

Global Optimal Image Reconstruction from Blurred Noisy Data by a Bayesian Approach

C. BRUNI,¹ R. BRUNI,² A. DE SANTIS,³ D. IACOVIELLO,⁴
AND G. KOCH⁵

Communicated by R. Conti

Abstract. In this paper, a procedure is presented which allows the optimal reconstruction of images from blurred noisy data. The procedure relies on a general Bayesian approach, which makes proper use of all the available information. Special attention is devoted to the informative content of the edges; thus, a preprocessing phase is included, with the aim of estimating the jump sizes in the gray level. The optimization phase follows; existence and uniqueness of the solution is secured. The procedure is tested against simple simulated data and real data.

Key Words. Image analysis, global constrained optimization, Bayesian modeling, wavelet processing.

1. Introduction

The issue of image reconstruction has received much attention in the last decades. Models proposed in the current literature include the blurring

¹Professor, Dipartimento di Informatica e Sistemistica and Centro Interdipartimentale di Ricerca per l'Analisi dei Modelli e dell'Informazione nei Sistemi Biomedici, Università di Roma La Sapienza, Rome, Italy.

²Postdoctoral Fellow, Dipartimento di Informatica e Sistemistica, Università di Roma La Sapienza, Rome, Italy.

³Professor, Dipartimento di Informatica e Sistemistica and Centro Interdipartimentale di Ricerca per l'Analisi dei Modelli e dell'Informazione nei Sistemi Biomedici, Università di Roma La Sapienza, Rome, Italy.

⁴Researcher, Dipartimento di Informatica e Sistemistica, Università di Roma La Sapienza, Rome, Italy.

⁵Professor, Dipartimento di Informatica e Sistemistica, Università di Roma La Sapienza, Rome, Italy.

by various types of point spread functions and/or noise corruption effects. In this context, a relevant aspect is the introduction of a suitable cost functional which allows us to define the image reconstruction as an optimization problem. Typically, the admissible image set is accounted for via hard constraints (Refs. 1–2) or via soft constraint penalty terms added to the cost functional (Refs. 3–6).

In this paper, following a comprehensive approach to the problem recently proposed in Refs. 1–2 and 7–9, we intend to design a procedure which allows us to reconstruct images starting from blurred noisy data. The procedure relies on a general Bayesian approach which makes proper use of all the available information. We attach special attention to the informative content of the edges; thus, information comes first from the prior knowledge of the set of possible edge locations. Over that set, proper data preprocessing provides us an estimate of the jump size, while inside the smoothness regions the information is carried by the available gray level data.

After the formulation of the general set up, we discuss the image reconstruction procedure with reference to a discretized model. In particular, we illustrate data preprocessing and carry out the Bayesian approach to image reconstruction until we get the overall cost function to be minimized.

Finally, some examples are reported with reference to both simulated data and real data.

2. General Set Up

In the sequel, we shall use the following notations: I_A is the indicator function of the set A , i.e.,

$$I_A(x) = \begin{cases} 0, & x \notin A, \\ 1, & x \in A; \end{cases}$$

∂A denotes the boundary of the set A , $\overset{\circ}{A}$ denotes its interior, while \bar{A} denotes its closure.

Let us first introduce the set \mathcal{D} of admissible images.

Definition 2.1. The class \mathcal{D} of admissible images is defined as the set of possibly discontinuous functions $f: D \rightarrow E$ which admit a representation

$$f = \sum_{k=1}^{N_f} \gamma_k^f I_{A_k^f}, \quad N_f < \infty, \quad (1)$$

where $E = [0, M]$ and D is a compact square subset of \mathbb{R}^2 ; here, $\{A_k^f, k = 1, 2, \dots, N_f\}$ is a finite partition of D such that the atoms A_k^f are

connected, with $\overline{A_k^f} = \widehat{A_k^f}$ and ∂A_k^f having zero 2D Lebesgue measure; also, each γ_k^f is continuous on D and its modulus of continuity is uniformly bounded over \mathcal{D} by $K < \infty$.

