TREIZIÈME COLLOQUE GRETSI - JUAN-LES-PINS DU 16 AU 20 SEPTEMBRE 1991

ESTIMATION PAR MAXIMUM DE VRAISEMBLANCE DES LAGRANGIENS DES DISTRIBUTIONS A MAXIMUM D'ENTROPIE

Ali Mohammad-Djafari

Laboratoire des Signaux et Systèmes (CNRS-ESE-UPS)
Plateau de Moulon, 91192 Gif sur Yvette Cédex, France

RÉSUMÉ

Le problème classique du Maximum d'Entropie (ME) est la détermination d'une loi de probabilité pour une variable aléatoire X connaissant les espérances $\mu_n = \mathrm{E}\{\phi_n(x)\}$, n=0,..., N de fonctions $\phi_n(x)$ connues. La solution dépend alors de N+1 multiplicateurs de Lagrange $\boldsymbol{\lambda}$ qui sont déterminés en résolvant un système d'équations non linéaires résultant de ces contraintes. Le problème abordé ici est différent. Nous ne disposons pas directement des valeurs de μ_n, mais d'un M-échantillons $x = [x_1, ..., x_M]$ de X. Nous montrons alors l'équivalence entre les deux chemins suivants: **i)** Déterminer les μ_n par leur estimateurs empiriques et utiliser ensuite l'approche ME pour déterminer les paramètres $\boldsymbol{\lambda}$, **ii)** Supposer que X suit une loi à ME $p(x; \boldsymbol{\lambda})$, ce qui fixe la forme et le nombre de paramètres de la loi, et utiliser l'approche du maximum de vraisemblance (MV) pour estimer ces paramètres. L'intérêt éventuel de ces développements réside dans la détermination des paramètres d'une loi *a priori* pour une image dans l'approche bayésienne pour la résolution des problèmes inverses que l'on rencontre en restauration et en reconstruction d'image.

ABSTRACT

The classical maximum entropy (ME) problem consists of determining a probability distribution function (pdf) of a random variable (r.v.) from a finit set of the expectations $\mu_n = \mathrm{E}\{\phi_n(x)\}$ of the known functions $\phi_n(x)$, n = 1, ..., N. The solution depends on N+1 Lagrange multipliers which are determined by solving the set of nonlinear equations formed by the N data constraints and the normality constraint. The problem we adresse here is different. We do not have des values of μ_n but a M-samples $\boldsymbol{x} =[x_1, ..., x_M]$ of the r.v. X. We show then the equivalency between the two following approaches : **i)** Determine μ_n, n=1, ..., N by their empirical estimates and use the ME approach to determine the parameters $\boldsymbol{\lambda}$, **ii)** Suppose that X has a ME pdf $p(x; \boldsymbol{\lambda})$. This hypothesis fixes the form and the number of its parameters $\boldsymbol{\lambda}$. Use then the maximum likelihood (ML) approach to estimate its parameters $\boldsymbol{\lambda}$. We show the interest of these developpements in determining the prior law of an image in a Bayesian approach to solving the inverse problems of image restoration and reconstruction.

1. INTRODUCTION

Les principes du Maximum d'Entropie (ME) et du Maximum de Vraisemblance (MV) sont deux approches différentes pour faire de l'inférence statistique. Dans les deux approches le but est de choisir une loi de probabilité qui représente le mieux le comportement du phénomène aléatoire observé. Quand on dispose de données qui sont les valeurs moyennes, on utilise l'approche ME et quand les données sont directement les échantillons observés de la variable aléatoire (v.a.) étudiée, on utilise l'approche MV. Mais dans ce dernier cas il faut émettre une hypothèse súpplémentaire qui est : "la v.a. suit une loi appartenant à une famille paramétrique connue avec un nombre fini de paramètres". L'approche MV permet alors d'estimer ces paramètres. L'objectif de cette communication est de montrer l'équivalence entre les deux chemins suivants :
i) Déterminer les valeurs moyennes par leur estimateurs empiriques et utiliser ensuite l'approche ME pour déterminer la loi,
ii) Supposer que la v.a. suit une loi appartenant à la famille des lois à ME, ce qui fixe la forme de la loi et le nombre de ses paramètres, et utiliser ensuite l'approche MV pour estimer ces paramètres.

