



Estimation des Hyperparamètres dans une Approche Bayésienne de la Résolution des Problèmes Inverses Linéaires

Ali Mohammad-Djafari

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

RÉSUMÉ

Dans cette communication nous proposons d'étudier et comparer deux méthodes pour l'estimation des hyperparamètres (paramètres de la loi *a priori*) dans une approche bayésienne de la résolution des problèmes inverses linéaires que l'on rencontre en reconstruction et en restauration des signaux et des images. Les lois *a priori* sont choisies parmi des lois de la famille exponentielles généralisées qu'on peut considérer comme des lois à maximum d'entropie. Nous nous sommes limités au cas des lois à deux paramètres qui satisfont de plus à une propriété d'invariance par changement d'échelle. Les deux méthodes sont : la méthode du maximum de vraisemblance généralisée (MVG) et la méthode du maximum de vraisemblance marginale (MVV) implantée par l'algorithme EM. Quelques résultats de simulation sont présentés pour montrer les performances de la méthode du MVG.

1. INTRODUCTION

Dans un grand nombre de problèmes inverses linéaires que l'on rencontre en reconstruction et restauration des signaux et des images, après discrétisation, on est amené à résoudre un système d'équations linéaires de la forme :

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}, \quad (1)$$

où \mathbf{x} est un vecteur contenant les grandeurs à estimer (par exemple les pixels d'une image), \mathbf{y} est un vecteur contenant les grandeurs observées (par exemple les pixels d'une image dégradée ou les projections ou les mesures du champ diffracté dans un problème de la tomographie à onde diffractée), \mathbf{b} est un vecteur contenant les erreurs de mesures et de modélisation que l'on suppose blanc et indépendant de \mathbf{x} , et \mathbf{A} est une matrice de très grandes dimensions qui lie \mathbf{y} et \mathbf{x} .

Dans une approche bayésienne pour estimation de \mathbf{x} , une fois les lois de probabilité $p(\mathbf{x}; \theta)$ et $p(\mathbf{y}|\mathbf{x})$ définies, on calcule la loi *a posteriori*

$$p(\mathbf{x}|\mathbf{y}; \theta) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}; \theta), \quad (2)$$

qui contient toute information que l'on peut connaître sur le vecteur des paramètres inconnus \mathbf{x} . Mais, en général, pour définir une estimation ponctuelle, on définit une fonction coût $C(\hat{\mathbf{x}}, \mathbf{x})$ et on minimise le coût moyen. Suivant le choix de la fonction coût on obtient, par exemple, l'estimation en moyenne quadratique ou à maximum *a posteriori* (MAP). Dans cette communication nous nous limiterons à l'estimée

ABSTRACT

In this paper we propose to study and to compare two methods for the estimation of the hyperparameters (the parameters of the prior law) when a Bayesian approach is used to solve the inverse problems which arise in signal and image reconstruction and restoration problems. The prior laws are choosed between the generalized exponential family probability density functions (*pdf*) which can be considered as the maximum entropy *pdf*'s. We limit ourselves to a family of two parameters laws which satisfy also a scale invariance property. The two methods are : the Generalized Maximum Likelihood (GML) method and the Marginalized Maximum Likelihood (MML) method implemented by the Expectation Maximization (EM) algorithm. Some simulation results are given to show the performances of the the GML method.

MAP qui est définie par

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \{p(\mathbf{x}|\mathbf{y}; \theta)\}. \quad (3)$$

Dans des travaux précédents [3]–[7], utilisant le principe du maximum d'entropie nous avons proposé de choisir une loi *a priori* $p(\mathbf{x}; \theta)$ de la forme

$$p(\mathbf{x}; \lambda, \mu) = \frac{1}{Z(\lambda, \mu)} \exp[-\lambda\phi_1(\mathbf{x}) - \mu\phi_2(\mathbf{x})], \quad (4)$$

avec

$$\phi_1(\mathbf{x}) = \sum_{n=1}^N H(x_n), \quad \text{et} \quad \phi_2(\mathbf{x}) = \sum_{n=1}^N S(x_n), \quad (5)$$

où $H(x)$ et $S(x)$ sont deux fonctions connues. Notons que ces lois appartiennent à la famille des lois exponentielles généralisées à deux paramètres. Un argument d'invariance par changement d'échelle [4], nous a permis de restreindre le choix de ces fonctions dans l'ensemble suivant :

$$\{(S(x), H(x))\} = \{(x^{r_1}, x^{r_2}), (x^{r_1}, \ln x), (x^{r_1}, x^{r_1} \ln x), (\ln x, \ln^2 x)\}, \quad (6)$$

où r_1 et r_2 sont deux nombres réels.