For any $f \in \mathcal{D}$, we denote by C_f its discontinuity set. Clearly, we have

$$C_f \subset S_f = \bigcup_{k=1}^{N_f} \partial A_k^f.$$

On \mathcal{D} , we consider the topology induced by an L_2 -norm so that two images f, g will be assumed to coincide whenever

$$f(x) = g(x), \quad \text{a.e. in } D.$$

The measured signal will be represented as

$$z = y + n, \tag{2}$$

where

$$y = f * N_{\Sigma_d} = H(f). \tag{3}$$

In (2), n denotes an additive white Gaussian noise, with zero mean value and variance σ_n^2 , while y denotes the blurred image. The blurring effect is described by (3) where the asterisk denotes the convolution operator H and N_{Σ_d} denotes the 2D Gaussian kernel with zero mean and covariance matrix Σ_d . That is, (3) amounts to

$$y(t) = \int_D f(\tau) N_{\Sigma_d}(t - \tau) d\tau, \quad t \in D. \tag{4}$$

The covariance matrix Σ_d will be assumed to be diagonal with

$$\Sigma_d = \text{diag}(\sigma_{d_1}^2, \sigma_{d_2}^2),$$

and $\sigma_{d_1}^2, \sigma_{d_2}^2 > 0$, so that the following factorization holds:

$$N_{\Sigma_d} = N_{d_1} N_{d_2},$$

where N_d is the 1D Gaussian kernel with zero mean and variance σ_d^2 . Here and in the following, the assumption that the covariance matrix is diagonal (that is, the 2D Gaussian kernel factorizes as the product of two 1D kernels) is introduced for sake of simplicity, but does not affect essentially the results.

If the set \mathcal{D} of admissible images is endowed with a suitable prior probability density $p(f)$, and if the available experimental information is exhibited, it is possible to evaluate the goodness of any given estimate of the unknown image via its posterior distribution.

Let $p(f|z)$, $f \in \mathcal{D}$, denote the posterior probability density over \mathcal{D} with respect to a reference distribution. Then, the Bayesian approach leads to assuming as an optimal estimate \hat{f} the value of f for which $p(\cdot|z)$ is maximum. The crucial problems in looking for the optimal solution are the existence and uniqueness of that solution. If $p(\cdot|z)$ is continuous and strictly convex, in turn these properties, call for the compactness and convexity of the admissible set. In this context, a general result is provided in Ref. 2, Theorem 6.4, where it is shown that these properties are secured for a subset \mathcal{D}' defined by the functions $f \in \mathcal{D}$ such that C_f belongs to a coarse grid G defined by a given Δ_C -size lattice L_C , where Δ_C is a positive real number,

$$L_C = \{t \in D: t = (x, y), x = i\Delta_C, y = j\Delta_C, i, j = 0, 1, 2, \dots\}, \quad (5)$$

C_f is the union of intervals whose lengths admit a uniform positive lower bound and ∂D belongs to the grid G itself.

In the following, we will consider the optimal estimation problem over such a subset \mathcal{D}' , which consequently will be the support of our probability distribution. As a matter of fact, when embedding the optimization problem in a discretized context, the assumption on \mathcal{D}' does not imply any real loss of generality.

As far as the prior distribution $p(f)$ is concerned, in the absence of further information, we assume it to be uniform over \mathcal{D}' .

On the other hand, in the context of the image reconstruction, a number of instances might occur in which the knowledge of C_f is of primary importance. Indeed, quite often most of the information carried by the signals concentrate in singularity points rather than in the regular part. This is understood easily if one considers the fact that discontinuities might account for the occurrence of rare events or abrupt changes in some underlying phenomenon. For instance, we can mention signals in the analysis of gravitational waves, shocks, earthquakes, structure or material failures, biomedical signals such as X-rays or NMRs, geophysical or meteorological data, airborne territory surveillance.

Taking the previous considerations into account, it appears convenient to preprocess the observed available data z so as to enhance the contribution of the discontinuous component over that of the regular part and damp down the noise.

