

Maximum d'entropie et problèmes inverses en imagerie

Maximum Entropy and Inverse Problems in Image Reconstruction

par Ali Mohammad DJAFARI

Laboratoire des Signaux et Systèmes (CNR-ESE-UPS)
Ecole Supérieure d'Electricité
Plateau de Moulon,
91192 Gif-sur-Yvette Cedex, France

Résumé

Dans un très grand nombre de problèmes d'imagerie on est amené à résoudre une équation intégrale de première espèce, ce qui constitue, dans la plupart des cas, 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.

Une information *a priori* utilisée dans beaucoup d'applications en l'imagerie est la positivité. Les méthodes basées sur le principe du maximum d'entropie (MaxEnt) sont classiquement employées pour prendre en compte cette information *a priori*. Cependant, l'utilisation de ce principe ne se limite pas à ce seul cas. Depuis ces dix dernières années, beaucoup de travaux sur l'utilisation de l'entropie dans ces problèmes ont été menés.

L'objectif principal de cet article est d'essayer de classer ces méthodes, de montrer leurs limites théoriques et pratiques, de montrer les liens qui peuvent exister entre elles, et, finalement, de fournir des *fiches synthétiques* pour chacune de ces méthodes. Pour ceci, nous distinguerons trois familles de méthodes que nous appellerons *MaxEnt classique*, *MaxEnt sur la moyenne*, et *MaxEnt bayésienne*. Nous étudierons les différentes variantes à l'intérieur de chaque famille et préciserons les liens qui existent entre ces différentes méthodes.

Mots clés : Maximum d'entropie, Problèmes inverses, Approche bayésienne, Maximum *a posteriori*, Lois *a priori*, Invariance d'échelle, Reconstruction d'image, Restauration d'image.

Abstract

In many imaging systems we have to solve integral equations of the first kind, which, in general, are ill-posed inverse problems. In these problems one cannot obtain a satisfactory unique and robust solution without introduction of some prior information on the solution.

A prior information often used in many imaging applications is the positivity of the solution. Methods based on the maximum entropy principle are classically developed to take this property into account. However, the use of the maximum entropy principle is not limited to this case. In the last decade, many works have been done on the ways of using maximum entropy principle in inverse problems. The main objective of this paper is to make a classification of these methods, to give explicitly the hypothesis, the practical and the theoretical limitations, to show the existing relations between them, and finally, to give some synthetic view of different implementations of these methods. First we distinguish three fundamentally different approaches which we call Classical MaxEnt, MaxEnt in mean and Bayesian MaxEnt. Then, in each approach, we describe the different methods and algorithms which are obtained when the nature and the amount of the available data change and we give explicit relations between all these methods.

Key words : Maximum entropy, Inverse problems, Bayesian approach, Maximum *a posteriori*, *a priori* laws, Scale invariance, Image reconstruction, Image restoration.

1. Introduction

Dans un grand nombre de problèmes d'imagerie tels que la tomographie à rayons X, la tomographie de diffraction par ultrasons ou par micro-ondes, l'imagerie par courants de Foucault, la tomographie à émission de positons (TEP), l'imagerie par résonance magnétique nucléaire (RMN), etc., on est souvent amené à résoudre

une équation intégrale de première espèce de la forme :

$$g(s_i) = \int_D f(\mathbf{r}) h(\mathbf{r}, s_i) d\mathbf{r} + b(s_i), \quad i = 1, \dots, M, \quad (1)$$

où

— g désigne les mesures (image observée en restauration d'image, projections en tomographie X, champ diffracté en tomographie de diffraction, force électromotrice aux

bornes d'une bobine en imagerie par courants de Foucault, etc...),

- f désigne l'objet à restaurer ou à reconstruire,
- D désigne le support de l'objet,
- h est le noyau de la transformation qui lie les mesures g et l'objet f , et enfin
- b représente les erreurs (ou l'incertitude) sur les mesures.

Pour montrer la généralité de cette équation nous précisons dans l'annexe 3 quelques-unes de ses applications comme la déconvolution des signaux, la restauration d'image, la reconstruction d'image en tomographie à rayons X, la synthèse d'ouverture en radioastronomie, etc.

L'inversion de cette équation est en général un problème mal posé car il ne satisfait pas aux trois conditions d'existence, d'unicité et surtout de stabilité [[Nashed81]], [[Demoment87]], [[Demoment89]]. Une manière de transformer un problème mal posé en un problème bien-posé consiste à le régulariser en introduisant une information *a priori* sur la solution. Cette information *a priori* peut nous être donnée soit sous une forme déterministe (limitation du support, positivité, etc.), soit sous une forme stochastique (loi de probabilité de l'image ou des 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 :

$$g = Hf + b, \quad (2)$$

où

- f est un vecteur de dimension N contenant l'ensemble des inconnues du problème,
- g est un vecteur de dimension M contenant l'ensemble des observations,
- H est une matrice de dimensions $(M \times N)$ connue qui ne dépend que du noyau du système de mesure (par exemple de la géométrie du problème en reconstruction d'image), et
- b est un vecteur contenant les termes d'incertitude qui correspondent au bruit sur les observations et aux erreurs de quadrature.

Une méthode naïve consisterait à choisir le pas de discrétisation afin que la matrice H soit carrée et à estimer la solution par $\hat{f} = H^{-1}g$ ou bien, même si la matrice n'est pas carrée, à rechercher une solution au sens des moindres carrés :

$$\hat{f} = \arg \min_f \{ Q(f) = [g - Hf]^t [g - Hf] \}.$$

Si la matrice $H^t H$ est inversible cette solution est donnée formellement par

$$\hat{f} = [H^t H]^{-1} H^t g.$$

Mais, en général, $M < N$ et les matrices H et $H^t H$ sont malconditionnées, voire même singulières. Ceci est la conséquence du caractère mal posé du problème initial. La solution inverse généralisée ou la solution au sens des moindres carrés de norme minimale ne permettent d'assurer que la condition de l'unicité de la solution quand elle existe, mais le problème de la stabilité (ou bien la sensibilité) de la solution vis-à-vis des erreurs de mesure ou de la discrétisation et de la quantification reste entier. 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 limiter à des contraintes sur les bornes (la positivité par exemple ($x_i > 0$)), et à un ou plusieurs paramètres statistiques (la moyenne, la variance, la moyenne à l'échelle logarithmique (moyenne géométrique), la longueur de corrélation, etc...).

Nous verrons aussi comment le principe du maximum d'entropie interviendra dans une étape d'inversion de ces problèmes inverses. Nous distinguerons trois approches que nous appellerons *MaxEnt classique*, *MaxEnt sur la moyenne*, et *MaxEnt bayésienne*.

Dans l'approche *MaxEnt classique*, à l'origine, l'objectif est de trouver une solution positive et douce, tout en satisfaisant exactement les contraintes imposées par les données, mais le bruit n'est pas pris en compte explicitement.

Dans l'approche *MaxEnt sur la moyenne* l'objectif est de satisfaire les contraintes imposées par les données sur la solution, en moyenne, mais le bruit d'observation n'est pas pris en compte explicitement.

Dans l'approche *MaxEnt bayésienne*, on peut prendre en compte à la fois le bruit de mesure proprement dit ainsi que toute autre incertitude (par exemple sur la modélisation du problème direct). Dans l'approche bayésienne avec *a priori* à maximum d'entropie, le principe du maximum d'entropie est utilisé pour l'attribution de la loi de probabilité *a priori* $p(f)$ et la loi conditionnelle $p(g|f)$.

L'article est organisé de la manière suivante :

- le chapitre 2 rappelle la définition de l'entropie et le principe du maximum d'entropie;
- le chapitre 3 décrit les principes des trois approches utilisant l'entropie pour la résolution des problèmes inverses;
- les chapitres 4, 5 et 6 fournissent des fiches synthétiques pour les différentes méthodes dans chaque famille;

— finalement, nous concluons ce travail de synthèse dans le chapitre 7.

2. Principe du maximum d'entropie

2.1. DÉFINITION DE L'ENTROPIE

Le principe du ME peut être approché de différentes manières. L'approche de la théorie de l'information [[Shannon48]] est sans doute la mieux adaptée à notre problème. Jaynes ([[Jaynes68], [Jaynes82], [Jaynes85]]) est parmi les premiers auteurs modernes à avoir introduit le formalisme du ME par l'approche de la théorie de l'information. Cette notion d'entropie est introduite de la manière suivante : considérons une variable aléatoire discrète X produisant des réalisations $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ et attribuons les probabilités $\mathbf{p} = \{p_1, p_2, \dots, p_n\}$ à ces réalisations pour représenter notre information partielle sur cette variable. On définit la quantité $I_i = \ln(1/p_i)$ comme la quantité d'information apportée par la réalisation x_i .

Le raisonnement intuitif conduisant à cette expression est que plus un événement est rare, plus le gain d'information obtenu par sa réalisation est grand. L'utilisation du logarithme rend additif le gain total d'information obtenu par la réalisation de 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 :

$$S(\mathbf{p}) = \sum_{i=1}^n p_i \ln \frac{1}{p_i} = - \sum_{i=1}^n p_i \ln p_i. \quad (3)$$

L'entropie S est une mesure d'incertitude de la distribution $\mathbf{p} = \{p_1, p_2, \dots, p_n\}$, déterminée uniquement par certaines règles élémentaires de cohérence logique et d'additivité ([[Jaynes82], [Jaynes85], [Balian82]]).

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

$$S(\mathbf{q}, \mathbf{p}) = - \sum_{i=1}^n q_i \ln \frac{p_i}{q_i} = \sum_{i=1}^n q_i \ln \frac{q_i}{p_i}, \quad (4)$$

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

Sous réserve de quelques précautions [[Shannon48], [Jaynes85], [Balian82]], on peut généraliser ce qui précède au cas de distributions continues, et on définit l'entropie [[Shannon48]] par :

$$S(p) = - \int p(x) \ln p(x) dx, \quad (5)$$

et l'entropie croisée par :

$$S(q, p) = - \int q(x) \ln \frac{p(x)}{q(x)} dx. \quad (6)$$

2.2. LOIS À MAXIMUM D'ENTROPIE

Voyons maintenant ce que signifie le choix d'une distribution de probabilité à maximum d'entropie contenant une information *a priori*, ou qui soit compatible avec des contraintes connues sur cette distribution. Pour cela, considérons 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. Considérons une variable aléatoire discrète X qui peut prendre des valeurs $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ avec une distribution de probabilité $\mathbf{p} = \{p_1, p_2, \dots, p_n\}$. Maintenant si on nous demande quelle est la meilleure estimée $\hat{\phi}$ d'une fonction $\phi(X)$ au sens du minimum de l'erreur quadratique moyenne. La solution est immédiate :

$$\hat{\phi} = E\{\phi\} = \sum_{i=1}^n p_i \phi(x_i).$$

C'est un problème direct et bien-posé. Inversement, si l'on se pose la question d'ajuster une distribution \mathbf{p} pour incorporer une information donnée sur la fonction $\phi(X)$, il faut comprendre par ceci :

connaissant une règle d'estimation précise, par exemple la minimisation de l'erreur quadratique moyenne, comment choisir \mathbf{p} pour que l'on ait $\hat{\phi} = E\{\phi\}$?

La réponse est qu'il existe, en général, beaucoup de distributions qui satisfont cette contrainte. Il s'agit d'un problème inverse mal posé au sens où la solution n'est pas unique. Le principe du maximum d'entropie nous permet alors de choisir une solution.

Considérons maintenant le cas plus général où nous considérons les m fonctions

$\{\phi_1(X), \phi_2(X), \dots, \phi_m(X)\}$ 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 :

$$E\{\phi_k(X)\} = \sum_{i=1}^n p_i \phi_k(x_i) = d_k, \quad k = 1, \dots, m.$$

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. En général, le nombre des contraintes m est inférieur à n et il y a une infinité de solutions à ce problème. En d'autres termes on peut trouver une infinité de distributions de probabilité $\{p_1, p_2, \dots, p_n\}$ qui satisfont ces contraintes. C'est ici que le principe du maximum d'entropie est mis en œuvre pour choisir, parmi ces solutions possibles, celle qui a l'entropie maximale, c'est-à-dire la loi qui satisfait toutes les contraintes (toute l'information connue) et qui est la moins compromettante vis-à-vis de toute autre information non connue. Le mot *information* devant être pris au sens de la définition de l'information moyenne ou de l'entropie de Shannon (équation 3).

Le problème se formule mathématiquement ainsi :

Problème P1 :

$$\text{maximiser } S(\mathbf{p}) = - \sum_{i=1}^n p_i \ln p_i,$$

$$\text{sous les contraintes } \sum_{i=1}^n p_i \phi_k(x_i) = d_k, \quad k = 1, \dots, m.$$

La solution est obtenue par une technique variationnelle de multiplicateurs de Lagrange (voir annexe 1) et est donnée par :

$$p_i = \frac{1}{Z(\lambda_1, \dots, \lambda_m)} \exp \left[- \sum_{k=1}^m \lambda_k \phi_k(x_i) \right], \quad i = 1, \dots, n, \quad (7)$$

où

$$Z(\lambda_1, \dots, \lambda_m) = \sum_{i=1}^n \exp \left[- \sum_{k=1}^m \lambda_k \phi_k(x_i) \right] \quad (8)$$

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

$$- \frac{\partial \ln Z(\lambda_1, \dots, \lambda_m)}{\partial \lambda_k} = d_k, \quad k = 1, \dots, m, \quad (9)$$

avec m données d_k et m inconnues λ_k . Notons que ce système d'équations n'est autre que le système d'équations des contraintes constituées par les données.

La valeur maximale de l'entropie est :

$$S_{\max} = \ln Z + \sum_{k=1}^m \lambda_k d_k$$

Si on note $\lambda_0 = \ln Z$ on peut vérifier facilement les propriétés suivantes :

$$- \frac{\partial \lambda_0}{\partial \lambda_k} = E\{\phi_k(x)\} = d_k, \quad (10)$$

$$- \frac{\partial^2 \lambda_0}{\partial \lambda_k^2} = E\{\phi_k^2(x)\} - d_k^2 = \text{Var}\{\phi_k(x)\}, \quad (11)$$

$$- \frac{\partial^2 \lambda_0}{\partial \lambda_k \partial \lambda_l} = E\{\phi_k(x)\phi_l(x)\} - d_k d_l = \text{Cov}\{\phi_k(x), \phi_l(x)\}, \quad (12)$$

2.3. LOIS À MINIMUM D'ENTROPIE RELATIVE

L'extension de ce qui précède au cas de l'entropie croisée se formule ainsi :

Problème P2 : Étant donné une distribution de probabilité *a priori* $\mathbf{p} = \{p_1, p_2, \dots, p_n\}$, déterminer la distribution de probabilité *a posteriori* $\mathbf{q} = \{q_1, q_2, \dots, q_n\}$ qui minimise

$$S(\mathbf{q}, \mathbf{p}) = - \sum_i q_i \ln \frac{p_i}{q_i} = \sum_i q_i \ln \frac{q_i}{p_i},$$

et qui satisfait les m contraintes

$$\sum_{i=1}^n q_i \phi_k(x_i) = d_k, \quad k = 1, \dots, m.$$

La solution est obtenue, ici aussi, par une technique variationnelle de multiplicateurs de Lagrange et donnée par :

$$q_i = \frac{1}{Z} p_i \exp \left[- \sum_{k=1}^m \lambda_k \phi_k(x_i) \right], \quad i = 1, \dots, n, \quad (13)$$

où

$$Z(\lambda_1, \dots, \lambda_m) = \sum_{i=1}^n p_i \exp \left[- \sum_{k=1}^m \lambda_k \phi_k(x_i) \right] \quad (14)$$

est la fonction de partition, et les $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ sont déterminés par le système d'équations (9).

2.4. EXTENSION AU CAS CONTINU

Dans le cas continu, les définitions (5) et (6) conduisent à :

Problème P3 :

$$\text{maximiser } S(p) = - \int p(x) \ln p(x) dx,$$

sous les contraintes

$$\int \phi_k(x) p(x) dx = d_k, \quad k = 1, \dots, m.$$

La solution est

$$p(x) = \frac{1}{Z} \exp \left[- \sum_{k=1}^m \lambda_k \phi_k(x) \right], \quad (15)$$

où

$$Z(\lambda_1, \dots, \lambda_m) = \int \exp \left[- \sum_{k=1}^m \lambda_k \phi_k(x) \right] dx \quad (16)$$

est la fonction de partition, et les $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ sont déterminés par le système d'équations (9).

Problème P4 : Étant donné une densité de probabilité *a priori* $p(x)$, déterminer la densité de probabilité *a posteriori* $q(x)$ qui minimise l'entropie croisée

$$S(q, p) = - \int q(x) \ln \frac{p(x)}{q(x)} dx,$$

et qui satisfait les m contraintes

$$\int \phi_k(x) q(x) dx = d_k, \quad k = 1, \dots, m.$$

La solution est

$$q(x) = \frac{1}{Z} p(x) \exp \left[- \sum_{k=1}^m \lambda_k \phi_k(x) \right], \quad (17)$$

où

$$Z(\lambda_1, \dots, \lambda_m) = \int p(x) \exp \left[- \sum_{k=1}^m \lambda_k \phi_k(x) \right] dx \quad (18)$$

est la fonction de partition, et les $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ sont déterminés par (9).

Notons également que le vecteur $\hat{\lambda} = (\lambda_1, \dots, \lambda_m)$ dans tous ces problèmes peut être considéré comme la solution du problème d'optimisation suivant :

Problème dual des problèmes P1-P4 :

$$\hat{\lambda} = \arg \min_{\lambda} \{ D(\lambda) = \lambda^t d + \ln Z(\lambda) \} \quad (19)$$

où $d = (d_1, \dots, d_m)$. Ce problème d'optimisation est appelé *problème dual* des problèmes P1-P4 (voir annexe 1). Notons que l'expression de $Z(\lambda)$ est différente suivant chaque problème.

