

# MÉTHODE DU MAXIMUM D'ENTROPIE SUR LA MOYENNE ET MÉLANGES DE DISTRIBUTIONS.

C. HEINRICH, J.-F. BERCHER, G. LE BESNERAIS et G. DEMOMENT.

Laboratoire des signaux et systèmes (CNRS-SUPÉLEC-UPS)  
Plateau de Moulon  
91192 Gif-sur-Yvette Cedex  
e\_mail : memm@lss.supelec.fr

## RÉSUMÉ

L'objet de cette communication est la résolution de problèmes inverses linéaires mal-posés. Nous proposons de résoudre ce type de problème par la Méthode du Maximum d'Entropie sur la Moyenne (MMEM). La MMEM permet de prendre en compte diverses informations *a priori*, contrairement à la méthode du maximum d'entropie classique. Ces informations seront incorporées non par la règle de BAYES comme dans les méthodes bayésiennes, mais par l'information de KULLBACK-LEIBLER.

Nous présentons dans un premier temps le principe de la méthode, avant d'étudier deux exemples d'application à des mélanges de distributions représentatifs de signaux piqués. Les cadres d'application seront respectivement le contrôle non-destructif et la spectroscopie.

## 1. INTRODUCTION

NOUS NOUS intéressons dans cette communication à la restauration de signaux  $\mathbf{x}$  observés par un système linéaire  $\mathbf{H}$  et dégradés par un bruit additif  $\mathbf{n}$ . Les observations  $\mathbf{z}$  satisfont l'équation  $\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{n}$ .

On supposera la matrice  $\mathbf{H}$  et la distribution  $\nu_N$  du bruit connues. La matrice  $\mathbf{H}$  est souvent mal conditionnée et ces seules connaissances ne suffisent donc pas pour obtenir une solution physiquement acceptable. Il faut introduire un supplément d'information. Nous modéliserons ici le problème en choisissant une distribution  $\nu$  de référence pour la classe d'objets étudiés. On prendra par exemple pour  $\nu$  un mélange de distributions, ce qui permettra de rendre compte des signaux piqués que nous recherchons. Dans le cadre des méthodes bayésiennes, cette hypothèse permet – par application de la règle de BAYES – d'obtenir la distribution *a posteriori*, dont on déduit un estimateur de  $\mathbf{x}$ . Dans le cadre de la MMEM, on minimise la distance de KULLBACK à  $\nu$  afin d'en déduire une distribution à maximum d'entropie, puis un estimateur de  $\mathbf{x}$ .

Par souci de simplicité, nous traiterons dans un premier temps le problème non bruité, le cas du problème bruité s'en déduisant aisément.

## 2. PRINCIPE DE LA MÉTHODE

La MMEM a été introduite initialement par NAVAZA [8] en cristallographie, puis étudiée par GAMBOA et GASSIAT [3, 4] en mathématiques et par LE BESNERAIS et BERCHER [1, 6] en traitement du signal. Afin d'introduire la méthode, nous proposons ici une analogie entre les signaux étudiés et

## ABSTRACT

The subject of this communication is the resolution of linear ill-posed inverse problems. We suggest to cope with this kind of difficulty using the maximum entropy on the mean method. This scheme allows to account for various priors, whereas classical maximum entropy doesn't. This prior information won't be incorporated via the BAYES' rule as in Bayesian methods, but via the KULLBACK-LEIBLER information.

We first present the principle of the method, and we then focus on its application to mixture of distributions standing for a spiky signal prior knowledge. Application field will be non-destructive evaluation and spectroscopy.

les états d'un système physique. D'autres arguments, s'appuyant sur des résultats de théorie de l'information, ou des résultats de grandes déviations, peuvent être trouvés dans [1, 4].

Considérons par exemple un ensemble de  $k$  particules d'un gaz discret dont nous étudions les vitesses  $\mathbf{u} = [u(1)\dots u(k)]^t$  dans une direction de l'espace,  $u(j)$  étant la vitesse de la particule  $j$ . Le vecteur  $\mathbf{u}$  est appelé microétat.