Clearly, preprocessing should be performed around the points of the grid G which contains C_f and should preserve all the information about f carried by the data. A preprocessing operator, which has been shown to comply with all above requirements (Ref. 2), may be found within the context of wavelet theory. In particular, we suggest here to convolute z via the

kernels $N_w^{(2),x}$ and $N_w^{(2),y}$, $w = 1, 2, \dots, W$, defined as second-order derivatives along the directions x, y respectively of a 2D Gaussian density with zero mean and covariance matrix

$$\Sigma_w = \text{diag}(\sigma_w^{2,x}, \sigma_w^{2,y}).$$

In Ref. 2, it is shown that these convolutions: (a) enjoy the property of continuous invertibility; (b) smooth out the polynomial terms of degree 0 and 1 in f , thus enhancing the relative weight of the rapidly varying components; and (c) contribute positively to the problem of identifying signal discontinuities in a noisy environment.

In particular, the third goal is achieved if more than one wavelet is used, so that one can exploit the well-known advantages of multiscale analysis (Refs. 10–11). Indeed, jump discontinuities do persist at any scale, while the spurious ones due to noise happen to be highly sensitive to scale and therefore can be smoothed out.

Clearly, a convolution by the second-order derivative of a Gaussian kernel will be sensitive mostly to the discontinuities along a subset of C_f locally orthogonal to the differentiation direction. Thus, the convolutions via $N_w^{(2),x}$ and $N_w^{(2),y}$ are expected to enhance the discontinuities along the directions x, y respectively. It follows that, in $N_w^{(2),x}$ and $N_w^{(2),y}$, the choice of $\sigma_w^{2,y}$ and $\sigma_w^{2,x}$ is inessential. Then, the simplest thing is to set $\sigma_w^{2,y}$ in $N_w^{(2),x}$ to zero and $\sigma_w^{2,x}$ in $N_w^{(2),y}$ to zero, so as to reduce the 2D convolution of z by $N_w^{(2),x}$ and $N_w^{(2),y}$ to the 1D convolution of z via the second-order derivative of the 1D Gaussian kernels $N_w^{(2)}$ with zero mean and covariance σ_w^2 .

3. Image Reconstruction Procedure

3.1. Discrete Model. The way discrete images are recorded suggests us to consider the values $f(i, j)$ that our functions f take on the finer lattice L_F defined as

$$L_F = \{t \in D : t = (x, y), x = i \Delta_F, y = j \Delta_F, i, j = 0, 1, 2, \dots, N\},$$

where

$$\Delta_F = \Delta_C / r$$

and r is a fixed integer, $r > 1$. For future convenience, we denote by F, C the following sets of integers:

$$F = \{0, 1, \dots, N\}$$

$$C = \{0, r, 2r, \dots, r \lfloor N/r \rfloor\},$$

where $[N/r]$ denotes the integer part of N/r .

From the $f(i, j)$, we derive the quantities

$$\Delta^x f_{ij} = f(i, j) - f(i, j-1), \quad i, j \in F, \quad (6)$$

$$\Delta^y f_{ij} = f(i, j) - f(i-1, j), \quad i, j \in F, \quad (7)$$

and we set

$$f(-1, j) = f(i, -1) = 0, \quad i, j \in F.$$

Within each $\Delta^x f_{ij}, \Delta^y f_{ij}$, we identify the components $\partial^x f_{ij}$ and $\partial^y f_{ij}$ due to the continuous variation of f and the components $\delta^x f_{ij}$ and $\delta^y f_{ij}$ due to the discontinuous variation. Thus,

$$\Delta^x f_{ij} = \partial^x f_{ij} + \delta^x f_{ij}, \quad i, j \in F, \quad (8)$$

$$\Delta^y f_{ij} = \partial^y f_{ij} + \delta^y f_{ij}, \quad i, j \in F. \quad (9)$$

The $f(i, j)$ are subject to the following constraints which are inherited from the constraints defining \mathcal{D}' :

$$0 \leq f(i, j) \leq M, \quad i, j \in F, \quad (10)$$

$$|\partial^x f_{ij}| \leq K\Delta_F, \quad i, j \in F, \quad (11)$$

$$|\partial^y f_{ij}| \leq K\Delta_F, \quad i, j \in F, \quad (12)$$

$$\delta^x f_{ij} = 0, \quad i \in F, j \in F \setminus C, \quad (13)$$

$$\delta^y f_{ij} = 0, \quad i \in F \setminus C, j \in F. \quad (14)$$

In a discretized context, the distinction between continuous and discontinuous variations of f is loose. Thus, to keep this distinction, we have to guarantee some difference in their relative size and add the following constraints:

$$|\delta^x f_{ij}| \leq \alpha M, \quad i \in F, j \in C, \quad (15)$$

$$|\delta^y f_{ij}| \leq \alpha M, \quad i \in C, j \in F, \quad (16)$$

$$K\Delta_F \leq \beta M, \quad (17)$$

with

$$0 < \beta \ll \alpha < 1.$$

The discretized versions of (2)–(3) are now

$$z(i, j) = y(i, j) + n(i, j), \quad i, j \in F, \quad (18)$$

$$y(i, j) = \sum_{h, k \in F} f(h, k) N_{\Sigma_d}(i-h, j-k), \quad i, j \in F, \quad (19)$$

where the samples $f(i, j), i, j \in F$, satisfy the constraints (10)–(17), while the samples $n(i, j), i, j \in F$, are independent Gaussian variables with zero mean and variance σ_n^2 .

3.2. Data Preprocessing. In Section 2, we stressed already the advantages coming from data preprocessing via convolution by a family of 1D Gaussian kernels $N_w^{(2)}$ with zero mean and variances $\sigma_w^2, w = 1, 2, \dots, W$. Of course, preprocessing is convenient only on the subset G of L_F where discontinuities are expected. Therefore, on G , we consider the following pre-processed data:

$$\tilde{z}_{ijw}^x = \sum_{k \in F} z(i, k) N_w^{(2)}(j - k), \quad i \in F, j \in C, w = 1, 2, \dots, W, \quad (20)$$

$$\tilde{z}_{ijw}^{x-} = \tilde{z}_{i(j-1)w}^x, \quad i \in F, j \in C, w = 1, 2, \dots, W, \quad (21)$$

$$\tilde{z}_{ijw}^y = \sum_{h \in F} z(h, j) N_w^{(2)}(i - h), \quad i \in C, j \in F, w = 1, 2, \dots, W, \quad (22)$$

$$\tilde{z}_{ijw}^{y-} = \tilde{z}_{(i-1)jw}^y, \quad i \in C, j \in F, w = 1, 2, \dots, W, \quad (23)$$

while on $F \setminus C$ we keep the original data z_{ij} . Let us consider the new vector of size v ,

$$\tilde{z} = \begin{bmatrix} \tilde{z} \\ \tilde{z}^x \\ \tilde{z}^{x-} \\ \tilde{z}^y \\ \tilde{z}^{y-} \end{bmatrix}; \quad (24)$$

v is given by

$$v = (N - [N/r])^2 + 4(N + 1)([N/r] + 1) W,$$

\tilde{z} is the vector of size $(N - [N/r])^2$, made up by $z(i, j), i, j \in F \setminus C$; \tilde{z}^x, \tilde{z}^y are vectors of size $(N + 1)([N/r] + 1) W$ made up respectively of the samples $\tilde{z}_{ijw}^x, \tilde{z}_{ijw}^y$ given respectively by (20), (22); and $\tilde{z}^{x-}, \tilde{z}^{y-}$ are vectors of size $(N + 1)([N/r] + 1) W$ made up respectively of the samples $\tilde{z}_{ijw}^{x-}, \tilde{z}_{ijw}^{y-}$ given respectively by (21), (23).

We note that

$$\tilde{z} = T'(z), \quad (25)$$

where z is the vector of size $(N + 1)^2$ made up of all the samples $z(i, j), i, j \in F$, and T' is a suitable linear function. Furthermore, despite the fact that $v > (N + 1)^2$, T' has an inverse over its range $\mathcal{R}(T')$; indeed, the

values $z(i, j)$, $i, j \in F \setminus C$, are already included in \bar{z} , while the remaining values $z(i, j)$ (for $i \in F, j \in C$ and for $i \in C, j \in F$) are uniquely recovered respectively for any fixed w from \bar{z}_{ijw}^x , $i \in F, j \in C$, and \bar{z}_{ijw}^y , $i \in C, j \in F$ (see Section 6, Appendix A).

