

RESTORATION OF SPIKY SIGNALS : A NEW OPTIMAL ESTIMATE AND A COMPARISON

Christian Heinrich, Jean-François Bercher, Guy Le Besnerais and Guy Demoment

Laboratoire des signaux et systèmes (CNRS-ESE-UPS)

Supélec

Plateau de Moulon

91192 Gif-sur-Yvette Cedex, France

E-mail : heinrich@brutus.lss.supelec.fr

ABSTRACT

The subject of this communication is the restoration of spiky sequences distorted by a linear system and corrupted by an additive noise.

A (now) classical way of coping with this problem is to use a Bayesian approach with a Bernoulli-Gaussian (BG) prior model of the sequence. We will refine this method using a Bernoulli-Gaussian plus Gaussian (BGG) prior model. This estimation method requires maximization of a posterior probability distribution, which cannot be performed optimally.

Thus we propose here a new non-Bayesian estimation scheme, derived from the KULLBACK-LEIBLER information or cross-entropy. This quite general method, called the Maximum Entropy on the Mean Method (MEMM) in [1] and [2], is firmly based on convex analysis and yields a unique solution which can be efficiently calculated in practice, and which is, in this sense, truly optimal.

As a conclusion, we present results obtained with both methods on a synthetic case.

1. PROBLEM STATEMENT

In this communication, we address the problem of restoring spiky sequences \mathbf{x} distorted by a linear system \mathbf{H} and corrupted by an additive noise \mathbf{n} . Let \mathbf{z} be the indirect and noisy observation vector

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{n}. \quad (1)$$

The linear system and noise distribution are supposed to be known. Since matrix \mathbf{H} is generally ill-conditioned, one has to complete the data with an *a priori* knowledge on the solution, so as to select a physically acceptable one.

When restoring spiky sequences, one is awaiting a mixture of large amplitude reflectors – modeling for example interfaces between layers of a stratified medium

– and small amplitude reflectors – modeling inhomogeneities inside layers. A convenient way to account for this knowledge is to use a stochastic modeling, which could be a mixture of large variance zero-mean Gaussian variables and of small amplitude zero-mean Gaussian variables. This is the so-called BGG distribution.

The BGG model can be summarized by :

$$\left\{ \begin{array}{l} \mathbf{y} = (\mathbf{q}, \mathbf{x}) : \text{joint white process,} \\ \mathbf{q}[k] : \text{Bernoulli random variable,} \\ \quad \left\{ \begin{array}{l} P(\mathbf{q}[k] = 1) = \lambda, \\ P(\mathbf{q}[k] = 0) = 1 - \lambda, \end{array} \right. \\ \mathbf{x}[k] : \text{Gaussian random variable given } \mathbf{q}[k], \\ \quad \left\{ \begin{array}{l} \mathbf{x}[k] \sim \mathcal{N}(0, \mathbf{q}[k]r_l + (1 - \mathbf{q}[k])r_s), \\ \text{with } r_l > r_s. \end{array} \right. \end{array} \right.$$

In the sequel, we will use this model to solve problem (1), first with a classical non-optimal Bayesian method and then with a new optimal entropic one.

2. A CLASSICAL BAYESIAN RESOLUTION METHOD

The Bayesian approach of the problem with a BG prior has been introduced by KORMYLO and MENDEL in the late 70's [3] : they could only account for strictly homogeneous layers ($r_s = 0$). MENDEL suggested an extension of their method relying on the aforementioned BGG prior.

An intuitive Bayesian estimator of the sequence \mathbf{y} would be the one given by the maximization of the joint posterior likelihood :

$$(\hat{\mathbf{q}}, \hat{\mathbf{x}}) = \arg \max_{(\mathbf{q}, \mathbf{x})} p(\mathbf{z}|\mathbf{x}, I)p(\mathbf{x}|\mathbf{q}, I)P(\mathbf{q}|I)$$

where I stands for prior informations such as matrix \mathbf{H} and noise statistics. The noise will be here modeled as the realization of a white Gaussian zero-mean

