} est, comme le bruit, une réalisation d'une variable aléatoire [28] [29]. Un préalable consiste à définir un système de voisinages $\{d_i\}$ dans lequel d_i est la collection des pixels qui sont supposés interagir directement avec le pixel i . Une caractéristique fréquente de nombreuses images est qu'il est très probable que les valeurs des intensités en des sites voisins soient elles-mêmes voisines. Pour traduire cette propriété attendue, on peut choisir, à titre d'exemple élémentaire, des énergies locales de la forme :

$$V(x_i, x_j) = |x_i - x_j|^2 \quad j \in d_i \quad (52)$$

$$= 0 \quad \text{sinon}$$

Ces énergies de liaison, calculées sur tous les pixels voisins, sont ajoutées pour définir l'énergie de l'image :

$$\mathcal{V}(\mathbf{x}) = \sum_i \sum_j V(x_i, x_j) = \sum_k V_k(\mathbf{x})$$

V_k est l'énergie associée au k^{e} ensemble de pixels voisins, et k indexe l'ensemble de toutes les paires de sites voisins. L'idée suivie dans cette construction est que $\mathcal{V}(\mathbf{x})$ doit être faible pour les images qui vérifient les propriétés que l'on veut utiliser pour définir *a priori*, c'est-à-dire dans cet exemple, les images dont les pixels voisins ont tendance à avoir les mêmes valeurs d'intensité.

Mais les fonctions d'énergie peuvent aussi traduire d'autres propriétés, telles que l'existence de régions presque uniformes séparées par des contours marqués. Pour modéliser ces ruptures d'homogénéité, l'image idéale est alors considérée comme une paire (\mathbf{x}, \mathbf{t}) associant un vecteur d'intensité des pixels \mathbf{x} , qui serait directement observable avec un instrument parfait, à un vecteur \mathbf{t} de variables supplémentaires cachées traduisant des caractéristiques inobservables telles que la présence d'un contour, d'une texture, etc. On définit ensuite une énergie globale de la forme :

$$\mathcal{F}(\mathbf{x}, \mathbf{t}) = \mathcal{V}_x(\mathbf{x}) + \mathcal{V}_t(\mathbf{t}) + \mathcal{V}_{xt}(\mathbf{x}, \mathbf{t})$$

composée de trois termes : un terme d'intensité $\mathcal{V}_x(\mathbf{x})$, un terme de contour $\mathcal{V}_t(\mathbf{t})$ et un troisième terme $\mathcal{V}_{xt}(\mathbf{x}, \mathbf{t})$ qui décrit les interactions nécessaires entre les contours et les pixels [29]. Ces énergies locales ont des liens étroits avec la physique statistique et les distributions de Gibbs qui ont la forme suivante : $p(\mathbf{x}, \mathbf{t}) = \exp \{-\mathcal{F}(\mathbf{x}, \mathbf{t})/T\}$ où T est un paramètre de température. Dans le domaine de la segmentation d'image, des formes spécifiques de variables de contour ont été introduites pour obtenir des contours fermés. Les variables cachées constituent un moyen très puissant d'incorporer une information *a priori* élaborée [29].

Un autre moyen de mieux préserver les discontinuités dans un objet que ne le font les méthodes de régularisation employant des critères quadratiques, consiste à employer des *énergies non quadratiques* [43]. Le principe est d'utiliser une fonction qui croît moins vite qu'une parabole, afin de moins pénaliser les variations importantes. Elles sont essentiellement de deux types :

- les fonctions L_{21} , c'est-à-dire les fonctions convexes, continûment différentiables, de comportement quadratique à l'origine, asymptotiquement linéaires, et dont un exemple typique est la branche d'hyperbole ;
- les fonctions L_{20} qui diffèrent des précédentes en étant asymptotiquement constantes, et donc non convexes.

Les premières offrent l'avantage d'être convexes, si bien que les techniques standards de minimisation sont assurées de converger (§ 5), et de fournir une certaine *robustesse* à la solution [30]. Les autres permettent de détecter réellement les discontinuités, mais

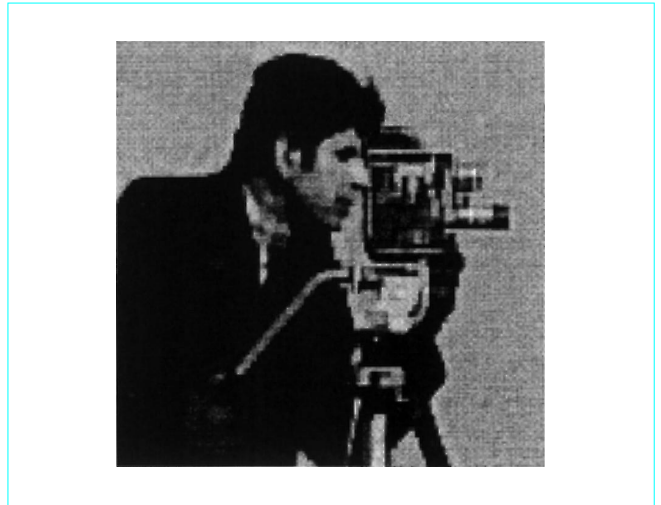


Figure 15 - Déconvolution convexe L_{21}

au prix d'une certaine instabilité et d'un coût calculatoire élevé [44].

Pour notre exemple en déconvolution d'images, la figure 15 présente un exemple typique de résultats obtenu à partir de pénalisation convexe L_{21} . Cette solution minimise un critère régularisé $\mathcal{F}(\mathbf{x})$ de la forme (25). La fonction \mathcal{G} est la norme quadratique, le terme d'adéquation aux données demeure donc un terme de moindres carrés, comme nous en avons déjà rencontré. En revanche, la fonction \mathcal{F} est bâtie sur la base d'une fonction L_{21} hyperbolique :

$$\mathcal{F}(\mathbf{x}) = \|\mathbf{y} - A\mathbf{x}\|^2 + \alpha \sum_{i,j} \varphi(x_i - x_j) \quad (53)$$

avec $\varphi(u) = \sqrt{s^2 + u^2}$. Ce résultat laisse apparaître sans équivoque une amélioration par rapport à la figure 10, obtenue par régularisation quadratique. Les contours du manteau et de la chevelure du personnage sont effectivement plus francs et les détails de la caméra mieux rendus.

4.9 Choix de critères

L'approche bayésienne ramène l'inversion à la détermination d'une distribution *a posteriori*. Comme il n'est pas envisageable de calculer complètement de telles distributions, on se contente de rechercher un estimateur ponctuel qui est souvent celui du maximum *a posteriori*. D'autres alternatives existent (*maximum a posteriori marginal*, *moyenne a posteriori*, etc.), mais il est important de bien évaluer les conséquences d'un tel choix et de proposer, si nécessaire, des alternatives.

La question peut être valablement soulevée de l'absolue nécessité d'imposer que la solution soit continue par rapport aux données et, par conséquent, de la convexité des critères de régularisation. Alors que les approches quadratiques et entropiques sont bien connues pour rendre les problèmes inverses bien posés, la minimisation d'une fonctionnelle non convexe ne peut assurer que la solution soit continue : une petite variation dans les données peut induire un « saut » d'une vallée à une autre, et donc une perte de continuité. Mais, dans beaucoup de problèmes, ces transitions sont non seulement désirables, mais nécessaires pour restaurer des discontinuités, des frontières, des interfaces, des points brillants, etc. Un éclairage différent peut être donné à ce problème en observant que certains critères non convexes introduits en imagerie possèdent une expression équivalente impliquant des *variables cachées*. Dans

ce cas, le problème quitte l'analyse convexe et incorpore une certaine dose d'analyse combinatoire, ou de test d'hypothèse, ce qui relève de la théorie de la décision plus que de l'estimation. L'analyse bayésienne reste pertinente dans ce contexte de détection-estimation combinées. Nombre de travaux récents ont été engagés dans cette voie, combinant plusieurs niveaux de variables, mélangeant descriptions à bas et haut niveau, ou des données recueillies par des modalités expérimentales différentes. C'est en ce sens que les concepts classiques de la régularisation, comme la continuité par rapport aux données, ne sont plus totalement appropriés et qu'un effort d'extension doit être fait.

5. Méthodes de minimisation de critère

La plupart des méthodes d'inversion reposent, d'une façon explicite ou implicite, sur la minimisation d'un critère. Suivant les propriétés de ce dernier, le coût du calcul effectif de la solution peut varier dans des proportions très importantes : typiquement, d'un facteur 1 000 entre la minimisation d'un critère quadratique par inversion d'un système linéaire et la minimisation d'un critère multimodal par une technique de relaxation comme le *recuit simulé*, toutes choses égales par ailleurs. *In fine*, le choix de la « bonne » méthode d'inversion doit donc dépendre des moyens de calculs disponibles. Encore faut-il savoir, pour un problème d'optimisation donné, quel algorithme mettre en œuvre ; ainsi, dans la comparaison précédente, il serait possible, mais totalement inefficace, d'utiliser la technique du recuit simulé pour minimiser un critère quadratique.

