Variational Methods in Image Restoration [1]

Dominikus Noll

LAO 96 – 11

Septembre 1996

LABORATOIRE APPROXIMATION ET OPTIMISATION
Université Paul Sabatier, 118 route de Narbonne – 31062 Toulouse cedex – France
Tél. 61.55.67.78 E-Mail : lao.at.cict.fr

[1] To appear in Proceedings of the 8th French-German Conference in Optimization

1 Introduction

Optimization has been playing a prominent role in any of the modern image processing techniques, and variational models are used in such different fields as image segmentation and edge detection, compression, and sharpening of blurred images or in noise reduction. The present paper is mainly concerned with the restoration problem, which arises when images have been severely degraded during recording.

Restoring a blurred image may be a crucial part of a successful processing when sharp contrasts have been smoothed away by the recording channel. As a consequence, sharpening has to be achieved before other techniques like segmentation, edge detection, or noise suppression could hopefully be brought into play.

One of the basic approaches towards enhancing degraded images is by direct filtering, and there exists a vast field of recipes, mostly heuristic, for designing such filters, cf. Jähne [14]. The overall advantage of these direct methods is their computational speed and the fact that they apply even to very large images found in many applications, where some of the more sophisticated inverse methods to be discussed here may be too slow. For instance, the maximum entropy restoration method to be discussed here works fairly well for an image up to a size of say 500×500 pixels, but does not seem to be attractive for images of size 1000×1000 or larger, unless some speeding up e.g. by parallelizing is obtained. Nevertheless, the field for applications of variational type restoration techniques is fairly wide, and they are used e.g. for photographic or satellite images, in tomography, but also for enhancing pictures obtained from electron microscopes, or for deblurring data from radar astronomy.

2 A recording model

Designing a sophisticated restoration technique requires modeling the recording channel. It is a widely accepted rule in the engineering literature that devoid some more detailed information on the nature of the degradation, the transmission may be represented by a linear model

$$v(x) = \int_\Omega q(x,y)u(y)\,dy + e(x), \qquad x \in \Omega \tag{1}$$

where $u(x) \geq 0$ represents the relative gray levels of the unknown original image over a region $\Omega \subset \mathbf{R}^2$, $v(x) \geq 0$ represents the relative gray levels of the recorded dirty image, the integral term represents signal dependent blurring, and $e(x)$ represents the signal independent (white or colored) channel noise. Usually, the channel neither absorbs nor generates optical energy, that is $\int_\Omega q(x,y)\,dy = 1$ for every $x \in \Omega$, and the noise process will then have zero mean. In applications, the signal dependent degradation may represent phenomena like optical system aberrations, defocussing, atmospheric turbulence, or relative object and camera motions, while the additive noise could represent a variety of noise sources such as film grain noise, scattering, random channel noise, but also nonphysical noise due to digitization or measurement errors, or uncertainties due to physical phenomena which have not been included among the channel model.

As is often the case, the underlying transmission is non-linear in nature, but exploiting the concrete physical model allows for retaining a linear model (1). As a typical example consider photographic images, where the relation between the original and the image planes, given by the characteristics of the exposure mapping, is in fact non-linear. However, when the exposure mapping is known in advance, it is possible to perform what is called an a priori gray level correction, which leads to a linear model (1). An alternative often used to justify a linear recording model

is introducing a logarithmic scale. See [23] for a discussion of both procedures for photographic images.

Nevertheless, in the context of mathematical morphology (cf. [1]), a variety of nonlinear approaches to describe blurring phenomena have been proposed. They expand on the diffusion process as the physical prototype of any degradation, replacing it with other dynamic equations (see [1] for an overview). While these approaches are mathematically appealing, it may be doubted whether identifying the correct dynamic equation describing a given blur is possible if restoration, as is often the case, has to start out with a sample v_{ij} of the degraded image $v(x)$ as the only source of information.

Accepting the linear recording model (1) for the time being, observe that in a practical situation the dirty image $v(x)$ would be known only partially, say sampled at the nodes x_{ij} of a grid over the region $\Omega \subset \mathbf{R}^2$. It is now a question of convention whether we choose the continuous model (1) or prefer to replace it with a discretized version

$$v_{ij} = \sum_{k=1}^{N} \sum_{\ell=1}^{M} q_{ijk\ell}\, u_{k\ell} + e_{ij}, \qquad i = 1, \ldots, N,\ j = 1, \ldots, M. \tag{2}$$

Here $u_{ij} = u(x_{ij})$ and $v_{ij} = v(x_{ij})$ represent the values of u, v at say the nodes x_{ij} of the grid with fixed mesh over the (usually rectangular) region Ω, $q_{ijk\ell} = h^2 q(x_{ij}; x_{k\ell})$, and the $e_{ij} = e(x_{ij})$ represent the noise terms. Our discussion will be using both types of approaches, but it will be clear how to make the transition from the continuous to the discrete model, and vice versa.

Restoring u_{ij} given v_{ij} requires estimating what we call the the model parameters: The blurring $q_{ijk\ell}$, and the statistical parameters of the noise process $e(x)$, represented by its covariance matrix $\mathrm{cov}(e)$. Assuming that noise is stationary over large parts of the region, it is often possible to estimate $\mathrm{cov}(e)$ by inspecting parts of the dirty image $v(x)$ with a homogeneous gray tone, where covariance of v should in fact represent noise covariance. The difficult task remaining is to identify the blurring parameters $q_{ijk\ell}$.