Dans ce dernier cas, nous mentionnons une autre méthode classique d'estimation des paramètres d'une loi de probabilité à partir d'un M-échantillon qui est la méthode des moments (MM). Nous verrons que cette méthode est équivalente à celle du MV seulement dans le cas particulier des lois exponentielles qui forment un sous-ensemble des lois à ME.

L'intérêt éventuel de ces développements réside dans la détermination des paramètres d'une loi *a priori* pour une image dans l'approche bayésienne pour la résolution des problèmes inverses que l'on rencontre en restauration et en reconstruction d'image. Notant que les paramètres de la loi *a priori* jouent le rôle des paramètres de régularisation, nous verrons comment ces développements peuvent nous proposer une méthode pour la détermination de ces paramètres.

Pour ceci, formalisons tout d'abord les deux approches du MV et du ME :

Approche du Maximum de Vraisemblance (MV) :

Supposons que l'on dispose d'un N-échantillon $\boldsymbol{x} = [x_1, ..., x_N]$ d'une variable aléatoire (v.a.) X dont on suppose connue la forme de sa fonction densité de probabilité (fdp) $p(x, \boldsymbol{\lambda})$ dépendant des paramètres $\boldsymbol{\lambda} \in \Lambda$. Notons :

$$p(\boldsymbol{x}) = \prod_{i=1}^{N} p(x_i) \quad \text{et} \quad p(\boldsymbol{x}; \boldsymbol{\lambda}) = \prod_{i=1}^{N} p(x_i; \boldsymbol{\lambda}) \qquad (1)$$

La méthode du MV choisit $\hat{p}$ de l'ensemble

$$\mathrm{P}_\lambda = \{\, p(\boldsymbol{x}) \mid p(\boldsymbol{x}) = p(\boldsymbol{x}; \boldsymbol{\lambda}),\ \boldsymbol{\lambda} \in \Lambda \,\}$$

par

$$\hat{p}(\boldsymbol{x}; \boldsymbol{\lambda}) = \arg\max_{p \in \mathrm{P}_\lambda} \mathrm{Ln}\, p(\boldsymbol{x}) \qquad (2)$$

Ce qui, d'une façon équivalente, revient à choisir les paramètre $\hat{\boldsymbol{\lambda}} \in \Lambda$ par

$$\hat{\boldsymbol{\lambda}} = \arg\max_{\boldsymbol{\lambda}\in\Lambda} \operatorname{Ln} p(x; \boldsymbol{\lambda}) \tag{3}$$

La fonction $p(x; \boldsymbol{\lambda})$ définie en (1), vue comme une fonction de $\boldsymbol{\lambda}$ est appelée la vraisemblance $l(\boldsymbol{\lambda})$, et son logarithme la log-vraisemblance $L(\boldsymbol{\lambda})$.

La principale limitation pratique de cette méthode est la nécessité de connaître la forme de $p(x; \boldsymbol{\lambda})$ et le nombre de paramètres (composantes de $\boldsymbol{\lambda}$), qui doit être nécessairement inférieur au nombre d'observations (composantes de $\boldsymbol{x}$). Pour un nombre d'observations fixé, plus nombreux sont les paramètres à estimer, plus grande est la vraisemblance.