Considérons  $m$  réalisations de ce microétat, ces différentes réalisations pouvant être simplement des mesures à des instants distincts. Nous observons le macroétat  $\mathbf{x} = \frac{1}{m} \sum_{i=1}^m \mathbf{u}_i$  au moyen d'un instrument de réponse  $\mathbf{H}$ , c'est-à-dire que nous observons  $\mathbf{z} = \mathbf{H}\mathbf{x}$ . L'ensemble des  $m$  microétats définit une distribution empirique  $p = \frac{1}{m} \sum_{i=1}^m \delta_{\mathbf{u}_i}$ , dont la moyenne est le macroétat :  $\mathbf{x} = E_p[\mathbf{u}]$ .

Le problème expérimental est le suivant : déterminer une valeur physiquement acceptable du macroétat  $\mathbf{x}$  à partir des observations  $\mathbf{z}$ , éventuellement bruitées, de ce macroétat dégradé par l'instrument de mesure. Il est en général possible de compléter cet énoncé par la connaissance d'une distribution typique ou le calcul d'une distribution théorique des vitesses, compte tenu des contraintes physiques. Considérons que ces microétats ont été engendrés sous cette distribution  $\nu$ . On peut dès lors chercher la distribution empirique la plus probable compte tenu des observations  $\mathbf{z}$ , le macroétat recherché étant alors identifié à la moyenne de cette distribution.

Remarquons dès à présent que nous pouvons ainsi aisément imposer des contraintes (par exemple une contrainte d'appartenance à un gabarit) au macroétat : en effet, si  $\nu$  mesure un certain ensemble  $\mathcal{C}$ ,  $\mathbf{u}$  prend ses valeurs

dans  $\mathcal{C}$  et le macroétat correspondant prend ses valeurs dans l'enveloppe convexe de  $\mathcal{C}$ .

La probabilité de réalisation de chaque distribution empirique est une quantité dépendant de  $\nu$ , que nous noterons  $\mathcal{Q}(\cdot|\nu)$ . Les distributions empiriques suivent un principe de grandes déviations de niveau 2 [2] dont la fonction de taux (ou entropie) est la distance (ou information) de KULLBACK-LEIBLER  $\mathcal{K}(\cdot, \nu)$  [5]. Ceci s'écrit :

$$\mathcal{Q}(p \in \mathcal{A}|\nu) \approx \exp(-m \inf_{p \in \mathcal{A}} \mathcal{K}(p, \nu))$$

avec  $\mathcal{K}(p, \nu) = \int \log \frac{dp}{d\nu} dp$ .

Notons  $\mathcal{C}_p$  l'ensemble des distributions satisfaisant les observations, c'est-à-dire satisfaisant  $\mathbf{z} = \mathbf{H}E_p[\mathbf{u}]$ . Pour de grandes valeurs de  $m$ ,  $\mathcal{Q}(\cdot|\nu)$  charge un seul point de  $\mathcal{C}_p$ , celui vérifiant

$$\hat{p} = \arg \min_{p \in \mathcal{C}_p} \mathcal{K}(p, \nu),$$

c'est-à-dire :  $\forall p \neq \hat{p} \quad \mathcal{Q}(\hat{p}|\nu) \gg \mathcal{Q}(p|\nu)$ .

La distribution  $\hat{p}$  a donc infiniment plus de chances de se réaliser, connaissant  $\mathbf{z}$  et  $\nu$ , que toute autre distribution de  $\mathcal{C}_p$ .

Puisque l'on observe indirectement des moyennes, ou macroétats, c'est par conséquent le macroétat  $\hat{\mathbf{x}} = E_{\hat{p}}[\mathbf{u}]$  qui a infiniment plus de chances d'être à l'origine des observations  $\mathbf{z}$  que tout autre macroétat. C'est lui que nous retiendrons donc pour estimateur.

