

Utilisation de l'entropie

dans les problèmes de restauration

et de reconstruction d'images

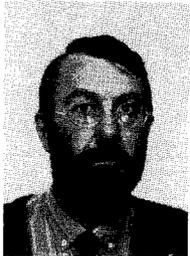
Using entropy in image reconstruction and restoration



Ali MOHAMMAD-DJAFARI

Laboratoire des Signaux et Systèmes (CNRS-ESE-UPS), École Supérieure d'Électricité, Plateau du Moulon, 91190 GIF-SUR-YVETTE et Gréco « Traitement du Signal et Image ».

Ali Mohammad-Djafari est né en Iran en 1952. Il est Ingénieur de l'École Polytechnique de Téhéran (1975), Ingénieur ESE (1977), Docteur-Ingénieur et Docteur ès Sciences Physiques (1987) de l'Université de Paris-Sud. Il travaille depuis 1977 dans le Laboratoire des Signaux et Systèmes au sein du groupe « Problèmes Inverses en Traitement du Signal et Imagerie ». Chargé de Recherche au CNRS, il s'intéresse à la résolution des problèmes inverses et plus particulièrement aux problèmes de reconstruction et restauration d'images en tomographie.



Guy DEMOMENT

Laboratoire des Signaux et Systèmes (CNRS-ESE-UPS), École Supérieure d'Électricité, Plateau du Moulon, 91190 GIF-SUR-YVETTE et Gréco « Traitement du Signal et Image ».

Guy Demoment est Ingénieur ESE, Docteur-Ingénieur, Docteur ès Sciences. Chargé de Recherche au CNRS et professeur de probabilités et théorie du signal à l'École Supérieure d'Électricité, il est également responsable du groupe « Problèmes Inverses en Traitement du Signal et Imagerie » au Laboratoire des Signaux et Systèmes (CNRS-ESE-UPS).

RÉSUMÉ

Dans un très grand nombre de problèmes de restauration et de reconstruction d'images on est amené à résoudre une équation intégrale de première espèce, ce qui est un problème inverse mal posé. Dans ces problèmes, l'obtention d'une solution unique et stable vis-à-vis des erreurs sur les données passe par l'introduction d'une information *a priori* sur la solution.

L'approche bayésienne est une approche cohérente pour la résolution d'un problème inverse car elle permet de prendre en compte et de traiter de la même manière l'information *a priori* sur la solution et celle sur les données. Cette approche peut se résumer aux étapes suivantes :

- (i) Attribuer une distribution de probabilité *a priori* aux paramètres à estimer pour traduire notre connaissance initiale sur ces paramètres.
- (ii) Attribuer une distribution de probabilité aux grandeurs mesurées pour traduire l'imprécision sur ces données (bruit de mesure).
- (iii) Utiliser la règle de Bayes pour transmettre l'information contenue dans les données aux paramètres. Autrement dit, calculer la distribution de probabilité *a posteriori* des paramètres.
- (iv) Définir une règle de décision pour déterminer les valeurs des paramètres à estimer.

Il faut noter cependant que :

- (i) Cette approche ne peut être utilisée que dans un problème qui est décrit par un nombre fini de paramètres (par exemple une fois que le problème a été discrétisé).
- (ii) La notion de probabilité dans cette approche n'est pas forcément liée à la fréquence de réalisation d'une variable aléatoire.
- (iii) Autant il est facile d'attribuer une distribution de probabilité aux grandeurs mesurées pour traduire l'existence du bruit sur ces grandeurs, autant il est plus difficile d'attribuer une distribution de probabilité *a priori* aux paramètres inconnus du problème. Le principe du maximum d'entropie permet de choisir une distribution de probabilité qui soit cohérente avec notre connaissance *a priori* sur les paramètres à estimer, et qui soit la moins compromettante, dans le sens où elle n'introduit pas d'information supplémentaire.

Dans cette communication nous allons utiliser cette approche pour établir une méthode de résolution d'équations intégrales de première espèce dans laquelle l'entropie de la solution joue le rôle d'une fonctionnelle de régularisation.

La méthode est ensuite utilisée pour la résolution de plusieurs problèmes inverses : la restauration d'images positives par déconvolution dans des situations de données manquantes, la reconstruction d'images en tomographie à rayons X et à ondes diffractées et la synthèse de Fourier multivariable. De nombreux résultats de simulation sont présentés et une comparaison est faite entre ces résultats et ceux que l'on obtient par des méthodes linéaires usuelles.

MOTS CLÉS

Maximum d'entropie, reconstruction d'image, déconvolution, problèmes inverses, statistique bayésienne.

SUMMARY

In a great number of image reconstruction and restoration problems we have to solve an integral equation of the first kind which is an ill-posed inverse problem. Therefore one cannot obtain a unique and stable solution without introducing an a priori information on the solution.

The Bayesian approach is a coherent one for solving inverse problems because it lets us to take into account and to process in the same way the a priori information on the solution and the data. This approach can be resumed as the following:

- (i) *Assign an a priori probability distribution to the parameters to translate our knowledge on these parameters.*
- (ii) *Assign a probability distribution to the measured data to translate the errors and the noise on the data.*
- (iii) *Use the Bayes' rule to transmit the information contained in the data to the parameters, i.e. calculate the a posteriori probability distribution of these parameters.*
- (iv) *Define a decision rule to determine the parameters values.*

One must note that:

- (i) *This approach can be used only in problems which can be described by a finite number of parameters (for example when the integral equation is discretized).*
- (ii) *The notion of probability in this approach is not always connected to the frequency of the realization of a random variable.*
- (iii) *While it is easy to assign a probability distribution to the measured data to translate the existence of noise on these data, it is more difficult to assign an a priori probability distribution on the unknown parameters of the problem. The maximum entropy principle permits us to choose a probability distribution which is coherent with our a priori knowledge on these parameters and which is less compromising in the sense that it does not introduce any supplementary information.*

In this paper we use this approach to establish a method for solving the integral equations of the first kind in which the entropy of the solution is used as a regularization functional.

This method is then used for solving many inverse problems: image restoration by deconvolution in the situation of missing data, image reconstruction in X ray tomography and diffraction tomography, and the multivariable Fourier synthesis problem. A great number of simulation results are showed and a comparison is made between these results and those obtained by other usual linear methods.

KEY WORDS

Maximum entropy, image reconstruction, deconvolution, inverse problems, Bayesian statistics.

1. Introduction

Dans les problèmes de restauration et de reconstruction d'images, on est souvent amené à résoudre une équation intégrale de première espèce de la forme :

$$(1) \quad g(x, y) = \iint_D f(x', y') h(x, x', y, y') dx' dy'$$

où g désigne l'image observée, f l'objet à restaurer ou à reconstruire, et h le noyau de la transformation qui les lie, exponentielle complexe par exemple [13]. L'inversion de la relation (1) est en général un problème *mal-posé* car il ne satisfait pas les trois conditions d'existence, d'unicité et surtout de stabilité [15]. Une manière de transformer le problème en un problème *bien-posé* consiste à le régulariser, c'est-à-dire à introduire une information *a priori* sur la solution.

Cette information *a priori* peut nous être donnée soit sous une forme déterministe (limitation du support, positivité, ...), soit sous une forme stochastique (loi de probabilité de l'image ou plutôt contraintes sur cette loi de probabilité).

La résolution numérique de tels problèmes passe par une étape de discrétisation qui peut être faite par une méthode de quadrature. On doit alors résoudre un système d'équations linéaires de la forme :

$$(2) \quad \mathbf{y} = \mathbf{A} \mathbf{x} + \mathbf{b}$$

où \mathbf{x} est un vecteur contenant l'ensemble des paramètres inconnus du problème, \mathbf{y} est un vecteur contenant l'ensemble des observations, \mathbf{A} est une matrice connue et \mathbf{b} est un vecteur contenant les termes d'incertitude correspondant au bruit sur les observations et aux erreurs de quadrature. Ce système d'équations linéaires est en général mal-conditionné, voire même singu-

lier. Ceci est la conséquence du caractère mal-posé du problème initial. Le problème est alors d'obtenir une solution unique et acceptable pour ce système d'équations linéaires, en exploitant l'information *a priori* dont on dispose sur la solution. Nous verrons dans la suite de l'article que cette information *a priori* peut être relativement grossière et se limitera à la positivité de la solution ($x_i > 0$) et au choix d'un paramètre de réglage important analogue à un coefficient de régularisation.

L'existence des erreurs et du bruit de mesure conduit souvent à adopter une approche stochastique. L'approche bayésienne est alors une approche cohérente pour la résolution d'un problème inverse car elle permet de prendre en compte et de traiter de la même manière l'information *a priori* sur la solution et celle sur les données. L'approche bayésienne de la résolution d'un problème inverse peut se résumer aux étapes suivantes :

- (i) Attribuer une distribution de probabilité *a priori* aux paramètres à estimer pour traduire notre connaissance initiale sur ces paramètres.
- (ii) Attribuer une distribution de probabilité aux grandeurs mesurées pour traduire l'imprécision sur ces données (bruit de mesure).
- (iii) Utiliser la règle de Bayes pour transmettre l'information contenue dans les données aux paramètres. Autrement dit, calculer la distribution de probabilité *a posteriori* des paramètres.
- (iv) Définir une règle de décision pour déterminer les valeurs des paramètres, et qui peut être, par exemple, le maximum *a posteriori*.