Spatial blurring is usually of local character, i.e. $q(x, y) \to 0$ rapidly as $|x - y| \to \infty$. In the discretized model (3) it is therefore typical to work with small support: $q_{ijk\ell} = 0$ if $|i - k| > p$ or $|j - \ell| > p$ for some $p \ll N, M$. Moreover, if blurring is spatially invariant, a hypothesis which in practice is met at least over large parts of the region, then $q_{ijk\ell} =: q_{i-k,j-\ell}$, reducing the degrees of freedom to $(2p+1)^2$, and the linear model (1) resp. (2) takes the convolutional form $v = q * u + e$, or more explicitly,

$$v_{ij} = \sum_{r=-p}^{p} \sum_{s=-p}^{p} q_{rs}\, u_{i-rj-s} + e_{ij}, \qquad i = 1, \ldots, N,\ j = 1, \ldots, M. \tag{3}$$

If blurring is to some extent symmetric, the number of degrees of freedom may be further reduced, hopefully to a manageable size. Approaches to solving the blind inversion problem (3) may now be roughly grouped into two types: techniques for random input signals, and methods using Kalman filtering.

3 Blind deconvolution

Kalman filtering is based on the hypothesis that the unknown signal and noise processes are realizations of mutually independent Gaussian random variables, and that moreover, the true image

$u(x)$ may be represented as a $2D$ autoregressive model of the form

$$u_{ij} = \sum_{(k,\ell)\in\mathcal{P}_{ij}} c_{ijk\ell} u_{k\ell} + w_{ij} \tag{4}$$

with an unknown transition filter c, and a driving random process w accounting for the uncertainty in the state model. Notice that (4) requires introducing a direction of recursion (a time axis) into the image, the usual one being left-to-right, top-to-bottom, and the indices $\mathcal{P}_{ij}$ represent the past of pixel (i,j).

Although the Kalman filter is linear, computing the filter gain is nonlinear and at each iteration $k = 1, \ldots, NM$ requires solving a matrix Ricatti equation with computational cost of order $\mathcal{O}(T_k^2)$, where T_k is the dimension of the state at iteration k. Since the state should in principle cover the past information, which would be of order k, i.e. $T_k = \mathcal{O}(k)$, the total cost of a naive implementation of the classical Kalman filter is of order $\mathcal{O}((NM)^3)$. Practical implementations therefore need to limit the memory of the state to a constant $T_k = \mathcal{O}(1)$, reducing the cost and storage to a manageable size. In the same vein, the transition filter c is usually chosen to be spatially invariant, a hypothesis which reduces the degrees of freedom, but contradicts the true nature of the recorded information. Engineers try to overcome this by embedding the image into a properly chosen larger frame. See for instance [2], where an approximative model (called ROMKF) is discussed and compared to other limited memory models.

While Kalman filtering – due to the dynamic (4) – requires introducing artificial parameters, $c_{ijk\ell}$ and the statistics of the driving noise w_{ij}, and a time axis not inherent to the image, its attractivity lies in the fact that estimating even non stationary blurring parameters $q_{ijk\ell}$ is possible (cf. [2]). Due to the assumption of Gaussian random processes $u(x)$, $e(x)$, this may be done by maximum likelihood parameter identification. The likelihood function is evaluated through intermediate steps of the Kalman filtering process, so no analytic form of the likelihood function is available, and gradients are hard to calculate, which in practice limits the use of the procedure. We refer to [15], where an approach based on the expectation maximization algorithm (EM) is presented: Using the $2D$ FFT, and a standard approximation of block Toeplitz matrices by block circulants, the optimization of the likelihood function may be carried out in the frequency domain, and the restoration may finally be obtained as the result of a Wiener filter with the dirty image as input.

As the Kalman filter is slow even with limited memory, and evaluating the likelihood objective is costly, it seems natural to investigate alternative strategies based on the following idea: Suppose that we have a tool which, for a fixed blur q, allows for a fast inversion of (3). Using the latter as a black box, blur identification could then be performed in some a posteriori procedure based on objectives other than the likelihood function, or even by using the eye for judgment. This idea will be further pursued in Sections 4 – 7.

A second type of methods used to solve the blind deconvolution problem which has first been popularized in geophysics is designed for white noise input signals. In geophysical exploitation, seismic data were interpreted as filtered versions of a white noise driving signal which needed to be restored. In this situation, optimal deconvolution techniques may in fact be developed. Donoho [7] compares several techniques like maximum Fisher information deconvolution, maximum curtosis deconvolution (using cummulants), and minimum entropy deconvolution. If the unknown input signal is a non Gaussian white noise, he is able to show that minimum entropy deconvolution is asymptotically optimal. While these techniques, originally designed for deconvolving time series, may in principle be generalized to the 2D imaging situation, there remain some obstacles due to the fact that (*i*) the procedure is sensitive to noise and fails if the noise contribution is sizable, and more seriously, (*ii*) the analysis in [7] is based on a white noise input signal, a hypothesis which is not met in our situation.

A more recent approach (also using cummulants) is presented in (cf. [9, 10]). The authors include the case of measurement noise, and they allow for more general types of input signals, which are now required to be a discrete stationary and ergodic. They report that stationarity of the input signal, which is clearly contradictory to the situation we face in image restoration, does not play an important role in numerical experiments, though is essential for their convergence analysis.

While experiments reported in [10] are encouraging in the time series case, it may be doubted that the approach (taken in [7, 9, 10] and many other approaches cited there) will work well in the imaging context because of the following observation: The general idea is to design an inverse linear filter θ which annihilates the effect of q up to scaling and phase shift. However, while the filter q is typically local, its left inverse θ in general is not, contrary to the situation for time series when causal processes are considered. Therefore, when the support of θ is large, we can hardly expect the procedure to be efficient. A compromise might be to confine oneself to just estimating the parameters $q_{ijk\ell}$, leaving the actual inversion to some of the more robust methods to be discussed below.

