European Signal Proc. Conf. (EUSIPCO), Sept. 1994, pp. 1062-1065

Discontinuity reconstruction from linear attenuating operators using the weak-string model

Mila Nikolova and Ali Mohammad-Djafari

Laboratoire des Signaux et Systèmes (CNRS–Supélec–UPS)
Supélec, Plateau du Moulon, 91192 Gif-sur-Yvette Cedex, France
Tel/Fax: +33 1 69 41 80 40/33 1 69 41 30 60
E-Mail: nikolova@lss.supelec.fr, djafari@lss.supelec.fr

Abstract. We consider the reconstruction of a piecewise smooth signal, observed through an attenuating operator, such as an incomplete Laplace transform. This inverse problem is very ill-posed, so a regularisation must be applied. The key question is the recovery of the breakpoints between smooth parts. The weak string nicely models this class of signals. But it happens that in presence of attenuation the relevant MAP estimator fails. Physically, the discontinuity recovery is only limited by the noise. We propose a reconstruction technique which works just at this limit and so exploits data very efficiently.

1 Introduction

Attenuation occurs in various applications: it may model the loss inside a medium while a wave propagates, or the signals received from an object moving away, *etc...* In this paper, we are concerned with models, which may be considered as being composed of a pure attenuation, applied on the signal of interest, and some subsequent non-attenuating, generaly non-injective, linear transformation (Section 2). Note that generally it is the *contribution* of samples to the data which is decaying, rather than the data. The noisy Laplace transform observed on a sparse set is a typical example.

The inverse problem is "doubly" ill-posed: *(a)* it is underdetermined, and *(b)* the attenuation, although mathematically invertible, has a too large condition number. So, a regularisation must be applied. The reconstructed signals (chains) are known *a priori* to be only piecewise smooth, and the regularisation should take this into account. A pertinent model is the *weak string*, introduced by Blake & Zisserman [1]. It turns out that the relevant MAP solution fails in presence of attenuation (Section 4). We derived an attenuation-adapted reconstruction technique, which in the noiseless case completely compensates for attenuation; in presence of noise, it may happen that the adaptation can only be introduced up to some level. This technique exploits data much better than the MAP estimator (Sections 5, 6).

The solution calculation involves optimising an energy function, which presents many local minima; finding the global one is non-trivial. In a recent paper [5] the authors extended the Blake & Zisserman's graduated non convexity (GNC) – a very effective in practice, though not theoretically convergent, deterministic global optimisation technique – in order to use it for ill-posed problems. The attenuation requires an adaptation of the GNC parameters as well (Section 7).

Simulation results are presented in Section 8 and concluding remarks in Section 9.

2 Attenuating models

The observation model is given by a linear integral equation whose discrete form relates noisy data $y \in C^N$ to the original chain $x \in R^M$ by:

$$y = \mathcal{A}x + n, \qquad (1)$$

where $\mathcal{A}$ is a known linear attenuating operator, which does not suppress the mean of x, and n is the observation noise, assumed additive, white and Gaussian $\mathcal{N}(0, \sigma^2 I)$. The observation operator $\mathcal{A}$ can be factorised into

$$\mathcal{A} = \mathcal{B}\mathcal{U}, \qquad (2)$$

where $\mathcal{B}$ is non-attenuating and its columns b_i verify $\|b_i\|_2 = 1\ \forall i$, and $\mathcal{U} = \mathrm{diag}(u_1, \ldots, u_M)$ is purely attenuating: $u_1 > u_2 > \ldots > u_M$. The contributions to y of the samples $x[i]$ along the chain are highly decreasing.

The likelihood is $\mathcal{L}(x) = \frac{1}{2\sigma^2}\|\mathcal{B}\mathcal{U}x - y\|^2$. Note that the Laplace transform on a regular set can be simplified to a purely attenuating observation: $\mathcal{B}$ is the DFT operator, and as $\mathcal{B}^{-1} = \mathcal{B}^\dagger$, the likelihood is simply $\mathcal{L}(x) = \frac{1}{2\sigma^2}\|\mathcal{U}x - \mathcal{B}^\dagger y\|^2$; $\dagger$ means transposed conjugate.

3 The weak-string model

The weak string corresponds to a locally Gaussian, nonstationnary first-order Markov chain with a Boolean line process. It models piecewise continuous signals and its regularisation role is to locally smooth the signal while preserving abrupt transitions.