Une présentation légèrement différente de ces relations peut être obtenue si on définit

$$\phi_0(x) = 1, \quad \text{et } d_0 = 1,$$

ce qui permet d'inclure la contrainte de la normalisation de $q(x)$ et formuler le problème précédent de la manière suivante :

$$\text{minimiser } S(q, p) = - \int q(x) \ln \frac{p(x)}{q(x)} dx,$$

sous les contraintes $\int \phi_k(x) q(x) dx = d_k, \quad k = 0, \dots, m.$

La solution peut être écrite sous la forme

$$q(x) = p(x) \exp \left[- \sum_{k=0}^m \lambda_k \phi_k(x) \right], \quad (20)$$

et les $\{\lambda_0, \lambda_1, \lambda_2, \dots, \lambda_n\}$ sont déterminés par :

$$G_k(\lambda_0, \lambda_1, \lambda_2, \dots, \lambda_n) = - \int \phi_k(x) p(x) \exp \left[- \sum_{l=0}^m \lambda_l \phi_l(x) \right] dx = d_k, \quad k = 0, \dots, m. \quad (21)$$

On remarque que λ_0 est relié à la fonction de partition Z par $Z = \exp[\lambda_0]$, et on a :

$$\lambda_0 = \ln Z = \ln \int p(x) \exp \left[- \sum_{k=1}^m \lambda_k \phi_k(x) \right] dx. \quad (22)$$

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

$$- \frac{\partial}{\partial \lambda_k} \lambda_0(\lambda_1, \dots, \lambda_n) = d_k, \quad k = 1, \dots, m. \quad (23)$$

Notons aussi que la valeur minimale $S_{\min}(q, p)$ peut être exprimée en fonction des λ_k par :

$$S_{\min}(q, p) = \lambda_0 + \sum_{k=1}^m \lambda_k d_k. \quad (24)$$

Il n'est en général pas possible d'obtenir une relation explicite pour les coefficients λ_k , et on résout le système d'équations (21) numériquement par des méthodes itératives. Cependant, dans certaines situations simples on peut résoudre le problème d'une façon analytique [[Jaynes82], [Johnson84]]. Le tableau 1 montre quelques exemples de lois, que l'on obtient sous forme analytique, en résolvant le problème P3 :

$$\text{maximiser } S(p) = - \int p(x) \ln p(x) dx,$$

sous les contraintes $\int \phi_k(x) p(x) dx = d_k, \quad k = 1, \dots, m.$

Tableau 1 : Exemples de lois à maximum d'entropie.

contraintes	domaine de x	loi de probabilité
$\phi_1(x) = x$	$x \in \mathbf{R}_+$	$p(x) = \mu \exp[-\mu x]$ loi exponentielle
$\phi_1(x) = x $	$x \in \mathbf{R}$	$p(x) = \frac{\mu}{2} \exp[-\mu x]$ loi de Laplace
$\begin{cases} \phi_1(x) = x \\ \phi_2(x) = x^2 \end{cases}$	$x \in \mathbf{R}$	$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]$ loi gaussienne
$\begin{cases} \phi_1(x) = x \\ \phi_2(x) = \ln x \end{cases}$	$x \in \mathbf{R}_+$	$p(x) \propto \exp[-\lambda \ln x - \mu x]$ $\propto x^{-\lambda} \exp[-\mu x]$ loi Gamma
$\begin{cases} \phi_1(x) = \ln x \\ \phi_2(x) = \ln(1-x) \end{cases}$	$x \in]0, 1[$	$p(x) \propto \exp[-\lambda \ln x - \mu \ln(1-x)]$ $\propto x^{-\lambda} (1-x)^{-\mu}$ loi Béta
$\begin{cases} \phi_1(x) = \ln x \\ \phi_2(x) = x^2 \end{cases}$	$x > 0$	$p(x) \propto \exp[-\lambda \ln x - \mu x^2]$ $\propto x^{-\lambda} \exp[-\mu x^2]$ loi de Rayleigh

2.5. EXTENSION AU CAS MULTIVARIABLE

Toutes ces relations peuvent être généralisées au cas multivariable. Par exemple le problème P4 devient :

Problème P5 : Étant donné une densité de probabilité *a priori* $p(\mathbf{x})$, déterminer la densité de probabilité *a posteriori* $q(\mathbf{x})$ qui minimise l'entropie croisée

$$S(q, p) = - \int q(\mathbf{x}) \ln \frac{p(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x},$$

et qui satisfait les m contraintes

$$\int \phi_k(\mathbf{x}) q(\mathbf{x}) d\mathbf{x} = d_k, \quad k = 1, \dots, m.$$

La solution est

$$q(\mathbf{x}) = \frac{1}{Z} p(\mathbf{x}) \exp \left[- \sum_{k=1}^m \lambda_k \phi_k(\mathbf{x}) \right], \quad (25)$$

où

$$Z(\lambda_1, \dots, \lambda_m) = \int p(\mathbf{x}) \exp \left[- \sum_{k=1}^m \lambda_k \phi_k(\mathbf{x}) \right] d\mathbf{x} \quad (26)$$

et les $\{\lambda_1, \lambda_2, \dots, \lambda_m\}$ sont déterminés par :

$$-\frac{\partial \ln Z(\lambda_1, \dots, \lambda_m)}{\partial \lambda_k} = d_k, \quad k = 1, \dots, m. \quad (27)$$

Ici aussi, dans certains cas on peut obtenir une relation explicite pour la solution. Par exemple, si $p(\mathbf{x})$ est une distribution exponentielle multivariée de la forme séparable

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

et si les contraintes sont des contraintes sur les moments d'ordre un

$$E \{X_i\} = \int x_i q(x_i) dx_i = m_i, \quad i = 1, \dots, n,$$

alors la solution $q(\mathbf{x})$ reste une fonction exponentielle multivariée séparable :

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

De même si $p(\mathbf{x})$ est une fonction gaussienne multivariée séparable :

$$p(\mathbf{x}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\alpha_i}} \exp \left[-\frac{1}{2} \left(\frac{x_i - m_i}{\alpha_i} \right)^2 \right],$$

et si les contraintes sont des contraintes du second ordre :

$$E \{(X_i - m_i)^2\} = \int (x_i - m_i)^2 q(x_i) dx_i = \sigma_i^2,$$

alors la solution $q(x)$ reste une fonction gaussienne multivariée séparable :

$$q(x) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_i} \exp \left[-\frac{1}{2} \left(\frac{x_i - m_i}{\sigma_i} \right)^2 \right].$$

Ce dernier résultat peut être généralisé au cas où $p(x)$ est une fonction gaussienne multivariée non dégénérée :

$$p(x) = \frac{|\Sigma|^{-1/2}}{(2\pi)^{n/2}} \exp \left[-\frac{1}{2} (x - m)^t \Sigma^{-1} (x - m) \right],$$

et si les contraintes sont des contraintes sur les moments d'ordre un et deux non singulières (Σ' inversible) :

$$E \{x\} = \int x q(x) dx = m'$$

$$E \{(x - m')(x - m')^t\} = \int (x - m')(x - m')^t q(x) dx = \Sigma'$$

alors la solution $q(x)$ reste une fonction gaussienne multivariée non dégénérée :

$$q(x) = \frac{|\Sigma'|^{-1/2}}{(2\pi)^{n/2}} \exp \left[-\frac{1}{2} (x - m')^t \Sigma'^{-1} (x - m') \right],$$

Arrivés à ce stade, nous avons les outils nécessaires pour comprendre comment le principe du maximum d'entropie peut être utilisé dans la résolution des problèmes inverses.

3. Classification principale

Nous avons distingué trois approches, qui nous semblent fondamentalement différentes, d'utilisation du principe du maximum d'entropie (PME) pour la résolution des problèmes inverses.

3.1. APPROCHE CLASSIQUE DU MAXIMUM ENTROPIE : MAXENT CLASSIQUE

Dans cette approche on fait l'hypothèse que la fonction $f(r)$ dans l'équation (2) est, ou a les propriétés d'une fonction densité de probabilité, c'est-à-dire

$$\begin{cases} f(r) \geq 0, \\ \int f(r) dr = 1. \end{cases} \quad (28)$$

Ainsi dans cette approche le problème inverse (2) est résolu en introduisant une fonction $f_0(r)$ a priori et en choisissant, parmi toutes les fonctions $f(r)$ satisfaisant aux contraintes (2), celle qui minimise l'entropie relative $S(f, f_0)$. Dans ce cas $S(f, f_0)$ est utilisée comme une mesure de distance entre f et f_0 . Le problème est ainsi résolu de manière déterministe. Nous choisissons, parmi toutes les solutions possibles, celle qui minimise $S(f, f_0)$. Notons que $S(f, f_0)$ peut être considérée comme une mesure de distance (de Kullback) [[Kullback59]] à une fonction de référence f_0 .

Notons que l'hypothèse selon laquelle $f(r)$ est une densité de probabilité peut correspondre à une réalité physique. Par exemple $f(r)$ peut représenter la distribution d'énergie des photons dans un volume, ou sa projection sur une surface ou sur une ligne droite [[Kikuchi77]]. Mais il ne s'agit pas ici d'inférer sur la variable aléatoire énergie des photons mais sur sa distribution. Ainsi $f(r)$ est une fonction déterministe supposée positive qui peut être assimilée, une fois normalisée, à une fonction densité de probabilité. La seule difficulté réside dans le fait que les données g_m (les contraintes) ne suffisent pas pour déterminer de manière unique cette fonction, et c'est là qu'intervient le problème du choix de cette fonction.

Le problème mathématique à résoudre, dans sa forme générale est :

Étant donné $h_m(r)$, g_m , $m = 1, \dots, M$, le domaine D et la fonction de référence $f_0(r)$ (qui doit aussi être une fonction densité de probabilité) ;

$$\text{maximiser} \quad - \int_D f(r) \ln \left(\frac{f(r)}{f_0(r)} \right) dr,$$

sous les contraintes

$$\int_D f(r) h_m(r) dr = g_m, \quad m = 1, \dots, M,$$

et les contraintes de normalisation (28).

Nous verrons dans la section 4 que la difficulté essentielle de cette approche est la prise en compte des erreurs de mesure qui peuvent mettre en cause l'existence d'une solution. Nous présenterons des extensions et des compléments à ce problème qui permettent de prendre en compte ces erreurs.

Les méthodes issues de cette approche sont parfois appelées *Maximum d'entropie structurelle*.

3.2. MAXIMUM ENTROPIE SUR LA MOYENNE : MAXENT SUR LA MOYENNE

Dans cette approche l'image est considérée comme la moyenne $\bar{f}(\mathbf{r}) = \langle f(\mathbf{r}) \rangle$ des réalisations possibles $f(\mathbf{r})$ d'un processus aléatoire $F(\mathbf{r})$ pour lequel nous supposons avoir défini une loi de probabilité $p(f)$. Notons que ceci nécessite tout d'abord de discrétiser l'espace engendré par les vecteurs \mathbf{r} et de représenter la fonction $f(\mathbf{r})$ par un vecteur aléatoire $\mathbf{f} \in \mathcal{C}$, et ensuite, de définir une mesure $d\mathbf{f}$ dans l'ensemble convexe \mathcal{C} des vecteurs \mathbf{f} de telle sorte que $p(\mathbf{f}) d\mathbf{f}$ puisse avoir un sens.

Les données \bar{g}_m sont supposées représenter des contraintes linéaires sur la moyenne $\bar{f}(\mathbf{r})$:

$$\bar{g}_m = \int_D \bar{f}(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} = \int_D \langle f(\mathbf{r}) \rangle h_m(\mathbf{r}) d\mathbf{r}, \quad m = 1, \dots, M, \quad (29)$$

que l'on peut aussi interpréter comme des valeurs moyennes d'une combinaison linéaire de $f(\mathbf{r})$:

$$\bar{g}_m = \langle g_m \rangle, \quad \text{avec} \quad g_m = \int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r}, \quad m = 1, \dots, M. \quad (30)$$

Notons que $\langle \cdot \rangle$ signifie l'espérance suivant la loi $p(f)$, et que dans le cas discret on a

$$\bar{\mathbf{g}} = \mathbf{H} \bar{\mathbf{f}} = \mathbf{H} \langle \mathbf{f} \rangle = \langle \mathbf{H} \mathbf{f} \rangle. \quad (31)$$

Le principe du maximum d'entropie dans cette approche est alors utilisé pour déterminer la loi de probabilité $p(f)$ qui satisfait ces contraintes et qui maximise l'entropie relative à la mesure $d\mathbf{f}$:

$$S(p(f)) = - \int_{\mathcal{C}} p(f) \ln p(f) d\mathbf{f}. \quad (32)$$

où \mathcal{C} est un ensemble convexe. Puisque ces contraintes sont satisfaites en moyenne, nous pouvons construire l'estimateur :

$$\hat{\bar{f}} = \langle f \rangle = E_{p(f)} \{ f \} = \int_{\mathcal{C}} f p(f) d\mathbf{f}. \quad (33)$$

Les méthodes issues de cette approche sont appelées *Maximum d'entropie en moyenne*.

Notons enfin que dans cette approche $f(\mathbf{r})$ n'a pas besoin d'être une fonction positive, mais que l'imposition d'une contrainte de positivité, ou de manière plus générale d'une contrainte de la forme $b(\mathbf{r}) > f(\mathbf{r}) > a(\mathbf{r})$ ne pose aucun problème. En effet, il suffit d'imposer à $p(f)$ d'être nulle en dehors de l'intervalle $]a(\mathbf{r}), b(\mathbf{r})[$, en spécifiant le domaine d'intégration dans les équations (32) et (33).

3.3. APPROCHE BAYÉSIENNE AVEC LES LOIS A PRIORI À MAXIMUM D'ENTROPIE : MAXENT BAYÉSIENNE

Dans cette approche $f(\mathbf{r})$ peut être une fonction quelconque. Notre état de connaissance incomplet sur cette fonction est traduit par une loi de probabilité, dans laquelle on inclut toute connaissance *a priori* I en utilisant le principe du maximum d'entropie. Cette loi est alors notée $p(f | I)$. Insistons sur le fait que l'attribution de cette loi de probabilité (subjective) ne signifie pas forcément que $f(\mathbf{r})$ soit une réalisation d'un processus aléatoire quelconque. Mais, il peut aussi correspondre à une réalité physique (probabilité objective). Notons encore que dans cette approche, comme dans l'approche précédente, nous devons échantillonner la fonction $f(\mathbf{r})$ pour la représenter par un vecteur \mathbf{f} et définir une mesure $d\mathbf{f}$ pour pouvoir définir la loi $p(\mathbf{f} | I)$.

En général, on suppose que les données g_m sont obtenues de manière indépendante; ceci permet de déterminer facilement les lois conditionnelles $p(g_m | f)$ puis $p(g_1, \dots, g_N | f)$. Cette loi n'est autre que la fonction de vraisemblance des mesures traduisant notre état de connaissance incomplète (ou notre incertitude) sur les erreurs du système de mesure.

L'étape suivante est alors de combiner, en utilisant la règle de Bayes, les lois $p(f | I)$ et $p(g_1, \dots, g_m | f)$ pour obtenir la loi *a posteriori* $p(f | g_1, \dots, g_m; I)$:

$$p(f | g_1, \dots, g_m; I) = \frac{1}{Z} p(g_1, \dots, g_m | f) p(f | I),$$

où Z est la constante de normalisation définie par

$$Z = \int p(g_1, \dots, g_m | f) p(f | I) d\mathbf{f}.$$

On peut ensuite utiliser cette loi *a posteriori* pour définir un estimateur ponctuel pour la solution. Notons que cette loi *a posteriori* contient toute notre connaissance sur la solution après les mesures. Mais, souvent, on peut se limiter à étudier le comportement de cette loi autour de son maximum (si elle est unimodale) ou de ses maxima locaux. Dans le cas où la loi est unimodale, on peut se permettre de l'approcher par une gaussienne autour de son maximum, et ainsi définir facilement une matrice de covariance *a posteriori*, qui nous permettra d'étudier plus précisément les caractéristiques de la solution du maximum *a posteriori* (MAP) (par exemple mettre des barres d'erreur sur la solution ou même étudier la corrélation entre les différents paramètres de la solution).

Notons cependant que, s'il est facile d'attribuer une distribution de probabilité aux grandeurs mesurées pour traduire l'existence du bruit sur ces grandeurs, il est plus difficile d'attribuer une distribution de probabilité *a priori* aux paramètres inconnus du problème. C'est ici qu'intervient à nouveau le principe du maximum d'entropie, car en général notre connaissance *a priori* sur

l'objet $f(\mathbf{r})$ ne suffit pas pour définir de manière unique une loi *a priori* $p(f|I)$. D'autre part, en général, notre connaissance *a priori* porte sur la moyenne de la classe des objets, par exemple on peut connaître la moyenne arithmétique ou géométrique de ces objets. Dans ce cas le principe du maximum d'entropie nous permet à nouveau de choisir une distribution de probabilité qui soit cohérente avec notre connaissance *a priori* et qui soit la moins compromettante, au sens où elle n'introduit pas d'information supplémentaire.

En résumé, dans cette approche, le principe du maximum d'entropie nous donne l'outil nécessaire pour l'attribution des lois de probabilités $p(f|I)$ et $p(g_1, \dots, g_m | f)$, alors que la règle de Bayes nous permet de conduire l'inférence sur les inconnues du problème. Nous avons appelé cette démarche *approche bayésienne avec les lois a priori à maximum d'entropie* ou bien, en résumé, *MaxEnt bayésienne*.

Dans chacune de ces trois familles de méthodes, suivant la manière de prendre en compte les erreurs et la manière d'implanter l'algorithme qui calcule effectivement la solution on aura différentes variantes des méthodes dites à *maximum d'entropie*. Par exemple on peut négliger entièrement les erreurs et vouloir satisfaire des contraintes exactement, ou bien prendre en compte ces erreurs en les considérant comme des variables aléatoires avec une loi uniforme sur un intervalle défini ou avec une loi gaussienne centrée et de variance connue. Ainsi on obtiendrait trois méthodes différentes. Dans ce qui suit, nous essayons de décrire ces différentes méthodes que nous appellerons successivement **MaxEnt 1.1**, **MaxEnt 1.2**, ..., **MaxEnt 2.1**, **MaxEnt 2.2**, ..., **MaxEnt 3.1**, **MaxEnt 3.2**, ..., le premier indice représentant la famille et le second la variante.