Ces principes peuvent se formuler comme suit :

$$\begin{cases} \hat{p} = \arg \min_p \mathcal{K}(p, \nu), \\ \hat{\mathbf{x}} = E_{\hat{p}}[\mathbf{u}], \\ \mathbf{z} = \mathbf{H}\hat{\mathbf{x}}. \end{cases} \quad (1)$$

La distribution  $\hat{p}$  sélectionnée est la distribution la moins informative par rapport à  $\nu$  et aux données. Tout autre choix incorporerait des informations supplémentaires puisque l'on choisirait alors une distribution ayant infiniment moins de chances de se réaliser que  $\hat{p}$ . Ce choix serait donc plus compromettant.

Au lieu de raisonner au niveau des distributions et de minimiser explicitement une information de KULLBACK, il est possible de reformuler le problème directement en termes de macroétats. Soit  $\mathcal{C}_{\mathbf{x}}$  l'ensemble des macroétats satisfaisant les données, et soit  $\mathbf{x} \in \mathcal{C}_{\mathbf{x}}$ . Parmi toutes les distributions de moyenne  $\mathbf{x}$ , la plus susceptible de se réaliser est celle qui maximise la probabilité  $\mathcal{Q}(\cdot|\nu)$ , c'est-à-dire celle qui minimise  $\mathcal{K}(\cdot, \nu)$ . La probabilité de réalisation  $\mathcal{R}$  du macroétat  $\mathbf{x}$  peut donc être quantifiée par

$$\mathcal{R}(\mathbf{x} \in \mathcal{B}) \approx \exp(-m \inf_{\mathbf{x} \in \mathcal{B}} \mathcal{F}(\mathbf{x})),$$

avec

$$\mathcal{F}(\mathbf{x}) = \inf_{\{p|\mathbf{x}=E_p[\mathbf{u}]\}} \mathcal{K}(p, \nu).$$

Ce passage du niveau des distributions empiriques au niveau des moyennes empiriques est appelé « principe de contraction » en théorie des grandes déviations. Il permet de passer d'une entropie  $\mathcal{K}(\cdot, \nu)$  de niveau 2 à une entropie  $\mathcal{F}(\cdot)$  de niveau 1. Une formulation (dite formulation primale) équivalente du problème est alors :

$$(\text{primal}) \quad \hat{\mathbf{x}} = \arg \min_{\{\mathbf{x}:\mathbf{z}=\mathbf{H}\mathbf{x}\}} \mathcal{F}(\mathbf{x}).$$

Cette dernière formulation consiste donc à retenir pour estimateur un élément de l'ensemble des macroétats compatibles avec les données : celui  $\hat{\mathbf{x}}$  associé à la distribution empirique  $\hat{p}$  de probabilité maximale.

Remarquons que si  $\mathcal{K}(p, \nu)$  est une mesure de l'information apportée par la probabilité  $p$  relativement à  $\nu$ , alors  $\mathcal{F}(\mathbf{x})$  est une mesure de l'information du macroétat  $\mathbf{x}$  déduite par contraction de  $\mathcal{K}(p, \nu)$ . Nous sélectionnons donc le macroétat le moins informatif au sens de l'information de KULLBACK, c'est-à-dire le moins compromettant.

### 3. RÉOLUTION DU PROBLÈME

La minimisation sous contrainte de l'information de KULLBACK (1) se résoud classiquement par mise en oeuvre de la méthode des multiplicateurs de LAGRANGE et de la théorie de la dualité [7]. Le vecteur  $\boldsymbol{\mu}$  de multiplicateurs de LAGRANGE permet de définir la famille exponentielle engendrée par  $\nu$

$$dp_{\boldsymbol{\mu}}(\mathbf{u}) = \exp(\boldsymbol{\mu}^t \mathbf{H} \mathbf{u} - \log Z(\boldsymbol{\mu})) d\nu(\mathbf{u})$$

dont  $\hat{p}$  est élément.  $Z(\boldsymbol{\mu})$  est la fonction de partition assurant la normalisation  $\int dp_{\boldsymbol{\mu}}(\mathbf{u}) = 1$ .