The relevant MAP energy is:

$$\mathcal{E}(x,l) = \|\mathcal{A}x - y\|^2 + \Phi(x,l), \tag{3}$$

$$\Phi(x,l) = \sum_{i=1}^{M-1} \lambda^2 (x[i+1] - x[i])^2 (1 - l[i]) + \alpha l[i], \boldsymbol{lmap} \tag{4}$$

where $l = \{l[1] \ldots l[M-1]\} \in \{0,1\}^{M-1}$ is the line process and (α, λ) are positive constants. Let $t[i]$ be the transition: $t[i] \doteq x[i+1] - x[i]$. Conditionally to x, the optimal line estimate $\hat{l}$ is

$$\begin{cases} \hat{l}[i] = 0 & \text{if} \quad |t[i]| < T \\ \hat{l}[i] = 1 & \text{if} \quad |t[i]| \geq T \end{cases}, \; T = \frac{\sqrt{\alpha}}{\lambda}. \tag{5}$$

T is the prior discontinuity detection threshold. The prior energy Φ is then a sum of interaction terms $\Phi(x) = \sum \phi(t[i])$ with

$$\phi(t[i]) = \min\{(\lambda t[i])^2, \alpha\}. \tag{6}$$

α is the cost of introducing a discontinuity; λ determines the local smoothing degree and typically $\lambda \gg 2\sigma^2$.

The solution $\hat{x}$, minimising the energy $\mathcal{E}$, shall be reffered to as the MAP solution. In presence of attenuation it is quite unsatisfactory and our objective is to find an adaptation of the model parameters (α_A, λ_A) as a function of the attenuation, leading a better reconstruction.

4 Behaviour in attenuation

A magnitude of crucial importance is the threshold amplitude H of a jump at position m in the orignal signal x^* above which the optimal solution $\hat{x}$ will have a discontinuity at m. H is in fact the posterior discontinuity detection threshold. Assuring a good thresholding assures a good estimation of the line process.

This threshold has been studied analytically and numerically in the case $\mathcal{A} = I$ [1], [3], *etc...* But the role of $\mathcal{B}$ and especially of $\mathcal{U}$ is really determinant.

In order to avoid any possible interferences, in this section we treated the simplest noiseless attenuating model: $y = \mathcal{U}x^*$. For the simulations $u_m = e^{-0,02m}$.

We established an expression for H as a function of the solution energies. Even when the solution can be written analytically, the energy expression is too complicated, so we computed thresholds numerically.

In the figures the exact optima are shown, which have been calculated using the Viterbi algorithm [4] for a very finely discretised state space.

The obtained results on the threshold behaviour in presence of attenuation remain true when $\mathcal{B}$ is an ill-conditioned non-attenuating operator.

4.1 The isolated step response

Let the original chain be

$$x^*(m;h) = hY_m, \quad Y_m[i] = \begin{cases} 0 & \text{if} \quad i < m \\ 1 & \text{if} \quad i \geq m \end{cases}, h \in R. \tag{7}$$

The optimal MAP solution is either $\hat{x}_c(m;h)$ – the minimiser of $\|\mathcal{A}x - y\|^2 + \|\lambda D x\|^2$, where D is the first-order difference operator, or $\hat{x}_d(m;h)$:

$$\begin{cases} \hat{x}_c = h(\mathcal{U}\Re(\mathcal{B}^\dagger\mathcal{B})\mathcal{U} + \lambda^2 D^\top D)^{-1}\mathcal{U}\Re(\mathcal{B}^\dagger)Y_m \boldsymbol{lsolc} \\ \hat{x}_d = hY_m \end{cases}; \tag{8}$$

$\Re$ stands for real part and $\top$ – for transposed. $\hat{x}_c$ corresponds to $\hat{l}_m = 0$ and $\hat{x}_d$ – to $\hat{l}_m = 1$.