Random Vector (RV) with variance r_n . This estimator usually gives unsatisfying reconstructions since too few large amplitude reflectors are detected, especially when r_s and r_l are of different orders of magnitude. The limit case when $r_s = 0$ corresponds to a BG prior, whose probability measure is known to be degenerated. In this case, the problem has been solved in [4] and [5] by using another Bayesian estimator. Therefore, our estimation algorithm will highly be based on those references. The selected estimator $\hat{\mathbf{y}}$ is :

$$\begin{cases} \hat{\mathbf{y}} &= (\hat{\mathbf{q}}, \hat{\mathbf{x}}), \\ \hat{\mathbf{q}} &= \arg \max_{\mathbf{q}} P(\mathbf{q}|\mathbf{z}, I) \\ &= \arg \max_{\mathbf{q}} p(\mathbf{z}|\mathbf{q}, I)P(\mathbf{q}|I), \\ \hat{\mathbf{x}} &= \arg \max_{\mathbf{x}} p(\mathbf{x}|\hat{\mathbf{q}}, \mathbf{z}, I). \end{cases}$$

We recognize here a detection step yielding $\hat{\mathbf{q}}$, followed by an estimation step yielding $\hat{\mathbf{x}}$. The detection step leads to maximizing the posterior log-likelihood L :

$$\begin{cases} \hat{\mathbf{q}} &= \arg \max_{\mathbf{q}} L(\mathbf{q}), \\ L(\mathbf{q}) &= -\log(\det(\mathbf{B}(\mathbf{q}))) - \mathbf{z}^t(\mathbf{B}(\mathbf{q}))^{-1}\mathbf{z} \\ &\quad - 2(\sum_k \mathbf{q}[k]) \log \frac{1-\lambda}{\lambda}, \end{cases}$$

where

$$\begin{cases} \mathbf{B}(\mathbf{q}) &= E[\mathbf{z}\mathbf{z}^t] \\ &= \mathbf{H}\mathbf{\Gamma}_x(\mathbf{q})\mathbf{H}^t + r_n\mathbf{I}, \\ \mathbf{\Gamma}_x(\mathbf{q}) &= r_s\mathbf{I} + (r_l - r_s)\mathbf{Q}, \\ \mathbf{Q} &= \text{diag}(\mathbf{q}[k]). \end{cases}$$

Knowing \mathbf{z} and \mathbf{q} , \mathbf{x} is Gaussian since we assume that the noise follows a Gaussian law. Hence $\hat{\mathbf{x}}$ is a linear function of the observations, which may be written :

$$\hat{\mathbf{x}} = \mathbf{\Gamma}_x(\hat{\mathbf{q}})\mathbf{H}^t(\mathbf{B}(\hat{\mathbf{q}}))^{-1}\mathbf{z}.$$

Nevertheless, detection of \mathbf{q} requires evaluation of criterion L for 2^n possible values of \mathbf{q} , which is computationally intractable for realistic values of the dimension n of vector \mathbf{x} . To estimate \mathbf{q} , KORMYLO and MENDEL introduced the Single Most Likely Replacement algorithm in the BG case [3], which has also been used in [4] and [5]. Knowing the current estimate \mathbf{q}_0 , L is computed on a set of neighboring sequences \mathbf{q}_k which differ from \mathbf{q}_0 at a location k only. The starting point of the next iteration will be the one minimizing L among this neighbors. This method is sub-optimal since it only guarantees convergence to a local optimum.

To perform the optimization step, we can take advantage of a simple relation linking $L(\mathbf{q}_0)$ and $L(\mathbf{q}_k)$. Since the BGG model can also be considered as a BG model with colored noise, we can derive this relation from [5], which gives :

$$\begin{aligned} L(\mathbf{q}_k) &= L(\mathbf{q}_0) + \frac{1}{\rho_k} \mathbf{w}_0^t \mathbf{v}_k \mathbf{v}_k^t \mathbf{w}_0 \\ &\quad - \log(\varepsilon_k (r_l - r_s) \rho_k) - 2\varepsilon_k \log \frac{1-\lambda}{\lambda}, \end{aligned}$$

where

$$\begin{cases} \mathbf{A}_0 &= \mathbf{H}^t(\mathbf{B}(\mathbf{q}_0))^{-1}\mathbf{H}, \\ \mathbf{w}_k &= \mathbf{H}^t(\mathbf{B}(\mathbf{q}_k))^{-1}\mathbf{z}, \\ \mathbf{v}_k &: k^{th} \text{ vector of the canonical basis,} \\ \varepsilon_k &= \pm 1 (\text{whether we add or remove a reflector}), \\ \rho_k &= \frac{\varepsilon_k}{r_l - r_s} + \mathbf{v}_k^t \mathbf{A}_0 \mathbf{v}_k. \end{cases}$$

Care has been taken to minimize the computational burden, which is not greater than $O(n^2)$ multiplications.