4 Stabilizing inversion

Let us now address the problem of inverting the linear system (3) in the case where the system parameters, $q_{ijk\ell}$ and the noise covariance $C = \mathrm{cov}(e)$ are known. Notice that this is still an unstable problem, as we are trying to solve a Fredholm integral equation of the first kind (1) with a sampled and noisy right hand side. As is well known, the solution may be highly sensitive to even small perturbations of the right hand side, and as a rule, the problem is ill-posed and needs being stabilized.

A common way to achieve this is to consider a variational approach, that is to introduce a cost functional $\mathcal{I}(u)$, and to choose as a valid restoration the solution of the program

$$(P_{\mathrm{pen}}) \quad \begin{array}{ll} \text{minimize} & \mathcal{I}(u) + \alpha \left\| C^{-1} e \right\|^2 \\ \text{subject to} & e = q * u - v \end{array} \tag{5}$$

Here $\mathcal{I}(u)$ serves as a regularization term which keeps the possible restorations u from being highly irregular, while $\|C^{-1}e\|^2 = \left\|C^{-1}(q * u - v)\right\|^2$ accounts for the model (3). The penalty parameter α may be used to mediate between both criteria. Notice that the scheme (P_{pen}) could be used for both the continuous (1) or the discretized model (2), (3), but in both cases, the dirty signal v should be understood as sampled at finitely many nodes x_{ij}, so the terms $C^{-1}e$ and $q * u - v$ are in discretized form.

Most objectives $\mathcal{I}(u)$ found in practice are quadratic functionals: $\mathcal{I}(u) = ||B^{-1}u||_2^2$, including in particular what might be called the default choices $\mathcal{I}(u) = ||u||_2^2$, $\mathcal{I}(u) = ||\nabla u||_2^2$, or $\mathcal{I}(u) = ||\Delta u||_2^2$ and their discretized versions. This preference is explained by the fact that for quadratic objectives the necessary optimality conditions rsp. the Euler equation are linear, and restoration is obtained as a linear inverse filter, at least when the positivity constraint $u(x) \geq 0$ is ignored. For instance, the choice of the Laplacian $\mathcal{I}(u) = ||\Delta u||_2^2$ leads to the inverse Wiener filter proposed by Hunt [13] as early as 1976, whose speed is further increased by a clever use of the FFT.

Non quadratic objectives are also encountered in practice, for instance the Boltzmann Shannon entropy $\mathcal{I}(u) = \int_\Omega u(x) \log u(x)\, dx$, or cross-entropy functionals including more specific prior information, the Burg entropy $\mathcal{I}(u) = -\int_\Omega \log u(x)\, dx$, or the Fisher's information $\mathcal{I}(u) = \int_\Omega \frac{|\nabla u(x)|^2}{u(x)}\, dx$. Similarly, Csiszar's class of statistical distances are used.

A general view for understanding the choice of $\mathcal{I}(u)$ is provided by the principle of maximum entropy in the mean MEM, as has recently been demonstrated in [16] (see also [17] for this). Let us briefly present the idea.

Returning to the model (3), and assuming that the image and noise contributions are realizations of independent random processes, the joint distribution of u and e in (3) is of the form $d\mu'(u,e) = d\mu(u)d\nu(e)$ for certain probability measures μ and ν. Here, ν should reflect our prior knowledge about what would be channel noise or measurement errors, responsible for the noise term $e(x)$, while μ should incorporate our prior knowledge about the expected signal u. The maximum entropy on the mean approach is now to find the joint probability density $p(u,e)$ which solves the program

$$(MEM)\qquad \begin{array}{ll} \text{maximize} & H[p] = -\iint p(u,e)\log p(u,e)\,d\mu(u)\,d\nu(e) \\ \text{subject to} & E_p(q*(u)+e) = v \\ & \iint p(u,e)\,d\mu(u)d\nu(e) = 1,\ p(u,e)\geq 0 \end{array}$$

where $E_p(\cdot)$ denotes expectation with respect to the probability measure $pd\mu d\nu$. Following [16] (MEM) is equivalent to a finite dimensional convex program, which is used to calculate the restoration:

$$(\mathcal{P})\qquad \begin{array}{ll} \text{minimize} & F_\mu^*(u) + F_\nu^*(e) \\ \text{subject to} & e = q*u - v \end{array}$$

Here F_μ^* denotes the Fenchel conjugate of $F_\mu(w) = \log\int \exp\langle u,w\rangle\, d\mu(u)$, and similarly for F_ν^*. The authors of [16] obtain the following result.

Proposition 4.1 *Suppose (1) the prior probabilities μ, ν are Gaussian with covariance matrices B, C. Then the Fenchel conjugates are quadratic functionals $F_\mu^*(u) = \mathcal{I}(u) = \frac{1}{2}\|B^{-1}u\|^2$, and $F_\nu^*(e) = \frac{1}{2}\|C^{-1}e\|^2$, and program $(\mathcal{P})$ coincides with the penalty model (P_{pen}) with $\alpha = 1$.*

On the other hand, (2) if the prior ν is Gaussian with covariance matrix C, but the prior μ is a Poisson measure of the form

$$\mu = \prod_{i=1}^{N}\prod_{j=1}^{M}\mu_{ij}, \qquad \mu_{ij} = \exp(-u_{0ij})\sum_{k=0}^{\infty}\frac{u_{0ij}^k}{k!}\,\delta_k$$

then the Fenchel conjugate $F_\mu^*(u)$ coincides with the discrete form of the Boltzmann-Shannon entropy resp. the cross entropy functional:

$$F_\mu^*(u) = \mathcal{I}(u) = \sum_{i=1}^{N}\sum_{j=1}^{M}\left(u_{ij}\log\frac{u_{ij}}{u_{0ij}} + (u_{0ij} - u_{ij})\right),$$

and hence $(\mathcal{P})$ is equivalent to the maximum entropy resp. minimum cross entropy form of model (P_{pen}) with $\alpha = 1$.