Utilisant une loi *a priori* $p(\mathbf{x}; \lambda, \mu)$ parmi cette famille, et faisant l'hypothèse que le vecteur du bruit \mathbf{b} est gaussien, centré avec la matrice de covariance $\mathbf{R}_b = \sigma_b^2 \mathbf{I}$, l'estimée



MAP devient :

$$\begin{aligned}\hat{\mathbf{x}} &= \arg \max_{\mathbf{x}} \{p(\mathbf{x} | \mathbf{y})\} \\ &= \arg \min_{\mathbf{x}} \{J(\mathbf{x}) = Q(\mathbf{x}) + \lambda_1 \phi_1(\mathbf{x}) + \lambda_2 \phi_2(\mathbf{x})\},\end{aligned}$$

avec

$$Q(\mathbf{x}) = [\mathbf{y} - \mathbf{A}\mathbf{x}]^t \mathbf{R}_b^{-1} [\mathbf{y} - \mathbf{A}\mathbf{x}]. \quad (7)$$

Une étude supplémentaires nous a permis de définir les conditions sur les paramètres (λ, μ) pour que la loi *a posteriori* soit unimodale, et par conséquence, le critère $J(\mathbf{x})$ à minimiser soit convexe, ce qui permet d'effectuer la minimisation de ce critère à l'aide d'un algorithme du gradient ou gradients conjugués.

2. PRÉSENTATION DU PROBLÈME

Dans une approche bayésienne il faut d'abord attribuer des lois de probabilités directes $p(\mathbf{x})$ et $p(\mathbf{y}|\mathbf{x})$. Ceci peut se faire soit par des considérations physiques soit d'une manière logique. Dans le premier cas ces lois sont imposées. Dans le second cas, ceci peut se faire soit d'une manière arbitraire mais de bon sens : on choisit une loi dans une famille paramétrique pour traduire une propriété, par exemple une loi gaussienne pour une variable appartenant à \mathbf{R} , une loi Gamma pour une variable appartenant à \mathbf{R}_+ ou une loi Beta pour une variable appartenant à $[0, 1]$, soit en utilisant le principe du maximum d'entropie (ME) quand on dispose des informations sous la forme d'un ensemble fini d'espérances, par exemple la moyenne, la variance, la moyenne géométrique, etc. C'est ce deuxième point de vue qui est utilisé dans ce travail. Supposons donc que notre connaissance *a priori* sur \mathbf{b} et sur \mathbf{x} est de la forme :

• sur \mathbf{b} on connaît la matrice de covariance $\mathbf{R}_b = \sigma_b^2 \mathbf{I}$. L'utilisant du principe du ME nous fournit la loi :

$$p(\mathbf{y}|\mathbf{x}) \propto \exp \left[-\frac{1}{2} Q(\mathbf{x}) \right], \quad (8)$$

avec $Q(\mathbf{x})$ donnée par l'équation (7),

• sur \mathbf{x} notre connaissance est supposée être de la forme :

$$\begin{cases} \mathbb{E} \{ \phi_1(\mathbf{x}) \} = \mu_1, \\ \mathbb{E} \{ \phi_2(\mathbf{x}) \} = \mu_2, \end{cases} \quad (9)$$

avec $\phi_1(\mathbf{x})$ et $\phi_2(\mathbf{x})$ définies dans (5).