Ce système peut être réécrit sous la forme dite duale :

$$(\text{dual}) \quad \begin{cases} \hat{\boldsymbol{\mu}} = \arg \max_{\boldsymbol{\mu}} D(\boldsymbol{\mu}), \\ D(\boldsymbol{\mu}) = \boldsymbol{\mu}^t \mathbf{z} - \log Z(\boldsymbol{\mu}), \\ \hat{\mathbf{x}} = E_{p_{\hat{\boldsymbol{\mu}}}}[\mathbf{u}], \end{cases}$$

$D$  étant une fonction concave, dite fonction duale, de  $\boldsymbol{\mu}$ . La résolution repose sur une bijection entre  $\mathbf{z}$  et  $\hat{\boldsymbol{\mu}}$  : à tout  $\mathbf{z}$  correspond un unique  $\hat{\boldsymbol{\mu}}$  tel que  $D(\hat{\boldsymbol{\mu}}) = \mathcal{K}(\hat{p}, \nu) = \mathcal{F}(\hat{\mathbf{x}})$ , et  $\hat{\boldsymbol{\mu}}$  est le vecteur de multiplicateurs de LAGRANGE associé aux données  $\mathbf{z}$  selon

$$\mathbf{z} = \left. \frac{\partial \log Z(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} \right|_{\boldsymbol{\mu}=\hat{\boldsymbol{\mu}}}.$$

Ce système permet donc de trouver un vecteur de paramètres de LAGRANGE  $\hat{\boldsymbol{\mu}}$  unique réalisant l'adéquation aux données. La détermination de l'estimateur, c'est-à-dire de la moyenne de  $p_{\hat{\boldsymbol{\mu}}}$ , se fait ensuite par relation primale-duale :

$$\hat{\mathbf{x}} = \mathbf{x}_{\hat{\boldsymbol{\mu}}} = E_{p_{\hat{\boldsymbol{\mu}}}}[\mathbf{u}] = \left. \frac{\partial \log Z(\boldsymbol{\mu})}{\partial (\mathbf{H}^t \boldsymbol{\mu})} \right|_{\boldsymbol{\mu}=\hat{\boldsymbol{\mu}}}.$$

Notons que le problème se résoud en maximisant la fonction duale car  $\mathcal{F}$  n'a, dans beaucoup de cas, pas d'expression explicite.

Prise en compte du bruit : il est possible de prendre en compte un bruit additif et sa statistique en appliquant la même méthodologie. Il suffit en effet d'appliquer le raisonnement précédent à l'objet étendu  $\tilde{\mathbf{x}} = [\mathbf{x}^t \mid \mathbf{n}^t]^t$  et à la matrice d'observation étendue  $\tilde{\mathbf{H}} = [\mathbf{H} \mid \mathbf{I}]$ , l'équation d'observation s'écrivant :  $\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{n} = \tilde{\mathbf{H}}\tilde{\mathbf{x}}$ .

La méthode de résolution exposée précédemment s'applique donc au problème bruité en choisissant une mesure de référence  $\tilde{\nu}(\tilde{\mathbf{x}}) = \nu(\mathbf{x}) \otimes \nu_{\mathbf{N}}(\mathbf{n})$ . Dans le cas d'une mesure de référence gaussienne pour le bruit, la nouvelle fonction duale s'écrit :  $\tilde{D}(\boldsymbol{\mu}) = D(\boldsymbol{\mu}) - \frac{1}{2} r_n \boldsymbol{\mu}^t \boldsymbol{\mu}$ .

### 4. APPLICATION

Nous avons étudié l'application de la MEM à la restauration de mélanges de populations en contrôle non-destructif (échographie ultra-sonore de milieux inhomogènes stratifiés) et en spectroscopie.