Un panorama non exhaustif des problèmes d'optimisation dans le contexte des problèmes inverses en traitement du signal et de l'image est ici proposé, ainsi que des outils algorithmiques associés. Elle ne prétend pas remplacer la littérature plus généralement dédiée à l'optimisation, [45] ou [46] par exemple.

5.1 Minimisation de critère pour l'inversion

Par *minimisation d'un critère*, on entend : trouver $\hat{\mathbf{x}}$ qui minimise $\mathcal{J}(\mathbf{x})$ parmi les éléments de \mathcal{X} . Dans la suite, on se place dans le cas de vecteurs réels : $\mathcal{X} \subset \mathbb{R}^M$.

Nota : le cas où \mathbf{x} est une fonction (plus précisément, le cas d'un espace \mathcal{X} de dimension infinie) pose des difficultés mathématiques qui sont du ressort de l'analyse fonctionnelle.

Le critère \mathcal{J} et l'ensemble \mathcal{X} peuvent dépendre : des données ; de propriétés structurelles (des termes additionnels dans l'expression de \mathcal{J} permettent de traduire des contraintes « molles », tandis que la spécification de \mathcal{X} est susceptible d'imposer des contraintes « dures ») ; d'hyperparamètres gérant le compromis entre adéquation aux données et régularité.

Ainsi, dans le cas de l'inverse généralisée (§ 2.2), $\mathcal{J}(\mathbf{x}) = \|\mathbf{x}\|$, et $\mathcal{X} = \{\mathbf{x}, \mathbf{A}^t \mathbf{A} \mathbf{x} = \mathbf{A}^t \mathbf{y}\}$ est l'ensemble des minimiseurs de $\|\mathbf{y} - \mathbf{A} \mathbf{x}\|$. Ou encore, dans le cas de la spécification de critères composites abordée au paragraphe 3.2 :

$$\mathcal{J}(\mathbf{x}) = \mathcal{G}(\mathbf{y} - \mathbf{A} \mathbf{x}) + \alpha \mathcal{F}(\mathbf{x}) \tag{54}$$

avec $\mathcal{X} = \mathbb{R}^M$ (cas non contraint) (55)

ou $\mathcal{X} = \mathbb{R}_+^M$ (contrainte de positivité) (56)

Dans le cas du *maximum a posteriori*, l'estimateur maximise (40), donc il minimise l'anti-log-vraisemblance *a posteriori* $-\ln p(\mathbf{x} | \mathbf{y}, \mathbf{A}, \boldsymbol{\theta})$, qui s'écrit encore (54), à une constante additive près. En revanche, d'autres estimateurs bayésiens classiques, tels que la moyenne *a posteriori*, résultent d'un calcul d'intégrale et non pas de la minimisation d'une fonction.

La définition formelle de $\hat{\mathbf{x}}$ comme minimiseur d'un critère cache principalement **trois niveaux de difficulté**, en termes de mise en œuvre, qui sont, par complexité croissante :

① \mathcal{J} est quadratique :

$$\mathcal{J} = \mathbf{x}^t \mathbf{M} \mathbf{x} - 2 \mathbf{v}^t \mathbf{x} + cte \quad \text{et} \quad \mathcal{X} = \mathbb{R}^M$$

ou bien \mathcal{X} est affine :

$$\mathcal{X} = \{\mathbf{x}_0 + \mathbf{B} \mathbf{u}, \mathbf{u} \in \mathbb{R}^P, P < M\};$$

② \mathcal{J} est une fonction convexe différentiable et $\mathcal{X} = \mathbb{R}^M$ ou un sous-ensemble convexe (fermé) de \mathbb{R}^M ;

③ \mathcal{J} n'a pas de propriété avérée.

5.2 Cas quadratique

Dans la situation ①, avec $\mathcal{X} = \mathbb{R}^M$ et en supposant \mathbf{M} symétrique et inversible, $\hat{\mathbf{x}}$ est la solution du **système linéaire** $\mathbf{M} \hat{\mathbf{x}} = \mathbf{v}$ de taille $M \times M$, qui exprime l'annulation du gradient $\nabla \mathcal{J}(\hat{\mathbf{x}}) = \mathbf{0}$. C'est, par exemple, le cas de l'équation (43) obtenue dans le cadre linéaire gaussien du paragraphe 4.6.

Dans la variante contrainte à un espace \mathcal{X} affine, il suffit de remplacer $\mathbf{x} \in \mathcal{X}$ par son expression en \mathbf{u} pour revenir à la minimisation non contrainte d'un critère quadratique, dans \mathbb{R}^P .

5.2.1 Techniques non itératives

Un nombre fini d'opérations suffit pour inverser tout système linéaire : de l'ordre de M^3 opérations (et M^2 casse-mémoire) pour un système $M \times M$. Si la *matrice normale* $\mathbf{M} = \{m_{ij}\}$ possède une structure particulière, le coût d'inversion du système est susceptible de diminuer.

Ainsi, en traitement de signal, la stationnarité se traduit par le caractère *Tœplitz* de la matrice normale (c'est-à-dire $m_{ij} = \mu_{j-i}$). En traitement d'image sous hypothèse stationnaire, la matrice normale est *doublement Tœplitz* (c'est-à-dire composée de sous-matrices de Tœplitz, l'agencement des sous-matrices étant lui-même Tœplitz). Dans ces deux cas, on trouve des algorithmes d'inversion coûtant de l'ordre de M^2 opérations et M cases-mémoire (algorithme de Levinson), et même des algorithmes encore plus rapides utilisant la transformée de Fourier rapide, coûtant seulement de l'ordre de $M \ln M$ opérations. L'expression spectrale du filtre de Wiener (44) se prête ainsi particulièrement bien à une mise en œuvre rapide, valable dans le cas d'une matrice normale *circulante* (c'est-à-dire $m_{ij} = \mu_{j-i \pmod{M}}$) [1].

Le caractère *creux* de la matrice normale peut également être mis à profit : si seulement ML coefficients de \mathbf{M} sont non nuls, on peut espérer ramener l'ordre du coût d'inversion à $M^2 L$ opérations, et à ML cases-mémoire. Par exemple, la matrice normale $\mathbf{A}^t \mathbf{R}_b^{-1} \mathbf{A} + \mathbf{R}_x^{-1}$ associée à (43) est *bande* ($m_{ij} = 0$ si $|j-i| \geq \ell < M$) si \mathbf{A} correspond au filtrage par une réponse impulsionnelle finie de petite taille, avec \mathbf{R}_b diagonale (hypothèse de bruit blanc) et $\mathbf{R}_x^{-1} = \mathbf{D}_k^t \mathbf{D}_k$. Une matrice bande est creuse, et L est du même ordre que ℓ .

5.2.2 Techniques itératives

Si le nombre d'inconnues M est très grand (par exemple, les pixels en restauration d'image, ou les voxels pour des objets volumiques), le coût de mémorisation des techniques non itératives devient souvent prohibitif. Il est alors préférable d'utiliser une méthode *à point fixe*, engendrant itérativement une suite $\hat{\mathbf{x}}^{(i)}$ de limite $\hat{\mathbf{x}} = \mathbf{M}^{-1} \mathbf{v}$. Toutes les variantes classiques sont des techniques de descente, c'est-à-dire qu'elles vérifient $\mathcal{J}(\hat{\mathbf{x}}^{(i+1)}) \leq \mathcal{J}(\hat{\mathbf{x}}^{(i)})$. On peut distinguer trois familles.

Parution : novembre 2001 - Dernière validation : juin 2019 - Ce document a été délivré pour le compte de 7200106152 - éditions ti // celine BLONDOU // 2.59.188.28

■ **Techniques « inconnue par inconnue »**

Une seule composante diffère entre $\hat{\mathbf{x}}^{(i)}$ et $\hat{\mathbf{x}}^{(i+1)}$. On balaye cycliquement les M composantes au cours des itérations. C'est le principe de la méthode de Gauss-Seidel, appelée encore *coordinate descent* [46]. Elle peut être généralisée à un traitement par bloc des composantes, et est d'autant plus intéressante, et partiellement parallélisable, que \mathbf{A} est creuse.