Là aussi, le principe du ME nous donne :

$$p(\mathbf{x}) = \prod_{i=1}^N p(x_i) \propto \exp [-\lambda \phi_1(\mathbf{x}) - \mu \phi_2(\mathbf{x})], \quad (10)$$

avec

$$p(x_i) = \frac{1}{Z(\lambda_1, \lambda_2)} \exp [-\lambda_1 H(x_i) - \lambda_2 S(x_i)], \quad (11)$$

et où

$$Z(\lambda_1, \lambda_2) = \int \exp [-\lambda_1 H(x) - \lambda_2 S(x)] dx \quad (12)$$

est la fonction de partition. Les paramètres (λ_1, λ_2) sont reliés aux (μ_1, μ_2) .

Un problème bien-posé et simple dans son principe dans cette approche bayésienne est :

“Étant donné la matrice \mathbf{A} , le vecteur des mesures \mathbf{y} , la matrice de covariance du bruit \mathbf{R}_b , les fonctions $H(x)$ et $S(x)$ et (μ_1, μ_2) ou d'une manière équivalente les paramètres (λ_1, λ_2) estimer la solution $\hat{\mathbf{x}}$.”

La solution est immédiate : utilisant la règle de Bayes on calcule la loi *a posteriori*

$$p(\mathbf{x} | \mathbf{y}) \propto \exp \{-Q(\mathbf{x}) - \lambda_1 \phi_1(\mathbf{x}) - \lambda_2 \phi_2(\mathbf{x})\}, \quad (13)$$

et utilisant la règle d'estimation MAP on a :

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \{p(\mathbf{x}|\mathbf{y})\} = \arg \min_{\mathbf{x}} \{J(\mathbf{x})\}, \quad (14)$$

avec

$$J(\mathbf{x}) = Q(\mathbf{x}) + \lambda_1 \phi_1(\mathbf{x}) + \lambda_2 \phi_2(\mathbf{x}). \quad (15)$$

Cependant, dans les applications réelles trois problèmes restent non-résolus :

- une fois les paramètres (λ_1, λ_2) fixés, comment résoudre d'une manière efficace le problème d'optimisation (14),
- comment choisir les fonctions $\phi_1(\mathbf{x})$ et $\phi_2(\mathbf{x})$, ou d'une manière équivalente comment choisir les fonctions $H(x)$ et $S(x)$, et surtout
- comment déterminer les hyperparamètres (λ_1, λ_2) à partir des données \mathbf{y} , car en pratique (μ_1, μ_2) ne sont pas connues.

Dans des travaux précédents [3]–[7], nous avons déjà proposé des solutions pour les deux premières questions. En effet, en ce qui concerne la deuxième question un argument d'invariance par changement d'échelle [4], nous a permis de restreindre le choix de ces fonctions dans l'ensemble suivant :

$$\{(S(x), H(x))\} = \{(x, x^2), (x, \ln x), (x, x \ln x)\}, \quad (16)$$

qui est un sous-ensemble de (6). En ce qui concerne la première question, nous avons défini les conditions sur les paramètres (λ, μ) pour que le critère $J(\mathbf{x})$ à minimiser soit convexe, ce qui permet d'utiliser un algorithme du gradient ou gradients conjugués pour son optimisation.

Dans ce travail, nous sommes intéressés au troisième problème qui est l'estimation des hyperparamètres. Plus précisément nous avons considéré le problème suivant :

“Étant donné la matrice \mathbf{A} , le vecteur des mesures \mathbf{y} et la matrice de covariance du bruit \mathbf{R}_b , les fonctions H et S , estimer (au sens du MAP) la solution $\hat{\mathbf{x}}$ et les hyperparamètres $\hat{\theta} = (\hat{\lambda}, \hat{\mu})$.”