$\mathcal{E}(\hat{x}_c)$ is easy to compute and $\mathcal{E}(\hat{x}_d) = \alpha$. Finally, the optimal solution $\hat{x}$ is determined by:

$$\begin{cases} \mathcal{E}(\hat{x}_c) < \mathcal{E}(\hat{x}_d) & \mapsto \quad \hat{x} = \hat{x}_c \\ \mathcal{E}(\hat{x}_c) \geq \mathcal{E}(\hat{x}_d) & \mapsto \quad \hat{x} = \hat{x}_d \end{cases} \tag{9}$$

h is the scale of the original chain; it is straightforward that $x^*(m;h) = hx^*(m;1)$ leads to $\hat{x}_c(m;h) = h\hat{x}_c(m;1)$ and hence $\mathcal{E}(\hat{x}_c(m;h)) = h^2\mathcal{E}(\hat{x}_c(m;1))$. The quantity of interest is $H_0(m)$ – the critical value of h, such that:

$$\begin{cases} |h| < H_0(m) & \mapsto \quad \hat{x} = \hat{x}_c(m;h) \\ |h| \geq H_0(m) & \mapsto \quad \hat{x} = hY_m \end{cases}. \tag{10}$$

$H_0(m)$ verifies $h^2\mathcal{E}(\hat{x}_c(m;1)) = \mathcal{E}(\hat{x}_d(m;h))$, so

$$H_0(m) = \sqrt{\frac{\alpha}{\mathcal{E}(\hat{x}_c(m,1))}}. \tag{11}$$

For a non-attenuating operator, $H_0(m)$ is independant of the position m of the jump along the chain (if m is not close to the ends) [1]. For an attenuating operator, $\mathcal{E}(\hat{x}_c(m,1))$ decreases when the step is shifted to the right and the detection threshold $H_0(m)$ increases jointly with the attenuation (see Fig. 1). That's why the MAP solution fails. In order to set $H_0(m)$ constant we shall substitute α_A for α:

$$\alpha_A[m] = H_0^2\mathcal{E}(\hat{x}_c(m,1)), \quad H_0 = H_0(1). \tag{12}$$

The role of this α-adaptation can be seen in Fig. 4.

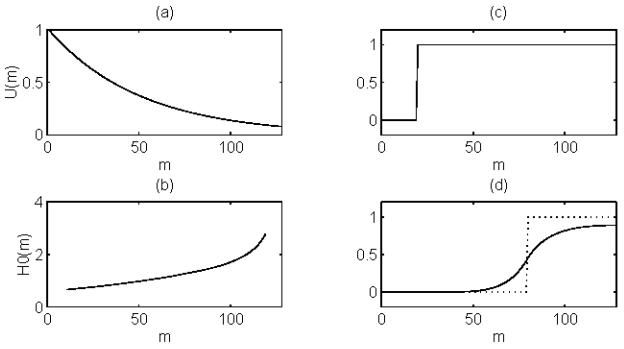

Figure 1: The isolated step response in attenuation. Original chain (...), reconstruction (–). **(a)** The attenuation function $u_m = e^{-0,02m}$. **(b)** The detection threshold $H_0(m)$. **(c)** The step is perfectly reconstructed when $m \leq 42$. **(d)** The same step but shifted cannot be detected any more.

4.2 Interacting discontinuities

Let the original chain be an a-width gate:

$$x^* = h(Y_m - Y_{m+a-1}), \; a \in N, \; h \in R. \tag{13}$$

In a basically similar way we studied the detection threshold $H_1(m;a)$ in x^*, leading to a discontinuity at

m in the solution $\hat{x}$. In order to determine the global minimum, we compared the potential optima: $\hat{x}_c$ – continuous, $\hat{x} = x^*$ – discontinuous, and $\hat{x}_i$ – discontinuous at m, continuous at $(m+a-1)$. The threshold $H_1(m;a)$ is the critical value of h such that the global minimum in $\hat{x}_c$ switches to either $\hat{x}_i$ or $\hat{x}_d$.

When $\mathcal{A} = I$, it is known that this threshold increases when a decreases. On the other hand, for a fixed, this increase is independant of the position m [1]. But in presence of attenuation, for each a fixed, $H_1(m;a)$ increases with the position m. And the smaller a, the stronger the effect is. It is produced for both models – the non-adapted (Fig. 2 (a)) and the α-adapted (α_A, λ) (Fig. 2 (b)), weaker in the second case.