■ **Techniques « donnée par donnée »**

Une seule composante de \mathbf{y} est prise en compte pour calculer $\mathbf{x}^{(i+1)}$ à partir de $\hat{\mathbf{x}}^{(i)}$. On balaye cycliquement les N données au cours des itérations, ce qui rend cette approche incontournable si les données sont trop nombreuses pour être traitées simultanément. Ainsi, les *techniques de reconstruction algébrique* (ART), historiques en imagerie médicale par tomographie à rayons X, suivent ce principe pour minimiser le critère des moindres carrés $\|\mathbf{y} - \mathbf{Ax}\|^2$ [47]. Elles sont généralisables au critère pénalisé $\|\mathbf{y} - \mathbf{Ax}\|^2 + \alpha \|\mathbf{x}\|^2$ [48], peuvent traiter des blocs de données, et sont d'autant plus intéressantes, et partiellement parallélisables, que \mathbf{A} est creuse.

■ **Techniques « globales »**

À chaque itération, toutes les inconnues sont remises à jour en fonction de toutes les données. Les algorithmes de *gradient* sont des prototypes d'approche globale : $\hat{\mathbf{x}}^{(i+1)} = \hat{\mathbf{x}}^{(i)} - \lambda(\hat{\mathbf{x}}^{(i)})\nabla\mathcal{F}(\hat{\mathbf{x}}^{(i)})$, avec $\nabla\mathcal{F}(\hat{\mathbf{x}}^{(i)}) = 2\mathbf{M}\hat{\mathbf{x}}^{(i)} - 2\mathbf{v}$. Remarquons que la méthode de Landweber (§ 3.1) est en fait une technique de gradient minimisant le critère non régularisé $\|\mathbf{y} - \mathbf{Ax}\|^2$. Les algorithmes de *gradient conjugué* ou *pseudo-conjugué* sont des variantes convergeant plus rapidement, dans lesquelles les directions successives de descente combinent les gradients précédemment calculés pour éviter la trajectoire en zig-zag du gradient simple [45]. Ces variantes sont du premier ordre, donc d'un encombrement mémoire réduit : elles utilisent seulement les M dérivées premières $\partial\mathcal{F}/\partial x_m$.

Quant aux **techniques du second ordre**, telles que les algorithmes de Newton et *quasi-Newton* [45], elles sont très rarement utilisables car elles nécessitent le calcul (ou l'approximation) des $M \times M$ dérivées secondes $\partial^2\mathcal{F}/\partial x_i \partial x_j$ et l'inversion d'un système linéaire à chaque itération, ce qui est prohibitif pour la plupart des problèmes d'inversion réalistes.

5.3 Cas convexe

Les critères quadratiques font partie de la famille plus vaste des fonctions f convexes, c'est-à-dire telles que :

$$\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}, \theta \in]0, 1[, \\ \mathcal{F}[\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2] \leq \theta\mathcal{F}(\mathbf{x}_1) + (1-\theta)\mathcal{F}(\mathbf{x}_2)$$

avec $\mathcal{X} \subset \mathbb{R}^M$. De même, \mathcal{X} est un ensemble convexe si :

$$\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}, \theta \in]0, 1[, \text{ on a } \theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in \mathcal{X}$$

La spécification de critères convexes, mais pas forcément quadratiques, offre un plus grand choix en terme de modélisation. Ainsi, la pseudo-distance de Kullback (28) est convexe sur \mathbb{R}_+^M ; les fonctions de pénalisation markoviennes $\mathcal{F}(\mathbf{x}) = \sum_{j=1}^M \varphi(x_j - x_{j+1})$ sont convexes sur \mathbb{R}^M si φ est une fonction scalaire convexe comme $\varphi(x) = \sqrt{s^2 + x^2}$.

La minimisation de critères convexes non quadratiques, bien que plus difficile et plus coûteuse que celle des critères quadratiques, reste tout à fait compatible avec les moyens de calculs modernes, d'où le recours de plus en plus fréquent aux fonctions de pénalisation convexes en restauration de signaux et

d'images [43]. Commençons par rappeler quelques propriétés fondamentales des critères convexes [46] :

- un critère continu convexe \mathcal{F} est *unimodal* : tout minimum local est global et l'ensemble de ses minimiseurs est convexe ;
- si $\mathcal{F}_1, \mathcal{F}_2$ sont convexes et $\alpha_1, \alpha_2 \geq 0$, alors $\alpha_1\mathcal{F}_1 + \alpha_2\mathcal{F}_2$ est convexe ;

Nota : cette propriété « explique » pourquoi on s'intéresse à la convexité plutôt qu'à l'unimodalité : par exemple, le critère pénalisé (54) est convexe (donc unimodal) si \mathcal{G} et \mathcal{F} sont convexes, alors que l'unimodalité de \mathcal{G} et de \mathcal{F} ne suffirait pas pour garantir l'unimodalité du critère.

- si \mathcal{F} est *strictement* convexe, il existe un et un seul minimiseur $\hat{\mathbf{x}}$ dans tout \mathcal{X} convexe *fermé* (c'est-à-dire que la frontière de \mathcal{X} lui appartient).

En revanche, si le critère est non quadratique, le minimiseur $\hat{\mathbf{x}}$ est une fonction des données qui n'est généralement **ni linéaire, ni explicite**. De ce fait, les techniques non itératives d'inversion de systèmes linéaires du paragraphe 2 n'ont plus cours. En revanche, les trois familles de techniques itératives, fondées sur la diminution successive du critère, engendrent des algorithmes convergeant vers $\hat{\mathbf{x}}$ si \mathcal{F} est convexe et différentiable si $\mathcal{X} = \mathbb{R}^M$. Le cas d'un critère convexe mais non différentiable est un peu plus délicat : les techniques modernes, dites *de point intérieur*, approchent la solution en minimisant une suite d'approximations convexes différentiables [46].

D'autres familles de techniques convergentes sont aussi envisageables : *moindres carrés repondérés*, baptisés aussi algorithmes *semi-quadratiques* [43], ou encore les approches fondées sur la maximisation d'un critère *dual* [46] [49] [50]. La présentation de ces approches dépasse le cadre de cet article.

Si \mathcal{X} est un sous-ensemble convexe fermé de \mathbb{R}^M , des adaptations sont nécessaires : *gradient projeté* ou *gradient conditionnel* dans la famille des techniques globales [46], techniques de *projection sur des ensembles convexes* [51] [52] dans la famille des techniques « donnée par donnée ». Quant aux techniques « inconnue par inconnue », elles restent particulièrement simples si les contraintes sont *séparables*, c'est-à-dire si \mathcal{X} est un produit cartésien ; par exemple, la positivité correspond à $\mathcal{X} = \mathbb{R}_+ \times \dots \times \mathbb{R}_+$. Enfin, certains problèmes contraints sont équivalents à un problème non contraint dans le domaine dual, ce qui justifie le recours aux méthodes duales.

5.4 Cas général

Dans le cas de critères non convexes, l'existence possible de minima locaux rend l'utilisation des techniques de descente hasardeuse, au sens où tout minimiseur local est un point fixe possible pour la plupart de ces techniques. La convergence vers $\hat{\mathbf{x}}$ plutôt que le piégeage par une solution locale dépend alors de l'initialisation. Plusieurs stratégies sont envisageables pour éviter les pièges ; sauf exception, elles sont notoirement plus coûteuses que les méthodes de descente, sans pour autant garantir la convergence vers le minimiseur global. Malgré l'absence de garanties mathématiques de convergence, certaines techniques sont cependant suffisamment robustes pour éviter les solutions aberrantes. Elles produisent alors des résultats qui ne pourraient être obtenus par minimisation d'un critère convexe, pour des applications telles que la segmentation d'image automatique ou la détection d'objet.

On peut distinguer deux types d'approche. D'une part, les méthodes déterministes privilégient la robustesse, faute de propriété mathématique de convergence. Ainsi, le principe de la **non convexité graduelle** [53] [54] consiste à minimiser de proche en proche une suite de critères, à l'aide d'une technique de descente classique, en commençant par un critère convexe et en terminant par le critère non convexe. La robustesse de cette technique provient de la qualité de la solution initiale, pour une mise en œuvre d'un coût et d'une complexité relativement faibles. D'autre part, les méthodes pseudo-aléatoires (recuit simulé [29], méthodes de crible [55]) exploitent la

Parution : novembre 2001 - Dernière validation : juin 2019 - Ce document a été délivré pour le compte de 72001061152 - éditions tñ // celine BLONBOU // 2.59.4188.28

génération d'un grand nombre d'aléas pour échapper aux pièges ; le recuit simulé possède des propriétés de convergence (de type probabiliste [29]), mais le coût de calcul élevé de ces techniques explique leur usage encore limité dans le cadre de la restauration de signaux et d'images.