Il faut noter cependant que :

- (i) Cette approche ne peut être utilisée que dans un problème qui est décrit par un nombre fini de paramètres (par exemple en imagerie une fois que le problème a été discrétisé et l'image réduite à un nombre fini de pixels).
- (ii) La notion de probabilité dans cette approche n'est pas forcément liée à la fréquence de réalisation d'une variable aléatoire.
- (iii) Autant il est facile d'attribuer une distribution de probabilité aux grandeurs mesurées pour traduire l'existence du bruit sur ces grandeurs, autant il est plus difficile d'attribuer une distribution de probabilité *a priori* aux paramètres inconnus du problème. C'est exactement ici que va intervenir le principe du maximum d'entropie qui permet de choisir une distribution de probabilité qui soit cohérente avec notre connaissance sur les paramètres à estimer, et qui soit la moins compromettante, dans le sens où elle est la moins informative.

Nous allons utiliser cette approche pour décrire une méthode de résolution d'équations intégrales de première espèce dans laquelle l'entropie de la solution joue le rôle d'une fonctionnelle de régularisation. L'article est organisé de la manière suivante : le chapitre 2 rappelle les principes de l'approche bayésienne, et le chapitre 3 montre l'intérêt de l'entropie pour la détermination des lois *a priori*. Le chapitre 4 décrit en détail l'application de ces deux principes à la résolution de notre problème. Après une description des problèmes algorithmiques au chapitre 5, la

méthode est utilisée au chapitre 6 pour la résolution de plusieurs problèmes inverses : la restauration d'images positives par déconvolution dans des situations de données manquantes, la reconstruction d'images en tomographie à rayons X et à ondes diffractées, et la synthèse de Fourier multivariable. Des résultats de simulation sont présentés et une comparaison est faite entre ces résultats et ceux que l'on obtient par des méthodes linéaires.

2. Approche bayésienne de la résolution d'un problème inverse

Dans l'approche bayésienne de la résolution d'un problème tel que celui décrit par l'équation (2), on suppose que l'on peut attribuer une loi *a priori* aux paramètres x et aux termes de bruit b . On peut de plus supposer que la matrice A n'est pas parfaitement connue et traduire ainsi le caractère partiel de la connaissance sur la relation théorique entre x et y par une loi conjointe sur x et y .

Notons par $p(x)$ la loi *a priori* des paramètres inconnus x , et par $p(y|x)$ la loi conditionnelle de y connaissant x , c'est-à-dire la loi de b . On peut alors appliquer la règle de Bayes pour obtenir la loi *a posteriori* $p(x|y)$ par :

$$(3) \quad p(x|y) = p(x) \frac{p(y|x)}{p(y)}$$

où

$$p(y) = \int p(y|x) p(x) dx$$

Le problème est ainsi formellement résolu car la loi *a posteriori* $p(x|y)$ définit entièrement la solution dans cette approche bayésienne.

Mais dans une application réelle deux difficultés majeures se présentent :

- (i) Comment déterminer ces lois *a priori* à partir d'une connaissance du problème qui ne nous donne pas ces lois d'une façon explicite? Cette connaissance est souvent donnée sous la forme de contraintes sur ces lois.
- (ii) Comment utiliser et interpréter la loi *a posteriori* résultante pour attribuer des valeurs numériques aux paramètres inconnus x et résoudre ainsi au sens usuel un problème inverse?

La deuxième difficulté est bien maîtrisée car assez familière. Il suffit de choisir une règle de décision pour effectuer une estimation ponctuelle des paramètres à partir de la loi *a posteriori* (par exemple le maximum *a posteriori* ou la moyenne *a posteriori*).

En ce qui concerne la première difficulté, on peut utiliser le principe du maximum d'entropie (ME) pour attribuer d'une façon unique une loi de probabilité *a priori* à partir des connaissances incomplètes dont l'on dispose. En effet, les contraintes sur la loi sont en général insuffisantes pour la définir d'une façon unique. C'est dans ce contexte que l'on peut utiliser le principe du ME pour choisir parmi toutes les lois

qui satisfont ces contraintes celle qui a l'entropie maximale. On est ainsi sûr d'avoir choisi la loi la moins compromettante dans le sens qu'elle est la moins informative (au sens de la théorie de l'information).

3. Probabilités *a priori* et maximum d'entropie

3.1. DÉFINITION DE L'ENTROPIE

Le principe du ME peut être approché de différentes manières. L'approche de la théorie de l'information est sans doute la mieux adaptée à notre problème. Jaynes [6-8] est parmi les premiers auteurs modernes à avoir introduit le formalisme du ME par l'approche de la théorie de l'information. Cette notion d'entropie est introduite de la manière suivante : considérons un processus aléatoire x produisant des réalisations $\{x_1, x_2, \dots, x_n\}$ et attribuons les probabilités $\{p_1, p_2, \dots, p_n\}$ à ces réalisations pour représenter notre information partielle sur ce processus. On définit la quantité $I_i = \log(1/p_i)$ comme la quantité d'information obtenue par la réalisation de x_i . Le raisonnement intuitif derrière cette expression est que plus un événement est rare, plus grand est le gain d'information obtenu par sa réalisation. L'utilisation du logarithme rend additif le gain total d'information obtenu par la réalisation de plusieurs événements indépendants. On définit alors l'entropie d'un processus par la somme pondérée des informations individuelles de chaque réalisation. C'est la définition de l'entropie donnée par Shannon :

$$(4) \quad H = \sum_i p_i \log \frac{1}{p_i} = - \sum_i p_i \log p_i$$

H est une mesure d'incertitude de la distribution $\{p_1, p_2, \dots, p_n\}$, déterminée uniquement par certaines règles élémentaires de cohérence logique et d'additivité [6-8].

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

$$(5) \quad H(p, q) = \sum_i p_i \log \frac{p_i}{q_i}$$

appelée entropie croisée de la distribution p_i par rapport à la distribution q_i . Il est clair que la minimisation de l'entropie croisée se réduit à la maximisation de l'entropie si l'*a priori* q_i est uniforme.

On peut généraliser ce qui vient d'être écrit au cas de distributions continues (cas continu), et on définit l'entropie par :

$$(6) \quad H(p) = - \int p(\mathbf{x}) \log p(\mathbf{x}) d\mathbf{x}$$

et l'entropie croisée par :

$$(7) \quad H(p, q) = \int p(\mathbf{x}) \log \frac{p(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x}$$

3.2. LOIS A ENTROPIE MAXIMALE

Voyons maintenant ce que signifie de choisir une distribution de probabilité qui a l'entropie maximale et qui contient une information *a priori*, ou qui est cohérente avec des contraintes connues sur cette distribution. Pour cela, voyons tout d'abord ce que signifie l'information contenue dans une distribution de probabilité $\{p_1, p_2, \dots, p_n\}$. A l'évidence il faut que l'on puisse extraire cette information de cette distribution. Supposons que l'on dispose d'une distribution de probabilité $\mathbf{p} = \{p_1, p_2, \dots, p_n\}$ des événements $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ et qu'on nous demande quelle est la meilleure estimée \hat{A} d'une fonction $A(x)$ au sens du minimum de l'erreur quadratique moyenne (EQM). La solution est immédiate :

$$(8) \quad \hat{A} = E[A] = \sum_i p_i A(x_i)$$

Inversement, si on se pose la question d'ajuster une distribution \mathbf{p} pour incorporer une information donnée sur la fonction $A(x)$, il faut comprendre par ceci « connaissant une règle d'estimation précise, par exemple $E[A]$, comment choisir p pour que l'on ait $\hat{A} = E[A]$? ». La réponse est qu'il existe, en général, beaucoup de distributions qui satisfont cette contrainte. Le principe du maximum d'entropie nous permet alors d'en choisir une.

Considérons maintenant les m fonctions $\{A_1(x), A_2(x), \dots, A_m(x)\}$ pour lesquelles nous avons un ensemble de données $\{d_1, d_2, \dots, d_m\}$ qui peuvent s'exprimer sous la forme des m contraintes simultanées suivantes :

$$(9) \quad \sum_i p_i A_k(x_i) = d_k, \quad k = 1, \dots, m$$

Dans chaque problème, ces données $\{d_1, d_2, \dots, d_m\}$ peuvent avoir des interprétations physiques différentes et la difficulté consiste à incorporer ces données (contraintes) dans notre distribution de probabilité. Nous voulons ajuster la distribution de probabilité $\{p_1, p_2, \dots, p_n\}$ à nos données. La solution est obtenue par une technique variationnelle de multiplicateurs de Lagrange :

$$(10) \quad p_i = \frac{1}{Z} \exp \left[\sum_k \lambda_k A_k(x_i) \right], \quad i = 1, \dots, n$$

où

$$(11) \quad Z(\lambda_1, \dots, \lambda_m) = \sum_i \exp \left[- \sum_k \lambda_k A_k(x_i) \right]$$

est la fonction de partition, et les $\{\lambda_1, \lambda_2, \dots, \lambda_m\}$ sont déterminés par le système d'équations :

$$(12) \quad d_k = - \frac{\partial}{\partial \lambda_k} \log Z, \quad k = 1, \dots, m$$

avec m données d_k et m inconnues λ_k .