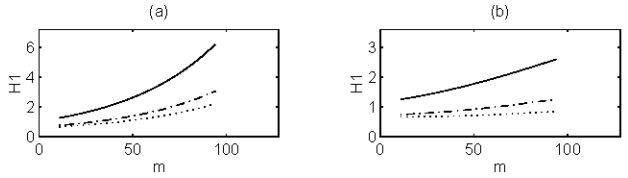

Figure 2: Interacting discontinuity detection threshold $H_1(m,a)$ for $a = 1$ (_), $a = 5$ (_ ._) and $a = 15$ (...). **(a)** Non-adapted model ($\alpha = C^{te}, \lambda = C^{te}$). **(b)** α-adapted model ($\alpha_A, \lambda = C^{te}$).

As a consequence, a gate and a shifted copy of it will be reconstructed differently (Fig. 3). The consequences for a more complicated chain can be seen in Fig. 4 (e).

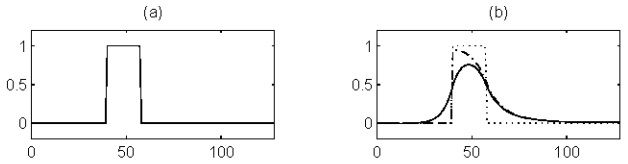

Figure 3: Inadequacy of the α-adaptation (α_A, λ). Original chain (...), reconstruction (_). **(a)** The gate signal is well recovered. **(b)** The same gate but delayed cannot be recovered.

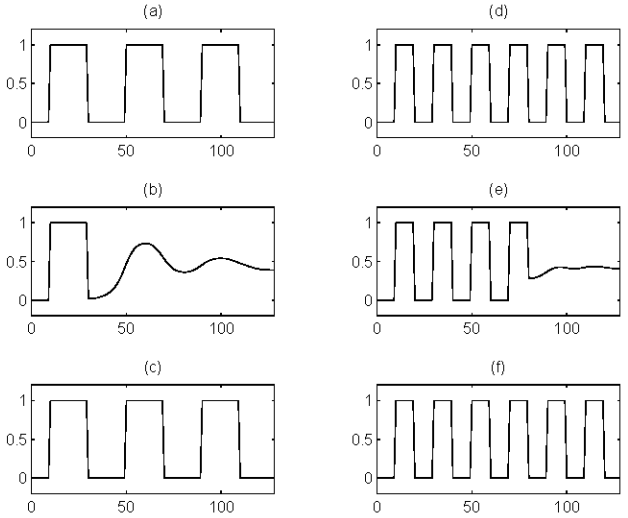

Figure 4: Discontinuity interaction. **(a)** Original chain with non-interacting discontinuities. **(b)** The MAP solution. **(c)** The α-adaptation (α_A, $\lambda = C^{te}$) is sufficient. **(d)** Original chain with interacting discontinuities. **(e)** The α-adaptation fails. **(f)** Reconstruction with (α_A, λ_A).

Under the hypothesis that $\forall m$, $(u_m - u_{m+1})$ is small enough, this effect can be suppressed by taking:

$$\lambda_A = \lambda \mathcal{U}. \tag{14}$$

The α-adaptation depends on λ, so we must first compute λ_A and then compute the adaptation α_A. The resultant energy function, say $\mathcal{E}_A$, reads:

$$\mathcal{E}_A(x) = \|\mathcal{A}x - y\|^2 + \sum_{i=1}^{M-1} \phi_i(x[i+1] - x[i]), \tag{15}$$

$$\phi_i(t[i]) = \min\{(\lambda_A[i]t[i])^2, \alpha_A[i]\}. \tag{16}$$

(α_A, λ_A) defines a non-homogeneous Markov chain, which in the noiseless case positions the attenuated problem solution on the solution of the equivalent non-attenuated problem (see Fig. 4).

5 The noise limit

This energy adaptation, aimed to improve recovery of discontinuities, reduces the regularisation parameter values jointly with the rise in attenuation: in presence of noise it is limited then by the stability requirement.

We shall apply the adaptation only until a level deduced from the false alarm probability P_{fa}. As the noise n is $\mathcal{N}(0, \sigma^2 I)$, the requirement: $P_{fa} \leq \epsilon$ is equivalent to $P(|n| \geq S)$, where S may be obtained numerically. The latter leads to $H_1(m;1) \geq S/u_m$, $m = 1, \ldots, M-1$. $H_1(m;1)$ has been set constant and S/u_m is increasing, so the adaptation can only be accomplished as long as the inequality holds: $\{\alpha_A[1], \ldots, \alpha_A[K], \alpha_A[K], \ldots\}$ and $\{\lambda_A[1], \ldots, \lambda_A[K], \lambda_A[K], \ldots\}$.