3. MÉTHODES PROPOSÉES

L'approche bayésienne nous propose deux voies :

- Maximum de vraisemblance généralisé (MVG):
- L'idée de base dans cette approche est de considérer les hyperparamètres $\theta = (\lambda, \mu)$ au même niveau que les paramètres \mathbf{x} et de les estimer par

$$(\hat{\mathbf{x}}, \hat{\theta}) = \arg \max_{(\mathbf{x}, \theta)} \{p(\mathbf{x}, \mathbf{y}; \theta) = p(\mathbf{x}; \theta)p(\mathbf{y}|\mathbf{x})\}. \quad (17)$$

Nous avons proposé de résoudre ce problème d'optimisation en le transformant en deux problèmes d'optimisation successives, l'un par rapport à θ et l'autre par rapport à \mathbf{x} :

$$\begin{cases} \hat{\theta}^{(k)} &= \arg \max_{\theta} \{p(\hat{\mathbf{x}}^{(k)}, \mathbf{y}; \theta)\} \\ \hat{\mathbf{x}}^{(k+1)} &= \arg \max_{\mathbf{x}} \{p(\mathbf{x}|\mathbf{y}; \hat{\theta}^{(k)})\} \end{cases} \quad (18)$$

Notons que la première équation se simplifie

$$\hat{\theta}^{(k)} = \arg \max_{\theta} \left\{ p(\hat{\mathbf{x}}^{(k)}; \theta) \right\}, \quad (19)$$

ce qui nous permettra d'interpréter $\hat{\theta}^{(k)}$ comme une estimation au sens de maximum de vraisemblance (*a priori*), si $\hat{\mathbf{x}}^{(k)}$ pouvait être considéré comme un vecteur d'échantillon effectivement observé de la loi *a priori* $p(\mathbf{x}; \theta)$. La deuxième équation n'est autre que l'estimé MAP une fois les paramètres fixés à l'itération précédente.

• **Maximum de vraisemblance marginale (MVM)**

L'idée de base dans cette approche est de ne pas considérer les hyperparamètres $\theta = (\lambda, \mu)$ au même niveau que les paramètres \mathbf{x} , ce qui est effectivement le cas. Il s'agit alors d'abord estimer θ par

$$\hat{\theta} = \arg \max_{\theta} \left\{ L(\theta) = \int p(\mathbf{x}, \mathbf{y}; \theta) d\mathbf{x} \right\}. \quad (20)$$

Ensuite, cet estimé peut être servi, dans une deuxième étape pour l'estimation de la solution au problème inverse considéré :

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \left\{ p(\mathbf{x}|\mathbf{y}; \hat{\theta}) \right\}. \quad (21)$$

Malheureusement $L(\theta)$ est rarement calculable (sauf dans le cas gaussien et certain autre cas d'école), c'est pourquoi on pourrait envisager l'utilisation de l'algorithme EM (Expectation–Maximization) qui dans notre cas s'écrit comme suit :

$$\begin{cases} \text{E: } Q(\theta; \hat{\theta}^{(k)}) = \mathbb{E}_{\mathbf{x}|\mathbf{y}; \hat{\theta}^{(k)}} \{ \ln p(\mathbf{x}, \mathbf{y}; \theta) \} \\ \text{M: } \hat{\theta}^{(k+1)} = \arg \max_{\theta} \left\{ Q(\theta; \hat{\theta}^{(k)}) \right\} \end{cases} \quad (22)$$

L'objet de cette communication est de comparer ces deux méthodes dans notre cas particulier où les lois *a priori* sont des lois choisies dans la famille (16).

Revenant à la notation $\theta = (\lambda_1, \lambda_2)$ et remplaçant $p(\mathbf{x}; \lambda_1, \lambda_2)$ de l'équation (10) dans les équations (19) et (22) on peut faire la comparaison suivante :

• **MVG** : À l'itération $(k + 1)$ de l'algorithme d'optimisation (19) on a à résoudre :

$$(\hat{\lambda}_1, \hat{\lambda}_2)^{(k+1)} = \arg \max_{(\lambda_1, \lambda_2)} \left\{ p(\hat{\mathbf{x}}_{MAP}^{(k)}; \lambda_1; \lambda_2) \right\}. \quad (23)$$

Pour ceci on aura à résoudre le système d'équations non-linéaires suivant :

$$\begin{cases} -\frac{\partial \ln Z(\lambda_1, \lambda_2)}{\partial \lambda_1} = \frac{1}{N} \phi_1(\hat{\mathbf{x}}_{MAP}^{(k)}) \\ -\frac{\partial \ln Z(\lambda_1, \lambda_2)}{\partial \lambda_2} = \frac{1}{N} \phi_2(\hat{\mathbf{x}}_{MAP}^{(k)}) \end{cases} \quad (24)$$