4. Algorithmes de MaxEnt classique

Comme nous l'avons vu, dans cette approche on fait l'hypothèse que la fonction $f(\mathbf{r})$ possède les propriétés d'une fonction densité de probabilité et les données sont des contraintes linéaires sur cette fonction. Ces contraintes ne suffisant pas à déterminer de manière unique cette fonction, nous choisirons comme solution la fonction qui a l'entropie maximale. Les différentes implantations algorithmiques de cette approche sont décrites ci-dessous.

4.1. MAXENT 1.1

Il s'agit là de l'implantation directe de la méthode du maximum d'entropie classique.

- **Hypothèse** : $f(\mathbf{r})$ est une fonction densité de probabilité ou en a les propriétés.

- **Données** : Les données \bar{g}_m sont des contraintes linéaires sur la fonction $f(\mathbf{r})$

$$\bar{g}_m = \int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r}, \quad m = 1, \dots, M. \quad (34)$$

Notons que la barre sur g_m signifie simplement que les données sont des valeurs exactes (sans bruit). Ceci nous facilitera l'écriture pour la prise en compte du bruit dans les méthodes qui suivront.

- **Problème mathématique** :

Étant donné $h_m(\mathbf{r})$, g_m , $m = 1, \dots, M$, le domaine D et une fonction de référence $f_0(\mathbf{r})$

$$\text{maximiser} \quad - \int_D f(\mathbf{r}) \ln \left(\frac{f(\mathbf{r})}{f_0(\mathbf{r})} \right) d\mathbf{r}.$$

sous les contraintes

$$\int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} = \bar{g}_m, \quad m = 1, \dots, M.$$

- **Solution** : Ce problème est directement résolu par

$$f(\mathbf{r}) = f_0(\mathbf{r}) \frac{1}{Z} \exp \left[- \sum_{m=1}^M \lambda_m h_m(\mathbf{r}) \right], \quad (35)$$

$$\text{avec} \quad Z(\lambda) = \int_D f_0(\mathbf{r}) \exp \left[- \sum_{m=1}^M \lambda_m h_m(\mathbf{r}) \right] d\mathbf{r}. \quad (36)$$

Les multiplicateurs de Lagrange $\lambda = [\lambda_1, \dots, \lambda_M]$ sont obtenus en résolvant le système d'équations non linéaires suivant :

$$- \frac{\partial \ln Z(\lambda)}{\partial \lambda_m} = \bar{g}_m, \quad m = 1, \dots, M. \quad (37)$$

La solution (35) peut aussi se mettre sous la forme

$$f(\mathbf{r}) = f_0(\mathbf{r}) \exp \left[- \sum_{m=0}^M \lambda_m h_m(\mathbf{r}) \right], \quad (38)$$

avec $h_0(\mathbf{r}) = 1$ et $\bar{g}_0 = 1$, ce qui inclut la contrainte de normalisation

$$\int_D f(\mathbf{r}) d\mathbf{r} = 1.$$

Les difficultés pratiques majeures dans cette méthode sont le calcul de la fonction de partition $Z(\lambda)$ et la résolution du système d'équations non linéaires (36) qui peut aussi s'écrire sous la forme

$$G_m(\lambda) = \int_D h_m(\mathbf{r}) f_0(\mathbf{r})$$

$$\exp \left[\sum_{m=0}^M -\lambda_m h_m(\mathbf{r}) \right] d\mathbf{r} = \bar{g}_m, \quad m = 0, \dots, M. \quad (39)$$

Ce système peut être résolu par une méthode du type Gauss-Newton-Raphson ou Newton-Raphson. Cette dernière consiste à développer en série de Taylor, au premier ordre, $G_m(\boldsymbol{\lambda})$ autour d'une solution initiale $\boldsymbol{\lambda}^0$ et à résoudre le système d'équations linéaires résultant qui nous fournit $\boldsymbol{\delta} = \boldsymbol{\lambda} - \boldsymbol{\lambda}^0$, ce qui permet de calculer une nouvelle solution $\boldsymbol{\lambda} = \boldsymbol{\lambda}^0 + \boldsymbol{\delta}$. En itérant ainsi, s'il existe une solution, on converge vers cette solution. Voir la référence [[Djafari91a]] pour une implantation numérique et des programmes écrits en Matlab.

— **Limitations :**

La limitation majeure de cette méthode pour la résolution des problèmes inverses pratiques est la non prise en compte du bruit et des erreurs sur les données. Ceci a pour conséquence que l'existence de la solution n'est pas assurée.

- **Références :** Voir ([[Agmon79]], [[Dusaussoy91]], [[Elfwing80]], [[Eriksson80]], [[Erlander81]], [[Jaynes82]], [[Jaynes85]], [[Johnson84]], [[Djafari91b]], [[Shore81]], [[Frieden80]], [[Frieden85]], [[Mukherjee84]], [[Noonan76]], [[Wragg70]], [[Verdugo78]], [[Zellner77]])

4.2. MAXENT 1.2

Dans cet algorithme on cherche à prendre en compte le bruit sur les données. Pour ceci on suppose que l'on dispose d'un gabarit pour les mesures, ce qui revient à faire l'hypothèse que le bruit est uniformément distribué dans un intervalle $[-\epsilon_m, +\epsilon_m]$.

- **Hypothèse :** $f(\mathbf{r})$ est une fonction densité de probabilité comme dans le cas MaxEnt 1.1.

- **Données :** Les données g_m sont supposées être bruitées :

$$g_m = \int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} + b_m = \bar{g}_m + b_m, \quad m = 1, \dots, M, \quad (40)$$

et les b_m sont supposés être uniformément distribués sur l'intervalle $[-\epsilon_m, +\epsilon_m]$

- **Problème mathématique :**

Étant donné $h_m(\mathbf{r}), g_m, \epsilon_m, m = 1, \dots, M$, le domaine D et une fonction de référence $f_0(\mathbf{r})$

$$\text{maximiser} \quad - \int_D f(\mathbf{r}) \ln \left(\frac{f(\mathbf{r})}{f_0(\mathbf{r})} \right) d\mathbf{r},$$

sous les contraintes

$$\left| \int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} - g_m \right| < \epsilon_m, \quad m = 1, \dots, M.$$

- **Solution :** La solution s'écrit

$$f(\mathbf{r}) = f_0(\mathbf{r}) \frac{1}{Z} \exp \left[- \sum_{m=1}^M \lambda_m h_m(\mathbf{r}) \right], \quad (41)$$

où $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_M]$ est la solution du système d'inégalités non linéaires suivant

$$\left| \frac{\partial \ln Z(\boldsymbol{\lambda})}{\partial \lambda_n} - g_m \right| < \epsilon_m, \quad m = 1, \dots, M, \quad (42)$$

qui peut aussi s'écrire sous la forme

$$\begin{aligned} & |G_m(\boldsymbol{\lambda}) - g_m| \\ &= \left| \int_D h_m(\mathbf{r}) f_0(\mathbf{r}) \exp \left[\sum_{m=0}^M -\lambda_m h_m(\mathbf{r}) \right] d\mathbf{r} - g_m \right| \\ &\leq \epsilon_m, \quad m = 0, \dots, M, \end{aligned} \quad (43)$$

avec $h_0(\mathbf{r}) = 1$ et $g_0 = 1$.

- **Limitations :**

Dans cette méthode, en choisissant les intervalles $[-\epsilon_m, +\epsilon_m]$ suffisamment large, l'existence d'une solution peut être garantie. Cependant l'implantation pratique de cette méthode n'est pas aisée.

- **Références :** Voir ([[Elfwing80]], [[Eriksson80]], [[Erlander81]])

4.3. MAXENT 1.3

Ici on fait l'hypothèse que le bruit est gaussien, centré et de variance connue. On ne cherche plus à satisfaire les contraintes de manière exacte une par une mais plutôt de façon globale en les remplaçant par une seule contrainte.

- **Hypothèse :** $f(\mathbf{r})$ est une fonction densité de probabilité comme dans le cas MaxEnt 1.1.

- **Données :** Les données g_m sont supposées bruitées :

$$g_m = \int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} + b_m = \bar{g}_m + b_m, \quad m = 1, \dots, M,$$

où les b_m sont supposés être indépendants, identiquement distribués, centrés, avec des variances σ_m^2 connues.

— **Problème mathématique :**

Étant donné $h_m(\mathbf{r}), g_m, \sigma_m^2, m = 1, \dots, M$, le domaine D et une fonction de référence $f_0(\mathbf{r})$

$$\text{maximiser} \quad - \int_D f(\mathbf{r}) \ln \left(\frac{f(\mathbf{r})}{f_0(\mathbf{r})} \right) d\mathbf{r},$$

$$\text{sous la contrainte} \quad \chi^2 = \sum_{m=1}^M \frac{1}{\sigma_m^2} (g_m - \bar{g}_m)^2 = c,$$

$$\text{avec} \quad \bar{g}_m = \int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r}, \quad m = 1, \dots, M.$$

c est une constante qui est choisie en fonction du bruit.

— **Solution :** La solution peut être déterminée en définissant le Lagrangien

$$\mathcal{L} = - \int_D f(\mathbf{r}) \ln \left(\frac{f(\mathbf{r})}{f_0(\mathbf{r})} \right) d\mathbf{r} - \frac{\lambda}{2} (\chi^2 - c), \quad (44)$$

et en résolvant $\delta\mathcal{L} = 0$. Le résultat a la forme suivante :

$$f(\mathbf{r}) = f_0(\mathbf{r}) \frac{1}{Z} \exp \left[-\lambda \sum_{m=1}^M \frac{1}{\sigma_m^2} h_m(\mathbf{r}) (g_m - \bar{g}_m) \right]. \quad (45)$$

Notons que cette équation est une équation implicite car $f(\mathbf{r})$ intervient des deux côtés de cette équation par l'intermédiaire de \bar{g}_m . La solution $f(\mathbf{r})$ peut être calculée de manière itérative (par une technique de Newton-Raphson, par exemple).

— **Limitations :**

La difficulté majeure dans cette méthode concerne le choix de c , dont dépend λ . Une autre difficulté est la convergence de la méthode itérative qui dépend aussi sensiblement du choix de $f_0(\mathbf{r})$, de λ et de l'algorithme d'optimisation choisi.

- **Références :** Voir ([[Bryan83]], [[Burch83]], [[Daniell80]], [[Navaza85]], [[Skilling84b]], [[Skilling84c]], [[Skilling84a]], [[Skilling85]], [[Djafari87a]], [[Djafari87b]], [[Djafari88b]], [[Djafari88a]], [[Djafari88c]], [[Wernecke77]], [[Zhuang87]], [[Trussell80]], [[Narayan86]])

4.4. MAXENT 1.4

Dans l'algorithme MaxEnt 1.1 on cherche à satisfaire les contraintes exactement, c'est pourquoi on associe un multiplicateur de Lagrange à chaque contrainte, et on calcule ces multiplicateurs de Lagrange afin que toutes les contraintes soient satisfaites. Ce calcul demande par ailleurs la résolution d'un système d'équations non-linéaires. Cette fois, on cherche à remplacer la résolution de ce système d'équations non-linéaires par le calcul du

minimum d'un critère quadratique. Ceci permet à la fois de relaxer les contraintes et de simplifier l'implantation.

— **Hypothèse :** comme dans MaxEnt 1.1

— **Données :** comme dans MaxEnt 1.1

— **Problème mathématique :** comme dans MaxEnt 1.1

— **Solution :** seule la manière de calculer les multiplicateurs de Lagrange change par rapport à MaxEnt 1.1. Dans MaxEnt 1.1 les multiplicateurs de Lagrange sont calculés en résolvant le système d'équations non-linéaires (37) ou (39) qui sont le reflet de la satisfaction exacte des contraintes imposées par les données. Dans cette variante on relaxe ceci, et on cherche le vecteur λ par

$$\hat{\lambda} = \arg \min_{\lambda} \chi^2(\lambda), \quad (46)$$

où

$$\chi^2(\lambda) = \sum_{m=0}^M (g_m - \bar{g}_m(\lambda))^2, \quad (47)$$

et

$$\begin{aligned} \bar{g}_m(\lambda) &= \int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} \\ &= \int_D f_0(\mathbf{r}) \exp \left[-\sum_{k=0}^M \lambda_k h_k(\mathbf{r}) \right] h_m(\mathbf{r}) d\mathbf{r}, \\ m &= 0, \dots, M, \end{aligned} \quad (48)$$

La minimisation peut être faite par une méthode du gradient conjugué, ce qui nécessite le calcul du gradient

$$\frac{\partial \chi^2(\lambda)}{\partial \lambda_k} = 2 \sum_{m=0}^M (g_m - \bar{g}_m) \left(\int_D h_k(\mathbf{r}) h_m(\mathbf{r}) f(\mathbf{r}) d\mathbf{r} \right) \quad (49)$$

$$k = 0, \dots, M,$$

— **Limitations :**

Notons que, en principe, les multiplicateurs de Lagrange λ calculés par cette méthode sont les mêmes que ceux calculés dans MaxEnt 1.1. Ainsi théoriquement les difficultés sont identiques. Cependant, en pratique, dans l'algorithme d'optimisation on peut définir une règle de décision $\chi^2(\lambda) < c$, c étant un réel positif. Ceci permet de régulariser la solution, mais le choix de c reste *ad hoc* et la solution peut être assez sensible à ce choix.

- **Références :** Voir ([[Bryan83]], [[Burch83]], [[Daniell80]])

5. Algorithmes de MaxEnt sur la moyenne

Comme nous l'avons vu, dans cette approche l'image est considérée comme la moyenne des réalisations possibles $f(\mathbf{r})$ d'un processus aléatoire $F(\mathbf{r})$ pour lequel nous supposons pouvoir définir une loi de probabilité $p(f) df$.

Notons que pour pouvoir définir cette loi de probabilité, nous devons contraindre $f(\mathbf{r})$ à appartenir à un ensemble convexe $\{f(\mathbf{r}) \in \mathcal{C}\}$ et munir cet ensemble d'une mesure $df(\mathbf{r})$. Soit \mathcal{P} l'ensemble des distributions de probabilités définies pour $\{f(\mathbf{r}) \in \mathcal{C}\}$ et soit $p(f(\mathbf{r})) \in \mathcal{P}$. La moyenne $\bar{\phi}$ d'une fonctionnelle $\phi(f(\mathbf{r}))$ est alors définie par :

$$\begin{aligned} \bar{\phi}(f(\mathbf{r})) &= \langle \phi(f(\mathbf{r})) \rangle = E\{\phi(f(\mathbf{r}))\} \\ &= \int_{\mathcal{C}} \phi(f(\mathbf{r})) p(f(\mathbf{r})) df(\mathbf{r}). \end{aligned}$$

Notons aussi qu'une fois le problème discrétisé, la fonction $f(\mathbf{r})$ est remplacée par un vecteur \mathbf{f} . Par la suite, nous ne distinguerons plus ces deux grandeurs.

Les mesures \bar{g}_m peuvent alors être interprétées, soit comme des valeurs moyennes des contraintes linéaires de l'ensemble des réalisations possibles du processus $F(\mathbf{r})$, soit comme des contraintes linéaires sur la moyenne de ces réalisations :

$$\begin{aligned} \bar{g}_m &= \langle g_m \rangle = \left\langle \int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} \right\rangle = \int_D \langle f(\mathbf{r}) \rangle h_m(\mathbf{r}) d\mathbf{r}, \\ & \quad m = 1, \dots, M. \end{aligned}$$

Dans cette approche le principe du maximum d'entropie est utilisé pour déterminer la loi de probabilité $p(f)$ qui satisfait ces contraintes et qui maximise

$$S(p(f)) = - \int_{\mathcal{C}} p(f) \ln p(f) df,$$

et puisque ces contraintes sont satisfaites exactement par l'image moyenne $\bar{f} = \langle f \rangle$, il est naturel de construire l'estimateur :

$$\hat{\bar{f}} = \langle f \rangle = E_{p(f)}\{f\} = \int_{\mathcal{C}} f p(f) df,$$

qui est la solution recherchée. Pour des discussions plus amples sur les propriétés de cette solution se référer à [[Gamboa89]].

Il est intéressant de noter que la solution $\hat{\bar{f}}$ dépend de la mesure de référence $df(\mathbf{r}) = \mu(f(\mathbf{r})) d\mathbf{r}$. En effet on montre que la solution $\hat{\bar{f}}$ se confond avec la solution du problème d'optimisation suivant :

$$\text{minimiser } \int_D F[\bar{f}(\mathbf{r})] d\mathbf{r}$$

sous les contraintes $\int_D \bar{f}(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} = \bar{g}_m, m = 1, \dots, M.$

où la fonction $F(\cdot)$ est la fonction convexe conjuguée de la transformée de Laplace (TL) de la mesure de référence $\mu(\cdot)$. D'un point de vue théorique on peut noter que suivant que l'on choisit une mesure gaussienne sur \mathbf{R} , poissonnienne sur \mathbf{R}_+ ou une mesure de Lebesgue sur $[0, \infty[$ on obtient :

— mesure gaussienne sur \mathbf{R} :

$$\text{TL}[\mu(f)] \propto \exp[t^2/2] \quad (50)$$

On obtient alors $F(\bar{f}) = \bar{f}^2$, ce qui nous ramène à l'estimation au sens des moindres carrés.

— mesure poissonnienne sur \mathbf{R}_+ :

$$\text{TL}[\mu(f)] \propto \exp[t - 1] \quad (51)$$

On obtient alors $F(\bar{f}) = \bar{f} \ln \bar{f}$, ce qui nous ramène à la minimisation de l'entropie de Shannon.

— mesure de Lebesgue sur $[0, \infty[$:

$$\text{TL}[\mu(f)] \propto \frac{-1}{t} \quad (52)$$

On obtient alors $F(\bar{f}) = -\ln \bar{f}$, ce qui nous ramène à la minimisation de l'entropie de Burg.

5.1. MAXENT 2.1

Il s'agit ici de la mise en œuvre classique de la méthode de maximum d'entropie sur la moyenne.

— **Hypothèse** : $f(\mathbf{r})$ est une réalisation d'un processus stochastique $F(\mathbf{r})$ ou, en d'autres termes, $f(\mathbf{r})$ est une fonction aléatoire avec une loi de probabilité $p(f) df$, et nous sommes seulement intéressés par l'estimation de sa moyenne $\bar{f}(\mathbf{r}) = \langle f(\mathbf{r}) \rangle$.