6. Questions plus avancées

6.1 Inversion myope

Dans les problèmes que nous avons étudiés jusqu' alors, nous avons supposé que le modèle direct est parfaitement connu. En pratique, il arrive souvent que ce modèle dépende de paramètres mal ou pas connus. On se trouve alors devant un problème dit « myope » ou « aveugle ». Un exemple type est celui de déconvolution à noyau inconnu. La **déconvolution aveugle** comporte un obstacle d'ordre informationnel encore beaucoup plus sévère que celui de la déconvolution à noyau connu. En effet, en l'absence de toute information *a priori* sur l'objet \mathbf{x} et sur la réponse impulsionnelle \mathbf{h} , il est évidemment impossible de discriminer le couple (\mathbf{x}, \mathbf{h}) de tout autre couple $(\mathbf{h}_1 \star \mathbf{x}, \mathbf{h}_2)$ tel que $\mathbf{h}_1 \star \mathbf{h}_2 = \mathbf{h}$. À l'extrême, la solution dégénérée $(\mathbf{h} \star \mathbf{x}, I)$ (où I est l'opérateur identité) joue le rôle d'une candidate indésirable mais « naturelle », dont il n'est pas forcément facile de se débarrasser. Il s'y ajoute, en général, une indétermination *sur la phase* : fréquence par fréquence, objet et filtre d'observation peuvent être déphasés d'une quantité arbitraire opposée sans affecter le résultat de la convolution.

Une revue de la littérature actuelle [56] [57] [58] [59] peut pourtant laisser penser que des solutions systématiques, « sans informations *a priori*, » existent.

Parmi les solutions souvent citées, [56] propose un estimateur de maximum de vraisemblance joint (noyau, hyperparamètre) sous hypothèse d'objet \mathbf{x} gaussien AR 2-D et de noyau RIF 2-D. Cette méthode repose sur l'identification des coefficients d'un processus ARMA, modèle résultant pour l'image observée. Mais le fonctionnement pratique d'une telle méthode semble très aléatoire, car il est manifeste que ce sont les hypothèses structurelles adoptées (l'objet est un AR 2-D, le noyau est un filtre RIF 2-D) qui permettent ici de lever l'indétermination. Or ces hypothèses correspondent à des modèles très grossiers. D'ailleurs, de nouvelles informations structurantes apparaissent en cours de route dans [56], pour pallier des difficultés qui sont faussement attribuées à des aspects numériques ou algorithmiques. La source de ces difficultés semble malheureusement beaucoup plus essentielle : l'indétermination fondamentale subsiste en l'absence d'informations suffisamment structurantes. Si de telles informations sont effectivement accessibles, il est excessif de parler de déconvolution *aveugle* ; on pourra parler, par exemple, de déconvolution *myope*.

On retrouve la même ambiguïté dans des contributions plus récentes. Ainsi, dans [59], un estimateur de maximum *a posteriori* joint $(\mathbf{x}, \mathbf{h})_{\text{MAP}}$ est proposé, et il est affirmé en introduction : « *our algorithm can recover both the images and psf's without any a priori information on the psf* ». Un peu plus loin, différents couples solutions $(\hat{\mathbf{x}}, \hat{\mathbf{h}})$ sont obtenus pour différentes valeurs d'un paramètre de régularisation α_2 , illustrant implicitement l'indétermination fondamentale avec le commentaire : « *... α_2 can be chosen proportionately according to the amount of desired deblurring* ». Finalement, des informations *a priori* sont nécessaires pour choisir α_2 , choix laissé à l'opérateur.

Selon notre propre expérience en déconvolution myope pour l'imagerie satellitaire, les résultats pratiques ne peuvent être que parcellaires, et les méthodes statistiques – maximum de vraisemblance $\hat{\mathbf{h}}_{\text{MV}}$, maximum *a posteriori* joint $(\mathbf{x}, \mathbf{h})_{\text{MAP}}$ – sont trop

fragiles car dépendantes de modèles *a priori* grossiers et d'hyperparamètres à régler. Deux voies de recherche réalistes sont en cours d'exploration :

- méthodes partiellement supervisées où, par exemple, on utilise une information pertinente sur l'objet (présence de quelques zones polygonales uniformes dans l'image) pour estimer avec une grande fiabilité le noyau qui permettrait de lever l'indétermination fondamentale de la déconvolution aveugle ;
- méthodes dans lesquelles on cherche à enrichir les données acquises, avec l'objectif explicite de lever l'indétermination. Il s'agit alors, par exemple de la déconvolution multicanal.

■ **Imagerie optique par diversité de phase en présence d'aberrations** (turbulence atmosphérique, défauts optiques). Les aberrations de phase inconnues ϕ se traduisent dans le plan focal par une convolution $\mathbf{z} = \mathbf{h}(\phi) \star \mathbf{x} + \text{bruit}$, avec $\mathbf{h}(\phi) = |\text{TF}(Pe^{i\phi})|^2$ (P est la fonction caractéristique de la pupille). Le principe de la diversité de phase est d'acquérir une seconde image hors du plan focal $\mathbf{z}' = \mathbf{h}(\phi + \delta\phi) \star \mathbf{x} + \text{bruit}'$, à déphasage $\delta\phi$ connu. Il faut souligner ici que la relation entre $\mathbf{h}(\phi)$ et $\mathbf{h}(\phi + \delta\phi)$ n'est pas linéaire. Les données complémentaires \mathbf{z}' peuvent donc permettre de lever l'indétermination fondamentale, y compris par les méthodes statistiques mentionnées précédemment [60].

■ **Imagerie en contrôle non destructif utilisant les micro-ondes** où l'idée est d'effectuer deux acquisitions avec deux balayages, horizontal et vertical, mais le même guide d'onde. Le problème se formule alors comme $g_i(x, y) = f(x, y) * h_i(x, y) + b_i(x, y)$, $i = 1, 2$ avec $h_1(x, y) = h(x, y)$ et $h_2(x, y) = h(y, x)$. L'objectif est d'estimer $h(x, y)$ et $f(x, y)$ à partir de $g_1(x, y)$ et $g_2(x, y)$. Il faut là aussi mentionner que la relation entre $h_1(x, y)$ et $h_2(x, y)$ n'est pas linéaire.

Nota : ces deux exemples sont tirés de notre propre expérience.

6.2 Choix de modèle

L'estimation des paramètres d'un modèle, supposé « vrai », n'est souvent qu'une étape dans un problème d'inversion, et l'on peut aussi avoir besoin de juger les mérites relatifs de différents modèles *a priori*.

Le problème consiste usuellement à choisir un modèle particulier $M_k \sim f_k(\mathbf{y} | \mathbf{x}_k)$, dans une famille candidate $(\mathbf{x}_k \in \mathbb{R}^{p_k}, k = 1, \dots, K)$, sur la base des données observées \mathbf{y} . Une des méthodes de choix les plus utilisées en traitement du signal, le critère d'information d'Akaike (AIC), conduit pourtant à une probabilité de surparamétrisation non nulle quand $N = \dim(\mathbf{y}) \rightarrow \infty$. Un autre critère populaire, celui de Rissanen-Schwarz (BIC), apporte une amélioration par une estimation convergeant presque sûrement. Il en existe bien d'autres, mais ils souffrent de défauts qui deviennent particulièrement évidents quand le *contraste* – rapport entre le nombre n de données disponibles et le nombre p_k de paramètres utilisés dans le modèle – est faible : ils tendent à surestimer considérablement l'ordre, à moins que de fortes restrictions ne soient mises à la dimension maximale des modèles candidats. De plus, les résultats obtenus avec un faible nombre de données sont très sensibles à la forme de la pénalisation apportée à la log-vraisemblance. Or, les problèmes inverses sont souvent caractérisés par un faible contraste. C'est une situation qui met en défaut les critères de choix les plus usités. Mais la difficulté du problème fait que les études à nombre fini de données sont très rares.

Dans un cadre bayésien [37] [61] [62], les modèles statistiques sont introduits pour représenter la probabilité des données en fonction de chacun des modèles (probabilités directes), et le théorème de Bayes est employé pour calculer la probabilité *a posteriori* que l'un des modèles soit correct. L'élément central est le *facteur de Bayes* dont le calcul, comme dans toute méthode bayésienne, implique des intégrales et donc, souvent, des intégrations numériques.

