

# THE MAXIMUM ENTROPY ON THE MEAN METHOD, NOISE AND SENSITIVITY

Jean-François BERCHER, Guy LE BESNERAIS and Guy DEMOMENT  
Laboratoire des Signaux et Systèmes (CNRS-ESE-UPS)  
École Supérieure d'Électricité,  
Plateau de Moulon, 91192 Gif-sur-Yvette Cédex, France

**ABSTRACT.** In this paper we address the problem of building convenient criteria to solve linear and noisy inverse problems of the form  $\mathbf{y} = \mathbf{Ax} + \mathbf{n}$ . Our approach is based on the specification of constraints on the solution  $\mathbf{x}$  through its belonging to a given convex set  $\mathcal{C}$ . The solution is chosen as the mean of the distribution which is the closest to a reference measure  $\mu$  on  $\mathcal{C}$  with respect to the Kullback divergence, or cross-entropy. This is therefore called the Maximum Entropy on the Mean Method (MEMM). This problem is shown to be equivalent to the convex one  $\mathbf{x} = \arg \min_{\mathbf{x}} \mathcal{F}(\mathbf{x})$  submitted to  $\mathbf{y} = \mathbf{Ax}$  (in the noiseless case). Many classical criteria are found to be particular solutions with different reference measures  $\mu$ . But except for some measures, these primal criteria have no explicit expression. Nevertheless, taking advantage of a dual formulation of the problem, the MEMM enables us to compute a solution in such cases. This indicates that such criteria could hardly have been derived without the MEMM. In order to integrate the presence of additive noise in the MEMM scheme, the object and noise are searched simultaneously for in an appropriate convex  $\mathcal{C}'$ . The MEMM then gives a criterion of the form  $\mathbf{x} = \arg \min_{\mathbf{x}} \mathcal{F}(\mathbf{x}) + \mathcal{G}(\mathbf{y} - \mathbf{Ax})$ , where  $\mathcal{F}$  and  $\mathcal{G}$  are convex, without constraints. The functional  $\mathcal{G}$  is related to the prior distribution of noise, and may be used to account for specific noise distributions. Using the regularity of the criterion, the sensitivity of the solution to variations of the data is also derived.

## 1. Problem statement

In many applications, one often faces the inverse problem  $\mathbf{y} = \mathbf{Ax} + \mathbf{n}$  which consists in estimating a vector  $\mathbf{x} \in \mathbb{R}^N$  from an *indirect* and *noisy* observation vector  $\mathbf{y}$ . The observation matrix  $\mathbf{A}$  is supposed to be known, together with some statistical characteristics of the noise  $\mathbf{n}$ . When the observation matrix  $\mathbf{A}$  is either not regular or ill-conditioned the problem is ill-posed and one has to complete the data with an *a priori* knowledge or constraints on the solution in order to select a physically-acceptable solution. Such information may be given in the form of the convex constraint

$$\mathbf{x} \in \mathcal{C}, \quad (1)$$

where  $\mathcal{C}$  is a convex set. Examples of this situation are plentiful, let us only cite the problem of imaging *positive* intensity distributions, which arises in spectral analysis, astronomy, spectrometry, etc... In other specific problems, such as crystallography or tomography, lower and upper bounds on the image are known, and have to be taken into account in the reconstruction process. Such constraints may be specified by the belonging of the object to the convex set  $\mathcal{C}$  (where the bounds  $a_k$  and  $b_k$  are given) and include the positivity constraint as a special case,

$$\mathcal{C} = \{ \mathbf{x} \in \mathbb{R}^N / \quad x_k \in ]a_k, b_k[, \quad k = 1..N \}. \quad (2)$$

## 2. Methods for solving linear inverse problems

In the case of an ill-posed problem, the generalized inverse solution is unsatisfactory because of the dramatic amplification of any observation noise. Quadratic regularization makes possible to get rid of ill-posedness effects, but it leads to linear estimates, and therefore cannot provide any guarantee with respect to the support constraint (2).

Possible answers are given with set theoretic estimation (for a review see [1]) and projection onto convex sets algorithms. Although good reconstructions can be obtained, they are often computationally expensive and do not lead to a unique and well-defined solution.