— **Données** : Les données sont

$$\begin{aligned} \bar{g}_m &= \left\langle \int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} \right\rangle = \int_D \langle f(\mathbf{r}) \rangle h_m(\mathbf{r}) d\mathbf{r}, \\ & \quad m = 1, \dots, M. \end{aligned} \quad (53)$$

où $\langle \cdot \rangle$ signifie l'espérance suivant la loi $p(f) df$. Les données peuvent être considérées, soit comme des valeurs moyennes des contraintes linéaires de l'ensemble des réalisations possibles du processus stochastique F , soit comme des contraintes linéaires sur la moyenne de ces réalisations $\langle f \rangle$.

— **Problème mathématique :**

Étant donné $h_m(\mathbf{r}), \bar{g}_m, m = 1, \dots, M$, et le domaine D ,

Problème P2.1 :

$$\text{maximiser} \quad - \int_C p(f) \ln p(f) df,$$

sous les contraintes

$$\bar{g}_m = \int_D \langle f(\mathbf{r}) \rangle h_m(\mathbf{r}) d\mathbf{r}, \quad m = 1, \dots, M$$

— **Solution :** La solution $p(f)$ est donnée par :

$$p(f) = \frac{1}{Z} \exp \left[- \sum_{m=1}^M \lambda_m \int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} \right] \quad (54)$$

où les multiplicateurs de Lagrange $\lambda = [\lambda_1, \dots, \lambda_M]$ sont obtenus par la résolution du problème d'optimisation suivant :

Problème dual du P2.1 :

$$\hat{\lambda} = \arg \min_{\lambda} \{ D(\lambda) = \lambda^t \bar{g} - \ln Z(\lambda) \}, \quad (55)$$

avec

$$Z(\lambda) = \int_C \exp \left[- \sum_{m=1}^M \lambda_m \int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} \right] df,$$

et où $\bar{g} = \{\bar{g}_1, \dots, \bar{g}_M\}$.

Une fois $p(f)$ calculée la solution \hat{f} est

$$\hat{f} = \langle f \rangle = \int_C f p(f) df. \quad (56)$$

Notons que \hat{f} satisfait, par construction, exactement les contraintes. Notons aussi que \hat{f} est la solution du problème d'optimisation suivant :

$$\text{maximiser} \quad F(\bar{f})$$

sous les contraintes

$$\int_D \bar{f}(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} = \bar{g}_m, \quad m = 1, \dots, M.$$

où $F(\bar{f})$, suivant le choix de la mesure df , prend une des formes suivantes :

— **Mesure gaussienne**

$$F(\bar{f}) = - \int_D \bar{f}^2(\mathbf{r}) d\mathbf{r} \quad (57)$$

— **Mesure poissonnienne**

$$F(\bar{f}) = - \int_D \bar{f}(\mathbf{r}) \ln \bar{f}(\mathbf{r}) d\mathbf{r} \quad (58)$$

— **Mesure de Lebesgue**

$$F(\bar{f}) = + \int_D \ln \bar{f}(\mathbf{r}) d\mathbf{r} \quad (59)$$

— **Limitations :**

La limitation majeure de cette méthode est de ne pas prendre en compte le bruit d'une manière explicite. Une autre difficulté dans cette méthode est l'interprétation des données g_m qui sont des valeurs moyennes. En effet, dans certaines applications, comme par exemple la tomographie à émission de positons (TEP), la grandeur mesurée est une valeur moyenne (du nombre de photons reçus dans un intervalle de temps τ), ce qui donne un sens physique à cette interprétation. Dans d'autres applications, il est plus difficile de donner un sens physique à cette interprétation.

— **Références :** Voir ([Borwein91], [Gamboa89], [Le Besnerais91b], [Le Besnerais93], [Leahy86], [Navaza85], [Navaza86], [Smith79], [Dacunha90], [Csiszar91])

5.2. MAXENT 2.2

Dans cet algorithme on cherche à prendre en compte le bruit en relaxant la satisfaction exacte des contraintes. En effet, on remplace l'ensemble convexe défini par les contraintes exactes par un ensemble convexe plus large qui le contient [[Gamboa89]].

— **Hypothèse :** comme dans MaxEnt 2.1

— **Données :** Les données sont

$$g_m = \int_D \langle f(\mathbf{r}) \rangle h_m(\mathbf{r}) d\mathbf{r} + b_m = \bar{g}_m + b_m, \quad m = 1, \dots, M \quad (60)$$

où les b_m sont supposés être indépendants, identiquement distribués, centrés, avec des variances σ_m^2 connues.

— **Problème mathématique :**

Étant donné $h_m(\mathbf{r}), g_m, \sigma_m^2, m = 1, \dots, M$, et le domaine D ,

$$\text{minimiser} \quad \int_D p(f) \ln p(f) df + \frac{1}{2\alpha} \sum_m \frac{1}{\sigma_m^2} (g_m - \bar{g}_m)^2.$$

α joue ici le rôle d'un paramètre de régularisation. Pour pouvoir comparer cette méthode avec **MaxEnt 2.1**, on remarque que ce problème peut être reformulé comme suit :

$$\text{minimiser} \quad \int_D p(f) \ln p(f) df + \frac{1}{2\alpha} \|\xi\|^2$$

sous les contraintes $g_m - \bar{g}_m = \xi_m \sigma_m^2, m = 1, \dots, M$,

avec $\xi = [\xi_1, \dots, \xi_M]^t$.

Le problème se formule donc comme un problème d'optimisation sous contraintes et on montre [[Luenberger69]] que la solution est de la même forme que celle de **MaxEnt2.1**.

— **Solution** : La solution $p(f)$ est de la même forme que dans le cas précédent :

$$p(f) = \frac{1}{Z} \exp \left[- \sum_{m=0}^M \lambda_m \int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} \right] \quad (61)$$

Seule la manière de calculer les multiplicateurs de Lagrange $\hat{\lambda} = [\lambda_1, \dots, \lambda_M]$ est différente. En effet, on montre que la fonction duale $D(\lambda)$ dans ce cas est :

$$D(\lambda) = \lambda^t \bar{g} - \ln Z(\lambda) - \frac{\alpha}{2} \lambda^t \lambda, \quad (62)$$

qui ne diffère de la fonction duale du **MaxEnt 2.1** que par le terme quadratique $\lambda^t \lambda$.

Ici aussi, une fois $p(f)$ calculée la solution \hat{f} est

$$\hat{f} = \langle f \rangle = \int_C f p(f) df. \quad (63)$$

Cette solution est aussi la solution du problème d'optimisation suivant :

$$\text{minimiser } F(\bar{f}) + \frac{1}{2\alpha} \|\xi\|^2$$

sous les contraintes

$$\int_D \bar{f}(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} - \bar{g}_m = \xi_m \sigma_m^2, \quad m = 1, \dots, M,$$

où $F(\bar{f})$ est, comme dans le cas précédent, une fonctionnelle du type quadratique, entropie de Shannon ou entropie de Burg.

— **Limitations** :

Comme dans **MaxEnt 2.1** la difficulté majeure dans cette méthode est l'interprétation des données g_m qui sont des valeurs moyennes. D'autre part, la manière de prendre en compte du bruit sur des mesures est un peu artificielle.

— **Références** : Voir ([[Borwein91]], [[Gamboa89]], [[Le Besnerais91b]], [[Navaza85]], [[Navaza86]], [[Le Besnerais91b]], [[Le Besnerais91a]], [[Le Besnerais93]], [[Bercher93]], [[Smith79]])

5.3. MAXENT 2.3

Ici on fait l'hypothèse que le système de mesure nous fournit pour chaque point de mesure à la fois la valeur moyenne et la variance

d'une quantité qui est liée par une équation intégrale à la grandeur recherchée. L'objectif est, comme dans les deux cas précédents, la détermination de cette valeur moyenne.

— **Hypothèse** : comme dans MaxEnt 2.1

— **Données** :

$$\begin{cases} \bar{g}_m = \left\langle \int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} \right\rangle \\ \sigma_m^2 = \left\langle \left(\int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} - g_m \right)^2 \right\rangle, \\ m = 1, \dots, M. \end{cases} \quad (64)$$

— **Problème mathématique** :

Étant donné $h_m(\mathbf{r}), g_m, \sigma_m^2, m = 1, \dots, M$, le domaine D et une loi *a priori* $p_0(f)$

$$\text{maximiser } - \int_D p(f) \ln \left(\frac{p(f)}{p_0(f)} \right) df,$$

sous les contraintes (64).

— **Solution** : La solution $p(f)$ est donnée par

$$p(f) = p_0(f) \frac{1}{Z} \left[- \sum_{m=1}^M \mu_m \int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} - \lambda_m \left(\int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} - g_m \right)^2 \right], \quad (65)$$

qui peut aussi se mettre sous la forme

$$p(f) \propto p_0(f) \exp[-Q(f)]$$

avec

$$Q(f) = \sum_{m=1}^M \frac{1}{2\sigma_m^2} \left(\int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} - \gamma_m \right)^2, \quad (66)$$

où

$$\sigma_m^2 = \frac{1}{2\lambda_m} \text{ et } \gamma_m^2 = g_m - \frac{\mu_m}{2\lambda}.$$

Notons que si $p_0(f)$ est uniforme ($p_0(f) = 1$) ou gaussienne, alors $p(f)$ est gaussienne.

Quand $p_0(f)$ est uniforme (sur un intervalle) on obtient

$$p(f) \propto \exp[-\chi^2(f)],$$

$$\text{avec } \chi^2(f) = \sum_{m=1}^M \frac{1}{\sigma_m^2} \left(\int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} - g_m \right)^2,$$

et

$$\hat{f} = E\{f\} = \arg \min_f \{\chi^2(f)\} \quad (67)$$

qui n'est autre que la solution au sens des moindres carrés.

Quand $p_0(f)$ est gaussienne $\mathcal{N}(f_0, \sigma_f^2)$, on a

$$p(f) \propto \exp \left[- \left\{ \chi^2(f) + \left(\frac{f - f_0}{\sigma_f} \right)^2 \right\} \right],$$

et, par conséquent

$$\hat{f} = \mathbf{E}\{f\} = \arg \min_f \left\{ \chi^2(f) + \left(\frac{f - f_0}{\sigma_f} \right)^2 \right\}. \quad (68)$$

qui n'est autre que la solution au sens des moindres carrés généralisé ou bien la solution régularisée.

— **Limitations :**

La difficulté majeure dans cette méthode réside dans le fait que, sauf dans les cas où $p_0(f)$ est uniforme ou gaussienne, la loi $p(f)$ n'a pas de forme analytique explicite.

— **Références :** Voir ([Macaulay85])

— La notion de probabilité dans cette approche n'est pas forcément liée à la fréquence de réalisation d'une variable aléatoire. On attribue une loi de probabilité à un paramètre pour traduire notre état de connaissance sur ce paramètre. Par exemple, si on connaît parfaitement la valeur d'un paramètre on lui attribue une loi de probabilité très concentrée sur sa valeur (une distribution de Dirac), en revanche pour représenter une connaissance plus diffuse on lui attribue une loi de probabilité plus uniforme (une gaussienne avec un grand écart-type par exemple).

— Alors qu'il est habituel d'attribuer une distribution de probabilité aux grandeurs mesurées pour traduire l'existence du bruit sur ces grandeurs, il est plus inhabituel d'attribuer une distribution de probabilité *a priori* aux paramètres inconnus du problème. Ceci peut être fait soit par une modélisation physique du problème, comme par exemple en mécanique quantique, soit de manière complètement subjective. Pour éclaircir ce point considérons les deux exemples suivants :

1. On veut attribuer une loi de probabilité à une variable qui représente la position X d'une molécule dans un tube de longueur L . Physiquement on sait qu'il est improbable que X soit en dehors de l'intervalle $[0, L]$. Si on ne connaît pas d'autre information sur X on peut lui attribuer une loi uniforme :

$$p(x) = \begin{cases} 1/L, & \text{si } x \in [0, L] \\ 0, & \text{sinon} \end{cases}$$

2. On veut attribuer une loi de probabilité à une variable X qui représente la valeur de la différence de potentiel entre deux bornes d'une prise (tension domestique). Ici, nous savons que la moyenne de X est 220 volts, et qu'il est peu probable que sa valeur s'écarte de cette valeur moyenne de 10 volts. Si on ne connaît pas d'autre information sur X on peut lui attribuer une loi gaussienne $\mathcal{N}(220, 10^2)$.

Une question peut se poser : pourquoi choisir une gaussienne plutôt qu'une autre loi avec la même moyenne et la même variance? La réponse à cette question peut être apportée par le principe du maximum d'entropie.

En effet, 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 (au sens de Shannon) supplémentaire.

6. Algorithmes de MaxEnt bayésienne

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 :

- Attribuer une distribution de probabilité *a priori* aux paramètres à estimer pour traduire notre connaissance initiale sur ces paramètres.
- Attribuer une distribution de probabilité aux grandeurs mesurées pour traduire l'imprécision sur ces données (bruit de mesure).
- 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.
- Définir une règle de décision pour déterminer les valeurs des paramètres à estimer.

Il faut noter cependant que :

- Cette approche ne peut être effectivement implantée et utilisée que dans un problème qui est décrit par un nombre fini de paramètres (par exemple une fois que le problème a été discrétisé).

Mathématiquement, le principe du maximum d'entropie peut être utilisé si notre connaissance *a priori* est de la forme

$$E \{ \phi_k(f) \} = \int \phi_k(f) p(f|I) df = \mu_k, \quad k = 1, \dots, K. \quad (69)$$

On a alors à résoudre

$$\text{maximiser} \quad - \int p(f|I) \ln p(f|I) df,$$

$$\text{sous les contraintes} \quad E \{ \phi_k(f) \} = \mu_k, \quad k = 1, \dots, K,$$

ce qui conduit, comme nous l'avons vu dans la première partie de cet article, à cette solution :

$$p(f|I) \propto \exp \left[- \sum_{k=1}^K \lambda_k \phi_k(f) \right]. \quad (70)$$

Les données g_m sont supposées être reliées à $f(\mathbf{r})$ de manière stochastique ce qui permet de définir des lois conditionnelles $p(g_m|f)$, mais mesurées de manière indépendante, ce qui permet de calculer

$$p(g_1, \dots, g_N | f) = \prod_{m=1}^M p(g_m | f).$$

Cette loi traduit ainsi notre connaissance incomplète sur le modèle (2) et notre incertitude sur les mesures (bruit de mesure). Dans le cas particulier d'un modèle linéaire déterministe avec bruit additif :

$$g_m = \int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} + b_m, \quad m = 1, \dots, M,$$

la loi conditionnelle $p(g_m|f)$ n'est autre que la loi de b_m translatée. Si les b_m sont indépendants, centrés, et de variances σ_m^2 , on obtient

$$p(g_1, \dots, g_N | f) \propto \exp [-\chi^2(f)]$$

avec

$$\chi^2(f) = \sum_{m=1}^M \frac{1}{\sigma_m^2} \left(\int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} - g_m \right)^2 \quad (71)$$

L'étape suivante est alors d'utiliser l'*a priori* $p(f|I)$, et la vraisemblance $p(g_1, \dots, g_m|f)$ pour calculer la loi *a posteriori* en utilisant la règle de Bayes :

$$p(f|g_1, \dots, g_m; I) \propto p(g_1, \dots, g_m|f) p(f|I).$$

Remplaçant les relations (70) et (71) dans cette dernière conduit à

$$p(f|g_1, \dots, g_N; I) \propto \exp [-J(f)]$$

avec

$$J(f) = \chi^2(f) + \sum_{k=1}^K \lambda_k \phi_k(f), \quad (72)$$

qui contient toute l'information objective sur la solution.

Cependant, en pratique, nous devons extraire de cette loi de probabilité une valeur pour la solution. Il convient alors de définir une règle de décision qui, partant de la loi *a posteriori*, conduise à une valeur ponctuelle. On peut ainsi choisir la moyenne *a posteriori* (MP) qui minimiserait la moyenne d'une fonction de coût quadratique, ou le maximum *a posteriori* (MAP). Dans le cas où la loi *a posteriori* est une loi gaussienne ces différents estimateurs ponctuels sont tous identiques. Mais dans le cas où cette loi n'est pas gaussienne on peut choisir parmi ces estimateurs en fonction de sa forme; par exemple si la loi est non-symétrique mais unimodale le MAP est plus justifié, par contre si la loi est très étalée la MP peut être préférée.

Dans les méthodes que nous allons présenter par la suite, la loi *a posteriori* est souvent unimodale et bien centrée autour de son maximum; c'est pourquoi le MAP convient bien, et dans ce cas on a :

$$\hat{f} = \arg \min_f \left\{ J(f(\mathbf{r})) = \chi^2(f(\mathbf{r})) + \sum_{k=1}^K \lambda_k \phi_k(f(\mathbf{r})) \right\}. \quad (73)$$

6.1. MAXENT 3.1

Il s'agit ici d'une interprétation bayésienne de la méthode de régularisation classique ou bien *le cas gaussien* dans l'approche bayésienne. Notons cependant que la propriété gaussienne est ici une conséquence de l'utilisation du principe du maximum d'entropie et non pas une hypothèse.

— **Hypothèse :** Notre connaissance *a priori* sur $f(\mathbf{r})$ est de la forme

$$\begin{cases} E \{ f(\mathbf{r}) \} & = f_0(\mathbf{r}) \\ E \{ (f(\mathbf{r}) - f_0(\mathbf{r}))^2 \} & = \sigma_f^2(\mathbf{r}) \end{cases} \quad (74)$$

— **Données :** Les données g_m sont reliées à $f(\mathbf{r})$ par

$$g_m = \int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} + b_m = \bar{g}_m + b_m, \quad m = 1, \dots, M,$$

où les b_m sont supposés indépendants, centrés et de variances σ_m^2 , $m = 1, \dots, M$ connues.

— **Problème mathématique :** Étant donné $f_0(\mathbf{r}), \sigma_f^2(\mathbf{r}); g_m, \sigma_m^2, m = 1, \dots, M$, déterminer $f(\mathbf{r})$.