En contrôle non-destructif, on cherche à caractériser des interfaces entre couches : on cherche donc des pics, positifs et négatifs, sur un fond continu. Une mesure de référence constituée d'un mélange de deux distributions de LAPLACE (mesure double-LAPLACE) rend bien compte de cet *a priori* [9] : les échantillons tirés avec la loi de grande variance rendent compte des pics, ceux tirés avec la loi de faible variance rendent compte du fond et des inhomogénéités du milieu.

En spectroscopie, on recherche des raies, c'est-à-dire des pics positifs sur un fond homogène. On rend compte de ces hypothèses en choisissant à nouveau pour mesure de référence un mélange de distributions, en l'occurrence un mélange de deux distributions gamma (mesure double-gamma), ce choix garantissant la positivité de l'estimateur.

Nous exposons ci-dessous les cinq points principaux de la résolution :

1. Choix de la mesure de référence :

On note  $\lambda$  le paramètre de mélange des lois.

$$\begin{aligned}\tilde{\nu}(\tilde{\mathbf{x}}) &= \nu(\mathbf{x}) \otimes \nu_{\mathbf{N}}(\mathbf{n}), \\ \nu(\mathbf{x}) &= \prod_i \nu_i(x(i)), \\ \nu_{\mathbf{N}}(\mathbf{n}) &= \prod_i \nu_{N_i}(n(i)), \\ \nu_{N_i}(n(i)) &= \frac{1}{\sqrt{2\pi r_n}} \exp\left(-\frac{n(i)^2}{2r_n}\right).\end{aligned}$$

• Mesure double-LAPLACE :

$$\begin{aligned}\nu_i(x(i)) &= \frac{\lambda}{2r_l} \exp\left(-\frac{|x(i)|}{r_l}\right) \\ &\quad + \frac{1-\lambda}{2r_s} \exp\left(-\frac{|x(i)|}{r_s}\right).\end{aligned}$$

• Mesure double-gamma ( $x(i) > 0$ ) :

$$\begin{aligned}\nu_i(x(i)) &= \frac{\lambda}{b_1^{a_1} \Gamma(a_1)} x(i)^{a_1-1} \exp\left(-\frac{x(i)}{b_1}\right) \\ &\quad + \frac{1-\lambda}{b_2^{a_2} \Gamma(a_2)} x(i)^{a_2-1} \exp\left(-\frac{x(i)}{b_2}\right).\end{aligned}$$

2. Calcul analytique de la fonction duale :

On pose :  $\mathbf{k} = \mathbf{H}^t \boldsymbol{\mu}$ .

• Mesure double-LAPLACE :

$$\begin{aligned}\tilde{D}(\boldsymbol{\mu}) &= \boldsymbol{\mu}^t \mathbf{z} - \frac{1}{2} r_n \boldsymbol{\mu}^t \boldsymbol{\mu} \\ &\quad - \sum_i \log\left(\frac{\lambda}{1-r_l^2 k(i)^2} + \frac{1-\lambda}{1-r_s^2 k(i)^2}\right)\end{aligned}$$

avec la contrainte :  $\sup_i (r_l k(i))^2 < 1$ .

• Mesure double-gamma :

$$\begin{aligned}\tilde{D}(\boldsymbol{\mu}) &= \boldsymbol{\mu}^t \mathbf{z} - \frac{1}{2} r_n \boldsymbol{\mu}^t \boldsymbol{\mu} \\ &\quad - \sum_i \log\left(\frac{\lambda}{(1-b_1 k(i))^{a_1}} + \frac{1-\lambda}{(1-b_2 k(i))^{a_2}}\right)\end{aligned}$$

avec la contrainte :  $\sup_i r_l k(i) < 1$ .