Other approaches use regularized criteria, which are usually written as a compound criterion made of two terms, one which enforces some fidelity of the solution to the data, the other which ensures that some desirable properties are met. Such regularized criteria will be noted under the generic form

$$\mathcal{J}(\mathbf{x}) = \mathcal{F}(\mathbf{x}) + \alpha \mathcal{G}(\mathbf{y} - \mathbf{A}\mathbf{x}) \quad \alpha \geq 0. \quad (3)$$

Many of these regularized criteria may be interpreted in a Bayesian setting. Indeed, if the functionals  $-\mathcal{F}$  and  $\mathbf{x} \mapsto -\alpha \mathcal{G}(\mathbf{y} - \mathbf{A}\mathbf{x})$  are respectively a log-prior and a log-likelihood, then the minimization of  $\mathcal{J}$  provides the maximum *a posteriori* (MAP) estimator. However, in a given problem, the *ab initio* choice of a good model is a difficult task, for which there is no general answer (see [6] for a discussion of the subject). Such situations are encountered when the only *a priori* knowledge is a convex constraint such as (1). Nevertheless, useful methods have been found in those cases: for instance, when reconstructing object with positivity as the only pre-requisite, several thought processes have lead different authors to the conclusion that the maximum entropy reconstruction method could be a useful answer. It consists in the optimization of a regularized criterion of the form (3) with

$$\mathcal{F}(\mathbf{x}) = \sum_{i=1}^N \{x_i \log \frac{x_i}{m_i} - x_i + m_i\}, \quad (4)$$

where  $\mathbf{m} = [m_1, m_2, \dots, m_N]$  is a prior guess. As far as the positivity constraint is concerned, criteria like (4), built upon logarithmic expressions, ensure positivity and are therefore said to be “positivity free”; an other well-known example is the “ $\log(x)$ ” or Burg entropy used in spectral analysis.

The several good properties of the maximum entropy reconstruction method have been studied by many authors (see axiomatic studies such as [2] or [5]). The MEMM construction generalizes in some way certain aforementioned “thought processes,” leading to the maximum entropy reconstruction method (4), in order to exhibit useful regularization functionals for a large class of convex constraints (1). The obtained regularizing functionals share many properties of the entropy (4).

## 3. The Maximum Entropy on the Mean Method

The foundations of the Maximum Entropy on the Mean Method originate from the work of J. Navaza [10], and some theoretical aspects of the method were further studied by F. Gamboa and D. Dacunha-Castelle [3]. We have also studied it with a special attention to its potential applications in signal and image reconstruction and restoration [8]. For the sake

of simplicity, this paragraph addresses the *noiseless* problem. Discussion of how to account for noise will take place in §5.2.

Much emphasis must be put on our only *a priori* information: the convex constraint of (2). The MEMM construction thus begins with the specification of the set  $\mathcal{C}$  and a reference measure  $d\mu(\mathbf{x})$  over it. The actual observations  $\mathbf{y}$  are considered as the mean of a process  $\mathbf{x}$  under a probability distribution  $P$  defined on  $\mathcal{C}$  (this idea comes from statistical physics where observations are average values or macrostates). The set  $\mathcal{C}$  being convex, the mean  $E_P\{\mathbf{x}\}$  under  $P$  is in  $\mathcal{C}$  and hence the convex constraint is automatically fulfilled by  $E_P\{\mathbf{x}\}$ .

### 3.1. Additional information principle

Since the constraint given by (2) does not lead to a unique distribution  $P$ , we have to invoke some additional information principle. For this purpose, we use the  $\mu$ -entropy  $K(P, \mu)$ , or Kullback-Leibler (K-L) information [7]. This information is defined for a reference measure  $\mu$  and a probability measure  $P$  by

$$K(P, \mu) = \int \log \frac{dP}{d\mu} dP \quad (5)$$

if  $P$  is absolutely continuous with respect to  $\mu$  ( $P \ll \mu$ ) and  $K(P, \mu) = +\infty$  otherwise.

The distribution  $P$  is selected as the minimizer of the  $\mu$ -entropy submitted to the constraints “on the mean”  $\mathbf{A} E_P\{\mathbf{X}\} = \mathbf{y}$ . In other words,  $P$  is the nearest distribution, with respect to the K-L divergence, to the reference measure  $\mu$  in the set of distributions such that  $\mathbf{A} E_P\{\mathbf{X}\} = \mathbf{y}$ . The maximum entropy on the mean problem then states as follows:

$$\text{MEMM problem} \quad \begin{cases} \hat{P} = \arg \min_P \int \log \frac{dP}{d\mu}(\mathbf{x}) dP(\mathbf{x}), \\ \text{such that } \mathbf{y} = \mathbf{A} \int \mathbf{x} dP(\mathbf{x}). \end{cases}$$

It is well known that the solution, if it exists, belongs to in the exponential family

$$dP_{\mathbf{s}}(\mathbf{x}) = \exp \left\{ \mathbf{s}^t \mathbf{x} - \log Z(\mathbf{s}) \right\} d\mu(\mathbf{x}), \quad (6)$$

and, more precisely, that its natural parameter is of the form  $\mathbf{s} = \mathbf{A}^t \boldsymbol{\lambda}$  for some  $\boldsymbol{\lambda}$ . In (6)  $\log Z$  is the log-partition function or the log-Laplace transform of the measure  $d\mu(\mathbf{x})$ ; this function will be noted  $\mathcal{F}^*$  in the sequel.