Parution : novembre 2001 - Dernière validation : juin 2019 - Ce document a été délivré pour le compte de 7200106152 - éditions ti // celine BLONBOU // 2.59.188.28

6.2.1 Facteurs de Bayes

Supposons pour commencer que les données \mathbf{y} aient été engendrées sous l'une ou l'autre de deux hypothèses M_1 et M_2 , selon une densité de probabilité $f(\mathbf{y} | M_1)$ ou $f(\mathbf{y} | M_2)$. Étant donné des probabilités *a priori* $P(M_1)$ et $P(M_2) = 1 - P(M_1)$, les données produisent les probabilités *a posteriori* $P(M_1 | \mathbf{y})$ et $P(M_2 | \mathbf{y}) = 1 - P(M_1 | \mathbf{y})$. L'évidence apportée par les données en faveur de l'une ou l'autre des hypothèses prend une forme particulièrement simple en utilisant l'échelle des *cotes* (cote = probabilité / (1 - probabilité)) :

$$\frac{P(M_1 | \mathbf{y})}{P(M_2 | \mathbf{y})} = \frac{P(M_1)}{P(M_2)} \frac{f(\mathbf{y} | M_1)}{f(\mathbf{y} | M_2)} \tag{57}$$

La cote *a priori* est donc simplement multipliée par :

$$B_{12} = \frac{f(\mathbf{y} | M_1)}{f(\mathbf{y} | M_2)} \tag{58}$$

qui est le *facteur de Bayes* [61]. Dans le cas le plus simple, quand les deux hypothèses sont des distributions sans paramètre libre (hypothèses *simples*), B_{12} est le rapport de vraisemblance usuel. Sinon, quand l'une ou l'autre des hypothèses implique des paramètres inconnus x_k , le facteur de Bayes a toujours la forme d'un rapport de vraisemblance, mais les densités $p(\mathbf{y} | M_k)$ ($k = 1, 2$) sont obtenues en *intégrant* (et non en maximisant) sur l'espace des paramètres :

$$p(\mathbf{y} | M_k) = \int p(\mathbf{y} | M_k, x_k) p(x_k | M_k) dx_k \tag{59}$$

Cette vraisemblance *marginale* n'est rien d'autre que le coefficient de normalisation dans l'expression de la densité de probabilité *a posteriori* pour le paramètre x_k donnée par la règle de Bayes, mais contrairement à beaucoup d'autres applications de la vraisemblance, tous les facteurs constants intervenant dans son expression doivent être conservés lors du calcul de B_{12} .

S'il y a K modèles concurrents M_1, \dots, M_K , et si l'on définit l'hypothèse *nulle* $M_0 : \mathbf{y} = \mathbf{b}$, la relation (57) montre que la probabilité *a posteriori* pour M_k est donnée par :

$$P(M_k | \mathbf{y}) = \frac{w_k B_{k0}}{\sum_{k'=0}^K w_{k'} B_{k'0}} \tag{60}$$

où $w_k = P(M_k) / P(M_0)$. Quand les différentes hypothèses M_k sont également probables *a priori*, les probabilités *a posteriori* sont proportionnelles aux facteurs de Bayes, avec pour toutes le même coefficient. Un passage au logarithme rend les relations (57) et (58) additives et montre le rôle essentiel de la log-vraisemblance marginale :

$$EV_k = \ln p(\mathbf{y} | M_k) \tag{61}$$

ou *evidence* pour le modèle candidat M_k .

6.2.2 Difficultés

On peut montrer [61] [63] que le critère de Rissanen-Schwarz est une approximation grossière de la log-vraisemblance marginale, ou *evidence* pour le modèle candidat, correspondant à des *a priori* variés. Il offre l'avantage appréciable d'éviter d'avoir à préciser l'*a priori*, au prix des inconvénients évoqués au paragraphe 6.2.1. Mais tout affinement soulève une batterie de difficultés.

L'emploi des distributions *a priori* impropres n'est pas permis ici, contrairement à l'estimation de paramètres où l'*a priori* est souvent choisi par commodité car on sait que si les données sont assez nombreuses, alors son effet est faible. Pour le test d'hypothèse, il n'en va pas de même car l'intégration sur des espaces différents introduit un problème de mise à l'échelle des facteurs de Bayes, conséquence du changement de la dimension du paramètre d'un modèle à l'autre.

L'emploi d'*a priori* diffus (de grande variance) n'est pas une solution. De plus, la dépendance de l'*a priori* ne décroît qu'asymptotiquement ($N \rightarrow \infty$).

Une difficulté d'un autre ordre provient du calcul analytique exact des intégrales des facteurs de Bayes qui est impossible la plupart du temps. On peut employer une expression approchée, fondée sur la méthode d'approximation analytique de Laplace. On peut aussi utiliser des méthodes de Monte Carlo, simples ou améliorées par échantillonnage préférentiel, mais elles sont plus lourdes à mettre en œuvre que les méthodes d'approximation analytique, et pas nécessairement plus précises dans le cas simple mais très répandu d'un modèle linéaire en paramètres [64].

6.2.3 Exemple

Pour illustrer ces problèmes de choix de modèle, nous utilisons l'exemple de l'analyse spectrale. Le premier modèle envisagé M_1 est le **modèle autorégressif long** du paragraphe 4.7 qui présente la particularité d'avoir une distribution directe $p(\mathbf{y} | \mathbf{a}, \theta)$ et une distribution *a priori* $p(\mathbf{a} | \theta)$ toutes deux normales. La distribution conjointe $p(\mathbf{y}, \mathbf{a} | \theta)$ est donc normale, elle aussi, et le calcul de l'intégrale de Gauss (46) (59) est faisable analytiquement. Reste à régler la question des hyperparamètres θ qui apparaissent dans la vraisemblance marginale $p(\mathbf{y} | \theta)$. Il faudrait, en toute rigueur, leur attribuer aussi une distribution *a priori*, diffuse mais propre, car ils ne sont pas nécessairement les mêmes pour tous les modèles à comparer, et les intégrer hors du problème pour obtenir $p(\mathbf{y} | M_1)$. Mais si la vraisemblance marginale $p(\mathbf{y} | \theta)$ est piquée autour de son maximum en $\hat{\theta}$, on peut écrire :

$$p(\mathbf{y} | M_1) \approx p(\mathbf{y} | \hat{\theta}) \tag{62}$$

ce qui donne, avec les données \mathbf{y} de la figure 5 et en utilisant le logarithme népérien :

$$EV_1 = \ln p(\mathbf{y} | M_1) = -80,3$$

Le deuxième modèle envisagé M_2 est le modèle **d'extrapolation de signal** introduit au paragraphe 4.8, avec un *a priori* gaussien circulaire, et dont on sait qu'il conduit à extrapoler le signal observé par des zéros, et donc à la solution du périodogramme pour le spectre. Cet exemple nécessite un traitement particulier puisqu'il n'y a pas de bruit (§ 4.8) et, partant, pas de distribution directe, sauf dans un sens rudimentaire. Le problème d'optimisation (51) s'écrit donc en fait :

$$\hat{\mathbf{X}} = \arg \max_{\mathbf{X} \in \mathbb{C}^M} p(\mathbf{X} | \sigma_g^2) \text{ s.c. } \mathbf{y} = \mathbf{F}\mathbf{X} \tag{63}$$

où σ_g^2 désigne la variance de la distribution gaussienne circulaire *a priori* pour \mathbf{X} . Le calcul de la vraisemblance marginale (59) passe par celui de l'intégrale :

$$p(\mathbf{y} | \sigma_g^2) = \int_{\mathcal{D}(\mathbf{y})} p(\mathbf{X} | \sigma_g^2) d\mathbf{X} \tag{64}$$

où le domaine $\mathcal{D}(\mathbf{y}) = \{\mathbf{X} \in \mathbb{C}^M / \mathbf{F}\mathbf{X} = \mathbf{y}\}$ est défini par la contrainte $\mathbf{y} = \mathbf{F}\mathbf{X}$. Le calcul est plus simple qu'il n'y paraît puisqu'il suffit de faire le changement de variable $\mathbf{z} = \frac{1}{M} \mathbf{G}^\dagger \mathbf{X}$, $\mathbf{z} \in \mathbb{R}^M$, où \mathbf{G} est la matrice de Fourier de dimensions $(M \times M)$ – \mathbf{z} est donc le vecteur des échantillons du signal extrapolé sur M points – et dont le jacobien vaut l'unité, pour obtenir :

$$p(\mathbf{y} | \sigma_g^2) = \int_{\mathbb{R}^{M-N}} p(\mathbf{z} | \sigma_g^2) dz(N+1) \dots dz(M) = (2\pi\sigma_g^2)^{-N/2} \exp \left[-\frac{1}{2\sigma_g^2} \|\mathbf{y}\|_2^2 \right] \tag{65}$$

L'estimateur au maximum de vraisemblance de l'hyperparamètre σ_g^2 est donc l'estimateur empirique $\hat{\sigma}_g^2 = \|\mathbf{y}\|_2^2 / N$, et un raisonnement analogue à celui fait pour le modèle précédent donne, avec les mêmes données bien sûr :

$$EV_2 = -88,3$$

On voit donc que si l'on se limite à la question de choisir entre le spectre du périodogramme et celui du modèle AR-long, et que ceux-ci sont *a priori* indifférents, le second a une évidence supérieure de 8 Néper à celle du premier, soit environ 35 dB. Autrement dit, leurs probabilités *a posteriori* sont dans un rapport d'environ 3 000 : la cote du modèle AR-long est suffisamment grande pour que l'on n'hésite pas à le retenir dans cet exemple. Mais attention aux erreurs d'interprétation ! Cela ne signifie absolument pas que le modèle AR-long soit le « vrai » modèle générateur du signal observé. Ce n'est que le modèle le plus vraisemblable dans l'ensemble des modèles concurrents examinés, et avec les données expérimentales disponibles. D'autres données ou d'autres modèles pourraient conduire à une conclusion différente.