3. Calcul analytique de la relation primale-duale :

• Mesure double-LAPLACE :

$$\begin{aligned}x_{\boldsymbol{\mu}}(i) &= 2k(i) \left( \frac{\lambda r_l^2}{(1-r_l^2 k(i)^2)^2} + \frac{(1-\lambda) r_s^2}{(1-r_s^2 k(i)^2)^2} \right) / \\ &\quad \left( \frac{\lambda}{1-r_l^2 k(i)^2} + \frac{1-\lambda}{1-r_s^2 k(i)^2} \right).\end{aligned}$$

• Mesure double-gamma :

$$\begin{aligned}x_{\boldsymbol{\mu}}(i) &= \left( \frac{\lambda a_1 b_1}{(1-b_1 k(i))^{a_1+1}} + \frac{(1-\lambda) a_2 b_2}{(1-b_2 k(i))^{a_2+1}} \right) / \\ &\quad \left( \frac{\lambda}{(1-b_1 k(i))^{a_1}} + \frac{1-\lambda}{(1-b_2 k(i))^{a_2}} \right).\end{aligned}$$

4. Calcul numérique de  $\hat{\boldsymbol{\mu}} = \arg \max_{\boldsymbol{\mu}} D(\boldsymbol{\mu})$ .

La fonction  $D$  est une fonction concave dont le gradient s'écrit simplement :  $\mathbf{grad}_D = \mathbf{z} - r_n \boldsymbol{\mu} - \mathbf{H} \mathbf{x}_{\boldsymbol{\mu}}$ .

Notons que la relation primale-duale pourra avantageusement être utilisée pour calculer ce gradient. Cette fonction se maximise donc par des algorithmes classiques, de type gradient ou gradient conjugué. Lors de la mise en oeuvre, il sera nécessaire de procéder à un réglage de deux hyperparamètres, à savoir la variance  $r_n$  du bruit qui détermine le compromis entre la confiance en les données et la solution par défaut, et le paramètre de mélange  $\lambda$ . Les autres paramètres de la loi de référence sont ajustés en fonction de l'histogramme des observations.

5. Calcul numérique de  $\hat{\mathbf{x}}$  par relation primale-duale :

$$\hat{\mathbf{x}} = \mathbf{x}_{\hat{\boldsymbol{\mu}}} = E_{\hat{p}}[\mathbf{u}].$$

Fonction de transfert du critère : un traitement par MEMM avec une mesure de référence donnée peut être caractérisé de façon simple par une « fonction de transfert » donnant la restauration  $\hat{\mathbf{x}}$  en fonction de l'observation  $\mathbf{z}$  pour  $\mathbf{H} = \mathbf{I}$ . Sous cette hypothèse, chaque échantillon de la reconstruction ne dépend que de l'échantillon correspondant des observations. Cette étude peut être complétée par l'étude du comportement de la relation primale-duale en fonction de  $\mathbf{k}$ .

## 5. SIMULATIONS

Nous présentons dans cette section quelques résultats de simulation.