Les parties droites de ces équations sont :

$$\begin{cases} \frac{1}{N} \phi_1(\hat{\mathbf{x}}_{MAP}^{(k)}) = \frac{1}{N} \sum_{i=1}^N H(\hat{\mathbf{x}}_{MAP_i}^{(k)}) \\ \frac{1}{N} \phi_2(\hat{\mathbf{x}}_{MAP}^{(k)}) = \frac{1}{N} \sum_{i=1}^N S(\hat{\mathbf{x}}_{MAP_i}^{(k)}) \end{cases} \quad (25)$$

• **MVM–EM** : Dans la mise en œuvre de l'algorithme EM (22), à l'itération $(k + 1)$ on a à estimer

$$(\hat{\lambda}_1, \hat{\lambda}_2)^{(k+1)} = \arg \max_{(\lambda_1, \lambda_2)} \left\{ Q((\lambda_1, \lambda_2), (\hat{\lambda}_1, \hat{\lambda}_2)^{(k)}) \right\}, \quad (26)$$

ce qui dans notre cas revient à résoudre le système d'équations non-linéaires suivant :

$$\begin{cases} -\frac{\partial \ln Z(\lambda_1, \lambda_2)}{\partial \lambda_1} = \frac{1}{N} \int \phi_1(\mathbf{x}) p(\mathbf{x}|\mathbf{y}) d\mathbf{x} \\ -\frac{\partial \ln Z(\lambda_1, \lambda_2)}{\partial \lambda_2} = \frac{1}{N} \int \phi_2(\mathbf{x}) p(\mathbf{x}|\mathbf{y}) d\mathbf{x} \end{cases} \quad (27)$$

Les parties droites de ces équations sont :

$$\begin{cases} \frac{1}{N} \int \phi_1(\mathbf{x}) p(\mathbf{x}|\mathbf{y}) d\mathbf{x} = \frac{1}{N} \sum_{i=1}^N \int H(x_i) p(x_i|\mathbf{y}) dx_i \\ \frac{1}{N} \int \phi_2(\mathbf{x}) p(\mathbf{x}|\mathbf{y}) d\mathbf{x} = \frac{1}{N} \sum_{i=1}^N \int S(x_i) p(x_i|\mathbf{y}) dx_i \end{cases} \quad (28)$$

Ainsi, en comparant les équations (24–25) aux (27–28), on peut remarquer que si $p(\mathbf{x}|\mathbf{y})$ est très concentrée autour de $\hat{\mathbf{x}}_{MAP}$, *i.e.*; $p(\mathbf{x}|\mathbf{y}; \theta) \approx \delta(\mathbf{x} - \hat{\mathbf{x}}_{MAP})$, alors les parties droites des équations (25) et (28) seront identiques et les deux méthodes fourniront les mêmes résultats numériques. Bien entendu, ceci ne sera pas le cas en général.

La méthode du MVG est plus facile à implanter et son coût de calcul reste raisonnable, tandis que le calcul des intégrales dans l'algorithme EM de la méthode MVM demandent un coût de calcul prohibitif. En effet si on veut appliquer cette méthode en restauration d'images on a à calculer N intégrales, N étant le nombre de pixels de l'image. Pour remédier à ceci on peut envisager deux solutions :

- faire une approximation gaussienne de la loi *a posteriori* en développant, en série de Taylor, le critère $J(\mathbf{x})$, autour de $\hat{\mathbf{x}}_{MAP}$, ou
- calculer ces intégrales par une méthode stochastique ce qui nous amène l'algorithme SEM (EM stochastique).

4. RÉSULTATS DE SIMULATIONS

Nous n'avons pas encore implanté la méthode MVM. Nous présentons donc les performances de la méthode MVG qui est implantée et utilisée dans des applications en restauration et reconstruction d'image. Ici, nous présentons seuls les résultats de simulation en restauration d'image.