### 3.2. The dual problem

Using results of duality theory, there is an equality between the optimum value of the previous problem and the optimum value of its dual counterpart (dual attainment):

$$\inf_{P \in \mathcal{P}_y} K(P, \mu) = \sup_{\boldsymbol{\lambda} \in \mathcal{D}_{\boldsymbol{\lambda}}} \left\{ \boldsymbol{\lambda}^t \mathbf{y} - \mathcal{F}^*(\mathbf{A}^t \boldsymbol{\lambda}) \right\}, \quad (7)$$

where  $\mathcal{P}_y = \{P : \mathbf{A} E_P\{\mathbf{X}\} = \mathbf{y}\}$  is the set of normalized distributions which satisfy the linear constraint on the mean, and  $\mathcal{D}_{\boldsymbol{\lambda}}$  is the set  $\{\boldsymbol{\lambda} \in \mathbb{R}^M : Z(\mathbf{A}^t \boldsymbol{\lambda}) < \infty\}$ , which is often the whole  $\mathbb{R}^M$ , in which case the dual problem is unconstrained.

Once the dual problem on the right side of (7) is solved, that is the maximization of the dual functional

$$D(\boldsymbol{\lambda}) = \boldsymbol{\lambda}^t \mathbf{y} - \mathcal{F}^*(\mathbf{A}^t \boldsymbol{\lambda}), \quad (8)$$

yielding an optimum value  $\hat{\lambda}$ , one has the expression of the density  $\hat{P} = P_{\mathbf{A}^t \hat{\lambda}}$  and can calculate the reconstructed object  $\hat{\mathbf{x}}$  by computing (numerically) the expectation  $E_{\hat{P}}\{\mathbf{X}\}$ . But this is not the more efficient way to compute the solution. Indeed, inside the exponential family (6) there is a one-to-one mapping between the natural parameter  $\mathbf{s}$  and the mean of the associated distribution  $\bar{\mathbf{x}}(\mathbf{s})$ :

$$\bar{\mathbf{x}}(\mathbf{s}) = \frac{d\mathcal{F}^*}{d\mathbf{s}}(\mathbf{s}). \quad (9)$$

Therefore, the solution  $\hat{\mathbf{x}}$  is simply obtained by calculating (9) at the optimal point  $\mathbf{A}^t \hat{\lambda}$ . Let us emphasize that the dual criterion is by construction a strictly concave functional. Efficient methods of numerical optimization, such as gradient, conjugate gradient, or second order methods (Gauss-Newton) can be used to compute the solution. They will use the gradient of  $D$  which is easily calculated to be just  $\mathbf{y} - \mathbf{A}\bar{\mathbf{x}}(\mathbf{A}^t \lambda)$ . During the algorithm, the primal-dual relation (9) is used to compute the current reconstruction from the dual vector  $\lambda$ .

### 3.3. Yet another primal problem

The previous development was done in the space of the dual parameters  $\lambda$ . The purpose of this paragraph is to come back to the natural “object space”. We will exhibit a new primal criterion, which we will call an entropy. This function, not surprisingly, is intimately related with the previous dual function and the K-L information.

For each  $\mathbf{x} \in \mathcal{C}$ , consider the MEMM problem when the constraint is  $E_P\{\mathbf{X}\} = \mathbf{x}$ . We define  $\mathcal{F}(\mathbf{x})$  to be the optimum value of the K-L information for this problem

$$\mathcal{F}(\mathbf{x}) = \inf_{P \in \mathcal{P}_{\mathbf{x}}} \mathcal{K}(P, \mu),$$

where  $\mathcal{P}_{\mathbf{x}} = \{P : E_P\{\mathbf{X}\} = \mathbf{x}\}$ .

As already seen, at the optimum, we have by dual attainment

$$\mathcal{F}(\mathbf{x}) = \sup_{\lambda \in \mathcal{D}_{\lambda}} \left\{ \lambda^t \mathbf{x} - \mathcal{F}^*(\lambda) \right\}, \quad (10)$$

The latter equation means that  $\mathcal{F}$  is the conjugate convex of  $\mathcal{F}^*$  and, as  $\mathcal{F}^*$  is the log-Laplace transform of  $\mu$ , the *Cramér transform* of  $\mu$ . Such transforms appear in various fields of statistics and in particular in the Large Deviations theory, which has important connections with the MEMM [9]. Properties of Cramér transforms are listed below [4]:

- $\mathcal{F}$  is continuously differentiable and strictly convex on  $\mathcal{C}$ ,
- $\mathcal{F}(\mathbf{x}) = +\infty$  for  $\mathbf{x} \notin \mathcal{C}$  and its derivative is infinite on the boundary of  $\mathcal{C}$ ,
- $\mathcal{F}(\mathbf{x}) \geq 0$  with equality for  $\mathbf{x} = \mathbf{m}$ , the mean value under the reference measure  $\mu$ .