Le modèle d'extrapolation du paragraphe 4.8, avec un *a priori* de Cauchy séparable, pourrait justement constituer un modèle M_3 à tester contre les deux précédents. Le principe du calcul de la vraisemblance marginale est le même que pour le modèle M_2 , mais cette fois, l'intégrale (65) n'est plus calculable explicitement, le changement de variable $\mathbf{z} = \frac{1}{M} \mathbf{G}^t \mathbf{X}$ conduisant à une distribution de probabilité dans \mathbb{R}^M non conventionnelle. Il n'est pas possible non plus d'utiliser des méthodes d'approximation analytique comme celle de Laplace puisque nous ne sommes pas dans des conditions asymptotiques, bien au contraire ($M \gg N$). Il reste les méthodes de calcul numérique par échantillonnage stochastique [64], mais nous sortons là définitivement du cadre limité donné à cet article.

7. Conclusion

Nous avons présenté brièvement les caractéristiques essentielles des problèmes inverses, de la régularisation et de son interprétation bayésienne. Nous nous sommes limités à l'estimation paramétrique et n'avons qu'évoqué le choix d'un modèle. Nous insistons à nouveau sur le fait que la plupart des méthodes d'inversion classiques peuvent être chacune établie, ou réinterprétée, dans plusieurs cadres théoriques, et qu'il n'y a pas de lien exclusif entre les méthodes (plus précisément les algorithmes de traitement des données) et leurs interprétations théoriques. Ces dernières, par contre, peuvent être classées en fonction de leur degré de généralité, c'est-à-dire leur capacité à affronter tous les problèmes soulevés lors de la résolution effective d'un problème inverse.

De ce point de vue, l'approche bayésienne offre un cadre remarquablement cohérent et complet pour traiter de problèmes

d'inférence dans une situation incertaine où l'on dispose de plusieurs sources d'information : données expérimentales, modèle direct, *a priori*. Outre le choix d'un critère régularisé, elle permet aussi le développement de techniques de calcul automatique des hyperparamètres et d'algorithmes d'optimisation pseudo-aléatoires, sans équivalents dans un cadre non probabiliste. Ses limitations sont bien connues, elles sont essentiellement d'ordre calculatoire. Il est clair que l'on ne peut pas espérer obtenir par cette approche de résultats satisfaisants dans des problèmes complexes, avec des modèles simplistes et des méthodes d'optimisation approximatives. C'est ce qui explique sans doute en grande partie l'échec partiel des tentatives faites dans les années 1980 en restauration d'image et en filtrage de bruit, avec des modèles gaussiens. Mais on a assisté, dans le domaine de la vision par ordinateur, et depuis l'article fondateur des frères Geman en 1984 [29], à une renaissance spectaculaire des méthodes de cette famille, associant champs markoviens, échantillonnage de Gibbs, techniques de recuit, parallélisation des calculs, dans des problèmes de segmentation, de détection de contours, d'extraction de textures, etc. Ces travaux ont fourni une base rationnelle très solide à ce domaine et influencent de plus en plus celui, beaucoup plus large, des problèmes inverses.

Notre conviction est qu'il y a encore beaucoup à faire dans cette voie, tant dans la construction de modèles probabilistes adaptés aux problèmes à résoudre, que dans la détermination des hyperparamètres, ou que dans l'optimisation de critère. Le travail se poursuit donc actuellement dans plusieurs directions.

■ Modélisation du problème direct

Les problèmes inverses sont souvent, par nature, non linéaires. C'est le cas, en particulier, de la tomographie de diffraction et de l'imagerie de résistivité. Il faudra donc introduire dans ce domaine des modèles non linéaires du problème direct.

■ Modélisation des objets étudiés

Les propriétés locales des images à reconstruire (homogénéité, contours, etc.) conduisent à utiliser des modèles *a priori* markoviens, mais cela soulève des difficultés nouvelles par rapport au cas de la vision par ordinateur, en raison du caractère non local des opérateurs linéaires issus du problème direct.

■ Choix des critères

L'approche bayésienne ramène l'inversion à la détermination d'une distribution *a posteriori*. Comme il n'est pas envisageable de calculer complètement de telles distributions, on se contente de rechercher un estimateur ponctuel qui est souvent celui du maximum *a posteriori*. Il est donc important d'évaluer les conséquences d'un tel choix et de proposer, si nécessaire, des alternatives. De même, la question du choix d'un critère de détermination des hyperparamètres nous semble encore largement ouverte.

■ Algorithmes d'extrémalisation de critères

Les critères étant choisis, se pose encore la question de leur extrémalisation, d'autant qu'ils sont souvent multimodaux. Il est donc important de continuer à développer des algorithmes stochastiques, en particulier du type « maximisation stochastique d'espérance conditionnelle ».

Références bibliographiques

- | | | |
|--|--|---|
| <p>[1] ANDREWS (H.C.) et HUNT (B.R.). – <i>Digital Image Restoration</i>. Prentice-Hall, Englewood Cliffs (1977).</p> <p>[2] HERMAN (G.T.), TUY (H.K.), LANGENBERG (K.J.) et SABATIER (P.). – <i>Basic Methods of Tomography and Inverse Problems</i>. Adam Hilgers, Bristol (1987).</p> | <p>[3] KAK (A.C.) et SLANEY (M.). – <i>Principles of Computerized Tomographic Imaging</i>. IEEE Press, New York (1988).</p> <p>[4] MARROQUIN (J.L.), MITTER (S.K.) et POGGIO (T.A.). – <i>Probabilistic solution of ill-posed problems in computational vision</i>. J. Amer. Stat. Assoc., 82, 76-89 (1987).</p> | <p>[5] BERTERO (M.), DE MOL (C.) et PIKE (E.R.). – <i>Linear inverse problems with discrete data : II. Stability and regularization</i>. Inverse Problems 4, 3 (1988).</p> <p>[6] KAY (S.M.) et MARPLE (S.L.). – <i>Spectrum analysis – a modern perspective</i>. Proc. IEEE, 69, 1380-1419 (1981).</p> |
|--|--|---|