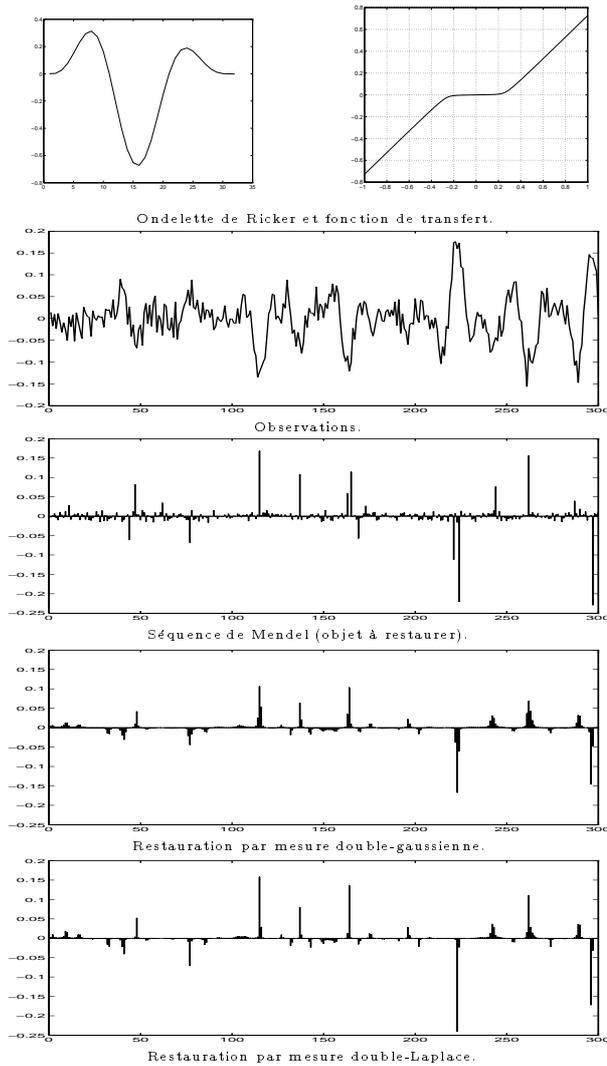
Dans le cadre du contrôle non-destructif (ici de la géophysique), on s'efforce de retrouver des réflecteurs amples et rares au milieu de petits réflecteurs plus denses. De très nombreux modèles probabilistes ont été employés pour y parvenir : Bernoulli-gaussiens, gaussiens généralisés (normes  $L_p$ ), etc... Certaines études du caractère non-gaussien des coefficients de réflexion des logs de puits [9, page 1045] ont montré qu'un mélange de lois de LAPLACE, offrant la possibilité d'avoir un coefficient d'aplatissement normalisé supérieur à 6, était de ce fait mieux adapté qu'un modèle gaussien généralisé.

Nous avons utilisé la séquence de test classique de MENDEL, convoluée par l'ondelette de RICKER au contenu spectral très pauvre. Le rapport signal à bruit est de 7 dB, ce qui représente également des conditions sévères. La MEMM a été appliquée en utilisant d'abord une mesure double-gaussienne, puis une mesure double-LAPLACE. L'allure de la fonction de transfert du second cas indique que nous allons restituer des pics positifs et négatifs tout en atténuant les échantillons de faible amplitude par effet de seuil, ce qui correspond à nos *desiderata*. On observe une restitution satisfaisante des pics dans les deux cas. La restauration par mesure double-LAPLACE est toutefois plus piquée, puisque cette mesure tolère plus de pics de grande amplitude qu'une mesure double-gaussienne. La méthode s'avère en outre robuste vis-à-vis des hyperparamètres de réglage et de la mesure de référence, ce qui laisse augurer un bon comportement pour un traitement de signaux expérimentaux.