The message of MEM is that any choice of a functional $\mathcal{I}(u)$ sould be understood as incorporating a priori information about the expected restoration u. For instance, choosing a Gaussian prior μ reflects the fact that the probability of high frequencies in the restored images is very low, meaning that in a so restored image high impulses are avoided. On the other hand choosing a Poisson

distribution μ means that we expect an image consisting of several high energy impulses, as found for instance in spectroscopic images. In the same vein, prior knowledge about edges in an image $u(x)$ might be incorporated in the prior μ, although, as a rule, this is more conveniently done by an appropriate modeling of the cost functional $\mathcal{I}(u)$.

Similarly, entropy type terms controlling the noise contribution have been proposed e.g. by Frieden [8], which may now be understood as a different bias about the type of noise to be expected. In any case, an important practical aspect of MEM) is that it provides a default choice ($\alpha = 1$) for the penalty constant in (5).

5 Tolerance model

There is yet another model which may replace the penalty objective from (5), and which we shall call the tolerance approach. To obtain this, notice that the process $e' = C^{-1}e$ is white noise with zero mean and unit variance. Hence the strong law of large numbers gives the estimate

$$||C^{-1}(q * u - v)||_2^2 = ||C^{-1}e||_2^2 = \sum_{i=1}^{N}\sum_{j=1}^{M} e_{ij}'^2 \approx NM \tag{6}$$

and this suggests that a valid restoration u should satisfy (6). This leads us to consider the following tolerance model, where we choose as a valid restoration the solution of

$$(P_{\text{tol}}) \qquad \begin{array}{ll} \text{minimize} & \mathcal{I}(u) = \displaystyle\int_\Omega h(u(x), \nabla u(x))\, dx \\ \text{subject to} & ||C^{-1}(q * u - v)||^2 \le \tau^2 \end{array} \tag{7}$$

Here (6) suggests the choice $\tau = NM$. Surprisingly, the following result shows that both approaches (5) and (7) are essentially equivalent. The proof is by inspecting the necessary optimality conditions for both programs (cf. [18]).

Proposition 5.1 *Suppose $\mathcal{I}(u)$ be convex. For a fixed $\tau > 0$ let $\bar{u}$ be the optimal solution of the tolerance model (P_{tol}), and suppose the inequality constraint is active at $\bar{u}$. Then there exists a choice $\alpha = \alpha(\tau)$ for which $\bar{u}$ is also the optimal solution of the penalty model (P_{pen}) with penalty constant $\alpha(\tau)$. Conversely, for a fixed $\alpha > 0$, let $\hat{u}$ denote the optimal solution of the corresponding penalty type program (P_{pen}). Then with $\tau = \tau(\alpha) := ||C^{-1}(q*\hat{u}-v)||$, $\hat{u}$ also solves the corresponding tolerance program (P_{tol}).*

We argue that the inequality constraint in (7) is *always* active in situations of practical relevance. Indeed, for objectives $\mathcal{I}(u)$ attaining their global minimum at the image without structure, $u_{ij}^0 = 1/(NM)$, u^0 could not be a feasible point for (P_{tol}), as $||C^{-1}(q*u^0-v)|| \le \tau$ is never satisfied for the relevant choices of τ. Take for instance the case of a white noise $e(x)$ with variance σ^2, and observe that $q * u^0 = u^0$. Then the feasibility of u^0 in (7) would imply

$$\text{Var}(v) = \frac{1}{NM}\sum_{ij}\left(\frac{1}{NM} - v_{ij}\right)^2 \le \frac{\sigma^2\tau}{NM}.$$

For the choice $\tau = NM$ suggested by (5), this meant that the variance of the dirty image was smaller than the noise variance, and therefore could be excluded.

Similar to what we have seen for the penalty approach in Section 4, the tolerance model comes along with a default choice ($\tau = NM$). Moreover, the law of large numbers shows that for large

$n = NM$, $\alpha(NM) \approx 1$ and $\tau(1) \approx NM$ for the functions $\alpha = \alpha(\tau)$ and $\tau = \tau(\alpha)$ in Proposition 5.1.

6 Maximum entropy deblurring

The Boltzmann-Shannon entropy functional $\mathcal{I}(u) = \int u(x) \log u(x) dx$, $u(x) \geq 0$ has been used to stabilize various types of inverse problems (5) or (7) arising in power spectrum estimation, moment problems, and also for deconvolution. Frieden [8] offers a theoretical justification for using entropy. Bayesian arguments in the spirit of [6] indicate that any prior estimate $u_0(x)$, if available, should be included among the model, leading to a minimum cross entropy functional

$$\mathcal{I}(u) = \int_\Omega u(x) \log \frac{u(x)}{u_0(x)} dx.$$

In absence of any prior information, the default choice $u_0 = 1$ is always possible, but an attractive alternative is to use the dirty image as a prior, as has been done e.g. in our experiments (see Figure 3 as compared to Figure 2).

The numerical treatment of the ME image restoration problem has been discussed by Frieden [8], Skilling et al. [6], Haralick et al. [12], and in Noll [20, 18]. While ref. [6] proposes as descent algorithm, the authors of [12] use a homotopy method based on a Lagrangian formulation. These relatively complicated approaches reflect the need for a feasible point method for this type of problems, where the objective $\mathcal{I}(u)$ behaves like a barrier functions at $u = 0$, and SQP methods are not applicable. Our own experiments reported in [20], which compare the cited methods with a convex duality approach, indicate that the latter should be given the preference, as the convex dual program is an unconstrained problem. As suggested by the barrier effect of $\mathcal{I}(u)$, interior point methods were also found to perform well for this kind of problem. See [21] for the polynomial bound, and for further references.