3. A NEW OPTIMAL ENTROPIC ESTIMATE

For the sake of simplicity, let us first study the noiseless problem. Our observation equation then reduces to $\mathbf{z} = \mathbf{H}\mathbf{x}$. The noisy case only needs a simple modification of the basic scheme.

An analogy can be made with statistical mechanics so as to derive the estimator introduced by the pioneering work of NAVAZA [6]. In this framework, \mathbf{x} is supposed to be the macrostate of some physical system. The macrostate is considered as being the average value of microstates, such as velocities of particles in a discrete gas. In a statistical point of view, those microstates are the realisation of a RV \mathbf{u} under a given law ν . Having performed no observation, the BGG distribution is a suitable measure ν related to a spiky signal. Performing an observation corresponds to observing the mean of a realization of the underlying microstates. Those microstates yield an empirical distribution p of the RV \mathbf{u} . We can state this as :

$$\begin{cases} \mathbf{x} &= E_p[\mathbf{u}], \\ \mathbf{z} &= \mathbf{H}\mathbf{x}. \end{cases} \quad (2)$$

Moreover, every empirical distribution has a given probability of occurrence. This probability depends on ν and will be written as $Q(p|\nu)$.

Given an observation, the most likely underlying macrostate is then the average value of the empirical distribution which is the most likely to occur. We can state this as :

$$\begin{cases} \hat{\mathbf{x}} &= E_{\hat{p}}[\mathbf{u}], \\ \mathbf{z} &= \mathbf{H}\hat{\mathbf{x}}, \\ \hat{p} &= \arg \max_p Q(p|\nu). \end{cases}$$

Since the empirical distributions follow a Large Deviations Principle, an exponential equivalent of our probability of occurrence $Q(p|\nu)$, assuming a large number m of microstates, is :

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

where

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

is the KULLBACK distance between distributions p and ν , and is called the rate function in the large deviations theory (see [7] for developments on statistical mechanics and large deviations).

We have obtained a new formulation of the problem :

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

In other words, we now have to find the distribution closest to ν according to the KULLBACK distance, and whose mean satisfies the observations, *i.e.* $\mathbf{z} = \mathbf{H} E_{\hat{p}}[\mathbf{u}]$. This is the less informative probability measure with respect to distribution ν and satisfying the constraints. The chosen estimator is the mean under that probability measure.

We face here a minimization problem under constraints, which is cumbersome to handle. A classical way of coping with such a difficulty is to use Lagrangians and duality theory [8] and to introduce an additional vector $\boldsymbol{\mu}$, leading to an unconstrained problem.

The sought distribution \hat{p} belongs to a family of distributions parameterized by $\boldsymbol{\mu}$

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

$\log Z(\boldsymbol{\mu})$ is the log-partition function, or the log-Laplace transform of the measure $d\nu(\mathbf{u})$, and is given by :

$$\log Z(\boldsymbol{\mu}) = \log \left(\int \exp(\boldsymbol{\mu}^t \mathbf{H} \mathbf{u}) d\nu(\mathbf{u}) \right). \quad (3)$$

The first optimization problem is then equivalent to the following dual and unconstrained problem :

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

Given distribution $p_{\hat{\boldsymbol{\mu}}}$, one has to derive its mean value. This is achieved using the so-called primal-dual relation :

$$E_{p_{\hat{\boldsymbol{\mu}}}}[\mathbf{u}] = \left. \frac{\partial \log Z(\boldsymbol{\mu})}{\partial (\mathbf{H}^t \boldsymbol{\mu})} \right|_{\boldsymbol{\mu}=\hat{\boldsymbol{\mu}}} \quad (4)$$

which no longer explicitly depends on $p_{\hat{\boldsymbol{\mu}}}$.

Actually, computation of $p_{\hat{\boldsymbol{\mu}}}$ appears to be an intermediary step which can be omitted using the primal criterion \mathcal{F} defined by :

$$\mathcal{F}(\mathbf{x}) = \min_{\boldsymbol{\mu}} \mathcal{K}(p, \nu).$$

NAVAZA, LE BESNERAIS and BERCHER have shown that an equivalent formulation of the previous problem was :

$$\text{(primal)} \quad \hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathcal{C}_z} \mathcal{F}(\mathbf{x})$$

where $\mathcal{C}_z = \{\mathbf{x} | \mathbf{z} = \mathbf{H} \mathbf{x}\}$, since, as a result of duality theory :

$$\min_{\mathbf{x} \in \mathcal{C}_z} \mathcal{F}(\mathbf{x}) = \sup_{\boldsymbol{\mu}} D(\boldsymbol{\mu}).$$