Dans le cas continu, si l'on dispose d'une fonction densité de probabilité (fdp) *a priori* des variables \mathbf{x} , $q(\mathbf{x})$, et d'un nombre fini de contraintes sur la fdp $p(\mathbf{x})$ sous la forme suivante :

$$(13) \quad \int A_k(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} = d_k, \quad k = 1, \dots, m$$

avec la contrainte de normalisation :

$$(14) \quad \int p(\mathbf{x}) d\mathbf{x} = 1$$

la loi *a posteriori* $p(\mathbf{x})$ qui minimise l'entropie croisée $H(p, q)$ sous ces contraintes s'obtient par :

$$(15) \quad p(\mathbf{x}) = q(\mathbf{x}) \exp \left[-\lambda_0 - \sum_{k=1}^m \lambda_k A_k(\mathbf{x}) \right]$$

La valeur minimale $H_{\min}(p, q)$ peut être exprimée en fonction de λ_k par :

$$(16) \quad H_{\min}(p, q) = -\lambda_0 - \sum_{k=1}^m \lambda_k A_k(\mathbf{x})$$

Les coefficients λ_k , $k=0, \dots, m$ sont choisis de façon à satisfaire les contraintes (13) et (14). On remarque que λ_0 est relié à la fonction de partition Z par $Z = \exp[\lambda_0]$, et on a :

$$(17) \quad \lambda_0 = \log \int q(\mathbf{x}) \exp \left[- \sum_{k=1}^m \lambda_k A_k(\mathbf{x}) \right] d\mathbf{x}$$

Les autres coefficients λ_k sont déterminés par le système d'équations :

$$(18) \quad - \frac{\partial}{\partial \lambda_k} \lambda_0(\lambda_1, \dots, \lambda_m) = d_k, \quad k=1, \dots, m$$

Il n'est pas possible d'obtenir en général une relation explicite pour les coefficients λ_k , et on résout le système d'équations (18) numériquement par des méthodes itératives. Cependant, dans certaines situations simples on peut résoudre le problème d'une façon analytique [8, 14]. Par exemple, si $q(\mathbf{x})$ est une exponentielle multivariées de la forme :

$$(19) \quad q(\mathbf{x}) = \prod_{i=1}^n \frac{1}{\alpha_i} \exp \left[- \frac{x_i}{\alpha_i} \right]$$

et si les contraintes sont de la forme :

$$(20) \quad E \{ x_i \} = \int x_i p(x_i) dx_i = \lambda_i, \quad i=1, \dots, n$$

la solution $p(\mathbf{x})$ reste une fonction exponentielle multivariées de la forme :

$$(21) \quad p(\mathbf{x}) = \prod_{i=1}^n \frac{1}{\lambda_i} \exp \left[- \frac{x_i}{\lambda_i} \right]$$

De même si $q(\mathbf{x})$ est une fonction gaussienne multivariées de la forme :

$$(22) \quad q(\mathbf{x}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\alpha_i}} \exp \left[- \frac{1}{2} \left(\frac{x_i - \lambda_i}{\alpha_i} \right)^2 \right]$$

et si les contraintes sont de la forme :

$$(23) \quad E \{ x_i^2 \} = \int x_i^2 p(x_i) dx_i = \sigma_i^2$$

la solution $p(\mathbf{x})$ reste une fonction gaussienne multivariées de la forme :

$$(24) \quad p(\mathbf{x}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_i}} \exp \left[- \frac{1}{2} \left(\frac{x_i - \lambda_i}{\sigma_i} \right)^2 \right]$$

Dans le chapitre suivant nous verrons comment utiliser ce principe dans le cadre des problèmes de restauration et de reconstruction d'images.

4. Application aux problèmes de restauration et de reconstruction d'images

Considérons le problème de la restauration d'images dégradées par un système d'imagerie et appliquons ces principes pour voir comment on aboutit à utiliser l'entropie de l'image comme une fonctionnelle de régularisation.

Supposons que le problème soit déjà discrétisé, ce qui veut dire que dans l'équation (2) $\mathbf{x} = [x_1, \dots, x_n]$ est un vecteur contenant les valeurs moyennes de l'intensité lumineuse dans chaque pixel de l'image, $\mathbf{y} = [y_1, \dots, y_n]$ est un vecteur contenant les observations (les valeurs moyennes des intensités lumineuses mesurées par les capteurs) et \mathbf{A} est une matrice connue et définie entièrement par la réponse impulsionnelle du système d'imagerie. Les observations y_j sont entachées des bruits b_j supposés non corrélés, indépendants des observations et centrés. La seule connaissance que l'on ait réellement sur ces bruits est leurs variances σ_j^2 .

Pour appliquer la règle de Bayes il faut d'abord déterminer la loi *a priori* de l'image et la loi des observations conditionnée par l'image.

4.1. DÉTERMINATION DE LA LOI A PRIORI DE L'IMAGE

C'est la partie la plus délicate car tout dépend de ce que représente x_i . Mais avant de discuter de la nature physique de l'image et des pixels, supposons que les pixels X_i soient des variables aléatoires indépendantes et de mêmes lois. Nous distinguons dans la suite deux situations :

- (i) les X_i sont des variables aléatoires continues (par exemple la variation continue d'un niveau de gris);
- (ii) les X_i sont des variables aléatoires discrètes (par exemple une image binaire ou ternaire, etc.).

Dans chacune de ces situations nous examinons les solutions ME obtenues pour des contraintes de différentes formes.

Notons par $p(x_i)$ la densité de probabilité associée au premier cas et par $\{p(X_i=k) = p_{ik}\}$ la distribution de probabilité associée au deuxième cas. Nous notons alors par H_i l'entropie associée à la densité $p(x_i)$ (ou la distribution p_{ik}). On a alors, pour un pixel :

dans le cas continu

$$(25) \quad H_i = - \int p(x_i) \log p(x_i) dx_i$$

et dans le cas discret

$$(26) \quad H_i = - \sum_k p_k \log p_k$$

Notons aussi par $p(\mathbf{x})$ la densité (distribution) de probabilité de l'image \mathbf{x} . Les pixels étant supposés indépendants on a, pour l'image :

dans le cas continu

$$(27) \quad p(\mathbf{x}) = p(x_1, \dots, x_n) = \prod_{i=1}^n p(x_i)$$

et dans le cas discret

$$(28) \quad p(\mathbf{x}) = p(x_1=k, \dots, x_n=k) = \prod_i p_{ik}$$

L'entropie de l'image $H = H[p(\mathbf{x})]$ est alors donnée par :

$$(29) \quad H = \sum_{i=1}^n H_i$$

4.1.1. Quelques exemples

Examinons maintenant les exemples suivants :

Exemple 1 : Les X_i sont des v.a. continues, leur domaine est \mathbb{R} , et on dispose d'une connaissance *a priori* sur les moments d'ordre un et deux :

$$(30) \quad \begin{aligned} E\{x_i\} &= \int x_i p(x_i) dx_i = \lambda_i \\ E\{x_i^2\} &= \int x_i^2 p(x_i) dx_i = \sigma_i^2 \end{aligned} \quad i=1, \dots, n$$

Maximisant l'entropie H_i sous ces contraintes, on obtient le résultat bien connu d'une distribution gaussienne pour chacun des pixels :

$$(31) \quad p(x_i) = N(\lambda_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{1}{2}\left(\frac{x_i - \lambda_i}{\sigma_i}\right)^2\right]$$

L'entropie maximale correspondante est :

$$(32) \quad H_i = \frac{1}{2} \log(2\pi e \sigma_i^2) = \frac{1}{2} \log 2\pi e + \log \sigma_i$$

e étant la base des logarithmes népériens.

L'image \mathbf{x} a alors la densité $p(\mathbf{x})$ donnée par :

$$(33) \quad p(\mathbf{x}) = N(\mathbf{m}, \mathbf{R})$$

avec

$$\mathbf{m} = [\lambda_1, \dots, \lambda_n]^T \quad \text{et} \quad \mathbf{R} = \text{diag}[\sigma_1^2, \dots, \sigma_n^2]$$

et l'entropie maximale de toute l'image est de la forme :

$$(34) \quad H = \sum_{i=1}^n H_i = \sum_{i=1}^n \log \sigma_i + \frac{n}{2} \log(2\pi e)$$

On retrouve une des deux expressions classique de l'entropie à un terme constant près.

Exemple 2 : Les X_i sont des v.a. continues, leur domaine est \mathbb{R} , et on dispose d'une connaissance *a priori* sur les moments d'ordre un seulement :

$$(35) \quad E\{x_i\} = \int x_i p(x_i) dx_i = \lambda_i, \quad i=1, \dots, n$$

La densité $p(x_i)$ qui maximise l'entropie H_i a cette fois la forme suivante :

$$(36) \quad p(x_i) = \frac{1}{\lambda_i + 1} \left(\frac{\lambda_i}{\lambda_i + 1}\right)^{x_i}$$

et son entropie maximale est :

$$(37) \quad H_i = (\lambda_i + 1) \log(\lambda_i + 1) - \lambda_i \log \lambda_i$$

L'image \mathbf{x} a alors la densité $p(\mathbf{x})$ donnée par :

$$(38) \quad p(\mathbf{x}) = \prod_{i=1}^n \frac{1}{\lambda_i + 1} \left(\frac{\lambda_i}{\lambda_i + 1}\right)^{x_i}$$

et l'entropie maximale de l'image est de la forme :

$$(39) \quad H = \sum_{i=1}^n H_i = \sum_{i=1}^n [(\lambda_i + 1) \log(\lambda_i + 1) - \lambda_i \log \lambda_i]$$

Deux approximations sont intéressantes :

si $\lambda_i \gg 1$

$$(40) \quad \text{alors} \quad H_i \approx \log \lambda_i \quad \text{et} \quad H \approx \sum_{i=1}^n \log \lambda_i$$

si $\lambda_i \ll 1$

$$(41) \quad \text{alors} \quad H_i \approx -\lambda_i \log \lambda_i \quad \text{et} \quad H \approx \sum_{i=1}^n -\lambda_i \log \lambda_i$$

Ici aussi on retrouve les deux expressions classiques de l'entropie.