An idea developed in Noll [18] is to combine the maximum entropy model with linear or nonlinear filtering techniques. In fact, if the signal-to-noise level of the degraded image is low, the image should be preprocessed with a linear noise suppression filter or a nonlinear median filter, before ME inversion is performed. Clearly, prior filtering will introduce an additional blur into the image, so these filters have to be chosen carefully. A theoretical justification for this combined strategy is offered in [18]. In a practical restoration, an appropriate noise suppression filter is usually found experimentally. The simulations reported in [20] suggest that using this combined procedure is worthwhile for severely degraded images. Naturally, the same combined approach could be used for other variational models as for instance those to be discussed below.

7 Restoring with minimum information

Working with the Boltzmann-Shannon entropy is perhaps not the final answer, as intuition tells us that our objectives $\mathcal{I}(u)$ should control derivative values of the unknown $u(x)$. In fact, these are measuring gray levels of neighboring pixels, and are therefore intimately related to inherent structure of our image $u(x)$ such as edges and sharp contrasts. The default choices for such objectives, often used if no specific information is available, are the energy integral $\int |\nabla u|^2 dx$, or Hunt's functional $\int |\Delta u|^2 dx$. While the energy integral is appropriate for control type situations, where we wish to find minimum energy solutions, it is known to produce unsatisfactory results in image restoration problems.

An idea we propose here is to use a class of statistical distances, which in some sense may be understood to measuring the distance of a possible restoration $u(x)$ from a reference function $u_0(x)$. Again, in absence of any apriori information, u_0 is chosen as the constant $u_0 \equiv 1$. In the latter case, these objectives are defined as

$$h(u,p) = \begin{cases} u\psi(\frac{p}{u}) & \text{for } u > 0 \\ 0 & \text{for } u = 0,\ p = 0 \\ +\infty & \text{else} \end{cases} \tag{8}$$

where $u \in \mathbf{R}$, $p \in \mathbf{R}^d$ and where ψ is some convex function on $\mathbf{R}^d$. For instance, the choice $\psi(t) = |t|^2$, with $|\cdot|$ the euclidean norm, leads to the *Fisher information*. In fact, in dimension $d = 1$, we obtain the classical Fisher information measure for probability densities, while in higher dimensions, the functional so obtained is the trace of the Fisher information matrix.

If some apriori guess u_0 is known, we may replace u by u/u_0 to obtain an objective measuring in some sense the distance from u to u_0 and in the absence of any data, returns u_0 as the correct answer. For simplicity, in the following, we shall restrict our discussion to the case $u_0 \equiv 1$.

Among Csiszar's call of statistical distances, the variational problem associated with Fisher's information is of particular interest and in dimension $d = 1$ has been discussed in [3]. Numerical results have been obtained in [4] and [5]. In particular, the important case of trigonometric weight functions was considered in [5], and a fast solver based on a heuristic was proposed and tested.

Using Fisher's information in the context of inference problems was proposed by Silver [22] in 1992. Since then, numerical results have been obtained for power spectrum estimation of time series and for moment problems [4], showing that the Fisher information often outperforms the maximum entropy method. For a deconvolution problem, this is demonstrated by the experiments presented here (see Section 8). These results strongly motivate the use of Fisher's information for real life deconvolution problems (1), (3) arising in image restoration.

While the numerical techniques presented in [4, 5] do not extend beyond the case of dimension $d = 1$, a numerical approach for higher dimensions was proposed in Noll [19]. Based on duality arguments, we showed how to transform the Fisher restoration program into an eigenvalue optimization problem for a second order elliptic PDE. When discretized via finite elements or finite differences, this problem could then be solved using semidefinite programming.

While the approach in [19] is technical, we presently show how a semidefinite program could be derived directly from the primal Fisher program, though at the cost of a slightly larger complexity.

Consider the discretized Fisher program in dimension d, where Ω has been discretized by a grid Ω_h of mesh size $h > 0$:

$$(P_{\text{tol}}) \qquad \begin{array}{ll} \text{minimize} & \displaystyle\sum_{i=1}^{n} \frac{|D_i u|^2}{u_i} \\ \text{subject to} & u_i \geq 0,\ \dfrac{1}{h^d} \displaystyle\sum_{i=1}^{n} u_i = 1, \\ & |Qu - v| \leq \epsilon \end{array}$$

Here $u = (u_i) \in \mathbf{R}^n$ denotes the unknown image sampled at the nodes x_i of Ω_h, $v = (v_j) \in \mathbf{R}^m$ denotes the observed dirty image, possibly sampled on a subset of the grid Ω_h. Here $Q : \mathbf{R}^n \to \mathbf{R}^m$ represents the discrete blurring operator, and for simplicity we have chosen the case of a white noise with variance $\sigma^2 > 0$. In this case the default choice (6) gives $\epsilon = \sigma\sqrt{n}$. The linear operators D_i mapping into $\mathbf{R}^d$ stand for the discretized form of the gradient at position i. In dimension $d = 1$,

$D_i u = u_{i+1} - u_i$ when a forward difference is used, and in dimension $d = 2$, and with the standard bijection $i \to (k, \ell)$, $u_i = u_{k\ell}$, $D_i = D_{k\ell}$ is $D_{k\ell}u = (u_{k+1,\ell} - u_{k\ell}, u_{k,\ell+1} - u_{k\ell})$, where again forward differences are used.

Let us show how program (P_{tol}) may be transformed to a semidefinite program

$$(SP) \qquad \begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & F_0 + \sum_{i=1}^{r} x_i F_i \geq 0, \end{array}$$

where $c, x \in \mathbf{R}^r$ and the F_i are symmetric matrices of size $s \times s$. We shall see that the size of the (SP) obtained from (P_{tol}) is $r = \mathcal{O}(n+m)$, $s = \mathcal{O}(n+m)$. This should be compared to the form we obtained in [19] via duality, which gave the slightly better complexity $r = \mathcal{O}(m)$, $s = \mathcal{O}(n)$.