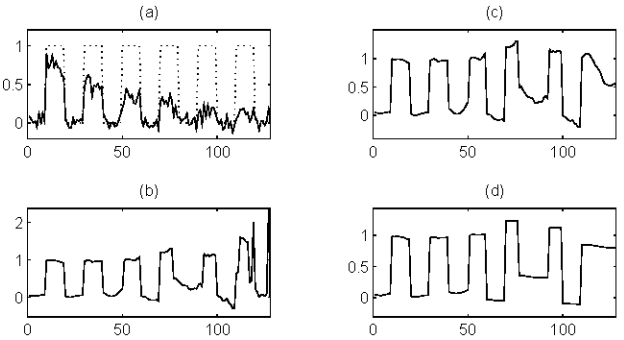

Figure 5: The noise influence. **(a)** Original chain x^* (...), noisy observation with SNR 10 dB (_). **(b)** Unlimited adaptation: fals alarms appear with the attenuation rise. **(c)** The minimum of $\mathcal{E}_A$ with limited adaptation. **(d)** Post-smoothed solution.

6 Local post-smoothing

As $\lambda_A[m]$ decreases with m, the continuous parts of the solution tend to be increasingly rough in presence of noise (Fig. 5 (b)). It is convenient to operate a local post-smoothing by computing the minimum of the true energy $\mathcal{E}$ conditionally to the discontinuity set, obtained via $\mathcal{E}_A$. Note that $\mathcal{E}$ is then convex in x and this calculation is very easy. The results in the Figs 5 (c), 7 (d) and 8 (b) have been post-smoothed.

7 GNC optimisation

For a real-world problem we are no longer able to compute the exact optimum: for a longer signal, or one having a greater variation, or when $\mathcal{B}$ is not diagonal, the Viterbi algorithm becomes numerically inapplicable. We then focused on the graduated non-convexity (GNC), which was developped by Blake & Zisserman [1] in order to compute the MAP estimate in the case when $\mathcal{A} = I$. The basic idea is the following. A family of continuously derivable functions $(\mathcal{F}_r)_{r\in R^+}$ is constructed, such that: *(a)* $\lim_{r\to\infty} \mathcal{F}_r = \mathcal{E}$ (the $\mathcal{F}_r$'s are approximations of $\mathcal{E}$), and *(b)* for some r_0, $\mathcal{F}_{r_0}$ is convex. Let $\{r_0, r_1, \ldots\}$ be the relaxation sequence. $\mathcal{F}_{r_0}$ has an unique minimum $\hat{x}_0$; starting with it, a sequence of minima $\hat{x}_i$ ($\hat{x}_i = \arg\min \mathcal{F}_{r_i}$) is tracked by local descent in the vicinity of the previously computed minimum $\hat{x}_{i-1}$. Practical convergence of GNC is very satisfactory [2].

When $\mathcal{A} = I$, the $\mathcal{F}_r$ are obtained by fitting a quadratic spline at $\pm T$; dropping indexes $[i]$, ϕ_r is [1]:

$$\phi_r(t) = \min\{(\lambda t)^2, \alpha - \frac{1}{2}r[t - p(r)]^2, \alpha\} \qquad (17)$$

where $p(r) = T\sqrt{(r + 2\lambda^2)/r}$. Clearly $\lim_{r\to\infty} \phi_r = \phi$. Transitions in the zone of the spline are undetermined. As r evolves to larger values, they leave this zone.

The convex energy function is found by checking the positive-definiteness of its Hessian matrix. When $\mathcal{A} = I$, convexity occurs for $r < \frac{1}{2}$ [1]. In a recent paper [5] the authors showed that an ill-posed problem does not allow the family $\{\mathcal{F}_r\}$ to admit any convex function. That's why they append a small auxiliary convex term, which is relaxed to zero afterwards: $\mathcal{F}_{r,s} = \mathcal{F}_r + s\sum(t[i])^2$ and $\mathcal{F}_{r,s}$ is convex when $s > \frac{1}{2}r$, $\forall\mathcal{A}$.