Approche du Maximum d'Entropie (ME) :

Supposons qu'il existe des fdp $p(x)$ appartenant à l'ensemble P défini par

$$\mathrm{P} = \{ p(x) \mid \mathrm{E}_p[\phi_n(x)] = \mu_n,\ n = 0, \ldots, N\}$$

avec $$\mu_0 = 1,\ \phi_0(x) = 1 \tag{4}$$

Si la seule connaissance dont on dispose sur la variable aléatoire X est la connaissance des valeurs moyennes μ_n, et si l'on cherche à déterminer $p(x)$ pour représenter la loi de probabilité de cette v.a., la méthode du ME nous suggère de choisir $p(x)$ telle que :

$$\hat{p} = \arg\max_{p\in\mathrm{P}} \{ -\int p(x) \operatorname{Ln} p(x)\, \mathrm{d}x \} \tag{5}$$

Ce qui, d'une façon équivalente, revient à résoudre :

$$\text{maximiser } H = -\int p(x) \operatorname{Ln} p(x)\, \mathrm{d}x \tag{6}$$

sous les contraintes :

$$\mathrm{E}\{\phi_n(x)\} = \int \phi_n(x)\, p(x)\, \mathrm{d}x = \mu_n \qquad n = 0, \ldots, N \tag{7}$$

La solution est classiquement donnée par :

$$p(x) = \exp\left[- \sum_{n=0}^{N} \lambda_n\, \phi_n(x) \right] \tag{8}$$

Les $(N+1)$ paramètres de Lagrange $\boldsymbol{\lambda} = [\lambda_0, \ldots, \lambda_n]$ sont obtenus en résolvant le système d'équations non linéaires suivant :

$$G_n(\boldsymbol{\lambda}) = \int \phi_n(x) \exp\left[- \sum_{n=0}^{N} \lambda_n\, \phi_n(x) \right] \mathrm{d}x = \mu_n,\ n=0 \ldots, N \tag{9}$$

La limitation majeure de cette méthode est dans le fait que si on disposait des données sous la forme d'une réalisation spécifique on devrait d'abord estimer les valeurs moyennes μ_n, $n = 0, \ldots, N$ à partir de cette réalisation, définir l'ensemble P et, ensuite, les utiliser dans la méthode du ME.

2. EQUIVALENCE ENTRE ME ET MV :

Le problème que nous étudions dans cette communication est différent de celui du ME classique. Nous ne disposons pas directement des valeurs de $\boldsymbol{\mu}_n$, $n=1, \ldots, N$, mais seulement de M échantillons $\boldsymbol{x} = [x_1, \ldots, x_M]$ de la variable aléatoire X. Nous montrons alors l'équivalence entre les deux chemins suivants:

i) Déterminer les valeurs moyenne $\boldsymbol{\mu}_n$, $n=1, \ldots, N$ par leur estimateurs empiriques :

$$\mu_n = \frac{1}{M} \sum_{i=1}^{M} \phi_n(x_i), \qquad n = 1, \ldots, N \tag{10}$$

et utiliser ensuite l'approche ME pour déterminer les paramètres $\boldsymbol{\lambda}$.

ii) Supposer que la famille des lois $p(x)$ est celle définie en (2), ce qui fixe la forme de la loi et le nombre de ses paramètres, et utiliser l'approche MV pour estimer ces paramètres.

Pour montrer l'équivalence entre les deux problèmes exposés plus haut revenons sur les détails de ces deux approches.

Maximum d'Entropie (ME) :

Nous avons vu quand les données du problème sont les valeurs moyennes données dans l'équation (7) que l'approche du ME fournit la solution (8) pour la loi $p(x, \boldsymbol{\lambda})$ qui dépend de $N+1$ paramètres $\boldsymbol{\lambda}$ (les multiplicateurs de Lagrange du problème du maximisation sous contrainte). Les valeurs de ces paramètres sont calculées en résolvant le système d'équations non linéaires (9).

En général, ces équations sont résolues par la méthode classique de Newton qui consiste à développer $G_n(\boldsymbol{\lambda})$ en série de Taylor, à l'ordre un, autour d'un point courant $\boldsymbol{\lambda}^0$, et résoudre l'équation linéaire résultante pour obtenir une nouvelle $\boldsymbol{\lambda}$ qui devient alors le point courant pour l'itération suivante. Cette méthode peut être résumée ainsi :