Our original MEMM problem can now be handled in a different way. If  $P$  is a candidate distribution with mean  $\mathbf{x}$ , its K-L information with respect  $\mu$  is greater or equal to  $\mathcal{F}(\mathbf{x})$ . Moreover, this lower bound can be decreased by searching a vector  $\hat{\mathbf{x}}$  minimizing  $\mathcal{F}$  over the set  $\mathcal{C}_{\mathbf{y}} = \{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{y}\}$ . Then the MEMM problem is reformulated as

$$\inf_{\mathbf{x} \in \mathcal{C}_{\mathbf{y}}} \left\{ \inf_{P \in \mathcal{P}_{\mathbf{x}}} \mathcal{K}(P, \mu) \right\}.$$

If we consider the reconstruction problem in the object space, we only need to solve

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathcal{C}_y} \mathcal{F}(\mathbf{x}). \quad (11)$$

Note that this problem has the same dual problem than that of (8). In fact, we have exhibited another primal problem associated to (8), directly in the object space  $\mathbb{R}^N$ . Its solution  $\hat{\mathbf{x}}$  is the mean of the optimal distribution in the MEMM problem, and a solution to our reconstruction problem. This swap between primal problems is referred to as a “contraction principle” in statistical physics [4]. In this context, functional  $\mathcal{F}$  appears as a level-1 entropy, therefore we will simply call it an entropy in the following.

Properties of the Cramér transform are useful for reconstruction purposes, when holding the entropy  $\mathcal{F}$  as the objective function, as in (11). Strict convexity enables a simple implementation and guarantees the uniqueness of the reconstruction. The second property shows that any descent method will provide a solution in  $\mathcal{C}$ , even if the constraint  $\mathbf{x} \in \mathcal{C}$  is not specified in the algorithm; this “ $\mathcal{C}$ -free property”, is here an analog of the “positivity free” property observed in conventional maximum entropy solutions (see above). The last property shows that  $\mathcal{F}$  may be considered as a discrepancy measure between  $\mathbf{x}$  and  $\mathbf{m}$ . In the sequel, we give some examples illustrating the different points developed above.

#### 4. A few examples of MEMM criteria

##### 4.1. Gaussian reference

Our first example consists in a problem where no constraint is known on the object, so that  $\mathcal{C} = \mathbb{R}^n$ . We choose the Gaussian measure  $\mathcal{N}(\mathbf{m}, \mathbf{R}_x)$  as our reference measure  $\mu$  on  $\mathcal{C}$ . A simple calculus then leads to the Cramér transform

$$\mathcal{F}(\mathbf{x}) = (\mathbf{x} - \mathbf{m})^t \mathbf{R}_x^{-1} (\mathbf{x} - \mathbf{m}), \quad (12)$$

which is recognized as a classical quadratic regularizing term.

##### 4.2. The positive case

###### • Poisson reference and the “Shannon entropy”

Let now  $\mathcal{C}$  be  $]0, +\infty[$ , and the reference distribution be a (separable) Poisson law, with expectation  $\mathbf{m}$ . Such a prior may correspond to the modeling of the fall of quanta of energy, following a Poisson process. This modeling may be encountered in astronomy (the speckle-images of optical interferometry) for instance. The reference measure is then

$$\mu(\mathbf{x}) = \prod_{j=1}^N \mu(x_j) = \prod_{j=1}^N \frac{m_j^{x_j}}{x_j!} \exp(-m_j).$$

The entropy functional  $\mathcal{F}$ , which measures the distance between any candidate solution  $\mathbf{x}$  and the prior mean  $\mathbf{m}$  is the Cramér transform of  $\mu$ , and works out to be

$$\mathcal{F}(\mathbf{x}) = \sum_{j=1}^N \left[ \frac{x_j}{m_j} \log \left( \frac{x_j}{m_j} \right) + m_j - x_j \right],$$

which is the generalized version of the Shannon entropy.

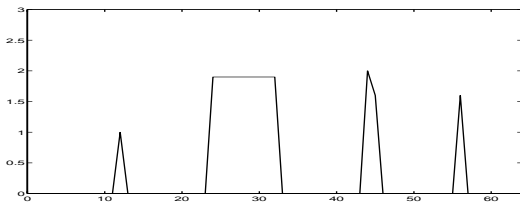
- Gamma reference and Itakura-Saito discrepancy measure

The presentation of [11] in spectrum analysis happens to be exactly a MEMM approach to a well known criterion: the Itakura-Saito discrepancy measure.