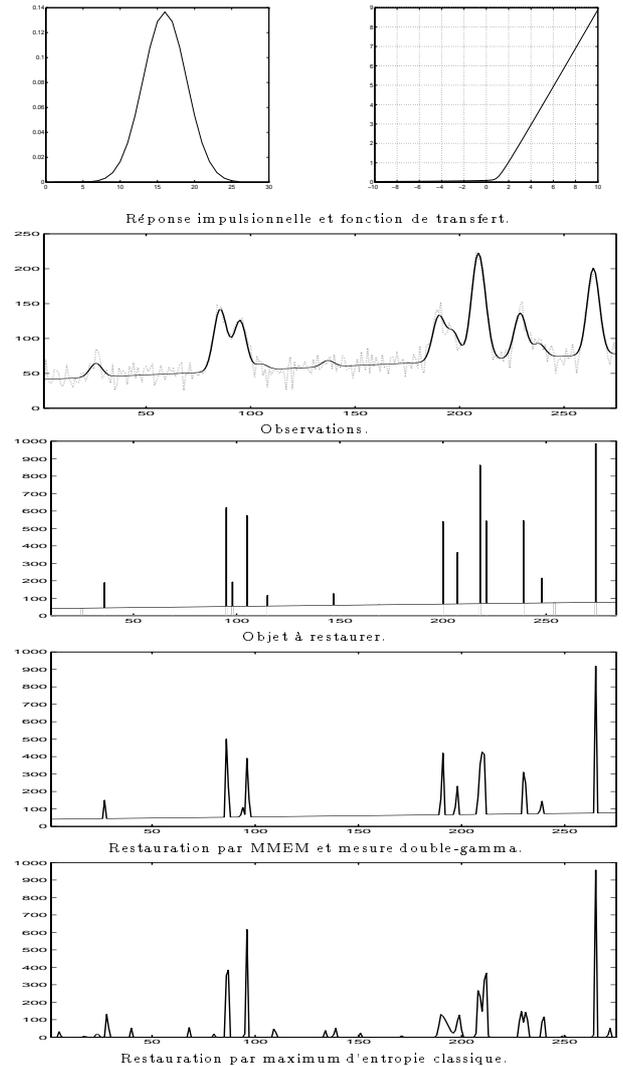
Dans le cadre de la spectroscopie, nous avons utilisé une ondelette et un signal d'entrée synthétiques, représentatifs de ce type d'application. La dérive du fond de l'objet restituée ici est couramment rencontrée en pratique. Nous nous sommes placés sous un rapport signal à bruit de 10 dB. Nous avons comparé la MMEM à la méthode du maximum d'entropie classique (MAXENT). L'allure de la fonction de transfert indique que nous allons restituer des pics positifs, avec un effet de seuil atténuant les observations de faible amplitude. MAXENT, qui est un cas particulier de la MMEM utili-

sant une mesure de référence de POISSON [3], manque ici de souplesse puisque l'on ne dispose que d'un seul paramètre (celui de la loi de POISSON) pour décrire à la fois le fond (qui est variable) et la présence de pics positifs. Il faut alors « soustraire le fond » (avec les difficultés que l'on imagine lorsqu'il y a beaucoup de pics) avant de traiter les données. L'utilisation, comme mesure de référence, d'un mélange de lois gamma permet d'obtenir la souplesse nécessaire pour à la fois suivre un fond variable et décrire des pics positifs dont les amplitudes sont indépendantes de ce fond.

### Application au contrôle non-destructif



### Application à la spectroscopie



### Références

[1] J.-F. BERCHER (1995), *Développement de critères de nature entropique pour la résolution des problèmes inverses linéaires*, thèse de doctorat, université de PARIS-SUD, ORSAY.

[2] R. S. ELLIS (1985), *Entropy, large deviations and statistical mechanics*, SPRINGER-VERLAG, NEW-YORK.

[3] F. GAMBOA (1989), *Méthode du maximum d'entropie sur la moyenne et applications*, thèse de doctorat, université de PARIS-SUD, ORSAY.

[4] F. GAMBOA & É. GASSIAT (1994), « Bayesian methods and maximum entropy for ill posed inverse problems », soumis à *The Annals of Statistics*.

[5] S. KULLBACK (1959), *Information theory and statistics*, J. WILEY, NEW-YORK.

[6] G. LE BESNERAIS (1993), *Méthode du maximum d'entropie sur la moyenne, critères de reconstruction d'image et synthèse d'ouverture en radio-astronomie*, thèse de doctorat, université de PARIS-SUD, ORSAY.

[7] D. LUENBERGER (1969), *Optimization by vector space methods*, J. WILEY, NEW-YORK.

[8] J. NAVAZA (1986), « The use of non-local constraints in maximum-entropy electron density reconstruction », *Acta Crystallographica*, pages 212-223.

[9] A. T. WALDEN & J. W. J. HOSKEN (1986), « The nature of the non-Gaussianity of primary reflection coefficients and its significance for deconvolution », *Geophysical Prospecting* 34, pages 1038-1066.