Exemple 3 : Les X_i sont des v.a. continues, leur domaine est \mathbb{R}^+ (les valeurs des pixels sont positives). Si l'on dispose d'une connaissance *a priori* globale limitée à des moments d'ordre un de la forme :

$$(42) \quad E\{g_i(\mathbf{x})\} = \int g_i(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} = s_i, \quad i=1, \dots, k$$

alors la loi *a priori* d'entropie maximale est de la forme suivante :

$$(43) \quad p(\mathbf{x}) = \exp\left[\lambda_0 + \sum_{i=1}^k \lambda_i g_i(\mathbf{x})\right]$$

Lorsque cette information *a priori* est locale, c'est-à-dire lorsque par exemple $g_i(\mathbf{x}) = x_i$, la contrainte s'écrit :

$$(44) \quad E\{x_i\} = \int x_i p(x_i) dx_i = \lambda_i, \quad i=1, \dots, n$$

La densité $p(x_i)$ qui maximise l'entropie H_i sous cette contrainte prend alors la forme classique :

$$(45) \quad p(x_i) = \frac{1}{\lambda_i} \exp \left[-\frac{x_i}{\lambda_i} \right]$$

et son entropie maximale est :

$$(46) \quad H_i = 1 + \log \lambda_i$$

L'image \mathbf{x} a alors la densité $p(\mathbf{x})$ donnée par :

$$(47) \quad p(\mathbf{x}) = \prod_{i=1}^n \frac{1}{\lambda_i} \exp \left[-\frac{x_i}{\lambda_i} \right] \\ = \frac{1}{\lambda} \left[-\frac{1}{\lambda} \sum_{i=1}^n \lambda_i x_i \right] \quad \text{avec} \quad \lambda = \prod_{i=1}^n \lambda_i$$

et l'entropie maximale de l'image est de la forme :

$$(48) \quad H = \sum_{i=1}^n (1 + \log \lambda_i) = n + \sum_{i=1}^n \log \lambda_i$$

Exemple 4 : Les X_i sont des v. a. discrètes, leur domaine est \mathbb{N} , c'est-à-dire que les X_i peuvent prendre des valeurs discrètes $\{0, 1, 2, \dots\}$, et on dispose d'une connaissance *a priori* sous la forme :

$$(49) \quad E\{x_i\} = \sum_{k=0}^{\infty} k p(x_i=k) = \sum_{k=0}^{\infty} k p_k = \lambda_i \\ i=1, \dots, n$$

Dans ce cas aussi la distribution p_k qui maximise l'entropie H_i sous cette contrainte est de la forme [20] :

$$(50) \quad p_k = p(X_i=k) = [1 + \lambda_i]^{-(\lambda_i+1)} \lambda_i^k$$

et son entropie maximale est :

$$(51) \quad H_i = \sum_{k=0}^{\infty} [1 + \lambda_i]^{-(\lambda_i+1)} \lambda_i^k \log([1 + \lambda_i]^{-(\lambda_i+1)} \lambda_i^k) \\ = (1 + \lambda_i) \log(1 + \lambda_i) - \lambda_i \log \lambda_i$$

L'image \mathbf{x} a alors la densité $p(\mathbf{x})$ donnée par :

$$(52) \quad p(\mathbf{x}) = \prod_{i=1}^n [1 + \lambda_i]^{-(\lambda_i+1)} \lambda_i^{x_i}$$

et l'entropie maximale de l'image est de la forme :

$$(53) \quad H = \sum_{i=1}^n [(1 + \lambda_i) \log(1 + \lambda_i) - \lambda_i \log \lambda_i]$$

Ici aussi deux situations sont intéressantes :

si $\lambda_i \gg 1$ alors

$$(54) \quad H_i \approx \log \lambda_i \quad \text{et} \quad H \approx \sum_{i=1}^n \log \lambda_i$$

si $\lambda_i \ll 1$ alors

$$(55) \quad H_i \approx -\lambda_i \log \lambda_i \quad \text{et} \quad H \approx \sum_{i=1}^n -\lambda_i \log \lambda_i$$

4.1.2. Choix d'une loi a priori pour notre problème

Dans notre problème, les valeurs des pixels X_i sont considérées comme des v. a. continues définies sur \mathbb{R}^+ . Nous choisissons dans toute la suite une seule contrainte globale de la forme :

$$(56) \quad E\{g(\mathbf{x})\} = s$$

avec

$$(57) \quad g(\mathbf{x}) = - \sum_{i=1}^n x_i \log x_i = S(\mathbf{x})$$

ce qui conduit à la loi *a priori* :

$$(58) \quad p(\mathbf{x}) = \exp[\lambda_0 + \lambda_1 S(\mathbf{x})] = \text{Cte.} \exp[\lambda_1 S(\mathbf{x})]$$

La quantité $S(\mathbf{x})$ est souvent appelée « entropie structurelle » de l'image et se confond avec l'entropie de Shannon (4) lorsque l'image, préalablement normalisée, est assimilée à une distribution de probabilité (ce qui n'est pas le cas ici). Le choix d'une telle contrainte globale (puisque'elle porte sur une somme étendue à toute l'image) se justifie par un calcul combinatoire et par un passage à la limite en utilisant la formule de Stirling [8, p. 37] et [7, p. 951]. En effet, si on note $p(\mathbf{x})$ la multiplicité de l'image, c'est-à-dire la probabilité d'obtenir une image $\mathbf{x} = [x_1, \dots, x_n]$ si les N particules élémentaires de l'image sont envoyées d'une manière aléatoire sur la surface de l'image, on obtient :

$$(59) \quad p(\mathbf{x}) = \frac{N!}{\prod_i x_i!}$$

Cette expression tend vers $\exp[NS(\mathbf{x})]$ quand N tend vers l'infini. La loi *a priori* résultante est donc la loi d'entropie maximale pour une image positive sous la seule contrainte de la connaissance de son entropie structurelle moyenne $E\{S(\mathbf{x})\}$. Nous verrons au paragraphe 4.4 que cette donnée $s = E\{S(\mathbf{x})\}$ est équivalente à celle d'un coefficient de régularisation λ .

4.2. DÉTERMINATION DE LA LOI DES OBSERVATIONS CONDITIONNÉE PAR L'IMAGE

La loi conditionnelle $p(\mathbf{y}|\mathbf{x})$ est, en raison du modèle (2), la loi du bruit \mathbf{b} à un décalage près. Il suffit donc de déterminer cette loi.

Tout d'abord, l'application du principe du ME au bruit b_j avec la seule connaissance de la variance σ_j^2 nous préconise une loi gaussienne $p(b_j) = N(0, \sigma_j^2)$ pour ce bruit. De plus, si on suppose que les b_j sont indépendants entre eux, on en déduit que le vecteur $\mathbf{b} = [b_1, \dots, b_m]$ a aussi une loi gaussienne avec une matrice de covariance diagonale $\mathbf{R} = \text{diag}[\sigma_1^2, \dots, \sigma_m^2]$. On en déduit donc immédiatement :

$$p(\mathbf{y}|\mathbf{x}) = \text{Cte.} \exp[-Q_1(\mathbf{x})] \\ \text{avec}$$

$$(60) \quad Q_1(\mathbf{x}) = \sum_{j=1}^m \frac{1}{\sigma_j^2} |y_j - \hat{y}_j|^2$$

où

$$\hat{y}_j = \sum_{i=1}^n A_{ji} x_i$$

Si, de plus, les variances sont égales $\sigma_1 = \sigma_2 = \dots = \sigma_m = \sigma$, on obtient :

$$(61) \quad p(\mathbf{y} | \mathbf{x}) = \text{Cte.} \exp \left[-\frac{Q(\mathbf{x})}{\sigma^2} \right]$$

avec

$$Q(\mathbf{x}) = [\mathbf{y} - \mathbf{A} \mathbf{x}]^T [\mathbf{y} - \mathbf{A} \mathbf{x}]$$

4.3. DÉTERMINATION DE LA LOI A POSTERIORI DE L'IMAGE

A ce stade, il suffit de remplacer les expressions (58) et (61) dans l'expression de la règle de Bayes (3) pour obtenir :

$$(62) \quad p(\mathbf{x} | \mathbf{y}) = \text{Cte.} \exp \left(\lambda_1 S(\mathbf{x}) - \frac{Q(\mathbf{x})}{\sigma^2} \right)$$