A crucial thing in GNC, which is not expressed explicitely, is that the first convex approximation must be "as close as possible" to the true energy $\mathcal{E}$. For small r a great part of the transitions are in the undetermined zone: let us look at the Hessian $\mathcal{H}$ in the extreme case:

$$\mathcal{H} = 2\mathcal{U}\Re(B^\dagger B)\mathcal{U} - (r - 2s)D^\top D. \qquad (18)$$

It suggests that the deformation imposed by the convex approximation upon undetermined transition at the end of the chain is much greater than at its beginning. That's why GNC fails, as it can be observed in the Fig. 6 (a). Supposing again that the u_m's vary slowly, we can counterbalance this effect by taking (r_A, s_A):

$$r_A = r\mathcal{U}^2 \text{ and } s_A = s\mathcal{U}^2. \qquad (19)$$

It is clear that initial convexity conditions remain unchainged. The obtained result is seen in Fig. 6 (b).

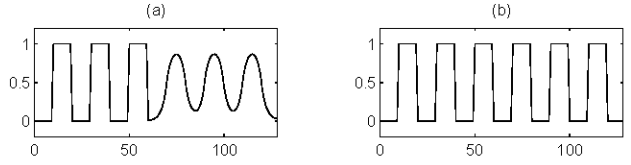

Figure 6: GNC adaptation. The original chain: see Fig. 4 (d). **(a)** Reconstruction with (α_A, λ_A) and non-adapted GNC **(b)** (α_A, λ_A) and adapted GNC

8 Simulations

We present the reconstruction, using the proposed technique, of a 128-point chain from its 64-point low-pass filtered and noisy Laplace transform, with attenuation $u_m = e^{-0,02m}$ and with a SNR of 20 dB.

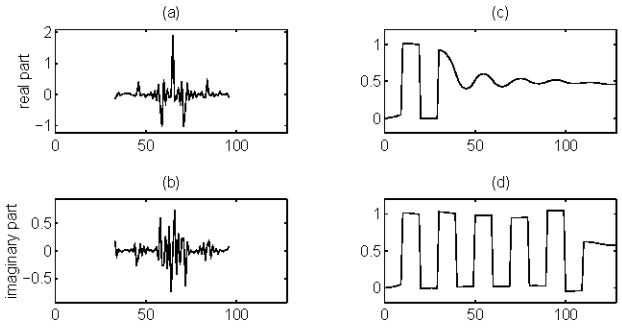

Figure 7: Incomplete noisy Laplace transform: SNR=20 dB, 64 data points for a 128-points signal. Original chain (...), reconstruction (_). **(a)** Data – real part. **(b)** Data – imaginary part. **(c)** The MAP solution, computed by GNC. **(d)** The reconstruction using the proposed technique.

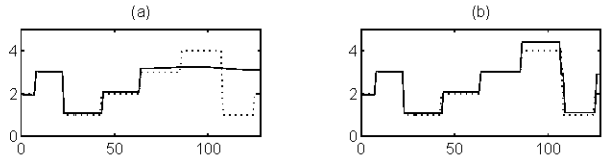

Figure 8: The same observation model as in Fig. 7. **(a)** MAP via GNC. **(b)** The proposed technique result.

9 Conclusion

The presented analysis explains why the weak-string MAP estimate fails when the linear observation model is attenuating. We established an attenuation-adapted processing and deduced its application domain in function of the noise variance. The obtained improvement in comparison with the MAP estimate is considerable: data is more efficiently exploited. The optimisation is carried out by GNC, whose relaxation parameters have also to be adapted to the attenuation. By way of application, we show the inversion, using the proposed technique, of a noisy and incomplete Laplace transform.

References

[1] Blake A. & Zisserman A., *Visual Reconstruction*. Cambridge, MA: MIT Press, 1987.

[2] Blake A., "Comparison of the Efficiency of Deterministic and Stochastic Algorithms for Visual Reconstruction," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-11, pp. 2—12, 1989.

[3] Mumford D. & Shah J., "Boundary Detection by Minimizing Functionals," *IEEE ICASSP 1985*, pp. 22-26.

[4] Rabiner L. & Juang B., "An Introduction to Hidden Markov models," *IEEE ASSP Magazine*, January 1986, pp. 4-16.

[5] Nikolova M., Mohammad-Djafari A. and Idier J., "Inversion of large-support ill-conditioned linear operators using a Markov model with a line process," *IEEE ICASSP 1994*, pp. V 357-360.