- [7] CIUCIU (P.). – *Méthodes markoviennes en estimation spectrale non paramétrique. Applications en imagerie radar Doppler.* Thèse de doctorat, université de Paris-Sud-Orsay, oct. 2000.
- [8] NASHED (M.). – *Operator-theoretic and computational approaches to ill-posed problems with applications to antenna theory.* IEEE Trans. Ant. Propag., 29, 220-231 (1981).
- [9] BRÉZIS (H.). – *Analyse fonctionnelle : théorie et applications.* Masson, Paris (1983).
- [10] NASHED (M.). – *Generalized Inverses and Applications.* Academic, New York (1976).
- [11] MURTAGH (F.), STARK (J.L.) et BIJAOUI (A.). – *Image restoration with noise suppression using a multiresolution support.* Astron. Astrophys. Suppl., 112, 179-189 (1995).
- [12] MOHAMMAD-DJAFARI (A.). – « *Model selection for inverse problems: Best choice of basis function and model order selection* », in *Maximum Entropy and Bayesian Methods.* J.R.G. Erickson et C. Smith, Kluwer Academic Publ., Eds. (1999).
- [13] KALIFA (J.), MALLAT (S.) et ROUGÉ (B.). – « *Restauration d'images par paquets d'ondelettes* » in *Actes 16^e coll. GRETSI*, Grenoble, 1339-1342, sept. 1997.
- [14] DEMOMENT (G.). – *Déconvolution des signaux.* Cours de l'École supérieure d'électricité, n° 3086 (1985).
- [15] LUCY (L.B.). – *An iterative technique for the rectification of observed distributions.* The Astronomical Journal, 79, 745-754 (1974).
- [16] LUCY (L.B.). – *Optimum strategies for inverse problems.* Astron. Astrophys., 289, 983-994 (1994).
- [17] TITTERINGTON (D.M.). – *Common structure of smoothing techniques in statistics.* Int. Statist. Rev., 53, 141-170 (1985).
- [18] TIKHONOV (A.) et ARSÉNINE (V.). – *Méthodes de résolution de problèmes mal-posés.* Éditions MIR, Moscou (1976).
- [19] TARANTOLA (A.). – *Inverse problem theory: Methods for data fitting and model parameter estimation.* Elsevier Science Publishers, Amsterdam (1987).
- [20] THOMPSON (A.), BROWN (J.C.), KAY (J.W.) et TITTERINGTON (D.M.). – *A study of methods of choosing the smoothing parameter in image restoration by regularization.* IEEE Trans. Pattern Anal. Mach. Intell., PAMI-13, 326-339 (1991).
- [21] HANSEN (P.). – *Analysis of discrete ill-posed problems by means of the L-curve.* SIAM Rev., 34, 561-580 (1992).
- [22] GOLUB (G.H.), HEATH (M.) et WAHBA (G.). – *Generalized cross-validation as a method for choosing a good ridge parameter.* Technometrics, 21, 215-223 (1979).
- [23] REEVES (S.J.) et MERSEREAU (R.M.). – *Optimal estimation of the regularization parameter and stabilizing functional for regularized image restoration.* Opt. Engng., 29, 446-454 (1990).
- [24] FORTIER (N.), DEMOMENT (G.) et GOUSSARD (Y.). – *GCV and ML methods of determining parameters in image restoration by regularization: Fast computation in the spatial domain and experimental comparison.* J. Visual Comm. Image Repres., 4, 157-170 (1993).
- [25] DEMOMENT (G.). – *Image reconstruction and restoration: Overview of common estimation structure and problems.* IEEE Trans. Acoust. Speech, Signal Processing, ASSP-37, 2024-2036 (1989).
- [26] JAYNES (E.T.). – *On the rationale of maximum-entropy methods.* Proc. IEEE, 70, 939-952 (1982).
- [27] COX (R.). – *The Algebra of Probable Inference.* Johns Hopkins University Press, Baltimore (1961).
- [28] BESAG (J.E.). – *On the statistical analysis of dirty pictures (with discussion).* J. R. Statist. Soc. B, 48, 259-302 (1986).
- [29] GEMAN (S.) et GEMAN (D.). – *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images.* IEEE Trans. Pattern Anal. Mach. Intell., PAMI-6, 721-741 (1984).
- [30] BOUMAN (C.A.) et SAUER (K.D.). – *A generalized Gaussian image model for edge-preserving MAP estimation.* IEEE Trans. Image Processing, 2, 296-310 (1993).
- [31] KITAGAWA (G.) et GERSCH (W.). – *A smoothness priors long AR model method for spectral estimation.* IEEE Trans. Automat. Contr., AC-30, 57-65 (1985).
- [32] DEMPSTER (A.P.), LAIRD (N.M.) et RUBIN (D.B.). – *Maximum likelihood from incomplete data via the EM algorithm.* J. R. Statist. Soc. B, 39, 1-38 (1977).
- [33] GASSIAT (E.), MONFRONT (F.) et GOUSSARD (Y.). – *On simultaneous signal estimation and parameter identification using a generalized likelihood approach.* IEEE Trans. Inf. Theory, 38, 157-162 (1992).
- [34] ROBERT (C.). – *L'analyse statistique bayésienne.* Économica, Paris (1992).
- [35] BERNARDO (J.M.) et SMITH (A.F.M.). – *Bayesian Theory.* Wiley, Chichester (1994).
- [36] KASS (R.E.) et WASSERMAN (L.). – *The selection of prior distributions by formal rules.* J. Amer. Statist. Assoc., 91, 1343-1370 (1996).
- [37] JEFFREYS. – *Theory of Probability.* Clarendon Press, Oxford (1939).
- [38] LE BESNERAIS (G.), BERCHER (J.F.) et DEMOMENT (G.). – *A new look at entropy for solving linear inverse problems.* IEEE Trans. Inf. Theory, 45, 1565-1578 (1999).
- [39] BROWN (L.D.). – *Foundations of exponential families.* IMS Lecture Notes, Monograph Series edition, vol. 9, Hayward (1986).
- [40] SACCHI (M.D.), ULRYCH (T.J.) et WALKER (C.J.). – *Interpolation and extrapolation using a high-resolution discrete Fourier transform.* IEEE Trans. Signal Processing, 46, 31-38 (1998).
- [41] DEMOMENT (G.) et IDIER (J.). – *Approche bayésienne pour la résolution des problèmes inverses en imagerie* in « *Problèmes inverses : de l'expérimentation à la modélisation* ». M. Bonnet, éd., coll. Arago, vol. 22, 59-77. Observatoire français des techniques avancées, Tec & Doc, Paris (1999).
- [42] CIUCIU (P.), IDIER (J.) et GIOVANNELLI (J.F.). – *Markovian high resolution spectral analysis* in Proc. IEEE ICASSP, Phoenix, 1601-1604, mars 1999.
- [43] IDIER (J.). – *Regularization tools and models for image and signal reconstruction* in 3rd Intern. Conf. Inverse Problems in Engng., Port Ludlow, 23-29, juin 1999.
- [44] GEMAN (D.) et REYNOLDS (G.). – *Constrained restoration and the recovery of discontinuities.* IEEE Trans. Pattern Anal. Mach. Intell., 14, 192, 367-383.
- [45] PRESS (W.H.), FLANNERY (B.P.), TEUKOLSKY (S.A.) et VETTERLING (W.T.). – *Numerical recipes, the art of scientific computing.* Cambridge Univ. Press (1986).
- [46] BERTSEKAS (D.P.). – *Nonlinear programming.* Athena Scientific, Belmont (1995).
- [47] GILBERT (P.). – *Iterative methods for the three-dimensional reconstruction of an object from projections.* J. Theor. Biol., 36, 105-117 (1972).
- [48] HERMAN (G.T.), HURWITZ (H.), LENT (A.) et LUNG (H.P.). – *On the Bayesian approach to image reconstruction.* Inform. Contr., 42, 60-71 (1979).
- [49] LUENBERGER (D.G.). – *Optimization by Vector Space Methods.* John Wiley, New York (1969).
- [50] HEINRICH (C.) et DEMOMENT (G.). – *Minimization of strictly convex functions: an improved optimality test based on Fenchel duality.* Inverse Problems, 16, 795-810 (2000).
- [51] YOULA (D.C.) et WEBB (H.). – *Image restoration by the method of convex projection: Part 1 – theory.* IEEE Trans. Medical Imaging, MI-1, 81-94 (1982).
- [52] SEZAN (M.I.) et STARK (H.). – *Image restoration by the method of convex projections: Part 2 – applications and numerical results.* IEEE Trans. Medical Imaging, MI-1, 95-101 (1982).
- [53] BLAKE (A.) et ZISSERMAN (A.). – *Visual reconstruction.* The MIT Press, Cambridge (1987).
- [54] NIKOLOVA (M.), IDIER (J.) et MOHAMMAD-DJAFARI (A.). – *Inversion of large support ill-posed linear operators using a piecewise Gaussian MRF.* IEEE Trans. Image Processing, 7, 571-585 (1998).
- [55] WALTER (E.) et PRONZATO (L.). – *Identification de modèles paramétriques à partir de données expérimentales.* Masson (1994).
- [56] LAGENDIJK (R.), BIEMOND (J.) et BOEKKEE (D.E.). – *Identification and restoration of noisy blurred images using the expectation minimization algorithm.* IEEE Trans. Acoust. Speech, Signal Processing, ASSP-38, 1180-1191 (1990).
- [57] YOU (Y.L.) et KAVEH (M.). – *A regularization approach to joint blur identification and image restoration.* IEEE Trans. Image Processing, 5, 416-428 (1996).
- [58] KUNDUR (D.) et HATZINAKOS (D.). – *Blind image deconvolution.* IEEE Signal Processing Magazine, 43-64 (1996).
- [59] CHAN (T.F.) et CHIU-KWONG (W.). – *Total variation blind deconvolution.* IEEE Trans. Image Processing, 7, 370-375 (1998).
- [60] BLANC (A.), MUGNIER (L.) et IDIER (J.). – *Identifiabilité et estimation myope d'aberrations optiques par diversité de phase.* Actes 17^e coll. GRETSI, Vannes, 63-66, sept. 1999.
- [61] KASS (R.E.) et RAFTERY (A.E.). – *Bayes factors.* Amer. Statist. Assoc., 90, 773-795 (1995).
- [62] ZELLNER (A.) et SIOW (A.). – *Posterior odds ratio for selected regression hypothesis.* In *Bayesian Statistics*, J.M. Bernardo et al., Eds. Valencia University Press, 585-648 (1980).
- [63] PAULER (D.K.). – *The Schwarz criterion and related methods for normal linear models.* Biometrika, 85, 13-27 (1998).
- [64] EVANS (M.) et SWARTZ (T.). – *Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems.* Statistical Science, 10, 254-272 (1995).

GAGNEZ DU TEMPS ET SÉCURISEZ VOS PROJETS EN UTILISANT UNE SOURCE ACTUALISÉE ET FIABLE

Techniques de l'Ingénieur propose la plus importante collection documentaire technique et scientifique en français !