4.4. DÉTERMINATION DE LA SOLUTION

Si l'on adopte comme critère de décision le maximum *a posteriori*, on obtient :

$$(63) \quad \hat{\mathbf{x}} = \underset{\mathbf{x}}{\text{Arg max}} p(\mathbf{x} | \mathbf{y}) \\ = \underset{\mathbf{x}}{\text{Arg max}} \left[\lambda_1 S(\mathbf{x}) - \frac{Q(\mathbf{x})}{\sigma^2} \right] \\ = \underset{\mathbf{x}}{\text{Arg min}} [Q(\mathbf{x}) - \lambda S(\mathbf{x})]$$

où $\lambda = \lambda_1 \sigma^2$. On est ainsi amené à minimiser un critère de la forme :

$$(64) \quad J(\mathbf{x}) = Q(\mathbf{x}) - \lambda S(\mathbf{x})$$

On voit donc que l'on se retrouve dans le schéma classique de la régularisation d'un problème mal-posé [15, 21]. Comme la solution inverse généralisée $\hat{\mathbf{x}}_0 = \mathbf{A}^+ \mathbf{y}$ construite à partir des seules données en minimisant la distance $\Delta_1(\mathbf{x}, \hat{\mathbf{x}}_0) = Q(\mathbf{x})$ est inacceptable en raison d'une variance excessive, on choisit d'introduire une information *a priori* pour stabiliser le problème, et on minimise un critère composite :

$$(65) \quad \hat{\mathbf{x}}(\lambda, \mathbf{y}) = \underset{\mathbf{x}}{\text{Arg min}} [\Delta_1(\mathbf{x}, \hat{\mathbf{x}}_0) + \lambda \Delta_2(\mathbf{x}, \hat{\mathbf{x}}_\infty)] \\ \lambda \in \mathbb{R}^+$$

dans lequel $\Delta_2(\mathbf{x}, \hat{\mathbf{x}}_\infty) = -S(\mathbf{x})$, opposée de l'entropie, est une distance de Kullback, et $\hat{\mathbf{x}}_\infty$ la solution *a priori*, uniforme dans notre cas. Les indices 0 et ∞ traduisent simplement le fait que les solutions correspondantes sont obtenues par (65) pour les valeurs extrêmes correspondantes de λ .

Lorsque l'on choisit de régulariser un problème mal-posé, deux questions se posent immédiatement : régulariser de quelle manière, et régulariser jusqu'à quel point ? La première question est relative au choix de Δ_1 et Δ_2 , c'est-à-dire au choix d'un chemin qui, dans l'espace des paramètres, relie la solution $\hat{\mathbf{x}}_0$ entière-

ment gouvernée par les données à la solution $\hat{\mathbf{x}}_\infty$ entièrement gouvernée par l'*a priori*. Cette question est donc ici résolue par notre approche bayésienne et par notre utilisation du principe du maximum d'entropie. La deuxième question est relative au choix de λ , c'est-à-dire au compromis entre la fidélité de la solution vis-à-vis des données (mesurée par Δ_1) et sa fidélité à l'*a priori* (mesurée par Δ_2). Ce problème est encore actuellement ouvert, bien que plusieurs méthodes aient été proposées [21]. Il est clair que la valeur à attribuer à λ doit être déterminée, comme celles de l'objet \mathbf{x} , à partir des données elles-mêmes.

La méthode la plus naturelle consiste à rechercher une certaine adéquation du modèle aux données. En effet, minimiser $Q(\mathbf{x}) - \lambda S(\mathbf{x})$ est équivalent à minimiser $S(\mathbf{x})$ sous la contrainte que $Q(\mathbf{x}) \leq c$. C'est là un autre aspect de la régularisation; on admet en régularisant qu'il soit impossible d'atteindre la solution exacte, entre autres à cause du bruit d'observation, et on recherche la « meilleure » solution (au sens du critère Δ_2) dans un ensemble de solution « acceptables » (au sens du critère Δ_1). Or c est une statistique et possède une loi qui découle directement de la loi conditionnelle $p(\mathbf{y} | \mathbf{x})$. Ainsi, dans notre exemple et compte tenu des hypothèses faites sur le bruit d'observation, c suit une loi $\chi^2(n)$ à n degrés de liberté, n étant le nombre de pixels de l'objet à reconstruire. Un choix naturel est donc $c = n \sigma^2$. Mais l'expérience montre qu'un tel choix est surrégularisant, la valeur de λ correspondante étant plus élevée que celle qui minimise l'erreur de reconstruction $\Delta_1(\mathbf{x}, \hat{\mathbf{x}}(\lambda, \mathbf{y}))$. Ceci s'explique par le fait que la régularisation a pour effet de « lisser » la solution en la « tirant » de $\hat{\mathbf{x}}_0$ vers $\hat{\mathbf{x}}_\infty$, et, qu'en général, cela s'accompagne d'un biais qui modifie la loi des erreurs résiduelles sur laquelle était fondé le calcul de la loi de c .

Une autre manière de procéder consiste à considérer $\Delta_1(\mathbf{x}, \hat{\mathbf{x}}(\lambda, \mathbf{y}))$ comme une fonction de perte à laquelle on peut associer le risque moyen suivant :

$$(66) \quad E \{ \Delta_1(\mathbf{x}, \hat{\mathbf{x}}(\lambda, \mathbf{y})) \} = \int \Delta_1(\mathbf{x}, \lambda, \mathbf{y}) p(\mathbf{y} | \mathbf{x}) d\mathbf{y}$$

Choisir pour λ la valeur qui minimise ce risque moyen est certainement raisonnable, malheureusement ce risque moyen dépend de la solution exacte \mathbf{x} qui est inconnue ! On peut estimer ce risque moyen par des méthodes de validation croisée. Mais ces méthodes semblent limitées à des problèmes dans lesquels la fonctionnelle de régularisation Δ_2 est quadratique [5]. Les deux méthodes précédentes sont déterministes par essence, mais on peut profiter du cadre bayésien dans lequel nous avons plongé le problème initial pour utiliser des méthodes probabilistes. L'objet étant estimé par maximum *a posteriori*, on peut, par exemple, étendre cette technique à l'estimation du coefficient de régularisation :

$$(67) \quad (\hat{\mathbf{x}}, \hat{\lambda}) = \underset{\mathbf{x}, \lambda}{\text{Arg max}} p(\mathbf{x} | \mathbf{y}, \lambda)$$

Mais ce problème est le plus souvent dégénéré, le coefficient λ partant aux frontières de son domaine qui est en général \mathbb{R}^+ .

On peut enfin minimiser un risque pseudo-bayésien, en choisissant pour λ la valeur qui maximise la vraisemblance des observations $p(y/\lambda)$. Ceci revient à considérer les paramètres x comme des paramètres de nuisance pour l'estimation de λ . Mais le calcul de cette vraisemblance nécessite alors d'intégrer ces paramètres hors du problème :

$$(68) \quad p(y/\lambda) = \int p(y/x, \lambda) p(x/\lambda) dx$$

ce qui soulève quelques difficultés de calcul dans notre cas.

Le problème de la détermination du coefficient de régularisation optimal à partir des seules données n'est pas résolu de manière satisfaisante et nous y travaillons actuellement. Dans toute la suite, ce coefficient est choisi empiriquement, ce qui est possible car la fonction de perte $\Delta_1(x, \hat{x}(\lambda, y))$ est évidemment calculable avec les exemples synthétiques que nous avons retenus. L'expérience que nous avons acquise à cette occasion montre que la solution ne dépend pas de manière très sensible de la valeur de λ , puisqu'une variation d'un ordre de grandeur autour de l'optimum ne l'affecte pas trop.

Ce problème étant (mal) réglé, il reste encore à résoudre celui du calcul de la solution elle-même. C'est l'objet du chapitre suivant.

5. Problèmes algorithmiques

Nous avons vu que la conséquence de l'application de la méthode que nous venons de présenter est qu'il faut rechercher le minimum d'un critère de la forme :

$$J(x) = Q(x) - \lambda S(x)$$

où $Q(x)$ est une fonction quadratique en x , mais où $S(x)$ ne l'est pas. On ne peut rechercher le minimum de $J(x)$ que par une méthode itérative.

Parmi ces méthodes, nous considérons celles qui recherchent le minimum par une suite de recherches monodimensionnelles dans l'espace des paramètres. Ces méthodes peuvent être classées selon les différentes informations locales dont l'on peut disposer sur la fonction pour la détermination des directions de recherche. Les méthodes d'ordre zéro, un et deux sont celles qui utilisent respectivement la fonction, la fonction et son gradient, et la fonction, son gradient et sa matrice hessienne des dérivées secondes, pour la détermination de ces directions.

Les méthodes du second ordre (Newton-Raphson par exemple) sont à exclure vu la très grande dimension de x dans des applications en imagerie. Parmi les méthodes du premier ordre (les méthodes du gradient), celles du gradient conjugué (GC) ont une vitesse de convergence asymptotiquement plus grande que les autres méthodes du premier ordre [2, 16].

Nous avons choisi un algorithme du GC bien que l'expression de $J(x)$ ne soit pas quadratique. Il s'agit du module ZXCGR de la bibliothèque mathématique IMSL qui met en œuvre l'algorithme proposé par Powell [16].

L'algorithme du GC a été utilisé par d'autres auteurs dans des problèmes similaires. En particulier on peut citer Wernecke et d'Addario [22] et Burch, Gull et Skilling [1, 19 et 20]. L'algorithme utilisé par les premiers est un algorithme classique du GC. Dans l'algorithme proposé par Burch, Gull et Skilling la recherche de la direction de descente se fait dans un sous-espace engendré par plusieurs vecteurs du gradient plutôt que sur une ligne. Mais le coût de calcul devient alors plus important car dans chaque étape il faut calculer plusieurs vecteurs du gradient.

