INVERSION OF LARGE-SUPPORT ILL-CONDITIONED LINEAR OPERATORS USING A MARKOV MODEL WITH A LINE PROCESS

Mila Nikolova, Ali Mohammad-Djafari and Jérôme Idier

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

ABSTRACT

We propose a method for the reconstruction of an image, only partially observed through a linear integral operator. As such an inverse problem is ill-posed, prior information must be introduced. We consider the case of a compound Markov random field with a non-interacting line process. In order to maximise the posterior likelihood function, we propose an extension of the Graduated Non Convexity principle pioneered by Blake & Zisserman which allows its use for ill-posed linear inverse problems. We discuss the role of the observation scale and some aspects of the implemented algorithm. Finally, we present an application of the method to a diffraction tomography imaging problem.

1. INTRODUCTION

In a wide variety of physical experiments a two-dimensional real object (an "image") is only partially observed through a linear integral operator (*e.g.*, Fourier transform on a sparse set, or convolution over a large window). Recovering the object from an incomplete set of noisy data is an ill-posed inverse problem. Such problems arise in various areas such as diffraction tomography, radio interferometry, crystallography, seismic exploration.

In a standard regularisation framework, the quality of the solution closely depends on the pertinence of prior information. In the following we assume that the unknown object is partitioned into nearly homogeneous regions. Since the pioneering work of Geman & Geman [2], such prior proved to be well accounted for by compound Markov random fields (MRFs). When the incorporated boolean line process is non-interacting, it is called a weak-membrane model by Blake & Zisserman [1].

The calculation of the MAP estimate bears all the difficulty of the problem: the energy function to be minimised generally exhibits numerous local minima. Simulated Annealing (SA) provides theoretical weak convergence towards the global minimum, and such probabilistic relaxation schemes have raised considerable attention since [2]. Unfortunately, SA is numerically intractable in our context (Section 2). This is the main reason why compound MRFs have only been used in restricted situations in similar contexts [3] [4].

In order to calculate a weak membrane MAP solution, we focus on a deterministic relaxation approach —Graduated Non Convexity (GNC)—initially developed by Blake & Zisserman for the purpose of image segmentation and noise cancellation [1] (Section 3). It generally finds nearly optimal minimum, and at least it avoids the many shallow local minima. However its generalization to inverse problems is not straightforward. In Section 4 we propose a modified GNC optimisation technique that manages any linear ill-conditioned observation operator.

For a non-linear estimator, invariance of the solution from the observation scale is not trivialy satisfied: we present the conditions for the scale-invariance of the weak-membrane estimator and of GNC relaxation (Section 5).

Tomographic imaging with diffracting sources is a typical case of ill-posed linear (or linearised) inverse problems where only restricted priors have been used (Section 6). Concluding remarks are presented in Section 7.

2. NEIGHBOURHOOD PROBLEM

The observation model is classically given by a linear integral equation relating noisy data y to the original image x. The discrete form is:

$$y = \mathcal{A}x + n, \tag{1}$$

where $\mathcal{A}$ is a known linear operator which stands for the experimental setting and n is the additive noise, assumed white and Gaussian. The reconstruction of

0-7803-1775-0/94 $3.00 © 1994 IEEE

x from y presents an intrinsic difficulty: both the ill-posedness of the continuous observation model and the physical data acquisition limitations make matrix $\mathcal{A}$ extremely ill-conditioned. Then satisfactory reconstruction of x needs some prior information. Markov modelling is an appealing way to introduce prior knowledge about x and further to provide Bayesian estimates of x such as the MAP.