Quand on développe $G_n(\boldsymbol{\lambda})$ dans l'équation (9) autour d'un point courant $\boldsymbol{\lambda}^0$ on obtient :

$$G_n(\boldsymbol{\lambda}) \cong G_n(\boldsymbol{\lambda}) + (\boldsymbol{\lambda} - \boldsymbol{\lambda}^0) \left[\operatorname{grad} G_n(\boldsymbol{\lambda})\right]_{\boldsymbol{\lambda} = \boldsymbol{\lambda}^0} = \mu_n \qquad n = 0, \ldots, N \tag{11}$$

Définissant les deux vecteurs $\boldsymbol{\delta}$ et $\boldsymbol{v}$

$$\boldsymbol{\delta} = \boldsymbol{\lambda} - \boldsymbol{\lambda}^0$$

$$\boldsymbol{v} = \left[G_0(\boldsymbol{\lambda}^0) - \mu_0, \ldots, G_N(\boldsymbol{\lambda}^0) - \mu_N \right]^t$$

et la matrice $\mathbb{G}$

$$\mathbb{G} = \begin{pmatrix} g_{nk} \end{pmatrix} = \begin{pmatrix} \frac{\partial G_n(\boldsymbol{\lambda})}{\partial \mu_k} \end{pmatrix}_{\boldsymbol{\lambda} = \boldsymbol{\lambda}^0} \qquad n, k = 0, \ldots, N, \tag{12}$$

l'équation (11) devient :

$$\mathbb{G}\, \boldsymbol{\delta} = \boldsymbol{v} \tag{13}$$

Une fois $\boldsymbol{\delta}$ calculé en résolvant cette équation, on en déduit $\boldsymbol{\lambda} = \boldsymbol{\lambda}^0 + \boldsymbol{\delta}$, qui devient alors le nouveau point courant et les itérations continuent jusqu'au moment où $\boldsymbol{\delta}$ devient suffisamment petit. Notons que la matrice $\mathbb{G}$ est une matrice symétrique et on a :

$$g_{nk} = g_{kn} = - \int \phi_n(x)\, \phi_k(x) \exp\left[-\sum_{n=0}^{N} \lambda_n\, \phi_n(x) \right] \mathrm{d}x = \mu_n \qquad n, k = 0, \ldots, N \tag{14}$$

Pour une implantation numérique de cette méthode, se référer à [7] dans laquelle on trouve aussi un programme écrit en langage MATLAB pour la résolution de ce problème.

Maximum de Vraisemblance (MV) :

Nous avons vu que, quand les données du problème sont les échantillons observés de la v.a. X, afin d'utiliser l'approche MV il faut d'abord définir la famille auquel appartient $p(x)$. Supposons que $p(x)$ appartient à la famille des lois définies par l'équation (8) qui dépendent de N paramètres $\boldsymbol{\lambda}$. L'équation (9) peut aussi être écrite de la forme :

$$p(x) = \frac{1}{Z(\boldsymbol{\lambda})} \exp\left[- \sum_{n=1}^{N} \lambda_n\, \phi_n(x) \right] \tag{15}$$

où $Z(\boldsymbol{\lambda})$ est la constante de normalisation. Le problème est maintenant celui de l'estimation des paramètres $\boldsymbol{\lambda}$. Nous avons vu que dans l'approche MV on définit la fonction de vraisemblance $l(\boldsymbol{\lambda})$:

$$l(\boldsymbol{\lambda}) = \prod_{i=1}^{M} p(x_i \mid \boldsymbol{\lambda}) = \prod_{i=1}^{M} \frac{1}{Z(\boldsymbol{\lambda})} \exp[-\sum_{n=1}^{N} \lambda_n \phi_n(x_i)] \quad (16)$$

ou plutôt la fonction de log-vraisemblance qui devient :

$$L(\boldsymbol{\lambda}) = -M \ \mathrm{Ln}\, Z(\boldsymbol{\lambda}) - \sum_{i=1}^{M} \sum_{n=1}^{N} \lambda_n \phi_n(x_i), \quad n = 1, \ldots, N \quad (17)$$