Dans toute méthode itérative trois difficultés majeures se présentent :

- (i) l'initialisation de l'algorithme;
- (ii) la convergence de l'algorithme;
- (iii) le dépassement du domaine de définition de la solution.

L'initialisation ne pose en général pas de difficulté pratique très sérieuse car il suffit de choisir une solution proche de la solution recherchée qui satisfait les contraintes (cela n'est pas obligatoire mais permet en pratique d'atteindre une solution acceptable avec un nombre plus faible d'itérations). Dans notre algorithme nous initialisons en général par une image calculée par une méthode d'inversion linéaire à laquelle on applique la contrainte de positivité.

En ce qui concerne la deuxième difficulté il n'existe malheureusement d'étude théorique de la convergence en un nombre fini d'itérations que pour les fonctions quadratiques. Dans le cas plus général il faut se contenter d'une convergence asymptotique.

La troisième difficulté apparaît quand par exemple dans notre cas on cherche une solution positive. Il s'agit là d'une minimisation sous contraintes. Une technique simple consiste à appliquer les contraintes à chaque itération. C'est celle que nous avons choisie pour assurer la positivité de la solution à chaque itération.

6. Simulations

Nous avons appliqué notre méthode de régularisation par ME à plusieurs problèmes de restauration et de reconstruction d'images. Nous présentons ici les principaux résultats de simulation.

6.1. DÉCONVOLUTION D'IMAGES POSITIVES AVEC MANQUE DE DONNÉES

Dans ce paragraphe nous présentons plusieurs résultats de l'application de la méthode à la déconvolution d'images positives en insistant sur des situations de données manquantes (c'est par exemple le cas d'un sous-échantillonnage ou d'un mauvais fonctionnement de certains capteurs) où le problème est très indéterminé. Dans de telles situations, les méthodes usuelles de filtrage linéaire (filtre de Kalman, etc.) ne peuvent pas être mises en œuvre facilement. De plus, la prise en compte du caractère positif de la solution n'est pas aisée dans ces méthodes.

Dans ces simulations, nous avons choisi deux objets, les avons convolués avec un noyau de la forme $\sin x/x$,

et avons ajouté un bruit sur les images ainsi obtenues. Ensuite, partant de ces données, nous avons effectué des restaurations par deux méthodes linéaires de filtrage de Kalman qui ont été développées dans notre équipe [17] ainsi que par notre méthode **ME**.

Par la suite nous appellerons :

– **MC** : notre algorithme du ME dans lequel nous avons forcé à zéro le terme d'entropie. Il s'agit en fait d'une méthode itérative dans laquelle on cherche à minimiser un critère de la forme $J(\mathbf{x})=Q(\mathbf{x})$. On peut considérer qu'il s'agit d'une méthode itérative des moindres carrés.

– **MCP** : notre algorithme du ME dans lequel on force à zéro le terme d'entropie mais en appliquant la contrainte de positivité à chaque itération. Il s'agit d'une méthode **MC** avec contrainte de positivité.

– **ME** : notre algorithme du ME en fonctionnement normal.

– **K1** : un filtre de Kalman optimal (voir référence [17]).

– **K2** : un filtre de Kalman sous-optimal (voir référence [17]). Ces deux dernières méthodes peuvent être considérées comme des méthodes des moindres carrés régularisés.

Dans la figure 1 sont présentés :

– **A** : objet, noyau, image et image bruitée avec un rapport signal-à-bruit de 10 dB;

– **B** : les résultats de déconvolution à partir des données sans bruit et

– **C** : ces mêmes résultats obtenus à partir des données bruitées.

La comparaison est donnée pour les méthodes suivantes : (a) méthode **MC**, (b) méthode de filtrage de Kalman optimal (**K1**), (c) méthode de filtrage de Kalman sous-optimal (**K2**) et (d) méthode **ME**.

On peut constater que :

– La méthode **MC** fournit une solution inacceptable en présence des données bruitées, ce qui n'est pas étonnant.

– Les deux méthodes de filtrages de Kalman (**K1** et **K2**) ont amélioré les images tout en filtrant le bruit, mais la résolution spatiale de ces images n'est pas très bonne. Ceci est dû au fait que l'on a fait l'hypothèse que la solution a une loi *a priori* gaussienne.

– La méthode **ME** donne des résultats qui ont à la fois une résolution spatiale et une précision en amplitude nettement supérieures aux autres méthodes. Ceci n'est pas étonnant car, premièrement, dans cette méthode nous avons utilisé une information supplémentaire sur la positivité de la solution, et, deuxièmement, cette méthode choisit la solution qui est à la fois cohérente avec les données et qui maximise l'entropie.

La figure 2 montre d'autres résultats dans des situations de données manquantes. Ici nous avons fait une comparaison entre les trois méthodes **ME**, **MC** et **MCP** pour effectuer des déconvolutions dans des situations où on ne dispose pas de tous les échantillons de l'image. Nous avons simulé 4 cas :

(i) on dispose de l'ensemble des données;

(ii) on ne dispose que de 50% des données (une ligne sur deux);

(iii) on ne dispose que de 25% des données (une ligne sur deux et une colonne sur deux), et, finalement;

(iv) on ne dispose que de 10% des données (une ligne sur trois et une colonne sur trois).

Dans cette figure sont présentées : en haut les images dégradées (avec bruit) et en bas les images restaurées par les méthodes **MC**, **MCP** et **ME** dans les quatre situations décrites plus haut. Sur ces images on peut constater que, même dans une situation où on ne

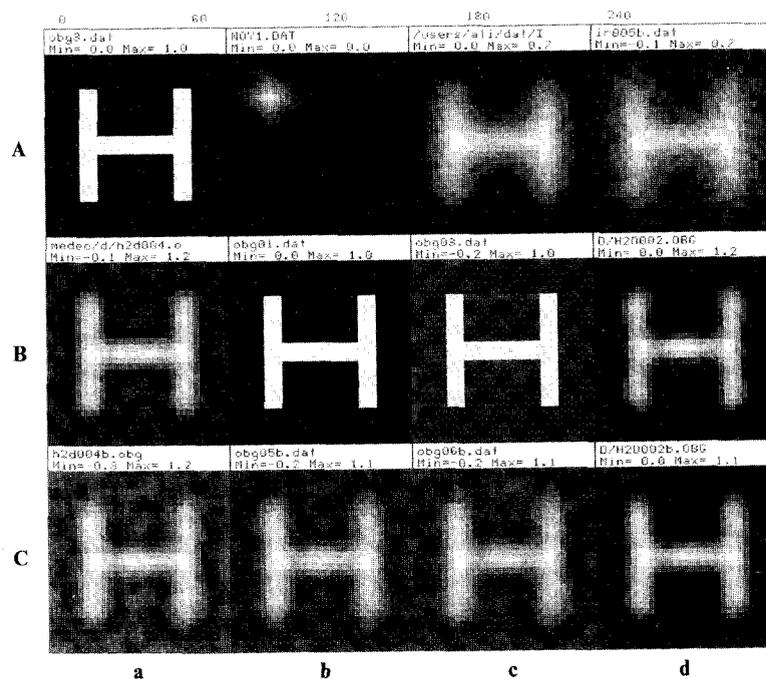


Fig. 1. — Un exemple de déconvolution d'images positives. Comparaison entre les méthodes : **MC** (moindres carrés), **K1** (filtre de Kalman optimal), **K2** (filtre de Kalman sous-optimal) et **ME** (maximum d'entropie). **A** : (a) objet, (b) noyau, (c) image, (d) image bruitée. **B** et **C** : (a) déconvolution par **MC**, (b) déconvolution par **K1**, (c) déconvolution par **K2**, (d) déconvolution par **ME**. **B** : données sans bruit, **C** : données avec bruit.

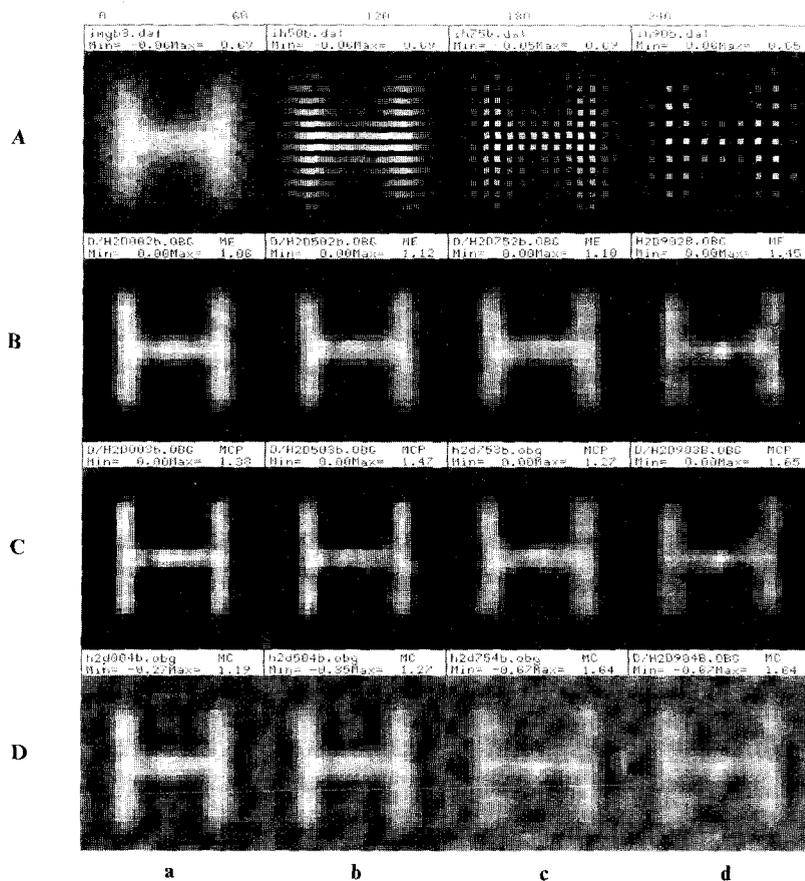


Fig. 2. — Déconvolution en présence de données manquantes : A : images dégradées, B : restauration par ME, C : restauration par MCP, D : restauration par MC, (a) avec toutes les données, (b) avec 50% des données (une ligne sur deux de l'image), (c) avec 25% des données (une ligne et une colonne sur deux de l'image), (d) avec 10% des données (une ligne et une colonne sur trois de l'image).