— **Solution :**

- L'utilisation du principe du maximum d'entropie pour les contraintes (74) conduit à

$$\begin{aligned} p(f | I) &\propto \exp[-\lambda_1 f(\mathbf{r}) - \lambda_2 f^2(\mathbf{r})] \\ &\propto \exp\left[-\frac{1}{2} \left(\frac{f(\mathbf{r}) - f_0(\mathbf{r})}{\sigma_f(\mathbf{r})}\right)^2\right] \\ &= \mathcal{N}(f_0(\mathbf{r}), \sigma_f^2(\mathbf{r})). \end{aligned} \quad (75)$$

- La connaissance de $\sigma_m^2, m = 1, \dots, M$ nous donne

$$p(g_1, \dots, g_N | f) \propto \exp\left[-\frac{1}{2} \chi^2(f)\right],$$

avec

$$\begin{aligned} \chi^2(f) &= \sum_{m=1}^M \frac{1}{\sigma_m^2} \left(\int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} - g_m \right)^2 \\ &= \sum_{m=1}^M \frac{1}{\sigma_m^2} (\bar{g}_m - g_m)^2. \end{aligned}$$

En notant

$$\bar{\mathbf{g}} = [\bar{g}_1, \dots, \bar{g}_M]^t \quad \text{et} \quad \Sigma = \text{diag}[\sigma_1^2, \dots, \sigma_M^2],$$

on obtient

$$\begin{aligned} p(g_1, \dots, g_N | f) \\ = \mathcal{N}(\bar{\mathbf{g}}, \Sigma) \propto \exp\left[-\frac{1}{2} [\mathbf{g} - \bar{\mathbf{g}}]^t \Sigma^{-1} [\mathbf{g} - \bar{\mathbf{g}}]\right]. \end{aligned}$$

- L'utilisation de la règle de Bayes permet d'obtenir la loi *a posteriori*

$$p(f | g_1, \dots, g_N; I) \propto \exp\left[-\frac{1}{2} J(f)\right]$$

avec
$$J(f) = \chi^2(f) + \left(\frac{f - f_0}{\sigma_f}\right)^2,$$

qui est une loi gaussienne.

- Finalement, la solution MAP s'obtient par

$$\hat{f} = \arg \max_f p(f | g_1, \dots, g_N; I) = \arg \min_f \{J(f)\}$$

ou bien encore

$$\hat{f}(\mathbf{r}) = \arg \min_f J(f)$$

avec
$$J(f) = \sum_{m=1}^M \frac{1}{\sigma_m^2} \left(\int_D f(\mathbf{r}) h_m(\mathbf{r}) d\mathbf{r} - g_m \right)^2 + \left(\frac{f(\mathbf{r}) - f_0(\mathbf{r})}{\sigma_f(\mathbf{r})}\right)^2. \quad (76)$$

On retrouve l'estimation en moyenne quadratique dans le cas gaussien. Notons que ce résultat est mathématiquement équivalent à celui obtenu en **MaxEnt 2.2** dans l'équation (68).

- **Limitations :** La principale limitation dans l'utilisation pratique de cette méthode est la connaissance *a priori* de $f_0(\mathbf{r})$ et $\sigma_f^2(\mathbf{r})$. Cependant en paramétrant ces deux fonctions (par exemple $f_0(\mathbf{r}) = f_0$ et $\sigma_f^2(\mathbf{r}) = \sigma_f^2$) on peut les estimer soit à partir d'une image type ou bien même à partir des données.

- **Références :** Voir ([Macaulay85], [Gull84], [Djafari89], [Djafari90a], [Djafari90b])

6.2. MAXENT 3.2

Il s'agit là d'une interprétation bayésienne de l'algorithme de maximum d'entropie le plus utilisé dans les applications grâce au logiciel commercialisé par Gull et Skilling et leurs collaborateurs.

- **Hypothèse :** Notre connaissance *a priori* sur $f(\mathbf{r})$ est de la forme

$$E \left\{ S(f, f_0) = - \int_D f(\mathbf{r}) \ln \left(\frac{f(\mathbf{r})}{f_0(\mathbf{r})} \right) d\mathbf{r} \right\} = s. \quad (77)$$

- **Données :** comme dans MaxEnt 3.1

- **Problème mathématique :** Étant donné $f_0(\mathbf{r}), s; g_m, \sigma_m^2, m = 1, \dots, M$, déterminer $f(\mathbf{r})$

- **Solution :**

- Le principe du maximum d'entropie nous donne :

$$p(f | I) \propto \exp\left[-\frac{\lambda}{2} S(f, f_0)\right]. \quad (78)$$

- La connaissance de $\sigma_m^2, m = 1, \dots, M$ fournit

$$p(g_1, \dots, g_N | f) \propto \exp\left[-\frac{1}{2} \chi^2(f)\right].$$

- Utilisant la règle de Bayes on obtient

$$p(f | g_1, \dots, g_N; I) \propto \exp\left[-\frac{1}{2} J(f)\right] \quad \text{avec} \quad J(f) = \chi^2(f) + \lambda S(f, f_0). \quad (79)$$

— Finalement, la solution MAP est

$$\hat{f} = \arg \min_f \{J(f) = \chi^2(f) + \lambda S(f, f_0)\}. \quad (80)$$

Notons que ce résultat est mathématiquement équivalent à celui obtenue en **MaxEnt 1.3**. Cependant, notons que dans **MaxEnt 1.3** le problème était d'obtenir une solution à :

$$\text{minimiser } S(f, f_0) \text{ sous la contrainte } \chi^2 = c,$$

qui est

$$\hat{f} = \arg \min_f \{\chi^2 + \lambda S(f, f_0)\}.$$

Il est important de noter que bien que les deux méthodes aboutissent au même algorithme et au même résultat numérique, l'interprétation de ces résultats n'est pas identique. Dans **MaxEnt 1.3**, λ est un multiplicateur de Lagrange dont la valeur dépend de c , tandis que dans **MaxEnt 3.2**, λ est un paramètre de la loi *a priori* $p(f | I)$ dont la valeur dépend de s . Dans **MaxEnt 1.3**, une fois la solution calculée, on ne peut rien dire de plus sur cette solution, tandis que dans **MaxEnt 3.2** nous disposons d'une loi de probabilité (*a posteriori*), ce qui nous permet théoriquement de calculer n'importe quelle grandeur statistique (la matrice de covariance, par exemple pour mettre des barres d'erreurs) sur la solution.

— **Limitations** : La limitation majeure ici est l'interprétation de s , et, par conséquent, la détermination du paramètre λ .

— **Références** : Voir ([Gull84]), [Djafari89], [Djafari90a], [Djafari90b])

6.3. MAXENT 3.3

Il s'agit ici d'une extension de ce qui précède pour permettre une famille plus riche de lois *a priori*. C'est la méthode développée et utilisée par l'auteur et ses collaborateurs dans un grand nombre d'applications en restauration et reconstruction d'image. En effet, la loi *a priori* obtenue dans le cas précédent est un peu trop restreinte. L'idée de base étant de chercher des lois ayant une propriété d'invariance par changement d'échelle comme nous le précisons dans l'annexe B. Ici nous nous limiterons au cas des lois avec deux paramètres.

— **Hypothèse** : Notre connaissance *a priori* sur $f(\mathbf{r})$ est de la forme

$$\begin{cases} E \left\{ \phi_1(f) = \int_D \phi_1(f(\mathbf{r})) d\mathbf{r} \right\} & = s \\ E \left\{ \phi_2(f) = - \int_D \phi_2(f(\mathbf{r})) d\mathbf{r} \right\} & = h \end{cases} \quad (81)$$

où $\phi_1(f)$ et $\phi_2(f)$ sont deux fonctionnelles connues. Nous précisons dans un instant comment choisir ces fonctionnelles pour que la loi *a priori* qui en découle, en utilisant le principe du maximum d'entropie (PME), ait cette propriété d'invariance par changement d'échelle.

— **Données** : comme dans MaxEnt 3.1

— **Problème mathématique** : Étant donné $\phi_1(f), \phi_2(f), s, h; g_m, \sigma_m^2, m = 1, \dots, M$, déterminer $f(\mathbf{r})$

— **Solution** :

— Utilisant le principe du ME, on a :

$$p(f | I) \propto \exp \left[-\frac{1}{2} (\lambda_1 \phi_1(f) + \lambda_2 \phi_2(f)) \right]. \quad (82)$$

— La connaissance de $\sigma_m^2, m = 1, \dots, M$ nous amène à

$$p(g_1, \dots, g_N | f) \propto \exp \left[-\frac{1}{2} \chi^2(f) \right].$$

— La règle de Bayes donne

$$p(f | g_1, \dots, g_N; I) \propto \exp \left[-\frac{1}{2} J(f) \right],$$

$$\text{avec } J(f) = \chi^2(f) + \lambda_1 \phi_1(f) + \lambda_2 \phi_2(f). \quad (83)$$

— Finalement, la solution MAP est

$$\hat{f} = \arg \max_f p(f | g_1, \dots, g_N; I) = \arg \min_f J(f) \quad (84)$$

Nous n'avons pas encore précisé les expressions de $\phi_1(f)$ et de $\phi_2(f)$. Dans une étude effectuée par l'auteur et ses collaborateurs [Djafari90b] nous avons discuté sur les différentes expressions de ces deux fonctionnelles si on désirerait avoir une loi *a priori* $p(f | I)$ qui soit invariante par changement d'échelle (voir l'annexe B et références [Djafari93a], [Djafari93b]). Nous avons trouvé que les différentes expressions possibles pour ces deux fonctionnelles $\{\phi_1(f), \phi_2(f)\}$ sont :

$$\begin{cases} \phi_1(f) = \int_D \phi_1(f(\mathbf{r})) d\mathbf{r} \\ \phi_2(f) = \int_D \phi_2(f(\mathbf{r})) d\mathbf{r} \end{cases}$$

avec

$$\begin{cases} \phi_1(f), \phi_2(f) \end{cases} = \begin{cases} (f^{r_1}, f^{r_2}), (f^{r_1}, \ln f), \\ (f^{r_1}, f^{r_1} \ln f), (\ln f, \ln^2 f) \end{cases}, \quad (85)$$

où r_1 et r_2 sont deux réels.

Considérons maintenant le cas où $\phi_1(f) = f$ et, par conséquent $\phi_2(f)$ est :

$$\{f^r, \ln f, f \ln f\}.$$

Parmi ces différents choix notons les trois cas suivants :

— $\phi_1(f) = f, \phi_2(f) = f^2 :$

Dans ce cas on a :

$$\begin{cases} \phi_1(f) = \int_D f(\mathbf{r}) \, d\mathbf{r} \\ \phi_2(f) = - \int_D f^2(\mathbf{r}) \, d\mathbf{r} \end{cases}$$

et on retrouve le cas gaussien.

— $\phi_1(f) = f, \phi_2(f) = \ln f :$

Dans ce cas on a :

$$\begin{cases} \phi_1(f) = \int_D f(\mathbf{r}) \, d\mathbf{r} \\ \phi_2(f) = \int_D \ln f(\mathbf{r}) \, d\mathbf{r} \end{cases}$$

et on retrouve le cas de la loi Béta (ou l'expression de l'entropie de Burg [[Burg67]]).

— $\phi_1(f) = f, \phi_2(f) = -f \ln f$

Dans ce cas on a :

$$\begin{cases} \phi_1(f) = \int_D f(\mathbf{r}) \, d\mathbf{r} \\ \phi_2(f) = - \int_D f(\mathbf{r}) \ln f(\mathbf{r}) \, d\mathbf{r} \end{cases}$$

et on retrouve le cas de la loi dite *entropique* de Gull et Skilling [[Gull84]].

Dans ce dernier cas la solution MAP est :

$$\hat{f} = \arg \min_f \left\{ J(f) = \chi^2(f) + \lambda_1 \int_D f(\mathbf{r}) \, d\mathbf{r} + \lambda_2 \int_D f(\mathbf{r}) \ln f(\mathbf{r}) \, d\mathbf{r}, \right\} \quad (86)$$

qui peut aussi être écrite sous la forme

$$\hat{f} = \arg \min_f \left\{ J(f) = \chi^2(f) + \lambda \int_D f(\mathbf{r}) \ln \left(\frac{f(\mathbf{r})}{f_0(\mathbf{r})} \right) \, d\mathbf{r}. \right\} \quad (87)$$

On peut aussi constater que ce dernier résultat est mathématiquement similaire au résultat de **MaxEnt 1.3** ou **MaxEnt 3.2**.

— **Limitations** : La limitation majeure ici, comme dans le cas de MaxEnt 3.1, est dans l'interprétation de s et de h , et, par conséquent, dans la détermination des paramètres λ et μ . Cependant l'approche bayésienne nous donne le moyen d'inférer sur ces paramètres. En effet, en considérant λ et μ comme les deux paramètres de la loi *a priori* dont la forme est connue et donnée par (82), on peut les estimer à partir d'une image type ou bien *a posteriori* à partir des données.

— **Références** : Voir ([[Djafari89]], [[Djafari90a]], [[Djafari90b]], [[Djafari91b]], [[Djafari91a]], [[Skilling84b]])

7. Conclusions

Dans ce travail nous avons essayé de fournir une vue synthétique sur l'usage du principe du maximum d'entropie dans la résolution des problèmes inverses. Nous avons tout d'abord distingué trois approches différentes :

- MaxEnt classique,
- MaxEnt sur la moyenne, et
- l'approche bayésienne avec les lois *a priori* à maximum d'entropie.

Dans l'approche MaxEnt classique, la fonction $f(\mathbf{r})$ est considérée comme une fonction densité de probabilité et les données g_m sont supposées être des contraintes (exactes) sur cette fonction. L'objectif est alors de choisir parmi toutes les fonctions (positives) possibles celle qui est la plus uniforme, c'est-à-dire celle qui a l'entropie maximale, ou bien celle qui minimise l'entropie relative entre cette fonction et une fonction de référence $f_0(\mathbf{r})$ (donnée *a priori*).

Dans l'approche MaxEnt sur la moyenne, le problème est l'estimation de la moyenne d'une fonction aléatoire $f(\mathbf{r})$ sur laquelle nous disposons des données g_m qui peuvent être considérées, soit comme des contraintes linéaires sur la moyenne $\langle f(\mathbf{r}) \rangle$, soit comme des valeurs moyennes des contraintes linéaires sur la fonction même. L'objectif est alors de déterminer la loi de probabilité $p(f)$ de cette fonction aléatoire, puis de calculer sa moyenne. La loi de probabilité $p(f)$ étant choisie parmi toutes les lois possibles comme celle qui a l'entropie maximale ou celle qui minimise son entropie relative par rapport à une loi de probabilité de référence $p_0(f)$ (donnée *a priori*). Ce qui est intéressant dans cette approche est que l'estimateur de la moyenne peut être considéré comme la solution d'un problème d'optimisation d'une fonction entropique de la moyenne sous des contraintes linéaires.

Dans l'approche bayésienne le problème est encore l'estimation d'une fonction sur laquelle nous disposons de deux types d'information; les données g_m qui sont les mesures bruitées, et une information *a priori* I sur la fonction sous forme des contraintes sur sa loi de probabilité (objective ou subjective).

Le principe du maximum d'entropie est utilisé seulement dans la première étape qui est la détermination de la loi *a priori* $p(f|I)$ représentant l'état de notre connaissance *a priori* I sur la fonction, et de la loi $p(g|f)$ représentant l'état de notre incertitude sur les données g .

Dans chacune de ces trois approches nous avons fait l'inventaire des différentes méthodes et les différents algorithmes d'implantation numérique, en essayant de décrire leurs hypothèses explicites et implicites et leurs limitations théoriques et pratiques. Nous renvoyons également aux principales références bibliographiques.

Pour conclure, lorsque nous sommes confrontés à un problème inverse, avant de choisir une démarche pour le résoudre, il est important de commencer par répondre à des questions élémentaires sur la **modélisation du problème** :

- Quelle quantité physique représente la grandeur inconnue?
- Quelle quantité physique représentent les données?
- Que savons-nous sur les erreurs de mesures?
- Que savons-nous sur la fonction inconnue?

Une fois que l'on dispose d'éléments de réponses sur ces questions, pour le **choix d'une méthode d'inversion**, nous devons aussi répondre à des questions du type :

- Quelles sont les hypothèses (implicites ou explicites) sur la fonction inconnue?
- Quelles sont les hypothèses sur le lien entre la fonction inconnue et les mesures?
- Que cherchons-nous sur la fonction inconnue?
- Que fournit la méthode? Est-elle capable de fournir une solution stable? Quelle est alors la sensibilité de cette solution vis-à-vis des erreurs?

Comme nous l'avons vu, il est possible que, partant de deux approches apparemment différentes, les algorithmes finaux soient absolument identiques; mais ceci ne signifie, en aucun cas, que l'interprétation que nous puissions donner à ces résultats soit la même.

Finalement, pour donner l'opinion de l'auteur sur ces trois approches, nous pensons que l'approche bayésienne peut répondre à toutes les questions bien-posées qu'on peut rencontrer lors de l'inversion d'un problème inverse.

8. Remerciements

Je tiens à remercier mes collègues J.-F. Bercher, G. Demoment, J. Idier, et G. Le Besnerais pour leurs lectures attentives et leurs remarques qui ont été bénéfiques pour la rédaction de cet article. Je tiens aussi à remercier les rapporteurs anonymes pour leurs travail en profondeur et pour leur critiques qui m'ont permis d'améliorer à la fois la présentation et le contenu de cet article, et qui m'ont montré aussi à quel point le sujet *entropie* reste passionnant. J'espère par ailleurs que la parution de cet article encourage les chercheurs qui travaillent sur ce sujet, à faire un effort de publication, surtout en utilisant la langue Française.

ANNEXES

1. Technique de multiplicateurs de Lagrange

Considérons le problème P1 qui est au cœur des problèmes de maximisation de l'entropie.