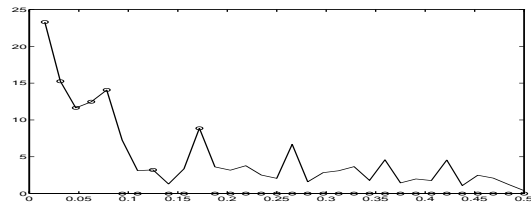
The periodogram having asymptotically a  $\chi^2$  distribution with two degrees of freedom, the corresponding reference measure  $\mu$  over the possible spectra is an exponential law with mean, *i.e. prior spectrum*  $\mathbf{m}$ . Using the Cramér transform definition, one easily obtains the entropy

$$\mathcal{F}(\mathbf{x}) = \sum_{j=1}^N \frac{x_j}{m_j} - \log \left( \frac{x_j}{m_j} \right) - 1, \quad (13)$$

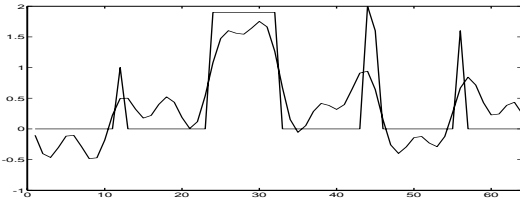
which is the Itakura-Saito distortion between  $\mathbf{s}$  and  $\mathbf{m}$ . With  $\mathbf{m} = \mathbf{1}$ , we measure a distance to a flat spectrum, and find out the so-called “log( $x$ )”, or Burg entropy.



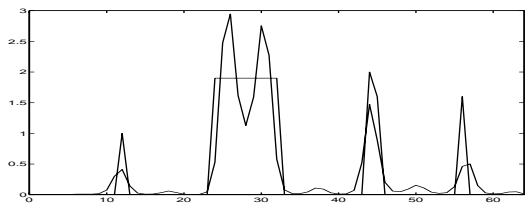
(a) Test object.



(b) Fourier transform of the object and (c) available data.

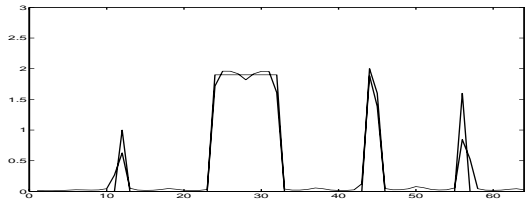


(c) Reconstruction with a Gaussian reference measure – no constraint.



(d) Reconstruction with a Poisson reference – positivity constraint.

**Figure 1:** this figure compares different reconstructions in a simple Fourier synthesis problem. The test object is in (a), its Fourier transform and the available data in (b). Then three reconstructions corresponding to different reference measures of the MEMM scheme, and also to different constraint sets, are given. They show the improvement with the reduction of the set of admissible solutions.



(e) Reconstruction with a uniform 0-2 measure –  $x \in [0, 2]^N$ .

#### 4.3. The bounded case

We consider here the case when  $\mathcal{C}$  has the general form of Eq. 2. Such constraints may be useful in many applied problems where the object is *a priori* known to lie between two bounds (tomography, filter design, crystallography).

Several reference measures can be used on the convex  $\mathcal{C}$ . A natural idea is indeed to use a product of uniform measures over each interval  $]a_j, b_j[$ :  $d\mu(\mathbf{x}) = \bigotimes_{j=1}^N \frac{1}{b_j - a_j} \mathbf{1}_{]a_j, b_j[}(x_j) dx_j$ . The calculus of the Cramér transform leads to implicit equations, therefore we have no analytic expression for  $\mathcal{F}$ . Nevertheless the primal-dual relation can be computed

$$x_j = -\frac{1}{s_j} + \frac{b_j e^{b_j s_j} - a_j e^{a_j s_j}}{e^{b_j s_j} - e^{a_j s_j}} \quad \text{with } s_j = [\mathbf{A}^t \boldsymbol{\lambda}]_j, \quad 1 \leq j \leq N$$

and the convex problem  $\inf_{\mathbf{x}} \mathcal{F}(\mathbf{x})$  subject to  $\mathbf{y} = \mathbf{A}\mathbf{x}$ , where  $\mathcal{F}$  is *not explicit*, can still be solved using its dual formulation (8) together with the aforementioned primal-dual relation. Other measures could be used in this case. The case of a Bernoulli measures product  $d\mu(\mathbf{x}) = \bigotimes_{j=1}^N \{\alpha_j \delta(x_j - a_j) + (1 - \alpha_j) \delta(x_j - b_j)\}$  (where  $\delta$  denotes the Dirac measure) is derived in a referenced work [9] and leads to a generalized version of Fermi-Dirac entropies.

Figure 1 compares reconstructions obtained with different entropies presented above.

## 5. Taking noise into account

So far MEMM criteria have been derived from the maximization of the  $\mu$ -entropy submitted to an *exact constraint*. Any observation noise will ruin our exact constraint, and as a consequence the two (primal-dual) formulations of the MEMM problem. The exact constraint was useful in interpreting observations as a linear transform of a mean, then enabling us to exhibit the discrepancy measure  $\mathcal{F}$ . Because of the good properties of  $\mathcal{F}$ , we will still consider the unknown object  $\mathbf{x}$  as a mean, in order to use its entropy  $\mathcal{F}(\mathbf{x})$ , but we have to modify the procedure.