dispose que de 10% des données, on arrive à la fois à interpoler pour les données manquantes, à améliorer la résolution spatiale et à diminuer le bruit. La méthode MC diverge très rapidement en présence du bruit. La contrainte de positivité dans la méthode MCP assure une meilleure stabilité bien que théoriquement cette contrainte seule ne puisse assurer dans tous les cas la stabilité. Pour les données sans bruit les deux méthodes MCP et ME fournissent des résultats identiques, mais en présence de bruit ME fournit des résultats plus stables.

6. 2. RECONSTRUCTION D'IMAGES EN TOMOGRAPHIE A RAYONS X

Dans ce paragraphe nous appliquons notre méthode de régularisation par maximum d'entropie au problème de la tomographie à rayons X, c'est-à-dire à la détermination d'une fonction $f(x, y)$ à partir d'un nombre limité de ses projections $p(r, \varphi)$. Dans ces simulations, nous avons choisi un fantôme proposé par Shepp et Logan et utilisé couramment dans la littérature sur la tomographie à rayons X [13]. Ensuite, nous avons calculé les projections en simulant les trois cas suivants :

- (1) $Np=12, \theta=180^\circ$,
- (2) $Np=8, \theta=180^\circ$, et
- (3) $Np=5, \theta=90^\circ$

où Np est le nombre des projections et θ est l'angle total balayé entre les projections extrêmes. Nous avons effectué des reconstructions directement à partir de ces projections, par les trois méthodes MC, MCP et ME.

La figure 3 montre les résultats de reconstruction obtenus pour les trois cas cités plus haut, dans les deux situations de données avec et sans bruit. Les conclusions de ces résultats sont les suivantes :

- (i) La méthode de régularisation par maximum d'entropie ME, telle que nous l'avons mise en œuvre, peut être utilisée pour la résolution d'un autre type d'équation intégrale qui est ici l'équation intégrale de Radon.
- (ii) Dans cette application ME fournit des images qui ont une meilleure résolution spatiale et une meilleure résolution en amplitude que les méthodes linéaires.
- (iii) La seule contrainte de positivité ne suffit pas pour obtenir une solution unique et stable dans ce cas. On peut constater ce fait sur le résultat de la reconstruction obtenue par la méthode MCP dans le cas des projections bruitées. De plus, on constate que ME n'a pas introduit de structure où il n'y en avait pas et que le support de l'objet a été fort bien estimé, bien que cette information n'ait pas été introduite *a priori*.

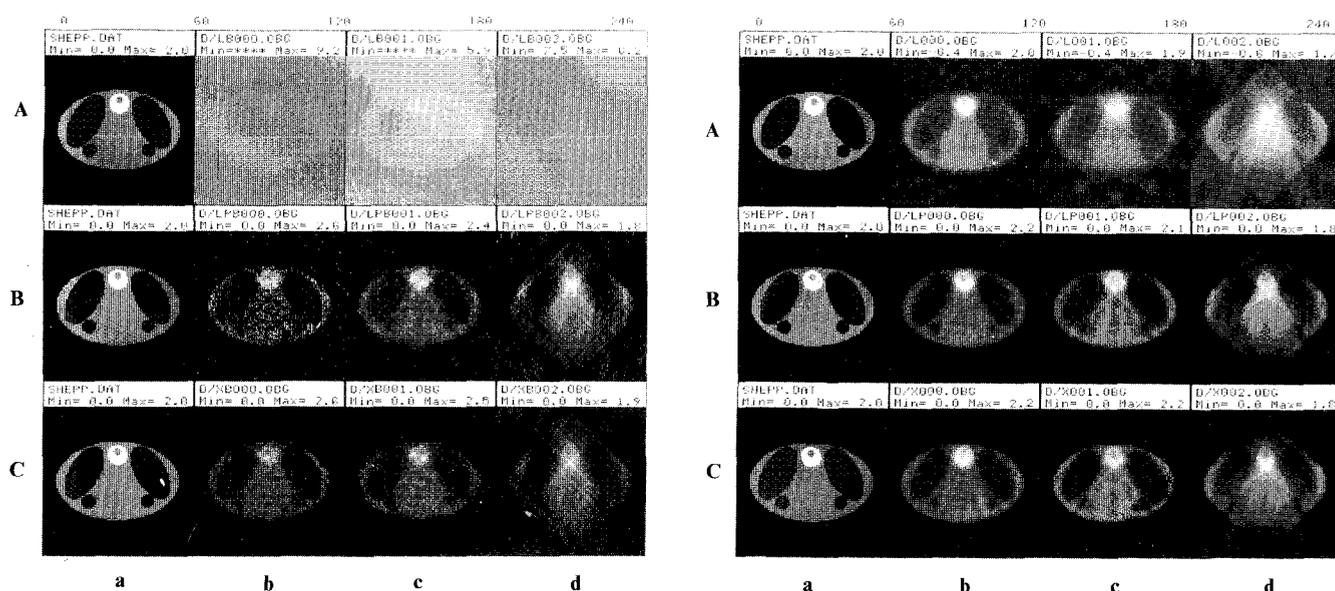


Fig. 3. — Reconstruction d'une image [fonction $f(x, y)$] à partir de ses projections en tomographie à rayons X. Comparaison entre les méthodes MC, MCP et ME : A : reconstruction par MC, B : reconstruction par MCP, C : reconstruction par ME. (a) objets, (b) $Np=12, \theta=180^\circ$, (c) $Np=8, \theta=180^\circ$, (d) $Np=5, \theta=90^\circ$; à gauche : données sans bruit, à droite : données avec bruit.

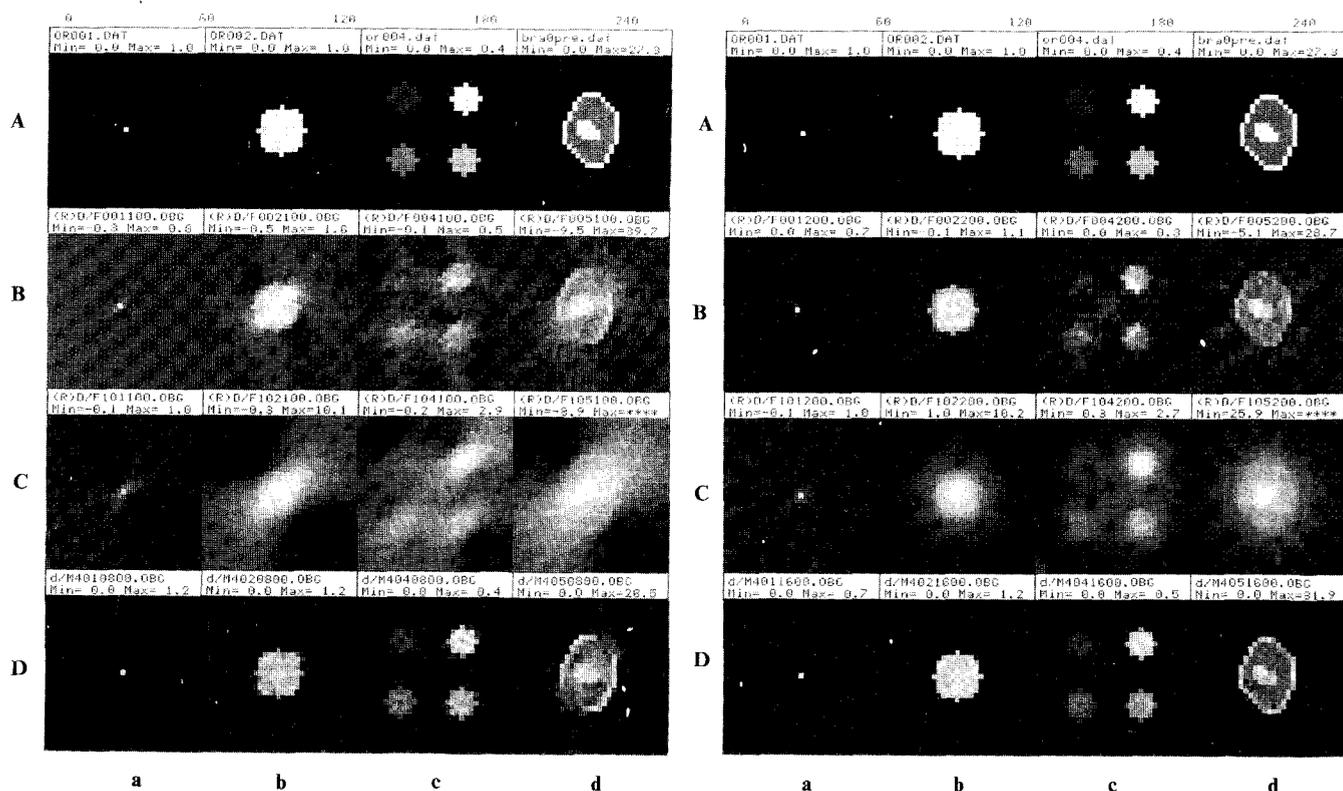


Fig. 4. — Reconstruction d'une image [fonction $f(x, y)$] à partir de la connaissance de sa TF sur des demi-cercles (simulant le problème de la synthèse de Fourier en tomographie à ondes diffractées), Comparaison entre les méthodes d'interpolation analytique (M1 et M2) et la méthode ME. (A) quatre objets différents, (B) reconstruction par M1, (C) reconstruction par M2, (D) reconstruction par ME; à gauche : $Np=16, \theta=360^\circ$, à droite : $Np=08, \theta=90^\circ$.