L'estimation $\hat{\boldsymbol{\lambda}}$ de $\boldsymbol{\lambda}$ par MV, est alors obtenue en recherchant l'argument $\boldsymbol{\lambda}$ qui maximise $L(\boldsymbol{\lambda})$, ce qui s'obtient en résolvant :

$$\frac{\partial L(\boldsymbol{\lambda})}{\partial \lambda_n} = 0, \quad n = 1, \ldots, N \quad (18)$$

ce qui donne :

$$\frac{\partial \mathrm{Ln}\, Z(\boldsymbol{\lambda})}{\partial \lambda_n} = -\frac{1}{M} \sum_{i=1}^{M} \phi_n(x_i), \quad n = 1, \ldots, N \quad (19)$$

où

$$Z(\boldsymbol{\lambda}) = \int \exp[-\sum_{n=1}^{N} \lambda_n \phi_n(x)]\, dx \quad (20)$$

Remplaçant (20) dans l'équation (19) nous obtenons :

$$\int \phi_n(x) \exp[-\sum_{n=1}^{N} \lambda_n \phi_n(x)]\, dx = \frac{1}{M} \sum_{i=1}^{M} \phi_n(x_i), \quad n = 0, \ldots, N \quad (21)$$

Comparant cette équation avec l'équation (9) on constate que si on remplace les μ_n par

$$\mu_n = \frac{1}{M} \sum_{i=1}^{M} \phi_n(x_i), \quad n = 1, \ldots, N \quad (22)$$

les équations seront équivalentes (CQFD).

A ce stade on peut se demander si d'autre méthodes d'estimation des paramètres d'une loi à partir des observations ont la même propriété. Nous considérons seule une autre méthode classique qui est la méthode des moments (MM) qui est décrite brièvement.

La méthode des moments pour estimer les paramètres d'une loi de probabilité $p(x, \boldsymbol{\lambda})$ consiste à identifier les N+1 moments de X :

$$G_n(\boldsymbol{\lambda}) = \mathrm{E}\{x^n\} = \int x^n p(x \mid \boldsymbol{\lambda})\, dx, \quad n = 0, \ldots, N \quad (23)$$

à leurs estimateurs empiriques :

$$\mu_n = -\frac{1}{M} \sum_{i=1}^{M} x_i^n \quad n = 0, \ldots, N \quad (24)$$

et à résoudre le système d'équations qui en résulte :

$$G_n(\boldsymbol{\lambda}) = \int x^n \exp[-\sum_{n=1}^{N} \lambda_n \phi_n(x)]\, dx = -\frac{1}{M} \sum_{i=1}^{M} x_i^n, \quad n = 0, \ldots, N \quad (25)$$

Comparant cette équation avec celle de l'équation (21) on constate leurs équivalence quand $\phi_n(x) = x^n$.

3. RESULTATS NUMERIQUES

L'objectif de ces expérimentations numériques est de montrer la différence entre les deux méthodes MV et MM en rapport avec l'approche ME. Pour cela nous avons considérés le cas de la distribution gamma :

$$p(x \mid \alpha, \beta) = \frac{\beta^{(1-\alpha)}}{\Gamma(1-\alpha)} x^{\alpha} \exp(-\beta x) \quad x>0,\ \alpha<1,\ \beta>0 \quad (26)$$

En effet cette loi peut être considérée comme une loi à maximum d'entropie sous les contraintes suivantes :

$$\begin{aligned} &\int p(x \mid \alpha, \beta)\, dx = 1 && \text{normalisation} \quad \phi_0(x) = 1 \\ &\int x\, p(x \mid \alpha, \beta)\, dx = \mu_1 && \phi_1(x) = x \quad (27) \\ &\int \mathrm{Ln}\,(x)\, p(x \mid \alpha, \beta)\, dx = \mu_2 && \phi_2(x) = \mathrm{Ln}\,(x) \end{aligned}$$

Notons que nous avons les relations analytiques suivantes entre (α, β) et $(m = \mathrm{E}\{x\},\ \sigma^2 = \mathrm{E}\{(x-m)^2\}$:

$$\begin{cases} m = (1-\alpha)/\beta \\ \sigma^2 = (1-\alpha)/\beta^2 \end{cases} \text{ou inversement} \begin{cases} \alpha = (\sigma^2 - m^2)/\sigma^2 \\ \beta = m/\sigma^2 \end{cases} \quad (28)$$

Considérons maintenant les problèmes suivants :

Problème 1 : Etant donné μ_1 et μ_2, déterminer α et β et en déduire la moyenne m et l'écart type σ.