First observe how the discrete Fisher objective may be transformed to the semidefinite form. Introduce dummy variables t_i, $i = 1, \ldots, n$, satisfying $|D_i u|^2 / u_i \leq t_i$. Then minimizing the Fisher objective is equivalent to minimizing $\sum_{i=1}^{n} t_i$ subject to the above constraints which taking into account that $u_i \geq 0$ may be transformed to a set of $(d+1) \times (d+1)$ matrix inequalities. For $d = 1$ this takes the form

$$\begin{pmatrix} t_i & D_i u \\ D_i u & u_i \end{pmatrix} = t_i \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + u_i \begin{pmatrix} 0 & -1 \\ -1 & 1 \end{pmatrix} + u_{i+1} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \geq 0, \quad i = 1, \ldots n-1 \quad (9)$$

Still in the case of dimension $d = 1$, define block diagonal matrices $\tilde{F}_i, \tilde{G}_i$ of size $2(n-1) \times 2(n-1)$ as follows:

$$\tilde{F}_i = \begin{pmatrix} C_{i1} & & \\ & \ddots & \\ & & C_{i,n-1} \end{pmatrix}, \qquad \tilde{G}_i = \begin{pmatrix} D_{i1} & & \\ & \ddots & \\ & & D_{i,n-1} \end{pmatrix}$$

with

$$C_{ij} = \begin{pmatrix} \delta_{ij} & 0 \\ 0 & 0 \end{pmatrix}, \quad D_{ii} = \begin{pmatrix} 0 & -1 \\ -1 & 1 \end{pmatrix}, \quad D_{i,i-1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad D_{ij} = 0 \text{ otherwise.}$$

(δ_{ij} = Kronecker symbol). Then (9) translates into the semidefinite inequality

$$\sum_{i=1}^{n-1} t_i \tilde{F}_i + \sum_{i=1}^{n} u_i \tilde{G}_i \geq 0.$$

A similar pattern is obtained in dimension $d = 2$, with C_{ij} and D_{ij} now of size 3×3, and $\tilde{F}_{ij}, \tilde{G}_{ij}$ of size $3(n - N - M + 1) \times 3(n - N - M + 1)$, if $n = NM$, and the discretization (3) is used.

Treating the constraint $|Qu - v| \leq \epsilon$ is standard. Writing $w = Qu - v$, it becomes $w^T w \leq \epsilon^2$, which is equivalent to the $(m+1) \times (m+1)$ matrix inequality

$$\begin{pmatrix} \epsilon^2 I & w \\ w^T & 1 \end{pmatrix} = \begin{pmatrix} \epsilon^2 I & 0 \\ 0 & 1 \end{pmatrix} + \sum_{j=1}^{m} w_j \begin{pmatrix} \mathcal{O} & e_j \\ e_j^T & 0 \end{pmatrix} \geq 0$$

($e_i = i$th unit vector). On substituting $w = Qu - v$ back, we obtain the semidefinite type inequality

$$\begin{pmatrix} \epsilon^2 I & -v \\ -v^T & 1 \end{pmatrix} + \sum_{i=1}^{n} u_i \begin{pmatrix} \mathcal{O} & q_i \\ q_i^T & 0 \end{pmatrix} \geq 0 \quad (10)$$

where $Q = (q_{ij})$, and q_i is the ith column of Q. Writing (10) as $\hat{F}_0 + \sum_{i=1}^{n} u_i \hat{G}_i \geq 0$, we finally get the following inequality ranging in the (SP) derived from (P_{pen}):

$$\sum_{i=1}^{n-1} t_i F_i + \sum_{i=1}^{n} u_i G_i \geq 0$$

where F_i, G_i are block diagonal of the form

$$F_i = \begin{pmatrix} \tilde{F}_i & & & \\ & \hat{F}_i & & \\ & & \mathcal{O} & \\ & & & \mathcal{O} \end{pmatrix} \quad G_i = \begin{pmatrix} \tilde{G}_i & & & \\ & \hat{G}_i & & \\ & & H & \\ & & & K \end{pmatrix} \quad \text{with } H = \begin{pmatrix} t_1 & & \\ & \ddots & \\ & & t_n \end{pmatrix}, \; K = \begin{pmatrix} u_1 & & \\ & \ddots & \\ & & u_n \end{pmatrix}$$

which for $d = 1$ is of size $2(n-1) + m + 1 + 2n = 4n + m - 1$.

Notice that the sparsity pattern of $\hat{G}_i$ in (10) is not block diagonal. Also recall that the multipliers Z arising in the convex programming duality for (SP) (see e.g. [24]) have the same block diagonal structure as the constraint matrices F_i, G_i. It is then clear that the blocks $\hat{G}_i$ are the only ones causing a fill-in of a relevant size in their corresponding multiplier block $\hat{Z}$. In toto, the multipliers Z have $\mathcal{O}(n + m^2)$ nonzero entries, which represents the problem size we face when solving (SP). An interesting feature is that the block structure of $\hat{G}_i$, and therefore of the corresponding $\hat{Z}$ is independent of the problem dimension d.

8 Experiments

Our experiments displayed in Figures 1 to 8 show a deconvolution problem in dimension $d = 1$, where we compared the Fisher based restoration with other types of objectives. For medium size problems up to several hundred variables, the dual approach of [19] and the primal model presented here work equally well. A third possibility for solving (P_{tol}), then consists in transforming it via the substitution $u(x) = p(x)^2$. This leads to a nonconvex problem with a particularly nice convex objective (see [19]), and solving via sequential quadratic programming produces quite satisfactory results.