6. 3. SYNTHÈSE DE FOURIER ET RECONSTRUCTION D'IMAGES EN TOMOGRAPHIE A ONDES DIFFRACTÉES

Dans ce paragraphe nous présentons les simulations effectuées dans le cadre d'une application de la méthode aux problèmes de reconstruction d'images en tomographie à ondes diffractées.

Une partie qui est commune à ce procédé tomographique ainsi qu'à de nombreuses autres techniques d'imagerie est la partie *synthèse de Fourier* qui consiste à déterminer une fonction à deux variables d'espace $f(x, y)$ à partir de la connaissance de sa TF donnée sur un ensemble fini de parcours algébriques [13]. Ces parcours sont des demi-cercles dans le cas de la

tomographie à ondes diffractées. Nous avons voulu simuler ce dernier cas. C'est pourquoi, dans ces simulations, nous avons choisi plusieurs objets et calculé la TFD de ces objets sur des demi-cercles. Ensuite, partant de ces données, nous avons effectué la reconstruction par notre méthode ME et par deux méthodes classiques d'interpolation dans ces applications que nous appellerons par la suite les méthodes M1 et M2. La méthode M1 consiste à supposer connu et limité le support de la fonction à déterminer et ainsi, profitant de l'analyticité de sa TF, effectuer une interpolation dans le domaine de Fourier pour remplir ce domaine et déterminer ensuite la fonction par une TF inverse. Dans la méthode M2, qui est désignée par rétroprojection filtrée en tomographie à rayons X et par rétropropagation filtrée en tomographie à ondes diffractées, on cherche une solution à bande limitée au problème. L'hypothèse d'une bande limitée permet d'effectuer l'interpolation dans le domaine spatial (pour plus de détails sur ces deux méthodes voir les références [11-13]).

Dans ces simulations nous avons choisi des situations où les projections (nous appelons projection les valeurs de la TF de la fonction $f(x, y)$ calculées sur un demi-cercle) sont très limitées en nombre ($Np=16$ ou 8) et en angle ($\theta=360$ ou 90°). Nous avons traité deux cas :

- (a) $Np=16, \theta=360^\circ$;
- (b) $Np=8, \theta=90^\circ$.

Dans ces simulations nous avons ajouté deux bruits gaussiens indépendants respectivement sur la partie réelle et la partie imaginaire des grandeurs dans le domaine de Fourier (valeurs de la TF sur des arcs de cercles) avec un rapport signal-à-bruit de 10 dB. Ceci peut s'interpréter comme un bruit sur les amplitudes des composantes en phase et en quadrature du champ diffracté.

La figure 4 montre une comparaison entre les trois méthodes M1, M2, et ME pour les quatre objets O1, O2, O4 et O5 dans les deux situations ($Np=16, \theta=360^\circ$) et ($Np=8, \theta=90^\circ$).

On constate que les résultats obtenus par les méthodes M1 et M2 ne sont pas acceptables, surtout dans le cas ($Np=8, \theta=90^\circ$). En effet, ni le support, ni les amplitudes ne sont bien reconstitués, tandis que la méthode ME fournit des résultats avec de meilleures résolution spatiale et précision en amplitude.

7. Conclusions

L'approche bayésienne permet de résoudre un problème inverse s'il peut être décrit par un nombre fini de paramètres (ceci est le cas pour un problème discrétisé). Elle permet de prendre en compte, d'une façon cohérente, les incertitudes sur les paramètres et sur les données ainsi que l'information *a priori* sur ces paramètres. Une difficulté majeure dans cette approche est la détermination des lois *a priori* dans un problème réel où la connaissance initiale nous est donnée plutôt sous la forme de contraintes sur les paramètres inconnus du problème. Le principe du ME permet de résoudre cette difficulté d'une manière

cohérente. Nous avons utilisé ces deux outils pour résoudre plusieurs problèmes inverses en imagerie. Il reste cependant quelques difficultés pour l'utilisation de l'entropie qui sont à la fois d'ordre pratique (le choix de λ dans un problème réel) et algorithmique (le coût de calcul), et qui doivent être résolues pour une utilisation courante de la méthode. Cependant, dans des situations où les méthodes linéaires ne peuvent qu'échouer, notre méthode fournit des résultats très intéressants.

Manuscrit reçu le 26 février 1988.

BIBLIOGRAPHIE

- [1] S. F. BURCH, S. F. GULL et J. SKILLING, Image restoration by a powerful maximum entropy method, *Comput. Vis. Graph. Im. Process.*, 23, 1983, p. 113-128.
- [2] R. FLETCHER et C. M. REEVES, Function minimization by conjugate gradients, *Comp. J.*, 7, 1964, p. 149-157.
- [3] B. R. FRIEDEN, Statistical models for the image restoration problem, *Comput. Graph. Im. Process.*, 12, 1980, p. 40-59.
- [4] B. R. FRIEDEN et C. K. ZOLTANI, Maximum bounded entropy: application to tomographic reconstruction, *Applied Optics*, 24, 1985, p. 201-207.
- [5] D. GIRARD, Les méthodes de régularisation optimale et leurs applications en tomographie, *Thèse de Docteur-Ingénieur*, INPG, Grenoble, 1984.
- [6] E. T. JAYNES, Prior Probabilities, *IEEE Trans.*, SSC-4, 1968, p. 227-241.
- [7] E. T. JAYNES, On the Rationale of Maximum Entropy Methods, *Proc. IEEE*, 70, 1982, p. 939-952.
- [8] E. T. JAYNES, Where do we go from here?, *Maximum-Entropy and Bayesian Methods in Inverse Problems*, C. R. SMITH et T. GRANDY Jr. éd., 1985, p. 21-58.
- [9] R. JOHNSON et J. SHORE, Which is Better Entropy Expression for Speech Processing: $-\text{SlogS}$ or logS ?, *IEEE Trans.*, ASSP-32, 1984, p. 129-137.
- [10] J. H. JUSTICE, *Maximum entropy and Bayesian methods in applied statistics*, Cambridge Univ. Press., 1986.
- [11] A. MOHAMMAD-DJAFARI et G. DEMOMENT, Tomographie de diffraction et synthèse de Fourier à maximum d'entropie, *Revue Phys. Appl.*, 22, 1987, p. 153-167.
- [12] A. MOHAMMAD-DJAFARI et G. DEMOMENT, Maximum entropy Fourier synthesis with application to diffraction tomography, *Applied Optics*, 26, 1987, p. 1745-1754.
- [13] A. MOHAMMAD-DJAFARI, Synthèse de Fourier multivariables à maximum d'entropie. Application à la reconstruction tomographique d'images, *Thèse de Doctorat d'État*, Université de Paris-XI, 1987.
- [14] D. MUKHERJEE et D. C. HURST, Maximum Entropy Revisited, *Statistica Neerlandica*, 38, 1984, p. 1-12.
- [15] M. Z. NASHED, Operator-Theoretic and Computational Approaches to Ill-Posed Problems with Applications to Antenna Theory, *IEEE Trans.*, AP-29, 1981, p. 220-231.
- [16] M. J. D. POWELL, Restart procedure for the conjugate gradient method, *Math. Prog.*, 12, 1977, p. 241-254.
- [17] D. SAINT-FÉLIX, X. C. DU et G. DEMOMENT, Image restoration using a non-causal state space model and a fast 2D Kalman filter, *Mathematics in Signal Processing*, T. S. DURRANI et al. éd., Clarendon Press, Oxford, 1987, p. 363-378.

- [18] J. E. SHORE et R. W. JOHNSON, Properties of Cross-Entropy Minimization, *IEEE Trans.*, IT-27, 1981, p. 472-482.
- [19] J. SKILLING et S. F. GULL, Maximum Entropy Method in Image Processing, *IEE Proceedings*, 131, Pt. F, 1984, p. 646-659.
- [20] J. SKILLING et S. F. GULL, The Entropy of an Image, *SIAM-AMS Proceedings*, 14, 1984, p. 167-189.
- [21] D. M. TITTERINGTON, Common Structure of Smoothing Techniques in Statistics, *Intern. Stat. Rev.*, 53, 1985, p. 141-170.
- [22] S. J. WERNECKE et L. R. D'ADDARIO, Maximum Entropy Image Reconstruction, *IEEE Trans.*, C-26, 1977, p. 351-364.