Ce problème est celui du ME classique. Pour calculer α et β il faut résoudre le système d'équations (9). Ceci est fait en utilisant le programme écrit en MATLAB dans la référence [7] qui met en oeuvre la méthode de Newton décrite par les équations (11) à (14).

Le tableau 1 montre les résultats obtenus pour différentes valeurs de μ_1 et μ_2.

Tableau 1: ME

μ_1	μ_2	α	β	m	σ^2
0.2000	-2.0000	-0.4124	-6.9968	0.2019	0.0289
0.3000	-1.3000	-4.3572	-17.8496	0.3001	0.0168
0.3000	-1.5000	-0.6969	-5.3493	0.3172	0.0593

Problème 2 : Etant donné les M échantillons x_i, $i = 1, \ldots, M$ déterminer (α, β) par la méthode MV et en déduire la moyenne m et l'écart type σ en utilisant les relations (28).

Le problème ici est exactement celui décrit plus haut. En utilisant la propriété d'équivalence démontrée plus haut, nous devons estimer μ_1 et μ_2 par leurs estimées empiriques :

$$\mu_1 = \frac{1}{M} \sum_{i=1}^{M} x_i \quad \text{et} \quad \mu_2 = \frac{1}{M} \sum_{i=1}^{M} \mathrm{Ln}\ x_i \quad (29)$$

Ensuite, utilisant la même procédure que celle utilisée dans le problème précédent on obtient α et β , à partir desquels on en déduit m et σ.

Le tableau (2) montre les résultats obtenus pour les trois tailles d'échantillons, M = 4096, 512 et 256. Ces échantillons sont obtenus en générant un vecteur aléatoire de longueur 4096 échantillons suivant une loi gamma avec les paramètres $\alpha = -0{,}2$ et $\beta = 0{,}3$. La moyenne et la variance théoriques sont respectivement, m= 4 et σ^2 = 13,3333.

Tableau 2: MV

Taille de l'échantillon	α	β	m	σ^2
M= 4096	-0.1872	0.2988	3.9732	13.2967
M= 512	-0.2029	0.2895	4.1547	14.3502
M= 256	-0.1412	0.2894	3.9433	13.6255

Problème 3 : Etant donnés les M échantillons x_i, $i = 1, \ldots, M$ déterminer (α, β) par la méthode MM.

Nous avons vu que dans la méthode MM il faut résoudre l'équation (25) qui est un cas particulier de l'équation (21) où les fonctions $\phi_n(x) = x^n$. Bien que l'on puisse résoudre ces équations numériquement, dans le cas de la loi Gamma nous avons une solution analytique. Les relations (28) résument la solution. La méthode consiste alors à estimer μ_1 et μ_2 par :

$$\mu_1 = \frac{1}{M} \sum_{i=1}^{M} x_i \quad \text{et} \quad \mu_2 = \frac{1}{M} \sum_{i=1}^{M} x_i^2 \quad (30)$$

Ensuite, on en déduit $m = \mu_1$ et $\sigma^2 = \mu_2 - \mu_1^2$ et, finalement, utilisant (28) on calcule α et β.

Le tableau (3) montre les résultats obtenus pour les trois tailles d'échantillons, M = 1024, 256 et 64. Nous avons utilisé les mêmes échantillons que dans le problème précédent.

Tableau 3: MM

Taille de l'échantillon	μ_1	μ_2	α	β
M= 4096	-0.1782	0.2966	3.9718	13.3888
M= 512	-0.0880	0.2620	4.1521	15.8454
M= 256	-0.0787	0.2737	3.9417	14.4036

On constate que les résultats obtenus par la méthode MV sont plus fidèles à la réalité. Nous avons aussi observé une plus grande sensibilité des estimations de α et β à la taille des échantillons.