### 5.1. The $\chi^2$ constraint

A classical way to account for noise is to construct a confidence region about the expected value of some statistic. For Gaussian noise, one usually uses the  $\chi^2$  constraint  $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 \leq \rho$ , where  $\rho$  is some constant. Then the problem becomes the minimization of  $\mathcal{F}$  submitted to the  $\chi^2$  constraint. There always exists a positive parameter  $\alpha$  (in fact it is a Lagrange parameter corresponding to the  $\chi^2$  constraint) such that the previous problem reduces to

$$\inf_{\mathbf{x}} \{\mathcal{F}(\mathbf{x}) + \alpha \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2\}. \quad (14)$$

### 5.2. Accounting for general noise statistic within the MEMM procedure

Thanks to a specific entropy function, more complicated penalizations than (14) can be performed in order to account for non-Gaussian noises. Such entropies can be derived directly in the same MEMM axiomatic approach as in the noiseless case. To this end, we only need to introduce an *extended object*  $\tilde{\mathbf{x}} = [\mathbf{x}, \mathbf{n}]$ , and consider the relation  $\mathbf{y} = \tilde{\mathbf{A}}\tilde{\mathbf{x}}$ , with  $\tilde{\mathbf{A}} = [\mathbf{A}, \mathbf{1}]$ . The vector  $\tilde{\mathbf{x}}$  evolves in the convex  $\tilde{\mathcal{C}}$  of  $\mathbb{R}^{N+M}$ , which separates on a product of the usual  $\mathcal{C}$  and of  $\mathcal{B}$ ,  $\tilde{\mathcal{C}} = \mathcal{C} \times \mathcal{B}$ , where  $\mathcal{B}$  is the convex hull of the state space of the noise vector  $\mathbf{n}$ .

We then use a reference measure  $\nu$  over the noise set. For instance, in the case of a Gaussian noise we take  $\mathcal{B} = \mathbb{R}^M$  and a centered Gaussian law with covariance matrix  $\mathbf{R}_\nu$  as  $\nu$ . With a Poisson noise we take  $\mathcal{B} = \mathbb{R}_+^M$  and a Poisson reference measure  $\nu$ .

Now we can define a new entropy functional by using a reference measure  $\tilde{\mu}$  on  $\tilde{\mathcal{C}}$ . If  $\nu$  is the distribution of the noise,  $\mu$  our object reference measure on  $\mathcal{C}$ , and if we assume that

the object and noise are independent, we obtain  $\tilde{\mu} = \mu \otimes \nu$ . The entropy function we are looking for is then the Cramér transform of  $\tilde{\mu}$  which is simply

$$\mathcal{F}_{\tilde{\mu}}(\tilde{\mathbf{x}}) = \mathcal{F}_{\mu}(\mathbf{x}) + \mathcal{F}_{\nu}(\mathbf{n}).$$

Estimation of the extended object is conducted through a constrained minimization of  $\mathcal{F}_{\tilde{\mu}}(\tilde{\mathbf{x}})$ , the constraint being  $\mathbf{y} = \tilde{\mathbf{A}}\tilde{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{n}$ . Therefore it reduces to the unconstrained minimization of the compound criterion

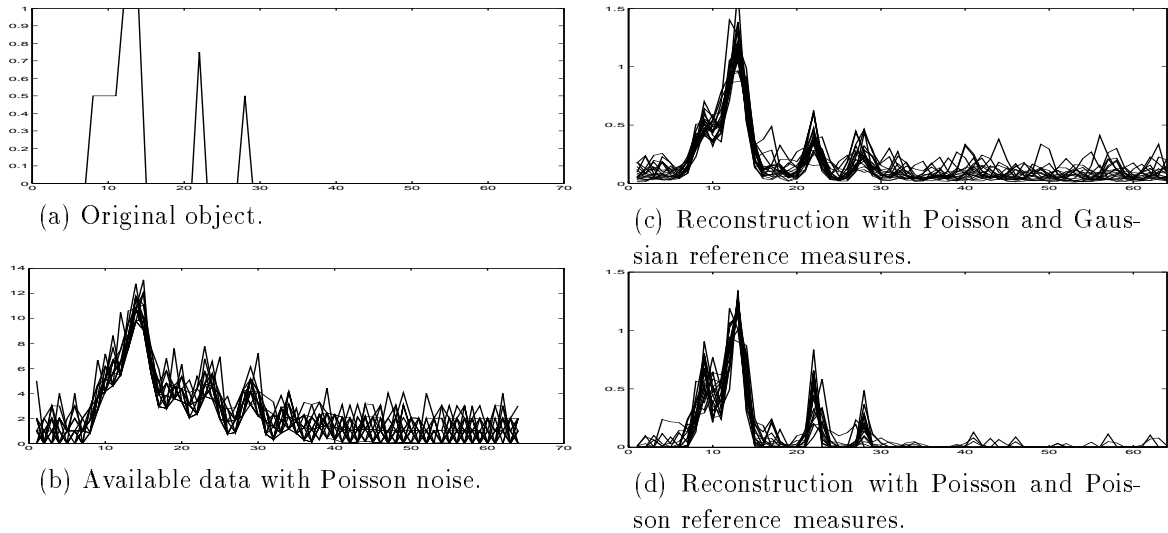
$$\mathcal{J}(\mathbf{x}) = \mathcal{F}_{\tilde{\mu}}([\mathbf{x}, \mathbf{y} - \mathbf{A}\mathbf{x}]^t) = \mathcal{F}_{\mu}(\mathbf{x}) + \mathcal{F}_{\nu}(\mathbf{y} - \mathbf{A}\mathbf{x}). \quad (15)$$