Pour ceci, nous avons simulés quatre images synthétiques (O1, O2, O3 and O4) (Fig. 1), et pour chacune nous avons calculé et présenté leurs histogrammes (Fig. 2). En fonction de l'allure de chaque histogramme, nous avons choisi une forme appropriée pour la loi *a priori* (choix des fonctions H et S) et estimé leurs paramètres associés par la méthode du maximum de vraisemblance (MV). Ainsi, à chaque objet est associée une loi de probabilité parmi les lois à ME mentionnées dans le texte (10)–(12), avec les fonctions H et S choisies parmi (16).

L'estimation par MV des paramètres (λ_1, λ_2) à partir de ces objets se fait en résolvant le système d'équations non-linéaires suivant :

$$\begin{cases} -\frac{\partial \ln Z(\lambda_1, \lambda_2)}{\partial \lambda_1} = \frac{1}{N} \sum_{i=1}^N H(x_i) \\ -\frac{\partial \ln Z(\lambda_1, \lambda_2)}{\partial \lambda_2} = \frac{1}{N} \sum_{i=1}^N S(x_i) \end{cases}$$



Ainsi, nous avons attribué :

- une loi gaussienne ($H(x) = x^2, S(x) = x$) à l'objet O1,
- une loi gaussienne tronquée ($H(x) = x^2, S(x) = x, x > 0$) à l'objet O2,
- une loi Gamma ($H(x) = \ln x, S(x) = x, x > 0$) à l'objet O3, et, finalement
- une loi *entropique* ($H(x) = -x \ln x, S(x) = x, x > 0$) à l'objet O4.

Le Tableau 1. résume le résultat de l'estimation des paramètres (λ_1, λ_2) à partir de ces objets par la méthode du MV.

Dans l'étape suivante, ces images ont été convoluées par une réponse impulsionnelle de forme gaussienne et bruitées avec un bruit gaussien de variance calculée de telle sorte que le rapport signal à bruit soit environ 20dB (Fig. 3).

Finalement, utilisant la méthode du MVG, ces images sont restaurées et simultanément les hyperparamètres sont estimés. Les résultats de l'estimation des hyperparamètres sont donnés dans le tableau 2, et les résultats de restauration sont montrés sur la figure 4.

La figure 5 montre les histogrammes empiriques et les lois *a priori* estimées (utilisant les paramètres λ, μ estimés par la méthode MVG).

images	$H(x)$	$S(x)$	domaine de x	λ_1	λ_2
O1	x^2	x	\mathbf{R}	18.1	-18.5
O2	x^2	x	$x > 0$	5.94	-0.78
O3	$\ln x$	x	$x > 0$	-0.56	14.3
O4	$-x \ln x$	x	$x > 0$	-11.1	9.46

Tableau 1. Les lois *a priori* et leurs paramètres estimés par la méthode du MV pour les images O1, O2, O3 et O4.

images	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$J(x)$	$Q(x)$	D
O1	18.7	-19.1	5.7e4	6.2e4	2.8e-2
O2	17.8	-7.58	2.4e4	2.5e4	1.4e-1
O3	-6.01	66.9	1.9e5	1.8e5	4.3e-1
O4	-64.8	64.2	9.8e5	1.0e6	2.3e-1

Tableau 2. Résultats de l'estimation des hyperparamètres ($\hat{\lambda}_1, \hat{\lambda}_2$), le critère $J(x)$, le terme $Q(x)$ et D la distance relative entre l'objet x et l'objet restauré \hat{x} .

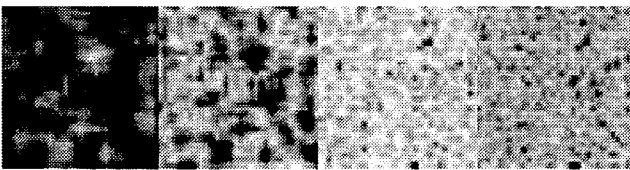


Figure 1. Les quatre images O1, O2, O3 et O4.