The purpose of the present experiments is to emphasize the merits of the Fisher objectives as compared to other functionals used in the field. While Figure 1 shows the problem setting, Figure 2 compares the Fisher restoration with maximum entropy. The latter obviously has difficulties, but as seen in Figure 3 catches up a little with Fisher if a minimum cross entropy functional is used, with the dirty signal used as a prior.

Figure 4 shows a restoration based on Burg's entropy, which is not competitive here. Figure 5 gives the result using the energy functional $\mathcal{I}(u) = \int |\nabla u|^2 dx$, which as expected is too smooth. The drawbacks of the energy functional in the context of image restoration being known, various remedies have been proposed. An idea in the spirit of Geman et al. [11] is to combine restoration with edge detection, which in the particularly simple case of dimension $d = 1$ could be done by minimizing a functional of the form

$$f(u, a) = \sum_{i=1}^{n-1} a_i (u_{i+1} - u_i)^2 + C \sum_{i=1}^{n-1} (1 - a_i)$$

subject to the constraint $|Qu - v| \leq \epsilon$. Here the binary variables $a_i \in \{0, 1\}$ serve as switches for the terms $(u_{i+1} - u_i)^2$ penalizing a strong oscillation at i. The state $a_i = 1$ corresponds to the

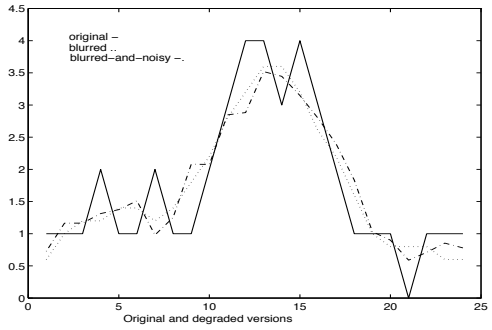

Figure 1.

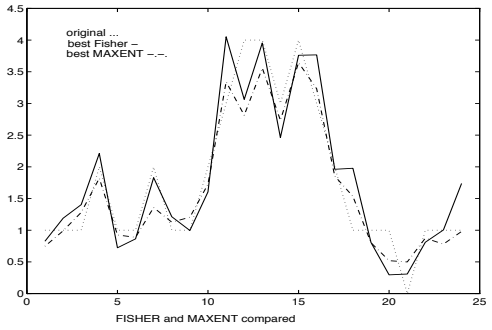

Figure 2.

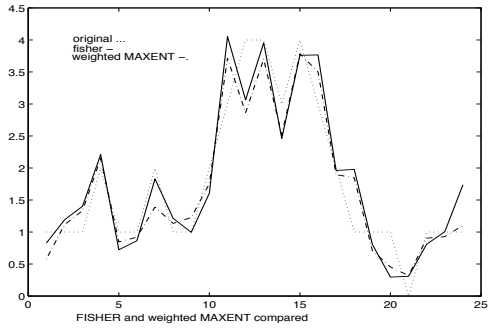

Figure 3.

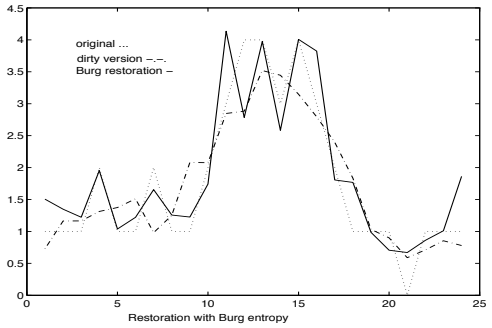

Figure 4.

switch being on, so that $(u_{i+1} - u_i)^2$ is fully penalized. The state $a_i = 0$ corresponds to the switch being turned off, which is meant to indicate an edge at position i, in which case a large jump at position i should no longer be penalized. Switching off is payed for through the second term, with C any constant ruling the trade-off between the two terms.

Similar models have been tested by several authors, who use simulated annealing type approaches to calculate solutions, with the expected poor results when images get larger. It appears to be a misconception to introduce integer variables into a problem which is continuous in nature, and we propose to replace switches by dimmers, which even allows to retain a convex program. In dimension $d = 1$, a quick shot would be to use an objective of the form

$$f(x, c) = \sum_{i=1}^{n-1} \frac{(u_{i+1} - u_i)^2}{c_i} + C \sum_{i=1}^{n-1} (c_i - 1)$$

under the same constraint aa above, where the dimmers $c_i \in [1, \infty)$ replace the switches used before. Here $c_i = 1$ means that a gap at position i is fully penalized, while larger values of c_i, gradually releasing penalization through the term $(u_{i+1} - u_i)^2$, indicate a higher probability for a structural gap at position i. Figure 6 shows a restoration obtained by this approach, with the stars displaying the states of the dimmer variables c_i.

Figure 7 shows a restoration obtained by Hunt's filter, while Figure 8 shows how badly a naive inversion ignoring noise and using nonnegative least squares may fail.

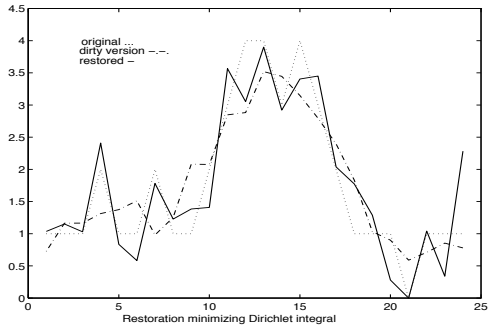

Figure 5.