Problème P1 :

$$\text{maximiser } S(\mathbf{p}) = - \sum_{j=1}^n p_j \ln p_j,$$

$$\text{sous les contraintes } \sum_{j=1}^n p_j \phi_i(x_j) = d_i, \quad i = 1, \dots, m.$$

Ce problème peut être considéré comme un cas particulier du problème suivant :

Problème P0 :

$$\text{minimiser } F(\mathbf{x})$$

$$\text{sous les contraintes } \mathbf{Ax} = \mathbf{y},$$

$$\text{où } \mathbf{x} \in \mathcal{C} \subset \mathbf{R}^n, \quad \mathbf{y} \in \mathbf{R}^m, \quad F : \mathcal{C} \mapsto \mathbf{R}.$$

On appelle Lagrangien de P0 la fonction

$$\begin{aligned} L(\mathbf{x}, \boldsymbol{\lambda}) &= F(\mathbf{x}) - \boldsymbol{\lambda}^t (\mathbf{Ax} - \mathbf{y}) = F(\mathbf{x}) - \boldsymbol{\lambda}^t \mathbf{Ax} + \boldsymbol{\lambda}^t \mathbf{y} \\ &= F(\mathbf{x}) - (\boldsymbol{\lambda}^t \mathbf{A})\mathbf{x} + \boldsymbol{\lambda}^t \mathbf{y}, \end{aligned}$$

où $\boldsymbol{\lambda}$ est appelé variable duale de \mathbf{x} ou vecteur des multiplicateurs de Lagrange du problème P0.

On appelle point-selle de $L(x, \lambda)$ le point $(\hat{x}, \hat{\lambda})$ tel que

$$(\hat{x}, \hat{\lambda}) = \sup_{\lambda} \inf_x L(x, \lambda) = \inf_x \sup_{\lambda} L(x, \lambda),$$

et on montre que sous certaines conditions sur F (voir [[Luenberger69]]), l'ensemble des premiers arguments des points-selles de L coïncide avec l'ensemble des solutions du problème P0.

Ce résultat nous permet de rechercher la solution du problème P0 de manière suivante :

Problème Q0 :

$$\begin{cases} \hat{\lambda} = \arg \max_{\lambda} \{D(\lambda) = \inf_x L(x, \lambda)\} \\ \hat{x} = \arg \min_x \{L(x, \hat{\lambda})\} \end{cases}$$

Le problème **Q0** est appelé le problème dual de **P0** et la fonction $D(\lambda)$ est appelée la fonction duale de $F(x)$.

Le gradient de L par rapport à x et par rapport à λ s'écrit :

$$\begin{cases} L'_x(x, \lambda) = F'(x) - \lambda^t A \\ L'_\lambda(x, \lambda) = Ax - y \end{cases}$$

En introduisant la fonction $G(s)$, que l'on appelle fonction convexe conjuguée de $F(x)$, et définie par :

$$G(s) = \sup_x \{x^t s - F(x)\},$$

on constate que, pour s fixé, G est maximale pour une valeur x_s telle que

$$s^t - F'(x_s) = 0,$$

et que

$$G'(s) = x_s.$$

Ainsi, en identifiant $s = \lambda^t A$, la solution au problème **P0** peut être obtenue par :

$$\begin{cases} \hat{\lambda} = \arg \max_{\lambda} \{D(\lambda) = \lambda^t y - G(\lambda^t A)\} \quad (\text{Problème Q0}) \\ \hat{x} = G'(\hat{\lambda}^t A) \end{cases}$$

Ainsi la procédure pour calculer la solution est :

— déterminer la fonction convexe conjuguée $G(s)$ de $F(x)$:

$$G(s) = \sup_x \{x^t s - F(x)\};$$

— résoudre le problème dual qui est un problème d'optimisation sans contrainte

$$\hat{\lambda} = \arg \max_{\lambda} \{D(\lambda) = \lambda^t y - G(\lambda^t A)\};$$

— calculer la solution

$$\hat{x} = G'(\hat{\lambda}^t A).$$

1.1. PROBLÈME BRUITÉ

Considérons maintenant le cas où nous ne souhaitons pas satisfaire exactement toutes les contraintes et considérons le problème suivant :

Problème P0b :

$$x = \arg \min_{x \in D} \left\{ F(x) + \frac{1}{2\alpha} \|y - Ax\|^2 \right\},$$

où α joue le rôle d'un coefficient de régularisation. Bien que ce problème soit sans contrainte, on peut le transformer en un problème sous contrainte, en notant qu'il est équivalent à

Problème P0b :

$$\text{minimiser } F(x) + \frac{1}{2\alpha} \|\xi\|^2,$$

sous les contraintes $y - Ax = \xi$.

Le Lagrangien \bar{L} du problème **P0b** s'écrit alors :

$$\begin{aligned} \bar{L}(x, \lambda, \xi) &= F(x) + \frac{1}{2\alpha} \|\xi\|^2 - \lambda^t (y - Ax - \xi) \\ &= L(x, \lambda) + \lambda^t \xi + \frac{1}{2\alpha} \|\xi\|^2, \end{aligned}$$

où $L(x, \lambda)$ est le Lagrangien du problème sans bruit P0. On montre alors facilement que la solution x_λ a la même expression que celle du problème sans bruit P0. Seule la valeur optimale de $\hat{\lambda}$ est calculée par

$$\hat{\lambda} = \arg \max_{\lambda} \{\bar{D}(\lambda)\},$$

où

$$\bar{D}(\lambda) = \inf_x L(x, \lambda) + \inf_{\xi} \left[\lambda^t \xi + \frac{1}{2\alpha} \|\xi\|^2 \right].$$

Le premier terme n'est autre que la fonction duale $D(\lambda)$ du problème sans bruit, tandis que la minimisation du second terme conduit à

$$\xi = y - Ax = \alpha \lambda.$$

Ceci conduit à une relation simple entre la fonction duale du problème avec bruit et celle du problème sans bruit :

$$\bar{D}(\lambda) = D(\lambda) + \frac{\alpha}{2} \|\lambda\|^2.$$

On peut généraliser ceci au cas du bruit coloré :

$$x = \arg \min_{x \in D} \left\{ F(x) + \frac{1}{2\alpha} [y - Ax]^t R_b^{-1} [y - Ax] \right\},$$

et montrer que la relation entre la fonction duale du problème avec bruit et celle du problème sans bruit est :

$$\bar{D}(\lambda) = D(\lambda) + \|\eta\|^2,$$

avec

$$\eta = \frac{\alpha}{2} \lambda^t R_b \lambda = \frac{\alpha}{2} \|\lambda^t \lambda\|_{R_b}^2.$$

1.2. CAS DES FONCTIONS ENTROPIQUES

Considérons maintenant le cas particulier où :

$$F(x) = \sum_j f(x_j).$$

Son Lagrangien s'écrit maintenant

$$\begin{aligned} L(x, \lambda) &= \sum_j f(x_j) - \lambda^t (Ax - y) \\ &= \sum_j [f(x_j) - x_j [\lambda^t A]_j] + \lambda^t y. \end{aligned}$$

En notant $g(s_j)$ les fonctions convexe conjuguées des $f(x_j)$, i.e.

$$g(s_j) = \sup_{x_j} \{x_j s_j - f(x_j)\},$$

on obtient :

$$D(\lambda) = \lambda^t y - \sum_j g([\lambda^t A]_j),$$

et

$$x_j = g'([\lambda^t A]_j).$$

On peut étendre ceci au cas où

$$F(x) = \int f(x(t)) dt,$$

avec des contraintes de la forme

$$\int a_i(t)x(t) dt = y_i, \quad i = 1, \dots, m.$$

Le Lagrangien s'écrit :

$$L(x, \lambda) = \int f(x(t)) - \sum_i \lambda_i \left(\int a_i(t)x(t) dt - y_i \right),$$

et on a

$$D(\lambda) = \lambda^t y - \int g \left(\sum_i \lambda_i a_i(t) \right) dt,$$

et

$$x(t) = g' \left(\sum_i \lambda_i a_i(t) \right),$$

où $g(\cdot)$ est la fonction convexe conjuguée de $f(\cdot)$:

$$g(s(t)) = \sup_{x(t)} \{x(t)s(t) - f(x(t))\}.$$

1.3. EXEMPLES

Exemple 1 :

$$\text{minimiser } \frac{1}{2} \sum_j x_j^2 \quad \text{s.c. } Ax = y.$$

Le Lagrangien s'écrit :

$$L(x, \lambda) = \frac{1}{2} x^t x - \lambda^t Ax + \lambda^t y.$$

Les fonctions $f(x)$, $g(s)$ et $g'(s)$ sont :

$$f(x) = \frac{1}{2} x^2,$$

$$g(s) = \sup_x \{sx - f(x)\} = \frac{1}{2} s^2, \quad \text{et } g'(s) = s$$

La fonction duale est :

$$D(\lambda) = \lambda^t y - \sum_j g([\lambda^t A]_j) = \lambda^t y - \frac{1}{2} [\lambda^t A^t A \lambda],$$

et la solution est :

$$\hat{\lambda} = \arg \max_{\lambda} \{D(\lambda)\} = [A^t A]^{-1} y.$$

La solution du problème d'optimisation est :

$$x_j = g'([\hat{\lambda}^t A]_j) = [[A^t A]^{-1} A^t y]_j,$$

ce qui, sous forme vectorielle s'écrit :

$$\hat{x} = [A^t A]^{-1} A^t y.$$

On note que ceci correspond à la solution inverse généralisée de l'équation $y = Ax$ si A est de rang plein.

Exemple 2 :

$$\text{minimiser } \sum_j \ln x_j \quad \text{s.c. } Ax = y.$$

Le Lagrangien s'écrit :

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \sum_j \ln x_j - \boldsymbol{\lambda}^t \mathbf{A} \mathbf{x} + \boldsymbol{\lambda}^t \mathbf{y}.$$

Les fonctions $f(x)$, $g(s)$ et $g'(s)$ sont :

$$f(x) = \ln x,$$

$$g(s) = \sup_x \{xs - f(x)\} = 1 - \ln s, \quad \text{et} \quad g'(s) = \frac{-1}{s}$$

La fonction duale est :

$$D(\boldsymbol{\lambda}) = \boldsymbol{\lambda}^t \mathbf{y} - \sum_j g([\boldsymbol{\lambda}^t \mathbf{A}]_j) = \boldsymbol{\lambda}^t \mathbf{y} - \sum_j (1 - \ln[\boldsymbol{\lambda}^t \mathbf{A}]_j),$$

et la solution s'obtient en résolvant :

$$\hat{\boldsymbol{\lambda}} = \arg \max_{\boldsymbol{\lambda}} \{D(\boldsymbol{\lambda})\}.$$

La solution s'écrit :

$$\hat{x}_j = g'([\hat{\boldsymbol{\lambda}}^t \mathbf{A}]_j) = \frac{1}{[\hat{\boldsymbol{\lambda}}^t \mathbf{A}]_j}$$

Notons que la solution $\hat{\boldsymbol{\lambda}}$ s'obtient en résolvant le système d'équations non linéaires suivant :

$$\sum_j a_{ij}^2 g''([\boldsymbol{\lambda}^t \mathbf{A}]_j) = y_i, \quad i = 1, \dots, M.$$

Exemple 3 :

$$\text{minimiser } \sum_j (x_j \ln x_j) \quad \text{s.c.} \quad \mathbf{A} \mathbf{x} = \mathbf{y}, \quad \text{et} \quad \sum_j x_j = 1.$$

Le Lagrangien s'écrit :

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \sum_j (x_j \ln x_j) - \boldsymbol{\lambda}^t \mathbf{A} \mathbf{x} + \boldsymbol{\lambda}^t \mathbf{y} - \lambda_0 \left(\sum_j x_j - 1 \right).$$

On montre facilement que ce problème est équivalent à

$$\text{minimiser } - \sum_j (x_j \ln x_j - x_j) \quad \text{s.c.} \quad \mathbf{A} \mathbf{x} = \mathbf{y}, \quad \text{et} \quad \sum_j x_j = 1,$$

dont le Lagrangien s'écrit :

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \sum_j (x_j \ln x_j - x_j) - \boldsymbol{\lambda}^t \mathbf{A} \mathbf{x} + \boldsymbol{\lambda}^t \mathbf{y} - \lambda_0 \left(\sum_j x_j - 1 \right).$$

Les fonctions $f(x)$, $g(s)$ et $g'(s)$ sont :

$$f(x) = -x \ln x + x,$$

$$g(s) = \sup_x \{xs - f(x)\} = \exp[s], \quad \text{et} \quad g'(s) = \exp[s]$$

La fonction duale est :

$$D(\boldsymbol{\lambda}) = \boldsymbol{\lambda}^t \mathbf{y} - \sum_j g([\boldsymbol{\lambda}^t \mathbf{A}]_j) = \boldsymbol{\lambda}^t \mathbf{y} - \sum_j \exp([\boldsymbol{\lambda}^t \mathbf{A}]_j),$$

et la solution est de la forme :

$$x_j = g'([\boldsymbol{\lambda}^t \mathbf{A}]_j) = \exp([\boldsymbol{\lambda}^t \mathbf{A}]_j).$$

On remarque que si on note

$$x_j = \frac{1}{Z(\boldsymbol{\lambda})} \exp([\boldsymbol{\lambda}^t \mathbf{A}]_j), \quad \text{avec} \quad Z(\boldsymbol{\lambda}) = \sum_j \exp([\boldsymbol{\lambda}^t \mathbf{A}]_j),$$

on a

$$D(\boldsymbol{\lambda}) = \boldsymbol{\lambda}^t \mathbf{y} - \ln Z(\boldsymbol{\lambda}),$$

et la solution

$$\hat{\boldsymbol{\lambda}} = \arg \max_{\boldsymbol{\lambda}} \{D(\boldsymbol{\lambda})\}$$

peut être calculée en résolvant le système d'équations

$$-\frac{\partial \ln Z(\boldsymbol{\lambda})}{\partial \lambda_i} = y_i, \quad i = 1, \dots, m.$$

Exemple 4 :

$$\text{minimiser } \int_C p(\mathbf{x}) \ln p(\mathbf{x}) \mu(\mathbf{x})$$

$$\text{s.c.} \quad \int_C p(\mathbf{x}) \mathbf{A} \mathbf{x} \mu(\mathbf{x}) = \mathbf{y}, \quad \text{et} \quad \int_C p(\mathbf{x}) \mu(\mathbf{x}) = 1.$$

Le Lagrangien s'écrit :

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \int_C p(\mathbf{x}) [\ln p(\mathbf{x}) - \boldsymbol{\lambda}^t \mathbf{A} \mathbf{x} - \lambda_0] \mu(\mathbf{x}) + \boldsymbol{\lambda}^t \mathbf{y}.$$

Les fonctions $f(x)$, $g(s)$ et $g'(s)$ sont :

$$f(p(\mathbf{x})) = p(\mathbf{x}) \ln p(\mathbf{x}),$$

$$g(s) = \sup_p \{ps - f(p)\} = \exp[-s - 1],$$

$$\text{et} \quad g'(s) = \exp[-s - 1].$$

La solution est de la forme

$$p(\mathbf{x}) = g'(\boldsymbol{\lambda}^t \mathbf{A} \mathbf{x}) = \frac{1}{Z(\boldsymbol{\lambda})} \exp[-\boldsymbol{\lambda}^t \mathbf{A} \mathbf{x}],$$

avec

$$Z(\boldsymbol{\lambda}) = \int_C \exp[-\boldsymbol{\lambda}^t \mathbf{A} \mathbf{x}] \mu(\mathbf{x}).$$

La fonction duale est :

$$D(\boldsymbol{\lambda}) = \boldsymbol{\lambda}^t \mathbf{y} - \ln Z(\boldsymbol{\lambda}),$$

et la solution

$$\hat{\lambda} = \arg \max_{\lambda} \{D(\lambda)\}$$

peut être calculée en résolvant le système d'équations :

$$-\frac{\partial \ln Z(\lambda)}{\partial \lambda_i} = y_i, \quad i = 1, \dots, m.$$

Pour illustrer le lien avec les méthodes de maximum d'entropie sur la moyenne, notons

$$\langle x \rangle = \int_{\mathcal{C}} x p(x) \mu(x).$$

On peut montrer alors la relation suivante :

$$-\frac{\partial \ln Z(\mathbf{A}^t \lambda)}{\partial (\mathbf{A}^t \lambda)} = \langle x \rangle.$$

Ceci nous permet de considérer $\langle x \rangle$ comme la solution du problème d'optimisation suivant

$$\text{maximiser } -F(\langle x \rangle) = -\sum_j f(\langle x_j \rangle),$$

$$\text{sous les contraintes } \mathbf{A} \langle x \rangle = \mathbf{y},$$

avec la fonction duale

$$D(\lambda) = \lambda^t \mathbf{y} - \ln Z(\mathbf{A}^t \lambda).$$

La fonction $F(\langle x \rangle)$ est la fonction convexe conjuguée de $\ln Z(\mathbf{A}^t \lambda)$.

$Z(\mathbf{A}^t \lambda)$ dépendant de la mesure $\mu(x)$, suivant que l'on choisit une mesure gaussienne, poissonnienne ou une mesure de Lebesgue on obtient :

— mesure gaussienne sur \mathbf{R}^n :

$$F(x) = \sum_{j=1}^n x_j^2,$$

qui est un critère quadratique;

— mesure poissonnienne sur \mathbf{R}_+^n :

$$F(x) = \sum_{j=1}^n x_j \ln x_j,$$

qui est un critère du type entropie de Shannon;

— mesure de Lebesgue sur $[0, \infty[^n$:

$$F(x) = -\sum_{j=1}^n \ln x_j,$$

qui est un critère du type entropie de Burg.

On peut ainsi considérer d'autres exemples. Le tableau qui suit résume et complète les résultats de ces exemples.