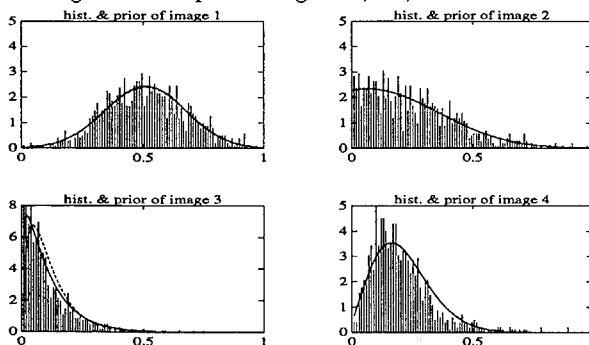


Figure 2. Les histogrammes et les lois *a priori* choisies des images O1, O2, O3 et O4.



Figure 3. Les images dégradées.

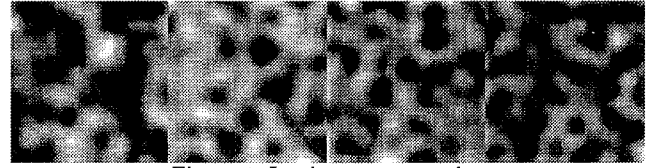


Figure 4. Les images restaurées.

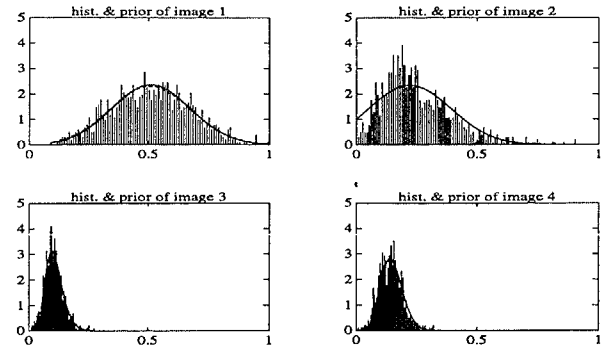


Figure 5. Les histogrammes et les lois *a priori* estimées.

5. CONCLUSIONS

Le principal objectif dans cette communication était de présenter deux méthodes pour l'estimation des hyperparamètres dans une approche bayésienne de la résolution des problèmes inverses. La mise en œuvre de la méthode MVM par l'algorithme EM nécessite le calcul numérique d'un grand nombre d'équations intégrales, ce qui limite son utilisation en pratique. Par contre, l'implantation de la méthode du MVG est plus facile et son coût de calcul est raisonnable. Cependant cette méthode n'a pas de bonnes propriétés, car, dans certain cas, le critère peut ne pas avoir de maximum. Une manière de surmonter à cette difficulté est de restreindre les paramètres. Cependant, l'estimée \hat{x} dépend beaucoup plus à la forme de la loi qu'à ses paramètres. Nous avons constaté, à travers quelques simulations, qu'en pratique, la méthode converge mais reste sensible à l'initialisation.

6. RÉFÉRENCES

- [1] Dempster A. P., Laird N. M., and Rubin D. B., "Maximum Likelihood from incomplete data via the EM algorithm.," *J. Royal Statist. Soc., B*, Vol. 39, pp: 1-38, 1977.
- [2] Miller M., Snyder D., "The Role of Likelihood and Entropy in Incomplete-Data Problems: Applications to Estimating Point-Process Intensities and Toeplitz Constrained Covariances," *Proceedings of the IEEE*, Vol.75, No.7 pp:892-906, 1987.
- [3] Mohammad-Djafari A. et Demoment G., "Estimating Priors in Maximum Entropy Image Processing," *Proc. of ICASSP 1990*, pp: 2069-2072
- [4] Mohammad-Djafari A. et Idier J., "Maximum entropy prior laws of images and estimation of their parameters," *Proc. of The 10th Int. MaxEnt Workshop, Laramie, Wyoming*, in Maximum-Entropy and Bayesian methods, T.W. Grandy ed., 1990.
- [5] Mohammad-Djafari, "Maximum Likelihood Estimation of the Lagrange Parameters of the Maximum Entropy Distributions," *Proc. of The 11th Int. MaxEnt Workshop, Seattle, USA*, 1991.
- [6] Mohammad-Djafari, "On the Estimation of Hyperparameters in Bayesian Approach of Solving Inverse Problems," *ICASSP-93, Minneapolis, USA*, 1993.