Grâce à vos droits d'accès, retrouvez l'ensemble des **articles et fiches pratiques de votre offre, leurs compléments et mises à jour,** et bénéficiez des **services inclus.**



RÉDIGÉE ET VALIDÉE
PAR DES EXPERTS



MISE À JOUR
PERMANENTE



100 % COMPATIBLE
SUR TOUS SUPPORTS
NUMÉRIQUES



SERVICES INCLUS
DANS CHAQUE OFFRE

- **+ de 350 000 utilisateurs**
- **+ de 10 000 articles de référence**
- **+ de 80 offres**
- **15 domaines d'expertise**

- Automatique - Robotique
- Biomédical - Pharma
- Construction et travaux publics
- Électronique - Photonique
- Énergies
- Environnement - Sécurité
- Génie industriel
- Ingénierie des transports
- Innovation
- Matériaux
- Mécanique
- Mesures - Analyses
- Procédés chimie - Bio - Agro
- Sciences fondamentales
- Technologies de l'information

**Pour des offres toujours plus adaptées à votre métier,
découvrez les offres dédiées à votre secteur d'activité**

Depuis plus de 70 ans, Techniques de l'Ingénieur est la source d'informations de référence des bureaux d'études, de la R&D et de l'innovation.

www.techniques-ingenieur.fr

CONTACT : Tél. : + 33 (0)1 53 35 20 20 - Fax : +33 (0)1 53 26 79 18 - E-mail : infos.clients@teching.com

LES AVANTAGES ET SERVICES compris dans les offres Techniques de l'Ingénieur

ACCÈS



Accès illimité aux articles en HTML

Enrichis et mis à jour pendant toute la durée de la souscription



Téléchargement des articles au format PDF

Pour un usage en toute liberté



Consultation sur tous les supports numériques

Des contenus optimisés pour ordinateurs, tablettes et mobiles

SERVICES ET OUTILS PRATIQUES



Questions aux experts*

Les meilleurs experts techniques et scientifiques vous répondent



Articles Découverte

La possibilité de consulter des articles en dehors de votre offre



Dictionnaire technique multilingue

45 000 termes en français, anglais, espagnol et allemand



Archives

Technologies anciennes et versions antérieures des articles



Impression à la demande

Commandez les éditions papier de vos ressources documentaires



Alertes actualisations

Recevez par email toutes les nouveautés de vos ressources documentaires

*Questions aux experts est un service réservé aux entreprises, non proposé dans les offres écoles, universités ou pour tout autre organisme de formation.

ILS NOUS FONT CONFIANCE



www.techniques-ingenieur.fr

CONTACT : Tél. : + 33 (0)1 53 35 20 20 - Fax : +33 (0)1 53 26 79 18 - E-mail : infos.clients@teching.com

Gagnez du temps et sécurisez vos projets en utilisant une source actualisée et fiable



RÉDIGÉE ET VALIDÉE
PAR DES EXPERTS




MISE À JOUR
PERMANENTE



100 % COMPATIBLE
SUR TOUS SUPPORTS
NUMÉRIQUES



SERVICES INCLUS
DANS CHAQUE OFFRE

- + de 340 000 utilisateurs chaque mois
- + de 10 000 articles de référence et fiches pratiques
- Des Quiz interactifs pour valider la compréhension 

SERVICES ET OUTILS PRATIQUES



Questions aux experts*

Les meilleurs experts techniques et scientifiques vous répondent



Articles Découverte

La possibilité de consulter des articles en dehors de votre offre



Dictionnaire technique multilingue

45 000 termes en français, anglais, espagnol et allemand



Archives

Technologies anciennes et versions antérieures des articles



Info parution

Recevez par email toutes les nouveautés de vos ressources documentaires

*Questions aux experts est un service réservé aux entreprises, non proposé dans les offres écoles, universités ou pour tout autre organisme de formation.

Les offres Techniques de l'Ingénieur

INNOVATION

- Éco-conception et innovation responsable
- Nanosciences et nanotechnologies
- Innovations technologiques
- Management et ingénierie de l'innovation
- Smart city – Ville intelligente

MATÉRIAUX

- Bois et papiers
- Verres et céramiques
- Textiles
- Corrosion – Vieillessement
- Études et propriétés des métaux
- Mise en forme des métaux et fonderie
- Matériaux fonctionnels. Matériaux biosourcés
- Traitements des métaux
- Élaboration et recyclage des métaux
- Plastiques et composites

MÉCANIQUE

- Frottement, usure et lubrification
- Fonctions et composants mécaniques
- Travail des matériaux – Assemblage
- Machines hydrauliques, aérodynamiques et thermiques
- Fabrication additive – Impression 3D

ENVIRONNEMENT – SÉCURITÉ

- Sécurité et gestion des risques
- Environnement
- Génie écologique
- Technologies de l'eau
- Bruit et vibrations
- Métier : Responsable risque chimique
- Métier : Responsable environnement

ÉNERGIES

- Hydrogène
- Ressources énergétiques et stockage
- Froid industriel
- Physique énergétique
- Thermique industrielle
- Génie nucléaire
- Conversion de l'énergie électrique
- Réseaux électriques et applications

GÉNIE INDUSTRIEL

- Industrie du futur
- Management industriel
- Conception et production
- Logistique
- Métier : Responsable qualité
- Emballages
- Maintenance
- Traçabilité
- Métier : Responsable bureau d'étude / conception

ÉLECTRONIQUE – PHOTONIQUE

- Électronique
- Technologies radars et applications
- Optique – Photonique

TECHNOLOGIES DE L'INFORMATION

- Sécurité des systèmes d'information
- Réseaux Télécommunications
- Le traitement du signal et ses applications
- Technologies logicielles – Architectures des systèmes
- Sécurité des systèmes d'information

AUTOMATIQUE – ROBOTIQUE

- Automatique et ingénierie système
- Robotique

INGÉNIERIE DES TRANSPORTS

- Véhicule et mobilité du futur
- Systèmes aéronautiques et spatiaux
- Systèmes ferroviaires
- Transport fluvial et maritime

MESURES – ANALYSES

- Instrumentation et méthodes de mesure
- Mesures et tests électroniques
- Mesures mécaniques et dimensionnelles
- Qualité et sécurité au laboratoire
- Mesures physiques
- Techniques d'analyse
- Contrôle non destructif

PROCÉDÉS CHIMIE – BIO – AGRO

- Formulation
- Bioprocédés et bioproductions
- Chimie verte
- Opérations unitaires. Génie de la réaction chimique
- Agroalimentaire

SCIENCES FONDAMENTALES

- Mathématiques
- Physique Chimie
- Constantes physico-chimiques
- Caractérisation et propriétés de la matière

BIOMÉDICAL – PHARMA

- Technologies biomédicales
- Médicaments et produits pharmaceutiques

CONSTRUCTION ET TRAVAUX PUBLICS

- Droit et organisation générale de la construction
- La construction responsable
- Les superstructures du bâtiment
- Le second œuvre et l'équipement du bâtiment
- Vieillessement, pathologies et réhabilitation du bâtiment
- Travaux publics et infrastructures
- Mécanique des sols et géotechnique
- Préparer la construction
- L'enveloppe du bâtiment
- Le second œuvre et les lots techniques

OFFRE



Le traitement du signal et ses applications

Maîtrisez les techniques fondamentales d'analyse et de traitement de données du monde du signal

Ref : TIP383WEB

PRÉSENTATION

Les connaissances indispensables pour **maîtriser les techniques de traitement, d'analyse et d'interprétation des signaux**:
séparation de source, traitements statistiques, analyse temps-fréquence,

Un panorama complet des dernières **technologies de codage, compression et restauration** des images fixes et vidéo,

Une analyse détaillée des dernières technologies de **radiodiffusion, télévision numérique et terrestre, radiolocalisation**,

La description des **techniques d'enregistrement et de restitution du son et de l'image**, et les évolutions vers les services multimédias.

VOTRE COMMANDE :

Référence	Titre de l'ouvrage	Prix unitaire H.T	Qté	Prix total H.T
TIP383WEB	Le traitement du signal et ses applications	1 670 €	1	1 670 €
Total H.T en €				1 670 €
T.V.A : 5,5%				91,85 €
Total TTC en €				1 761,85 €

VOS COORDONNÉES :

Civilité M. Mme

Prénom _____

Nom _____

Fonction _____

E-mail _____

Raison sociale _____

Adresse _____

Code postal _____

Ville _____

Pays _____

Date :

Signature et cachet obligatoire

CONDITIONS GÉNÉRALES DE VENTE

Conditions générales de vente détaillées sur simple demande ou sur www.technique-ingenieur.fr

Si vous n'êtes pas totalement satisfait, vous disposeriez d'un délai de 15 jours à compter de la réception de l'ouvrage pour le retourner à vos frais par voie postale. Livraison sous 30 jours maximum.