Nevertheless, the primal problem is constrained and \mathcal{F} has no explicit expression in the case considered here. Though \mathcal{F} has a nice convexity property, computation is practically performed using the dual unconstrained formulation : $D(\boldsymbol{\mu})$ is a convex function of $\boldsymbol{\mu}$ and its maximization, using classical optimization methods such as gradient or conjugate-gradient, yields a unique solution. The estimate is optimal because of the convexity of $D(\boldsymbol{\mu})$. Notice that $\sup_{\boldsymbol{\mu}} D(\boldsymbol{\mu})$, viewed as a function of \mathbf{z} , is the LEGENDRE transform of $\log Z(\boldsymbol{\mu})$ and also the CRAMÉR transform of the reference measure $d\nu(\mathbf{u})$.

Let k_i be : $k_i = (\mathbf{H}^t \boldsymbol{\mu})[i]$. In practice, resolution of the MEMM requires five steps :

1. Choice of the reference measure $d\nu$ accounting for our prior information on the sequence [2]. In the case of spiky sequences, we used :

$$\begin{aligned} \nu(\mathbf{x}) &= \prod_i \nu_i(x_i), \\ \nu_i(x_i) &= \frac{\lambda}{\sqrt{2\pi r_l}} \exp\left(-\frac{x_i^2}{2r_l}\right) + \frac{1-\lambda}{\sqrt{2\pi r_s}} \exp\left(-\frac{x_i^2}{2r_s}\right). \end{aligned}$$

2. Analytical computation of the log-partition function $\log Z(\boldsymbol{\mu})$ using (3), from which we derive the dual function $D(\boldsymbol{\mu})$:

$$D(\boldsymbol{\mu}) = \boldsymbol{\mu}^t \mathbf{z} - \sum_i \log \left(\lambda \exp\left(\frac{r_l k_i^2}{2}\right) + (1-\lambda) \exp\left(\frac{r_s k_i^2}{2}\right) \right).$$

3. Analytical computation of the primal-dual relation (4) :

$$\left(E_{p_{\hat{\boldsymbol{\mu}}}}[\mathbf{u}] \right)[i] = \frac{\lambda r_l k_i \exp\left(\frac{r_l k_i^2}{2}\right) + (1-\lambda) r_s k_i \exp\left(\frac{r_s k_i^2}{2}\right)}{\lambda \exp\left(\frac{r_l k_i^2}{2}\right) + (1-\lambda) \exp\left(\frac{r_s k_i^2}{2}\right)}$$

4. Numerical computation of $\hat{\boldsymbol{\mu}} = \arg \max_{\boldsymbol{\mu}} D(\boldsymbol{\mu})$.
5. Numerical computation of our MEMM estimator of the sequence, as $E_{p_{\hat{\boldsymbol{\mu}}}}[\mathbf{u}]$.

Up to this point, we merely considered the noiseless problem. How can we account for the noise ? Let us introduce an extended object $\tilde{\mathbf{x}} = [\mathbf{x}^t | \mathbf{n}^t]^t$. If we let $\tilde{\mathbf{H}}$ be the corresponding extended matrix $[\mathbf{H} | \mathbf{I}d]$, the noisy problem can now be written as $\mathbf{z} = \tilde{\mathbf{H}} \tilde{\mathbf{x}}$ and the

MEMM still applies to the extended object. The reference measure is now : $\tilde{\nu}(\tilde{\mathbf{x}}) = \nu_1(\mathbf{x}) \otimes \nu_2(\mathbf{n})$, where ν_2 is a Gaussian law of variance r_n in the case considered here. The new dual function is :

$$\tilde{D}(\boldsymbol{\mu}) = D(\boldsymbol{\mu}) - \frac{1}{2}r_n\boldsymbol{\mu}^t\boldsymbol{\mu}.$$

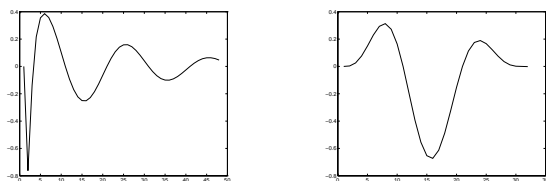
4. SIMULATION RESULTS AND COMPARISONS

In this section, we present some simulation results. We convolved the so-called MENDEL sequence either with a KRAMER or a RICKER wavelet with a 7 dB SNR, which are realistic conditions.