When $\mathcal{A}$ is of large support, a new difficulty arises when a Markovian prior is used. Let $\mathcal{V}_i$ be the prior neighbourhood of pixel i. Let $\mathcal{A}_{x_i}$ denote the restriction of $\mathcal{A}$ to pixel i and $\mathcal{C}_i$ its support; let $\mathcal{A}^{y_j}$ have its range restricted to the data point y_j and let $\mathcal{R}_j$ be its support. Then the neighbours of i in the posterior distribution, say $\mathcal{V}_i^P$, are elements of

$$\mathcal{V}_i^P = (\bigcup_{j \in \mathcal{C}_i} \mathcal{R}_j) \bigcup \mathcal{V}_i. \qquad (2)$$

So, when the operator $\mathcal{A}$ has a large support, prior locality is lost in the posterior model. *E.g.*, for a Fourier transform, each $\mathcal{C}_i$ is the whole range of $\mathcal{A}$ and each $\mathcal{R}_j$ is the whole domain of $\mathcal{A}$, so in the posterior distribution the neighbourhood of each pixel is all the image.

SA [2] is based on local updates of each pixel i according to its posterior conditional distribution which involves only the neighbouring pixels of i in the posterior. The computational cost of each update obviously depends on the posterior neighbourhood size. It becomes enormous when the posterior neighborhood involves the whole image. In such conditions, SA is intractable and another optimisation technique should be seeked.

3. GNC AND THE WEAK-MEMBRANE MODEL

In some cases GNC is a very effective deterministic alternative to SA [5] which does not require the locality of the posterior distribution. Let $E(x/y)$ denote the energy corresponding to the posterior likelihood. The principle of the GNC consists in the following. A sequence of continuously derivable functions $(F_p)_{p \in N}$ is constructed. $(F_p)_{p \in N}$ approximate E at different levels as functions of a parameter p in such a way that for some p_0, F_{p_0} is convex and $F_p \to E$ as $p \to \infty$. F_{p_0} has a unique minimum, say $\hat{x}_0$, and each $\hat{x}_p$—a minimum of F_p—is obtained by local descent in the vicinity of $\hat{x}_{p-1}$. The relaxation is controlled by p. There is no proof of convergence of the final GNC solution $\hat{x}_\infty$ towards the global minimum. Nevertheless, its practical use is very satisfactory.

The main limitation of GNC originates in the necessity to construct a sequence $(F_p)_{p \in N}$ as just described, because this leads to some severe restrictions on what prior can be used. In particular, application of GNC to a model with an interacting line process is not straightforward; an attempt to carry it out can be found in [6], but it is done at the expence of numerical complexity. For a model with a non-interacting line process, the astuteness is that the regularisation term in the MAP energy function can be equivalently stated as a sum of only pair interaction functions, which subsequently can be approximated separatedly. But it is not always easy to find a convenient approximation of the interaction functions.

The weak membrane model was the first compound MRF optimised by GNC [1]. It corresponds to a locally Gaussian, non-stationnary first-order MRF with a boolean non-interacting line process. The relevant MAP energy function, after the elimination of the line process, is given by:

$$E(x/y) = \|\mathcal{A}x - y\|^2 + \sum_{\{i;j \in \mathcal{V}_i\}} \phi(x_i - x_j), \qquad (3)$$

$$\phi(t) = \lambda^2 t^2 I_{(-T,T)}(t) + \alpha I_{(-\infty,-T] \cup [T,+\infty)}(t); \qquad (4)$$

ϕ is the interaction function, t is a first-order difference, $T = \frac{\sqrt{\alpha}}{\lambda}$ and $I_{(a,b]}(t)$ is the indicator function for $(a, b]$.

Such prior locally smooths the image while it can preserve significant abrupt transitions. $|t| \geq T$ is equivalent to a discontinuity between the relevant pixels; for $|t| < T$, the pixels belong to the same homogeneous zone. T is the threshold of discontinuity detection: clearly the solution is not everywhere a continuous function of the data (for α, λ fixed), nor of the model parameters (α, λ) (for a given y).