A ce stade nous allons voir l'intérêt de ces développements pour la détermination des paramètres d'une loi *a priori* pour une image dans une approche bayésienne pour la résolution des problèmes inverses que l'on rencontre en restauration et en reconstruction d'image.

4. CONCLUSIONS ET PERSPECTIVES

En traitement d'image (restauration ou reconstruction) on est souvent amené à résoudre le problème suivant :

Déterminer $x \in \mathbb{R}_+^n$ à partir des données $y \in \mathbb{R}^m$:

$$y = A\,x + b \tag{31}$$

où $b \in \mathbb{R}^m$ englobe les erreurs de modélisation et le bruit des mesures, et A est une matrice singulière sinon très mal-conditionnée. On suppose connaître la matrice A et $w_i\,\sigma^2$ les variances du bruit b_i.

Notant par $p(x)$ la loi de probabilité *a priori* de X et par $p(y \mid x)$ la loi de Y conditionnellement à X, la solution bayésienne à ce problème s'obtient par :

$$\hat{x} = \text{Arg} \max_{x \in \mathbb{R}_+^n} p(x \mid y) = \text{Arg} \max_{x \in \mathbb{R}_+^n} \{p(y \mid x)\, p(x)\} \tag{32}$$

Le principe du maximum d'entropie (ME) nous permet d'attribuer $p(y \mid x)$ et $p(x)$. Par exemple connaissant seulement $w_i\,\sigma^2$ sur le bruit b_i, le principe du ME nous donne :

$$p(y \mid x) \propto \exp[-\,Q(x)] \tag{33}$$

avec $Q(x)=[y - Ax]^t\, W\, [y - Ax]$ et $W = \frac{1}{\sigma^2}\,\text{diag}[\frac{1}{w_1{}^2}, \ldots, \frac{1}{w_M{}^2}]$

Une étape importante dans cette approche est la détermination d'une loi de probabilité *a priori* pour l'image. Cette loi doit être la traduction en langage de probabilité de notre connaissance *a priori* sur l'image. Dans le cas où cette connaissance peut être formulée sous forme de l'espérance d'un nombres fini de fonctions dépendant des valeurs des pixels de l'image l'approche ME nous permet d'attribuer une loi *a priori* qui reflète seulement cette connaissance. Dans les travaux précédents [4-8] nous avons étudiés le cas où notre information *a priori* consiste en deux contraintes globales sur $p(x)$:

$$\begin{cases} E\ \{\phi_1(x)\} = \mu_1 \\ E\ \{\phi_2(x)\} = \mu_2 \end{cases} \quad \text{avec} \quad \begin{cases} \phi_1(x) = \sum_{i=1}^{N} H(x_i) \\ \phi_2(x) = \sum_{i=1}^{N} S(x_i) \end{cases}$$

Dans ce cas on a :

$$p(x) = \prod_{i=1}^{N} p(x_i) = \frac{1}{Z(\lambda_1, \lambda_2)} \exp\,[-\lambda_1\, \phi_1(x) - \lambda_2\, \phi_2(x)] \tag{34}$$

Nous avions aussi montré que, si on impose une propriété d'invariance par changement d'échelle sur $p(x)$, les seules choix possibles pour les fonctions $S(x)$ et $H(x)$ sont les suivantes:

$$\{(S(x), H(x))\} = \{(x^{r1}, x^{r2}), (x^{r1}, \text{Ln}\, x), (x^{r1}, x^{r1}\, \text{Ln}\, x), (\text{Ln}\, x, \text{Ln}^2 x)\}$$

Remplaçant (33) et (34) dans (32), la solution de l'équation (32) devient équivalent à :

$$\hat{x} = \text{Arg} \min_{x>0} \{J(x) = Q(x) + \lambda_1\, \phi_1(x) + \lambda_2\, \phi_2(x)\} \tag{35}$$

Les difficultés pratiques sont à deux niveaux :

1) Une fois les hyper paramètres (λ_1, λ_2) déterminés, résoudre d'une manière efficace le problème de la minimisation (35).