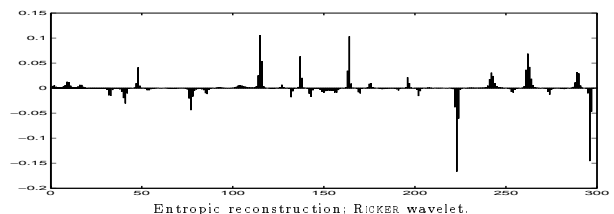
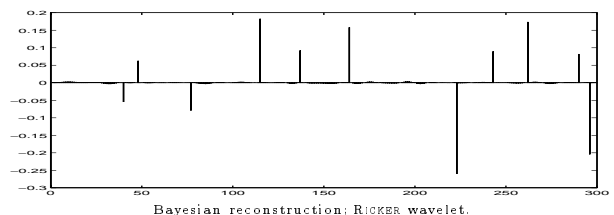
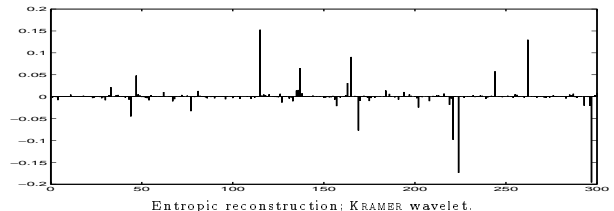
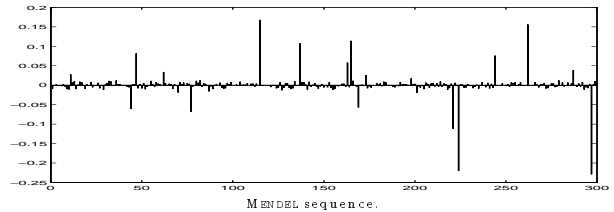
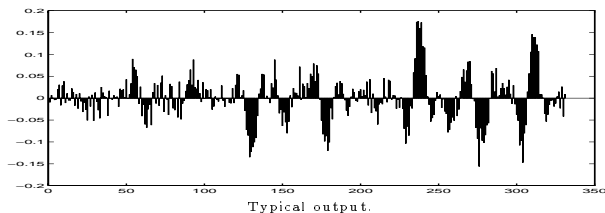
The entropic method provides outstanding reconstructions with the KRAMER wavelet. With the RICKER wavelet, reconstructions obtained with the Bayesian method are of better quality than entropic ones on this synthetic example because we have fewer non-detections and because the reconstructed signal is more spiky.

Nevertheless, Bayesian reconstruction often fails for actual signals because (i) of a lack of robustness with respect to the BGG modeling and (ii) a quick degradation of the detection step with low SNR values.

We have good reasons to think that the flexibility of the MEMM introduced by our general form of prior should improve the reconstruction. In the MEMM, the absence of any comparison to a threshold as in the detection step of the Bayesian method should improve the robustness with respect to noise. Work is currently in progress to compare both approaches on actual data measured either in reflection seismology or ultrasound imaging.



KRAMER and RICKER wavelets.



5. REFERENCES

- [1] F. GAMBOA, *Méthode du maximum d'entropie sur la moyenne et applications*. PhD thesis, Univ. de Paris-Sud, Paris XI., Orsay, France, 1989.
- [2] 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*, 1995.
- [3] J. KORMYLO and J. MENDEL, "Maximum-likelihood detection and estimation of Bernoulli-Gaussian processes," *IEEE Transactions on Information Theory*, vol. IT-28, 1982.
- [4] Y. GOUSSARD, G. DEMOMENT, and J. IDIER, "A new algorithm for iterative deconvolution of sparse spike trains," in *Proceedings of IEEE ICASSP*, 1990.
- [5] F. CHAMPAGNAT, J. IDIER, and G. DEMOMENT, "Deconvolution of sparse spike trains accounting for wavelet phase shifts and colored noise," in *Proceedings of IEEE ICASSP*, 1993.
- [6] J. NAVAZA, "The use of non-local constraints in maximum-entropy electron density reconstruction," *Acta Crystallographica*, vol. A-42, pp. 212-223, 1986.
- [7] R. S. ELLIS, *Entropy, Large Deviations, and Statistical Mechanics*. Springer-Verlag, New York, 1985.
- [8] D. LUENBERGER, *Optimization by Vector Space Methods*. J. Wiley, New York, 1969.