4. CONVEX APPROXIMATION FOR AN ILL-POSED INVERSE PROBLEM

In order to find if F_{p_0} is a convex function, positive definiteness of its Hessian matrix $\mathcal{H}\{F_{p_0}(x/y)\}$ can be alternately checked. As the second derivative of ϕ is $\phi''(t) =: \lambda^2[I_{(-T,T)}(t) - \delta_{-T} - \delta_T]$, δ being the Dirac delta function, the first step is to approximate ϕ by a continuously derivable function ϕ_c, depending on a parameter c, and such that: $\forall t$, $\phi_c''(t) \geq -B(c)$ and $\lim_{c \to 0} B(c) = 0$, $B(c) > 0$. An easy approximation is given in [1]:

$$\begin{aligned} \phi_c(t) \quad = \quad & \lambda^2 t^2 I_{(-q,q)}(t) + \alpha I_{(-\infty,-r] \cup [r,+\infty)}(t) + \\ & [\alpha - \frac{1}{2} c(t-r)^2] I_{(-r,-q] \cup [q,r)}(t), \end{aligned} \qquad (5)$$

where $q = T(c/(c + 2\lambda^2))^{\frac{1}{2}}$, $r = T((c + 2\lambda^2)/c)^{\frac{1}{2}}$, and $B(c) = c$. Transitions between neighbouring pixels may be continuous ($|t| < q(c)$), discontinuous ($|t| \geq r(c)$), or

else undetermined. The final solution is reached when there is no more undetermined transitions.

If $\mathcal{A}$ is well-conditioned, the sought approximation sequence is $F_c = \|\mathcal{A}x - y\|^2 + \sum \phi_c(.)$, whose Hessian matrix satisfies $\forall x$, $\mathcal{H}\{F_c\} \geq 2\Re(\mathcal{A}^\dagger\mathcal{A}) - cQ^\top Q$; $\Re$ stands for real part, $\dagger$ - for hermitian, $\top$ - for transposed and Q is the first order difference operator. Let $[\mu_1, \ldots, \mu_n]$ be the eigenvalue system of $\Re(\mathcal{A}^\dagger\mathcal{A})$, then it is easy to check that

$$F_c \text{ is convex } \Longleftrightarrow c < c_0 = \frac{\mu_{min}}{4}. \quad (6)$$

This is a generalisation of the convexity condition furnished in [1], which is valid for any well-posed linear inverse problem.

The last relation is interesting mainly because it highlights what happens when $\mathcal{A}$ is singular, or nearly so. In this case $\mu_{min} = 0$, or at least $\mu_{min} \simeq 0$, the convexity condition is either never satisfied, or is satisfied for so small a value of c, that it is of no practical interest.

In order to guarantee the initial convexity for an ill-conditionned operator, we propose to append an auxiliary convex term, which we rapidly relax to zero afterwards. Since the initial concavity is characterised by $-\frac{1}{2}c_0t^2$, it is naturally compensated by a_0t^2 when $a_0 \geq \frac{1}{2}c_0$. An approximation of E finally reads as follows:

$$F_{a,c} = \|\mathcal{A}x-y\|^2 + \sum_{\{i,j\in\mathcal{V}_i\}} [\phi_c(x_i-x_j)+a(x_i-x_j)^2] \quad (7)$$

and F_{a_0,c_0} is convex. It is convenient to first relaxe a from a_0 to 0, c remaining constant (c_0), and afterwards to relax c alone.

5. SCALE INVARIANT SOLUTION

A physically meaningful requirement for a real-world linear inverse problem is for the solution to be scale-invariant [7]: if an object x is observed simultaneously at two different scales, $y = \mathcal{A}x + n$ and $y^s = \mathcal{A}^s x + n^s$, the last being referred to as "the scaled observation model", and defined by $y^s = sy$ and $\mathcal{A}^s = s\mathcal{A}$, where s is the scale factor, both observations should yield the same estimate $\hat{x}$. For a non-linear estimator and for any given s, the prior model parameters corresponding to the scaled observation ought to be corrected appropriately in function of s. For some priors, this is not always possible. For the weak-membrane posterior energy function, a sufficient condition for the scale-invariance is:

$$E^s(x/y^s; \alpha^s, \lambda^s, \mathcal{A}^s) = s^2 E(x/y; \alpha, \lambda, \mathcal{A}). \quad (8)$$

Obviously E^s and E have the same set of minima. It is easy to deduce that Eq. (8) is assured when $\lambda^s = s\lambda$ and $\alpha^s = s^2\alpha$. The correction comes from the multiplication by s of the observation noise.