A dual approach is again useful, in particular if  $\mathcal{F}_{\mu}$  or  $\mathcal{F}_{\nu}$  or both are not explicit. It is easy to show that the dual criterion is

$$\tilde{\mathcal{D}}(\boldsymbol{\lambda}) = \boldsymbol{\lambda}^t \mathbf{y} - \mathcal{F}_{\mu}^*(\mathbf{A}^t \boldsymbol{\lambda}) - \mathcal{F}_{\nu}^*(\boldsymbol{\lambda}).$$

Having solved the dual problem, we come back to the primal solution by the primal-dual relation which is, thanks to the separability of the log-Laplace transform of  $\tilde{\mu}$ , the same as in the noiseless case:

$$\bar{\mathbf{x}}(\mathbf{A}^t \boldsymbol{\lambda}) = \frac{d\mathcal{F}_{\mu}^*}{ds}(\mathbf{A}^t \boldsymbol{\lambda}).$$



**Figure 2:** Different accounts for noise. The original object (a) is convolved and corrupted by Poisson noise (b). Twenty different realizations of the noise are represented. Reconstructions are in (c) and (d): reconstructions in (c), where the noise is considered as Gaussian are outperformed by reconstructions in (d), where the real nature of the noise is accounted for.

We are then able to account for specific noise distributions, without loss in the nice properties of our criteria: the global criterion of (15) is always convex, and the convex



constraint is automatically satisfied. Concerning the case of a Gaussian noise, it can easily be checked using result of (12), that a Gaussian reference measure for the noise term leads exactly to the problem of (14), which was obtained by statistical considerations.

### 5.3. Sensitivity of the reconstruction

Within the limit of small variations, we can also study the stability of the reconstruction with a sensitivity analysis. This enables to study the importance of a given data point on the reconstruction and quantify the amount of change resulting from a perturbation of the data. The sensitivity analysis is based on the determination of the derivative  $d\mathbf{x}/d\mathbf{y}$ . Although an expression can be obtained in the direct domain, the derivation is done in the dual domain, because the primal functions may be not explicit.

The stationary point  $\bar{\lambda}$  of the dual function  $\tilde{D}(\lambda)$  verifies

$$\mathbf{y} = \mathbf{A}\mathbf{x}_{\bar{\lambda}} + \mathcal{F}_{\nu}^{*'}(\bar{\lambda}).$$

Let us note  $\mathcal{F}_{\nu}^{*''}$  and  $\mathcal{F}_{\mu}^{*''}$  the diagonal matrix of the second derivatives  $\mathcal{F}_{\nu}^{*''}(\bar{\lambda})$  and  $\mathcal{F}_{\mu}^{*''}(\mathbf{A}^t\bar{\lambda})$ . Then, we have

$$d\mathbf{y} = \frac{\partial \mathbf{y}}{\partial \bar{\lambda}} d\bar{\lambda} = \left[ \mathbf{A} \frac{\partial \mathbf{x}_{\bar{\lambda}}}{\partial \bar{\lambda}} + \mathcal{F}_{\nu}^{*''} \right] d\bar{\lambda}.$$

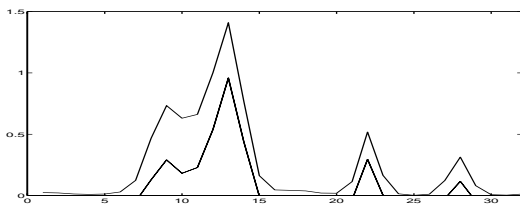
With the primal-dual relation  $\mathbf{x}_{\bar{\lambda}} = \mathcal{F}_{\mu}^{*'}(\mathbf{A}^t\bar{\lambda})$ , the partial derivative of  $\mathbf{x}_{\bar{\lambda}}$  with respect to  $\bar{\lambda}$  is simply  $\mathcal{F}_{\mu}^{*''}\mathbf{A}^t$ . Thus

$$d\mathbf{y} = [\mathbf{A}\mathcal{F}_{\mu}^{*''}\mathbf{A}^t + \mathcal{F}_{\nu}^{*''}]d\bar{\lambda}.$$