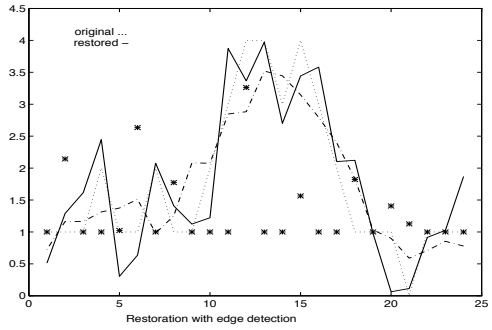

Figure 6.

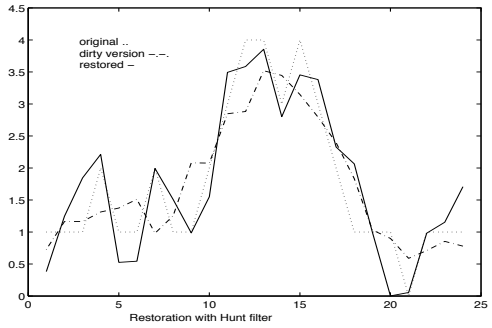

Figure 7.

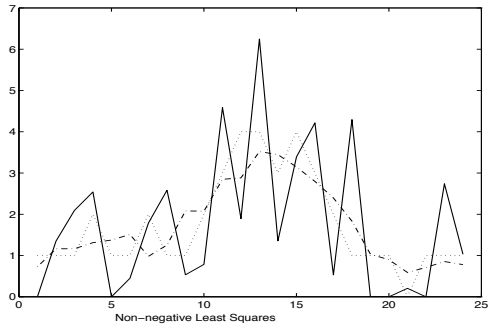

Figure 8.

References

[1] Alvarez, L., F. Guichard, P.L. Lions, J.-M. Morel. Axioms and fundamental equations of image processing. *Archive for Rational Mechanics and Analysis*, pages 199 – 257, 1993.

[2] Angwin, D.L., H. Kaufman. Nonhomogeneous image identification and restoration procedures. *Springer Series in Information Sciences* , 23:177 – 208, 1991.

[3] Borwein, J.M., A.S. Lewis, D. Noll. Maximum entropy reconstruction using derivative information I: Fisher information and convex duality. *Mathematics of Operations Research*, 21:442 – 468, 1996.

[4] Borwein, J.M., A.S. Lewis, M.N. Limber, D. Noll. Maximum entropy reconstruction using derivative information II: Computational results. *Numerische Mathematik*, 69:243 – 256, 1995.

[5] Borwein, J.M., M.N. Limber, D. Noll. Fast heuristic methods for function reconstruction using derivative information. *Applicable Analysis*.

[6] Burch, S.F., S.F. Gull, J.K. Skilling. Image restoration by a powerful maximum entropy method. *Computer Vision, Graphics, and Image Processing*, 23:113 – 128, 1983.

[7] Donoho, D. On minimum entropy deconvolution. *Applied Time Series Analysis*, II:565 – 608, 1981.

[8] FRIEDEN, B.R. Restoring with maximum likelihood and maximum entropy. *J. Opt. Soc. Amer.*, G2:511 – 518, 1972.

[9] GAMBOA, F., E. GASSIAT. Blind deconvolution of discrete linear systems. *to appear.*

[10] GASSIAT, E., E. GAUTHERAT. Identification of noisy linear systems with discrete random input. *to appear.*

[11] GEMAN, S., D. GEMAN. Stochastic relaxation, gibbs distribution, and the bayesian restoration of images. *IEEETrans. Pattern Analysis and Machine Intelligence*, 6:721 – 741, 1984.

[12] HARALICK, R.M., E. ØSTEVOLD, X. ZHUANG. The principle of maximum entropy in image recovery. *Springer Series in Information Sciences*, 23:157 – 193, 1991.

[13] HUNT, B.R. Deconvolution of linear systems by constrained regression and its relationship to the Wiener theory. *IEEE Trans. Autom. Contr.*, pages 703 – 705, 1972.

[14] JAEHNE, B. Digital image processing: concepts, algorithms, and scientific applications. *Springer Verlag, Berlin, New York*, 1995.

[15] KATSAGGELOS, A.K., K.-T. LAY. Maximum likelihood identification and restoration of images using the expectation-maximization algorithm. *Springer Series in Information Sciences*, 23:143 – 176, 1991.

[16] MARÉCHAL, P., A. LANNES. Towards a better understanding of a wide class of inverse problems via the principle of maximum entropy on the mean. *Inverse Problems*, To appear.

[17] NAVAZA, J. The use of non-local constraints in maximum-entropy electron density reconstruction. *Acta Crystallographica*, A42:212 – 223, 1986.

[18] NOLL, D. Consistence of a nonlinear deconvolution method with applications in image restoration. *Advances in Mathematical Sciences and Applications*, To appear.

[19] NOLL, D. Reconstruction with noisy data: An approach via eigenvalue optimization. *SIAM Journal of Optimization*, To appear.

[20] NOLL, D. Restoration of degraded images with maximum entropy. *Journal of Global Optimization*, To appear.

[21] POTRA, F. Interior-point algorithm for quadratically constrained entropy minimization problems. *J. Optimiz. Theory Appl.*, 77:79 – 95, 1993.

[22] SILVER, R.N. Quantum statistical inference. *IEEE Report, Kluwer, Dordrecht, The Netherlands*, 1992.

[23] TEKALP, A.M., G. PAVLOVIĆ. Restoration of scanned photographic images. *Springer Series in Information Sciences*, 1991.

[24] VANDENBERGHE, L., S. BOYD. Semidefinite programming. *SIAM Review*, 38:49 – 95, 1996.

Université Paul Sabatier, Laboratoire Approximation et Optimisation, 118 route de Narbonne, 31062 Toulouse Cedex, France

e-mail: noll.at.dumbo.ups-tlse.fr

http://web.cict.fr:8200/lao/LAO.html