The scale also plays an important role in the GNC relaxation schedule itself. Let $F^s_{a^s,c^s}$ be the approximation of E^s for some relaxation parameter values a^s and c^s and $\hat{x}^s_{a^s,c^s}$ its minimum obtained by GNC with a continuous relaxation; let $\hat{x}_{a,c}$ be the minimum of $F_{a,c}$ obtained analogously. Then: $\hat{x}_{a,c} = \hat{x}^s_{a^s,c^s}$ provided that $a^s = s^2a$ and $c^s = s^2c$. Hence the relation between the evolutions of the scaled estimator solution and of the non-scaled one, is simply a dilatation.

In practice, a valid way to get rid of scale considerations is to normalise $\mathcal{A}$ and y.

6. A 2D DIFFRACTION TOMOGRAPHY APPLICATION

In transmission diffraction tomography, the cross-section of an object has to be recovered from some transmitted diffraction field data. In our example, eight sets of measures are collected by illumination of the object from eight different directions in the range of $[0, \frac{7}{8}\pi]$ *rad*. Under the standard Born approximation, the observation model linearly relates the 1D Fourier transform of each set of measures to the 2D Fourier transform of the object, calculated along the relevant half-circle in the frequency domain [8]. Data point repartition in the Fourier domain (Fig. 1(a)) is rather irregular and the inverse problem is very ill-posed. Noisy data were simulated with a signal to noise ratio (SNR) of 20 dB (see Fig. 1(c)).

Because of the support of A, non-convex regularisation is not standard. Instead, the following methods are currently used [9]:

- Analytic methods using Fourier transformation are computationally inexpensive, but they break down in the presence of noise (see Fig. 1(d)).
- Quadratic regularisation is also inexpensive, but it is well-known to allow noise cancellation only, at the expense of oversmoothing the object (Fig. 1(e)).
- Maximum Entropy reconstruction is very efficient when the image has a spiky appearance over a homogeneous background, but it is misled by flat regions over the background (Fig. 1(f)).

In contrast, the weak membrane solution presented in Fig. 1(g) provides a very good estimate. Discontinuities are neat and homogeneous zones are well smoothed.

We first normalised $\mathcal{A}$ and y so that $\|\mathcal{A}\|^2_F = N$, where F stands for Frobenius norm and N is the number of pixels in x. In Eq. (7) a degree of freedom subsists in the choice of (c_0, a_0). Numerically, the fi-

nal solution does not appear to be very sensitive to the exact values of (c_0, a_0); we worked with $(0.2, 0.1)$. The reconstruction was computed for $\alpha = 1$ and $\lambda = 2.3$; the choice of (α, λ) is independant of the data. We relaxed a in two steps. Then c was relaxed similarly to [1].

We compared solutions computed with an "almost continuous" relaxation (a and c varying extremely slowly) and solutions obtained using different relaxation schedules. It may happen that the "continuous" relaxation and a three-step relaxation $((a_0, c_0), (0, c_0), (0, +\infty))$ furnish the same solution: in these cases the intermediate solution $\hat{x}_{0,c_0}$ is alredy in the vicinity of the final GNC solution. Unfortunately, there is no way to know, at a given step, if the actual intermediate solution is or is not in the final solution valley. Nevertheless, a more dense relaxation schedule would permit to obtain the best GNC-performances more often.

7. CONCLUDING REMARKS

The proposed extension allows proper application of the GNC to Markovian reconstruction of ill-posed inverse problems.