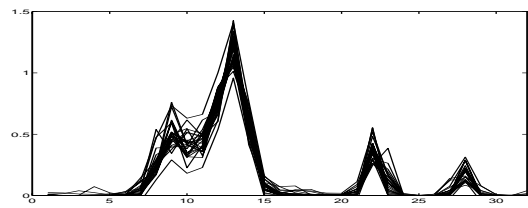
Using  $d\mathbf{x}_{\bar{\lambda}} = \frac{\partial \mathbf{x}_{\bar{\lambda}}}{\partial \bar{\lambda}} d\bar{\lambda}$ , one gets finally the relation

$$d\mathbf{x}_{\bar{\lambda}} = \mathbf{H}d\mathbf{y} = \mathcal{F}_{\mu}^{*''}\mathbf{A}^t [\mathbf{A}\mathcal{F}_{\mu}^{*''}\mathbf{A}^t + \mathcal{F}_{\nu}^{*''}]^{-1} d\mathbf{y}.$$

With  $E\{d\mathbf{y}d\mathbf{y}^t\} = \mathbf{R}_y$ , the noise covariance matrix, we use the “sensitivity matrix”  $\mathbf{H}\mathbf{R}_y\mathbf{H}^t$  whose (square root) diagonal terms may serve as “sensitivity bars”.



(a) Sensitivity bars.



(b) Monte Carlo study and sensitivity analysis.

**Figure 3:** Sensitivity of the reconstruction. Sensitivity “bars” are plotted in (a) and (b). In (b), we have reported 20 reconstructions of a Monte Carlo study. The variations of reconstructions are in good agreement with the sensitivity analysis.

## 6. Conclusion

It is always possible to modify our reference measures to balance the two terms of the global criterion (15) which should therefore be written as

$$\mathcal{J}(\mathbf{x}) = \mathcal{F}_\mu(\mathbf{x}) + \alpha \mathcal{F}_\nu(\mathbf{y} - \mathbf{A}\mathbf{x}),$$

where  $\alpha$  is a regularization parameter. The Maximum Entropy on the Mean procedure enables us to find the generic form of regularized criteria, and to solve the problem even if primal criteria  $\mathcal{F}_\mu$  and  $\mathcal{F}_\nu$  have no analytical expression.

Such an approach provides a new general framework for the interpretation and derivation of these criteria. Many other criteria as those presented in §4 have been derived [9]. In particular, reference measures defined as mixture of distributions (Gaussian, Gamma) have been successfully used for the reconstruction of blurred and noisy sparse spike trains. Poissonized sums of random variables also lead to interesting regularized procedure in connection with the general class of Bregman divergences. Work is also in progress concerning the *quantification* of the quality of MEMM estimates, the links with the Bayesian approach, especially with correlated *a priori* models such as Gibbs random fields.

## References

- [1] P. L. Combettes. The foundation of set theoretic estimation. *Proceedings of the IEEE*, 81(2):182–208, Feb. 1993.
- [2] I. Csizsár. Why least-squares and maximum entropy ? An axiomatic approach to inference for linear inverse problems. *The Annals of Statistics*, 19(4):2032–2066, 1991.
- [3] D. Dacunha-Castelle and F. Gamboa. Maximum d’entropie et problème des moments. *Annales de l’Institut Henri Poincaré*, 26(4):567–596, 1990.
- [4] R. S. Ellis. *Entropy, Large Deviations, and Statistical Mechanics*. Springer-Verlag, New York, 1985.
- [5] L. K. Jones and C. L. Byrne. General entropy criteria for inverse problems, with applications to data compression, pattern classification and cluster analysis. *IEEE transactions on Information Theory*, 36(1):23–30, Jan. 1990.
- [6] R. E. Kass and L. Wasserman. Formal Rules for Selecting Prior Distributions: A Review and Annotated Bibliography. Technical report, Department of Statistics, Carnegie Mellon University, 1994. Submitted to Jnl. of American Statistic Association.
- [7] S. Kullback. *Information Theory and Statistics*. Wiley, New York, 1959.
- [8] G. Le Besnerais. *Méthode du maximum d’entropie sur la moyenne, critères de reconstruction d’image et synthèse d’ouverture en radio-astronomie*. PhD thesis, University of Paris-Sud, 1993.
- [9] G. Le Besnerais, J.-F. Bercher, and G. Demoment. A new look at the entropy for solving linear inverse problems. *submitted to IEEE transactions on Information Theory*, 1994.
- [10] J. Navaza. The use of non-local constraints in maximum-entropy electron density reconstruction. *Acta Crystallographica*, pages 212–223, 1986.
- [11] J. E. Shore. Minimum cross-entropy spectral analysis. *IEEE transactions on Acoustics, Speech and Signal Processing*, (2):230–237, Apr. 1981.