2. Invariance par changement d'échelle

2.1. PRINCIPE GÉNÉRAL

Considérons le problème linéaire suivant :

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{b}.$$

Nous avons vu que dans l'approche bayésienne pour résoudre ce problème nous devons commencer par l'attribution des lois $p(\mathbf{x})$ et $p(\mathbf{y}|\mathbf{x})$. Ensuite la règle de Bayes nous permet de calculer $p(\mathbf{x}|\mathbf{y})$ et, finalement, un estimateur ponctuel peut être défini en utilisant cette loi. D'une manière générale, on définit une fonction coût $C(\hat{\mathbf{x}}, \mathbf{x})$ et on minimise le coût moyen. C'est ainsi que l'on obtient, par exemple, les estimateurs

— bayésien (cas général)

$$\begin{aligned} \hat{\mathbf{x}} &= \arg \min_{\mathbf{z}} \{E_{X|Y} \{C(\mathbf{z}, \mathbf{x})\}\} \\ &= \arg \min_{\mathbf{z}} \left\{ \int C(\mathbf{z}, \mathbf{x}) p(\mathbf{x}|\mathbf{y}) d\mathbf{x} \right\}, \end{aligned}$$

— moyenne *a posteriori* (MP)

$$\hat{\mathbf{x}} = E_{X|Y} \{\mathbf{x}\} = \int \mathbf{x} p(\mathbf{x}|\mathbf{y}) d\mathbf{x},$$

en choisissant une fonction coût quadratique $C(\hat{\mathbf{x}}, \mathbf{x}) = (\hat{\mathbf{x}} - \mathbf{x})^t (\hat{\mathbf{x}} - \mathbf{x})$, et

— maximum *a posteriori* (MAP)

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \{p(\mathbf{x}|\mathbf{y})\},$$

en choisissant une fonction coût $C(\hat{\mathbf{x}}, \mathbf{x}) = 1 - \delta(\hat{\mathbf{x}} - \mathbf{x})$.

Dans le cas général ces estimateurs ne sont pas des fonctions linéaires des observations, ce qui a pour conséquence que la solution du problème (l'estimé $\hat{\mathbf{x}}$) dépend d'une manière non linéaire des observations \mathbf{y} . Le cas gaussien est une exception. En effet dans ce cas la solution $\hat{\mathbf{x}}$ est linéaire en \mathbf{y} . L'idée de l'invariance par changement d'échelle consiste à chercher les lois $p(\mathbf{x})$ et $p(\mathbf{y}|\mathbf{x})$ telle que la loi *a posteriori* $p(\mathbf{x}|\mathbf{y})$ et, par conséquent, l'estimateur ponctuel $\hat{\mathbf{x}}$ qui en dépend soit invariant par changement d'échelle.

Rappelons que la linéarité est une propriété plus forte que l'invariance d'échelle. En effet $\hat{\mathbf{x}}$ est linéaire en \mathbf{y} ssi

$$\begin{cases} \mathbf{y}_1 \mapsto \hat{\mathbf{x}}_1 \\ \mathbf{y}_2 \mapsto \hat{\mathbf{x}}_2 \end{cases} \implies \forall k_1, k_2, \quad k_1 \mathbf{y}_1 + k_2 \mathbf{y}_2 \mapsto k_1 \hat{\mathbf{x}}_1 + k_2 \hat{\mathbf{x}}_2,$$

$F(x) = \sum_j f(x_j)$	$g(s), g'(s)$ et $g''(s)$	solution : $x_j = g'([\lambda^t A]_j)$	fonction duale : $D(\lambda) = \lambda^t y - \sum_j g([\lambda^t A]_j)$
$\frac{1}{2} x_j^2$	$\begin{cases} g(s) = \frac{1}{2} s^2 \\ g'(s) = s \\ g''(s) = 1 \end{cases}$	$x_j = \sum_i \lambda_i a_{ij} = [A^t \lambda]_j$	$D(\lambda) = \lambda^t y - \frac{1}{2} \lambda^t A^t A \lambda$
$\frac{1}{2} \alpha x_j^2 + \beta x_j$	$\begin{cases} g(s) = \frac{1}{2\alpha} (s - \beta)^2 \\ g'(s) = \frac{1}{\alpha} (s - \beta) \\ g''(s) = \frac{1}{\alpha} \end{cases}$	$x_j = \frac{1}{\alpha} (\sum_i \lambda_i a_{ij} - \beta)$	$D(\lambda) = \lambda^t y - \frac{1}{2\alpha} \sum_j ([A^t \lambda]_j - \beta)^2$
$\ln x_j$	$\begin{cases} g(s) = 1 - \ln s \\ g'(s) = -\frac{1}{s} \\ g''(s) = \frac{1}{s^2} \end{cases}$	$x_j = \frac{1}{\sum_i \lambda_i a_{ij}}$	$D(\lambda) = \lambda^t y - \sum_j \ln [A^t \lambda]_j$
$\alpha \ln x_j + \beta x_j$	$\begin{cases} g(s) = \alpha(1 - \ln \alpha) + \alpha \ln(s - \beta) \\ g'(s) = \frac{\alpha}{s - \beta} \\ g''(s) = \frac{-\alpha}{(s - \beta)^2} \end{cases}$	$x_j = \frac{\alpha}{\sum_i \lambda_i a_{ij} - \beta}$	$D(\lambda) = \lambda^t y - \alpha \sum_j \ln ([A^t \lambda]_j - \beta)$
$x_j \ln x_j$	$\begin{cases} g(s) = \exp[s - 1] \\ g'(s) = \exp[s - 1] \\ g''(s) = \exp[s - 1] \end{cases}$	$x_j = \exp[-1 + \sum_i \lambda_i a_{ij}]$	$D(\lambda) = \lambda^t y - \frac{1}{e} \sum_j \exp [A^t \lambda]_j$
$\alpha x_j \ln x_j + \beta x_j$	$\begin{cases} g(s) = \alpha \exp[\frac{s}{\alpha} - \frac{\beta}{\alpha} - 1] \\ g'(s) = \exp[\frac{s}{\alpha} - \frac{\beta}{\alpha} - 1] \\ g''(s) = \frac{1}{\alpha} \exp[\frac{s}{\alpha} - \frac{\beta}{\alpha} - 1] \end{cases}$	$x_j = \exp[-1 - \frac{\beta}{\alpha} + \frac{1}{\alpha} \sum_i \lambda_i a_{ij}]$	$D(\lambda) = \lambda^t y - \alpha \sum_j \exp [\frac{1}{\alpha} A^t \lambda]_j - \frac{\beta}{\alpha} - 1]$
$x_j \ln x_j + (1 - x_j) \ln(1 - x_j)$	$\begin{cases} g(s) = \ln(1 + \exp[s]) \\ g'(s) = \frac{\exp[s]}{1 + \exp[s]} \\ g''(s) = \frac{\exp[s]}{(1 + \exp[s])^2} \end{cases}$	$x_j = \frac{\exp[\sum_i \lambda_i a_{ij}]}{1 + \exp[\sum_i \lambda_i a_{ij}]}$	$D(\lambda) = \lambda^t y - \sum_j \ln(1 + \exp [A^t \lambda]_j)$

ou bien sous une autre forme, ssi

$$\begin{cases} \mathbf{y}_1 \mapsto \hat{\mathbf{x}}_1 \\ \mathbf{y}_2 \mapsto \hat{\mathbf{x}}_2 \end{cases} \implies$$

$$\begin{cases} \mathbf{y}_1 + \mathbf{y}_2 \mapsto \hat{\mathbf{x}}_1 + \hat{\mathbf{x}}_2, & \text{(additivité), et} \\ \forall k > 0, k\mathbf{y}_1 \mapsto k\hat{\mathbf{x}}_1, & \text{(invariance par changement d'échelle).} \end{cases}$$

Tandis que, $\hat{\mathbf{x}}$ est invariant par changement d'échelle en \mathbf{y} ssi

$$\mathbf{y}_1 \mapsto \hat{\mathbf{x}}_1 \implies \forall k > 0 \quad k\mathbf{y}_1 \mapsto k\hat{\mathbf{x}}_1.$$

Pour être plus précis, notons par :

— θ l'ensemble des paramètres qui définissent notre connaissance sur le système de mesure (variance du bruit et paramètres de la loi *a priori* par exemple), et par

— $p_1(\mathbf{x}_1 | \mathbf{y}_1; \theta_1)$ et $p_k(\mathbf{x}_k | \mathbf{y}_k; \theta_k)$ les deux expressions de la loi *a posteriori* à l'échelle 1 et à l'échelle k avec

$$\mathbf{x}_k = k\mathbf{x}_1, \quad \mathbf{y}_k = k\mathbf{y}_1, \quad \theta_k = f(\theta_1, k).$$

Théorème 1 : Si

$$\exists \theta_k \mid \forall k > 0, \forall \mathbf{x}_1, \mathbf{y}_1, \quad p_k(\mathbf{x}_k | \mathbf{y}_k; \theta_k) = \frac{1}{k^n} p_1(\mathbf{x}_1 | \mathbf{y}_1; \theta_1),$$

et si $\theta_k = f(\theta_1, k)$ est une fonction qui ne dépend que de θ_1 et de k , alors tout estimateur bayésien avec une fonction coût $C(\hat{\mathbf{x}}, \mathbf{x})$ telle que

$$C(\hat{\mathbf{x}}_k, \mathbf{x}) = a_k + b_k C(\hat{\mathbf{x}}_1, \mathbf{x}),$$

est invariant par changement d'échelle :

$$\hat{\mathbf{x}}_k(\mathbf{y}_k; \theta_k) = k\hat{\mathbf{x}}_1(\mathbf{y}_1; \theta_1).$$

Preuve : En effet, on peut écrire

$$\begin{aligned} & \hat{\mathbf{x}}_k(\mathbf{y}_k; \theta_k) \\ &= \arg \min_{\mathbf{z}_k} \left\{ \int C(\mathbf{z}_k, \mathbf{x}_k) p_k(\mathbf{x}_k | \mathbf{y}_k; \theta_k) d\mathbf{x}_k \right\} \\ &= k \arg \min_{\mathbf{z}_1} \left\{ \int [b_k C(\mathbf{z}_1, \mathbf{x}_1) + a_k] \frac{1}{k^n} p_1(\mathbf{x}_1 | \mathbf{y}_1; \theta_1) k^n d\mathbf{x}_1 \right\} \\ &= k \arg \min_{\mathbf{z}_1} \left\{ b_k \int C(\mathbf{z}_1, \mathbf{x}_1) p_1(\mathbf{x}_1 | \mathbf{y}_1; \theta_1) d\mathbf{x}_1 + a_k \right\} \end{aligned}$$

$$= k \arg \min_{z_1} \left\{ \int C(z_1, x_1) p_1(x_1 | y_1; \theta_1) dx_1 \right\} = k \hat{x}_1(y_1; \theta_1)$$

Pour faciliter la compréhension examinons les deux cas particuliers suivant :

– Cas de l'estimateur MAP

$$\begin{aligned} \hat{x}_k(y_k; \theta_k) &= \arg \max_{x_k} \{ p_k(x_k | y_k; \theta_k) \} = \arg \max_{x_k} \left\{ \frac{1}{k^n} p_1\left(\frac{x_k}{k} | y_1; \theta_1\right) \right\} \\ &= k \arg \max_{x_1} \left\{ \frac{1}{k^n} p_1(x_1 | y_1; \theta_1) \right\} = k \hat{x}_1(y_1; \theta_1). \end{aligned}$$

– Cas de l'estimation en moyenne quadratique

$$\begin{aligned} \hat{x}_k(y_k; \theta_k) &= \int x_k p_k(x_k | y_k; \theta_k) dx_k = \int k x_1 \frac{1}{k^n} p_1(x_1 | y_1; \theta_1) k^n dx_1 \\ &= k \int x_1 p_1(x_1 | y_1; \theta_1) dx_1 = k \hat{x}_1(y_1; \theta_1). \end{aligned}$$

Notons l'importance de cette propriété, car bien que l'estimateur $\hat{x}(y; \theta)$ soit une fonction non linéaire de y il reste invariant par changement d'échelle.

Maintenant la tâche de trouver les lois $p(x)$ et $p(y|x)$ telle que la loi *a posteriori* $p(x|y)$ ait cette propriété dans le cas général n'est pas aisée. La famille des lois exponentielles généralisées est une famille suffisamment riche. Nous limiterons alors notre recherche dans cette famille, c'est-à-dire à des lois de la forme :

$$p(x; \lambda) \propto \exp \left[- \sum_{i=1}^r \lambda_i \phi_i(x) \right],$$

et

$$p(y|x; \sigma^2) \propto \exp [-\chi^2(x, y)],$$

$$\text{avec} \quad \chi^2(x, y; \sigma^2) = \frac{1}{2\sigma^2} [y - Hx]^t [y - Hx].$$

Notons que ces lois peuvent être considérées comme des lois à maximum d'entropie si notre connaissance *a priori*

— sur x est de la forme :

$$E \{ \phi_i(x) \} = d_i, \quad i = 1, \dots, r,$$

— et sur b est de la forme :

$$\begin{cases} E \{ b \} = 0, \\ E \{ bb^t \} = R_b = \sigma^2 I, \end{cases}$$

où R_b est la matrice de covariance de b .

Dans ce cas on a :

$$p(x|y; \lambda, \sigma^2) \propto \exp \left[-\chi^2(x, y; \sigma^2) - \sum_{i=1}^r \lambda_i \phi_i(x) \right].$$

Maintenant en notant

$$\theta = (\sigma^2, \lambda_1, \dots, \lambda_r), \quad \lambda = (\lambda_1, \dots, \lambda_r),$$

et

$$\phi(x) = (\phi_1(x), \dots, \phi_r(x)),$$

on a

$$p(x|y; \theta) \propto \exp [-\chi^2(x, y) - \lambda^t \phi(x)],$$

et la condition pour assurer l'invariance par changement d'échelle devient :

$$\forall k > 0, \forall x_1, y_1,$$

$$\chi_k^2(x_k, y_k) + \lambda_k^t \phi(x_k) = k (\chi_1^2(x_1, y_1) + \lambda_1^t \phi(x_1)) + cte.$$

Mais avec le choix d'une loi gaussienne pour le bruit il est évident que l'on a

$$\forall k > 0, \forall x_1, y_1,$$

$$\begin{aligned} \chi_k^2(x_k, y_k; \sigma_k^2) &= \frac{1}{2\sigma_k^2} \|y_k - Hx_k\|^2 \\ &= \frac{1}{2k^2\sigma_1^2} k^2 \|y_1 - Hx_1\|^2 = \chi_1^2(x_1, y_1; \sigma_1^2). \end{aligned}$$

(Ceci peut être généralisé, mais pour l'instant limitons-nous à ce cas).

La condition devient alors

$$\forall k > 0, \forall x, \quad \lambda_k^t \phi(x_k) = \lambda_1^t \phi(x_1) + cte,$$

où d'une manière équivalente

$$p(x_k; \lambda_k) = \frac{1}{k^n} p_1(x_1; \lambda_1) \quad \text{avec} \quad \lambda_k = f(\lambda_1, k).$$

Ainsi, dans le cas d'un bruit gaussien, il suffit que la loi *a priori* soit invariante par changement d'échelle pour que la loi *a posteriori* le soit. Considérons maintenant quelques cas particuliers.

2.2. LOIS AVEC UN SEUL PARAMÈTRE

Prenons maintenant le cas d'une variable scalaire X avec une loi de la forme

$$p(x; \lambda) \propto \exp [-\lambda \phi(x)].$$

La condition d'invariance par changement d'échelle devient

$$\exists \lambda_k | \forall k > 0, \forall x, \quad \lambda_k \phi(kx) = \lambda_1 \phi(x) + cte.$$

Nous avons montré [[Djafari90b]] que dans ce cas les seules fonctions $\phi(x)$ qui satisfont cette condition sont :

$$\left\{ x^\alpha, \ln x \right\}$$

où α est un réel.

Avec

— $\phi(x) = x^\alpha$ on obtient

$$p(x) \propto \exp[-\lambda x^\alpha], \quad \alpha > 0, \lambda > 0,$$

qui est une loi gaussienne généralisée, et avec

— $\phi(x) = \ln x$ on obtient

$$p(x) \propto \exp[-\lambda \ln x],$$

qui est un cas particulier de la loi Béta.

Notons que la loi dite *entropique* :

$$p(x) \propto \exp[-\lambda x \ln x],$$

de Gull et Skilling ne vérifie pas la propriété de l'invariance d'échelle. Mais par contre, comme nous allons voir dans le paragraphe suivant, la loi

$$p(x) \propto \exp[-\lambda x \ln x + \mu x],$$

est une loi avec deux paramètres qui vérifie cette propriété.

2.3. LOIS AVEC DEUX PARAMÈTRES

Considérons maintenant le cas d'une variable scalaire X avec une loi de la forme

$$p(x; \lambda) \propto \exp[-\lambda \phi_1(x) - \mu \phi_2(x)]$$

La condition l'invariance par changement d'échelle devient

$$\exists \lambda_k, \mu_k \mid \forall k > 0, \forall x,$$

$$\lambda_k \phi_1(kx) + \mu_k \phi_2(kx) = \lambda_1 \phi_1(x) + \mu_1 \phi_2(x) + cte.$$

Nous avons montrés [[Djafari90b]] que dans ce cas les seules fonctions $(\phi_1(x), \phi_2(x))$ qui satisfont cette condition sont :

$$\left\{ (x^{\alpha_1}, x^{\alpha_2}), (x^{\alpha_1}, \ln x), (x^{\alpha_1}, x^{\alpha_1} \ln x), (\ln x, \ln^2 x) \right\}.$$

où α_1 et α_2 sont deux réels.

Si de plus on impose $\phi_2(x) = x$ alors $\phi_1(x)$ ne peut prendre que les expressions suivantes :

$$\{x^\alpha, \ln x, x \ln x\}.$$

Dans ce cas avec

— $\phi_1(x) = x^2$ on a

$$p(x) \propto \exp[-\lambda x^2 - \mu x] \propto \exp\left[-\lambda \left(x + \frac{\mu}{2\lambda}\right)^2\right],$$

qui est une loi gaussienne $\mathcal{N}(m = -\frac{\mu}{\lambda}, \sigma^2 = \frac{1}{2\lambda})$.

— $\phi_1(x) = \ln x$ on a

$$p(x) \propto \exp[-\lambda \ln x - \mu x] = x^{-\lambda} \exp[-\mu x],$$

qui est une loi gamma, et finalement,

— $\phi_1(x) = x \ln x$ on a

$$p(x) \propto \exp[-\lambda x \ln x - \mu x].$$

Une étude pour la généralisation de ces idées est en cours (voir références [[Djafari93a]], [[Djafari93b]]).