When the prior is suited, the quality of the reconstructed images is highly satisfactory while the computational cost remains moderate.

The advantages of compound MRFs are well-known in image processing. But their use in ill-posed inverse problems with large-support operators strongly depends on the possibility to implement a non-local optimisation technique.

In spite of the limitations that GNC imposes to the prior, it is also conceivable for other compound MRFs, in order to solve ill-posed and large-support-operator inverse problems.

8. REFERENCES

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

[2] Geman S. & Geman D., "Stochastic relaxation, Gibbs distribution, and Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721—741, 1984.

[3] Doerschuk P., "Signal reconstruction from Fourier transform magnitude using Markov random fields in X-ray crystallography," *Proc. of ICASSP*, vol. IV, pp. 141-144, 1992

[4] Dinten J.-M., "Tomographic reconstruction of axially symmetric objects: regularization by a Markovian modelisation," *Proc. of the Int. Conf. on Pattern Recog.*, pp. 153-158, 1990.

[5] 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.

[6] Rangarajan A. & R. Chellappa, "Generalized Graduated Non-Convexity Algorithm for Maximum A Posteriori Image Estimation," USC-SEPIA Report No. 149, December 1989.

[7] A. Mohammad-Djafari and J. Idier, "Scale invariant Bayesian estimators for linear inverse problems," *ISBA meeting*, Aug. 1993, San-Fransisco, to appear in *American Mathematical Society*

[8] Kak A. & M. Slaney, *Principles of Computerized Tomographic Imaging*. IEEE Press, 1987.

[9] Mohammad-Djafari A. & Demoment G, "Maximum Entropy Image Reconstruction in X-Ray and Diffraction Tomography," *IEEE Trans. on Medical Imaging*, vol. 7, pp. 345—353, 1988.

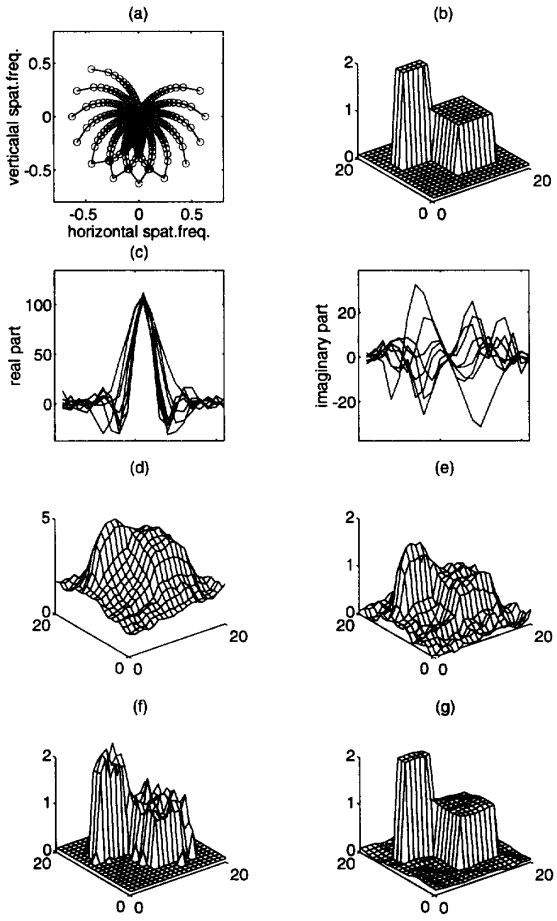

Figure 1: **Diffraction tomography synthetic example** **(a)** 8 x 21 data points in the Fourier domain; each half-circle corresponds to one projection. **(b)** Original image of size 21 x 21. **(c)** Data, 20 dB signal-to-noise ratio.
Reconstruction results: **(d)** Inverse Fourier transform. **(e)** Quadratic regularization. **(f)** Maximum Entropy. **(g)** The proposed Markovian reconstruction.