2) Estimer ces hyper paramètres (λ_1, λ_2), qui jouent le rôle des paramètres de régularisation, à partir des données y.

Nous n'abordons pas ici la première. Pour la deuxième, nous proposons la méthode itérative suivante :

Nous supposons disposer d'une bonne estimation initiale de la solution $x^{(0)}$. Nous proposons de choisir pour l'estimation initiale

$$x^{(0)} = \frac{1}{\det(A^t A)} A^t y$$

On peut alors, à chaque itération, estimer k, $x^{(k)}$ et $(\lambda_1^{(k)}, \lambda_2^{(k)})$ par:

1) Estimer $\mu_1^{(k)} = \text{E}\{\phi_1(x^{(k)}\}$ et $\mu_2^{(k)} = \text{E}\{\phi_2(x^{(k)})\}$ par :

$$\mu_1^{(k)} = \sum_{j=1}^{N} S(x_j^{(k)}) \qquad \text{et} \qquad \mu_2^{(k)} = \sum_{j=1}^{N} H(x_j^{(k)})$$

2) Déterminer $(\lambda_1^{(k)}, \lambda_2^{(k)})$ en solvant les équations (9)

4) Résoudre (35) avec cette nouvelle estimation pour obtenir $x^{(k+1)}$ et retour à l'étape 1.

Notons que les étapes 1 et 2 correspondent à l'estimation par MV des paramètres (λ_1, λ_2) à partir des échatillons $x^{(k)}$, ce qui suppose que ces échantillons soient des réalisations possibles de la loi *a priori* (34).

D'autre part, bien que nous n'avons pas de propriété théorique de la convergence de cette méthode, en pratique, son utilisation dans plusieurs applications en restauration et reconstruction d'image montre que la méthode converge. Une étude sur le lien de cette méthode avec la technique de l'estimation par l'algorithme EM est en cours.

REFERENCES BIBLIOGRAPHIQUES

[1] A. Zellner and R. Highfiled, "Calculation of Maximum Entropy Distributions and Approximation of Marginal Posterior Distributions", Journal of Econometrics 37, 1988, 195-209, North Holland.

[2] Gokhale, "Maximum Entropy Characterizations of some distributions", Statistical distributions in Scientific work, vol. 3, 299-304 (G.P. Patil et al., Eds., Reidel, Dordrecht, Holland, 1975).

[3] Jaynes, "Papers on probability, statistics and statistical physics", Reidel Publishing Company, Dordrecht, Holland, 1983.

[4] Mohammad-Djafari A. et Demoment G., "*Maximum Entropy and Bayesian Approach in Tomographic Image Reconstruction and Restoration,*" Maximum Entropy and Bayesian Methods, J. Skilling (ed.) by Kluwer Academic Publishers, pp:195-201, (1989).

[5] Mohammad-Djafari A. et Demoment G., "*Estimating Priors in Maximum Entropy Image Processing,*" ICASSP, Albuquerque, USA, Avril 1990.

[6] Mohammad-Djafari A. et Idier J., "*Maximum Entropy Priors Laws of Images and Estimation of their parameters,*" Proceedings of the 10th Int. Conf. on Maximum Entropy and Bayesian Methods, Laramie, USA, August 1990.

[7] Mohammad-Djafari, "*A Matlab program to calculate the Maximum Entropy Distributions,*" Rapport interne LSS, Communication à : The 11th Int. Conf. on Maximum Entropy and Bayesian Methods, Seattle, USA, June 1991.

[8] Mohammad-Djafari, "*Maximum Likelihood Estimation of the Lagrangian Parameters of the Maximum Entropy Distributions,*" Rapport interne LSS, Communication à : The 11th Int. Conf. on Maximum Entropy and Bayesian Methods, Seattle, USA, June 1991.