3. Exemple d'application aux problèmes inverses linéaires

Dans un grand nombre de problèmes d'imagerie on est amené à résoudre une équation intégrale de première espèce de la forme :

$$g(s_i) = \int_D f(\mathbf{r}) h(\mathbf{r}, s_i) d\mathbf{r} + b(s_i), \quad i = 1, \dots, M, \quad (88)$$

où g désigne les mesures (image observée en restauration d'images, projections en tomographie X, champ diffracté en tomographie de diffraction, force électromotrice aux bornes d'une bobine en imagerie par courants de Foucault, etc...), f désigne l'objet à restaurer ou à reconstruire, D désigne le support de l'objet, h est le noyau de la transformation qui lie les mesures g et l'objet f , et enfin b représente les erreurs (ou l'incertitude) sur les mesures.

Pour montrer la généralité de cette équation nous précisons ici quelques unes de ses applications.

— Déconvolution de signaux 1-D :

$$g(t) = \int_0^t f(t') h(t-t') dt' + b(t),$$

où $g(t)$ est le signal mesuré, $f(t)$ est le signal recherché, et $h(t)$ est la réponse impulsionnelle du système de mesure.

— Restauration d'image :

$$g(x, y) = \iint_D f(x', y') h(x-x', y-y') dx' dy' + b(x, y),$$

où $g(x, y)$ est l'image observée, $f(x, y)$ est l'image recherchée, et $h(x, y)$ est la réponse impulsionnelle du système d'imagerie.

— Reconstruction d'image en tomographie X :

$$g(r, \phi_i) = \iint_D f(x, y) \delta(r - x \cos \phi_i - y \sin \phi_i) dx dy + b(r, \phi_i), \quad i = 1, \dots, M,$$

où $g(r, \phi_i)$ représente la projection suivant l'angle ϕ_i , et $f(x, y)$ est l'objet à reconstruire.

— Radiographie des objets cylindriques :

$$g(r_i) = \int_r^R f(\rho) \frac{\rho}{\sqrt{\rho^2 - r^2}} d\rho + b(r_i), \quad i = 1, \dots, M,$$

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

— Synthèse d'ouverture en radio astronomie ou imagerie radar :

$$g(u_i, v_i) = \iint_D f(x, y) \exp[-j(u_i x + v_i y)] dx dy, \quad i = 1, \dots, M,$$

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

— Synthèse de Fourier en reconstruction tomographique d'image :

$$g(\Omega, \phi_i) = \iint_D f(x, y) \exp\{-j[T_1(\Omega, \phi_i)x + T_2(\Omega, \phi_i)y]\} dx dy, \quad i = 1, \dots, M,$$

où

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

définissent un ensemble de contours algébriques dans le plan (u, v) .

La synthèse de Fourier se trouve au cœur d'un grand nombre d'applications en imagerie. Parmi celles-ci, on trouve :

— Tomographie à rayons X :

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

où les contours algébriques sont des lignes droites concentriques, et où $g(\Omega, \phi_i)$ représente la TF spatiale de $p(r, \phi_i)$.

Cette formulation du problème se trouve aussi dans le cas de l'imagerie RMN. Dans ce cas $g(\Omega, \phi_i) = s(t, \phi_i)$, où $s(t, \phi_i)$ est directement le signal de précession libre mesuré quand le gradient du champ magnétique est dans la direction ϕ_i .

— Tomographie à ondes diffractées

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

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

— Synthèse de Fourier-Laplace :

$$g(u, s) = \iint_D f(x, y) \exp[jux + sy] dx dy,$$

où u est une variable réelle, tandis que $s = \gamma + jv$ est une variable complexe. Cette équation intervient en tomographie de diffraction dans des objets dissipatifs ou bien en imagerie tomographique par courants de Foucault en contrôle non destructif. (u, v) représente un point dans le domaine de Fourier spatial de l'objet $f(x, y)$. Ceci correspond au phénomène de propagation dans les deux directions, tandis que γ représente un phénomène de diffusion ou d'atténuation.

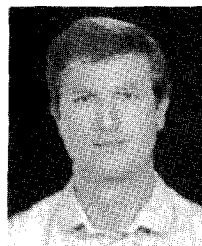
4. BIBLIOGRAPHIE

- [Agmon79] N. Agmon, Y. Alhassid et D. Levine. – An Algorithm for Finding the Distribution of Maximal Entropy. *Journal of Computational Physics*, vol. 30, 1979, pp. 250–258.
- [Balian82] R. Balian. – *Du microscopique au macroscopique*. – 1982, Ellipses, Paris édition.
- [Bercher93] J.F. Bercher et G. Le Besnerais. – Barres d'erreurs sur les solutions de méthodes de type entropique. *Rapport interne LSS, 20p.*, 1993.
- [Borwein91] J.M. Borwein et A.S. Lewis. – Duality relationships for entropy-like minimization problems. *SIAM J. Control and Optimization*, vol. 29, March 1991.
- [Bryan83] R.K. Bryan, M. Bansal, W. Folkhard., C. Nave et D.A. Marvin. – Maximum-entropy calculation of the electron density at 4 Å resolution of Pfl filamentous bacteriophage. *Proc. Natl. Acad. Sci. USA*, vol. 80, 1983, pp. 4728–4731.

- [Burch83] S.F. Burch, S.F. Gull et J. Skilling. – Image Restoration by a Powerful Maximum Entropy Method. *Comput. Vis. Graph. Im. Process.*, vol. 23, 1983, pp. 113–128.
- [Burg67] J.P. Burg. – Maximum Entropy Spectral Analysis. *Proc. 37th Meet. Soc. Exploration Geophysicists.*, vol. Stanford Thesis, 1967.
- [Csiszar91] Csiszár. – Why least-squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *The Annals of Statistics*, no4, 1991, pp. 2032–2066.
- [Daniell80] G.J. Daniell et S.F. Gull. – Maximum Entropy algorithm applied to image enhancement. *Proc. IEE*, vol. 127E, 1980, pp. 170–172.
- [Dacunha90] D. Dacunha-Castelle et F. Gamboa. – Maximum d'entropie et problème des moments. *Ann. de l'Institut Poincaré*, vol. 26, 1990, pp. 567–596.
- [Demoment87] G. Demoment. – Déconvolution des signaux. *Polycopié de l'ESE*, 1987.
- [Demoment89] G. Demoment. – Image Reconstruction and Restoration : Overview of Common Estimation Structure and Problems. *IEEE Trans. on Acoustics Sounds and Signal Processing*, vol. 37, 1989, pp. 2024–2036.
- [Dusaussouy91] N.J. Dusaussouy et I.E. Abdou. – The extended MENT algorithm : A maximum entropy type algorithm using prior knowledge for computerized tomography. *IEEE Trans. on Signal Processing*, vol. 39, 1991, pp. 1164–1180.
- [Elfwing80] T. Elfving. – On some Methods for Entropy Maximization and Matrix Scaling. *Linear Algebra and its Applications*, vol. 34, 1980, pp. 321–339.
- [Eriksson80] J. Eriksson. – A note on Solution of Large Sparse Maximum Entropy Problems with Linear Equality Constraints. *Mathematical Programming*, vol. 18, 1980, pp. 146–154.
- [Erlander81] S. Erlander. – Entropy in linear programs. *Mathematical Programming*, vol. 21, 1981, pp. 137–151.
- [Frieden80] B.R. Frieden. – Statistical Models for the Image Restoration Problem. *Comput. Graph. Im. Process.*, vol. 12, 1980, pp. 40–59.
- [Frieden85] B.R. Frieden et C.K. Zoltani. – Maximum bounded entropy : application to tomographic reconstruction. *Applied Optics*, vol. 24, 1985, pp. 201–207.
- [Gamboa89] F. Gamboa. – Méthode du maximum d'entropie sur la moyenne and applications. *Thèse, Dépt. de Mathématiques, Univ. de Paris-Sud*, 1989.
- [Gull84] S.F. Gull et J. Skilling. – Maximum entropy method in image processing. *IEE Proceedings*, vol. 131, Pt. F, 1984.
- [Jaynes68] E.T. Jaynes. – Prior Probabilities. *IEEE Trans.*, vol. SSC-4, 1968, pp. 227–241.
- [Jaynes82] E.T. Jaynes. – On the Rationale of Maximum-Entropy Methods. *Proceedings of the IEEE*, vol. 70, 1982, pp. 939–952.
- [Jaynes85] E.T. Jaynes. – Where do we go from here? *Maximum-Entropy and Bayesian Methods in Inverse Problems*, vol. C.R. Smith and T. Grandy, Jr. (eds.), 1985, pp. 21–58.
- [Johnson84] R. Johnson et J. Shore. – Which is Better Entropy Expression for Speech Processing :-SlogS or logS? *IEEE Trans. on Acoustics Sounds and Signal Processing*, vol. ASSP-32, 1984, pp. 129–137.
- [Kikuchi77] R. Kikuchi et B.H. Soffer. – Maximum entropy image restoration. I. The entropy expression. *J. Opt. Soc. Am.*, vol. 67, 1977, pp. 1656–1665.
- [Kullback59] S. Kullback. – Information theory and statistics. *Wiley, New York*, 1959.
- [Le Besnerais91a] G. Le Besnerais, J. Navaza et G. Demoment. – Aperture synthesis using maximum entropy on the mean in radio-astronomy. *GRETSI*, 1991.
- [Le Besnerais91b] G. Le Besnerais, J. Navaza et G. Demoment. – Synthèse d'ouverture en radio-astronomie par maximum d'entropie sur la moyenne. *Actes du 13ème colloque GRETSI, Juan-les-pins, 16–20 septembre 1991*, 1991, pp. 217–220.
- [Le Besnerais93] G. Le Besnerais, J.F. Bercher et G. Demoment. – Probabilistic issues in Fourier Synthesis. *Rapp. intern LSS*, 1993.
- [Le Besnerais93b] G. Le Besnerais. – Méthodes du maximum d'entropie sur la moyenne, critère de reconstruction d'image et synthèse d'ouverture en radioastronomie. *Thèse de l'Université de Paris-Sud, Orsay*, décembre 1993.
- [Leahy86] R.M. Leahy et C.E. Goutis. – An optimal Technique for constrained-based image restoration and reconstruction. *IEEE Trans. on Acoustics Sounds and Signal Processing*, vol. ASSP-34, 1986.
- [Luenberger69] D.G. Luenberger. – Optimization by Vector Space Methods. *1st ed. New York : Wiley*, 1969.
- [Macaulay85] V.A. Macaulay et B. Buck. – Linear inversion by the method of maximum entropy. *Maximum-Entropy and Bayesian Methods in Inverse Problems*, vol. C.R. Smith & T. Grandy, Jr.(eds.), 1985.
- [Djafari87a] A. Mohammad-Djafari et G. Demoment. – Maximum entropy Fourier synthesis with application to diffraction tomography. *Applied Optics*, vol. 26, 1987, pp. 1745–1754.
- [Djafari87b] A. Mohammad-Djafari et G. Demoment. – Tomographie de diffraction and synthèse de Fourier à maximum d'entropie. *Revue Phys. Appl.*, vol. 22, 1987, pp. 153–167.
- [Djafari88a] A. Mohammad-Djafari et G. Demoment. – Image restoration and reconstruction using entropy as a regularization functional. *Maximum Entropy and Bayesian Methods in Science and Engineering*, vol. 2, 1988, pp. 341–355.
- [Djafari88b] A. Mohammad-Djafari et G. Demoment. – Maximum entropy reconstruction in X ray and diffraction tomography. *IEEE Trans. on Medical Imaging*, vol. 7, 1988, pp. 345–354.
- [Djafari88c] A. Mohammad-Djafari et G. Demoment. – Utilisation de l'entropie dans les problèmes de restauration and de reconstruction d'images. *Traitement du Signal*, vol. 5, 1988, pp. 235–248.
- [Djafari89] A. Mohammad-Djafari et G. Demoment. – Maximum Entropy and Bayesian Approach in Tomographic Image Reconstruction and Restoration. *Maximum Entropy and Bayesian Methods*, 1989, pp. 195–201.
- [Djafari90a] A. Mohammad-Djafari et G. Demoment. – Estimating Priors in Maximum Entropy Image Processing. *Proc. of ICASSP 1990*, 1990, pp. 2069–2072.
- [Djafari90b] A. Mohammad-Djafari et J. Idier. – Maximum entropy prior laws of images and estimation of their parameters. *T.W. Grandy (ed.), Maximum-entropy and Bayesian methods*, 1990.
- [Djafari91a] A. Mohammad-Djafari. – A Matlab Program to Calculate the Maximum Entropy Distributions. *Proc. of The 11th Int. MaxEnt Workshop, Seattle, USA*, 1991.

- [Djafari91b] A. Mohammad-Djafari. – Maximum Likelihood Estimation of the Lagrange Parameters of the Maximum Entropy Distributions. *Maximum-entropy and Bayesian methods*, 1991.
- [Djafari93a] A. Mohammad-Djafari. – Bayesian Approach with Maximum Entropy Priors to Imaging Inverse Problems, Part I : Foundations. *submitted to IEEE Trans. on Image Processing*, August 1993.
- [Djafari93b] A. Mohammad-Djafari. – Bayesian Approach with Maximum Entropy Priors to Imaging Inverse Problems, Part II : Applications. *submitted to IEEE Trans. on Image Processing*, August 1993.
- [Mukherjee84] D. Mukherjee et D.C. Hurst. – Maximum Entropy Revisited. *Statistica Neerlandica*, vol. 38, 1984.
- [Narayan86] R. Narayan et R. Nityananda. – Maximum entropy image restoration in astronomy. *Ann. Rev. Astron. Astrophys.*, vol. 24, 1986, pp. 127–170.
- [Nashed81] M.Z. Nashed. – Operator-Theoretic and Computational Approaches to Ill-Posed Problems with Applications to Antenna Theory. *IEEE Trans. on Antenna and Propagation*, vol. AP-29, 1981, pp. 220–231.
- [Navaza85] J. Navaza. – On the maximum entropy estimate of electron density function. *Acta Cryst.*, vol. A-41, 1985, pp. 232–244.
- [Navaza86] J. Navaza. – Use of non-local constraints in maximum entropy electron density reconstruction. *Acta Cryst.*, vol. A-42, 1986, p. 212.
- [Noonan76] J.P. Noonan, N.S. Tzannes et T. Costello. – On the Inverse Problem of Entropy Maximizations. *IEEE Trans. on Nuclear Sciences*, 1976, pp. 120–123.
- [Shannon48] C.E. Shannon et W. Weaver. – The Mathematical Theory of Communication. *Bell System Technical Journal*, vol. 27, 1948, pp. 379–423.
- [Shore81] J.E. Shore et R.W. Johnson. – Properties of Cross-Entropy Minimization. *IEEE Trans. on Nuclear Sciences*, vol. IT-27, 1981, pp. 472–482.
- [Skilling84a] J. Skilling et R.K. Bryan. – Maximum entropy image reconstruction : general algorithm. *Mon. Not. R. astr. Soc.*, vol. 211, 1984, pp. 111–124.
- [Skilling84b] J. Skilling et S.F. Gull. – The Entropy of an Image. *SIAM-AMS Proceedings*, vol. 14, 1984.
- [Skilling84c] J. Skilling et S.F. Gull. – Maximum entropy method in image processing. *IEE Proceedings*, vol. 131, 1984, pp. 646–659.
- [Skilling85] J. Skilling et S.F. Gull. – Algorithms and Applications. *Maximum-Entropy and Bayesian Methods in Inverse Problems*, vol. C.R. Smith & T. Grandy, Jr.(eds.), 1985, pp. 83–132.
- [Smith79] C.B. Smith. – A dual method for maximum entropy restoration. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, 1979, pp. 411–414.
- [Trussell80] H.J. Trussell. – The Relationship Between Image Restoration by the Maximum A Posteriori Method and Maximum Entropy Method. *IEEE Trans. on Acoustics Sounds and Signal Processing*, vol. ASSP-28, 1980, pp. 114–117.
- [Verdugo78] L.A. Verdugo et P.N. Rathie. – On the entropy of continuous probability distributions. *IEEE Trans. on Nuclear Sciences*, vol. IT-24, 1978, pp. 120–122.
- [Wernecke77] S.J. Wernecke et L.R. D'Addario. – Maximum Entropy Image Reconstruction. *IEEE Trans. on Communications*, vol. C-26, 1977, pp. 351–364.
- [Wragg70] A. Wragg et D.C. Dawson. – Fitting continuous probability distributions over $(0, \infty)$ using information theory ideas. *IEEE Trans. on Nuclear Sciences*, vol. IT-16, 1970, pp. 226–230.
- [Zellner77] Zellner. – Maximal data information prior distributions. *in New developments in the applications of bayesian methods*, A. Aykac and C. Brumat diteurs associés, North-Holland, Amsterdam, 1977, pp. 211–232.
- [Zhuang87] X. Zhuang, K.B. Yu et R.M. Haralick. – A Differential Equation Approach to Maximum Entropy Image Reconstruction. *IEEE Trans. on Acoustics Sounds and Signal Processing*, vol. ASSP-35, 1987, pp. 208–218.

L'AUTEUR



Ali Mohammad-Djafari est né en Iran en 1952. Il est Ingénieur de l'École Polytechnique de Téhéran (1975), Ingénieur de l'École Supérieure d'Électricité (1977), Docteur-Ingénieur (1981) et Docteur-ès-Sciences Physiques (1987) de l'Université de Paris-Sud, Orsay. 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 depuis 1983, il s'intéresse à la résolution des problèmes inverses en utilisant des méthodes d'inférence probabilistes bayésiennes et le principe du maximum d'entropie. Ses thèmes de recherche portent sur les sujets suivants : restauration des signaux mono- ou multi-variables, reconstruction d'image en tomographie à rayons X ou à ondes diffractées (ultrasons ou micro-ondes), synthèse de Fourier, imagerie tomographique par courants de Foucault en contrôle non destructif (CND), synthèse de Fourier-Laplace, problèmes mal-posés, régularisation, théorie de l'information, attribution des lois de probabilité a priori, principe du maximum d'entropie, modélisation markovienne des images, algorithmes du gradient, du gradient conjugué et de non convexité graduelle (GNC), détermination des paramètres de régularisation, estimation des hyperparamètres, maximisation de vraisemblances généralisée et marginales.

Manuscrit reçu le 23 juin 1993.