Similarly, for future convenience, we consider the vector of size μ ,

$$\tilde{f} = \begin{bmatrix} \bar{f} \\ \Delta^x f \\ \Delta^y f \end{bmatrix}, \quad (26)$$

where μ is given by

$$\mu = (N - [N/r])^2 + 2(N + 1)([N/r] + 1),$$

\bar{f} is the vector of size $(N - [N/r])^2$ made up by $f(i, j)$, $i, j \in F \setminus C$; $\Delta^x f, \Delta^y f$ are the vectors of size $(N + 1)([N/r] + 1)$ made up respectively of the samples $\Delta^x f_{ij}$, $i \in F, j \in C$, and $\Delta^y f_{ij}$, $i \in C, j \in F$, as in (6)–(7).

Again, we note that

$$\tilde{f} = T''(f), \quad (27)$$

where f is the vector of size $(N + 1)^2$ made up of all the samples $f(i, j)$, $i, j \in F$ and T'' is a suitable linear function. Also, T'' is shown easily to have an inverse on $\mathcal{R}(T'')$.

3.3. Bayesian Approach to Image Reconstruction. As already mentioned in Section 2, we shall now assess the probability densities of relevant quantities, so as to proceed to the formulation of the Bayesian estimation problem. In the following, the probability densities will be taken implicitly with respect to the Lebesgue measure of the corresponding dimension.

To proceed, we consider first the conditional densities

$$p(\bar{f} | \Delta^x f, \Delta^y f), p(\bar{z} | f), p(\Delta^x f | \bar{z}^x), p(\Delta^y f | \bar{z}^y),$$

in the case of no information about the subject represented in the image.

Now, $p(\bar{f} | \Delta^x f, \Delta^y f)$ is a flat density over the admissible set defined by (8)–(14); that is,

$$0 \leq f(i, j) \leq M, \quad i, j \in F \setminus C, \quad (28)$$

$$|f(i, j) - f(i - 1, j)| \leq K\Delta_F, \quad i, j \in F \setminus C, \quad (29)$$

$$|f(i, j) - f(i, j - 1)| \leq K\Delta_F, \quad i, j \in F \setminus C. \quad (30)$$

In (29)–(30), the values of $f(i, j)$ out of \bar{f} are provided by conditioning on $\Delta^x f, \Delta^y f$.

$p(\bar{z}|f)$ is the Gaussian $N(H(f), \sigma_n^2 I)$, where $H(f)$ is the vector of size $N - [N/r]$ made up by the samples $y(i, j), i, j \in F \setminus C$, as in (19) and I is the identity matrix of size $(N - [N/r])^2$.

$p(\Delta^x f | \bar{z}^x, \bar{z}^{x-})$ is the joint density of

$$\Delta^x f_{ij} = \partial^x f_{ij} + \delta^x f_{ij}, \quad i \in F, j \in C,$$

conditioned on \bar{z}^x, \bar{z}^{x-} . Because any two entries in $\Delta^x f$ do not operate on any common value of f , we assume them to be independent. Since \bar{z}^x, \bar{z}^{x-} yield virtually no information about $\partial^x f_{ij}, i \in F, j \in C$, we may further assume $\partial^x f_{ij}, \delta^x f_{ij}$ to be independent from each other, conditioned on \bar{z}^x, \bar{z}^{x-} . Thus, we approximate

$$\begin{aligned} p(\Delta^x f | \bar{z}^x, \bar{z}^{x-}) &\simeq \prod_{\substack{i \in F \\ j \in C}} p(\Delta^x f_{ij} | \bar{z}^x, \bar{z}^{x-}) \\ &= \prod_{\substack{i \in F \\ j \in C}} p(\partial^x f_{ij} | \bar{z}^x, \bar{z}^{x-}) * p(\delta^x f_{ij} | \bar{z}^x, \bar{z}^{x-}) \\ &= \prod_{\substack{i \in F \\ j \in C}} p(\partial^x f_{ij}) * p(\delta^x f_{ij} | \bar{z}^x, \bar{z}^{x-}). \end{aligned} \tag{31}$$

Now, $p(\partial^x f_{ij})$ is a flat density over the admissible set

$$|\partial^x f_{ij}| \leq K \Delta_F. \tag{32}$$

As far as $p(\delta^x f_{ij} | \bar{z}^x, \bar{z}^{x-})$ is concerned, it is possible to exploit the information provided by \bar{z}^x, \bar{z}^{x-} (actually $\bar{z}_{ijw}^x, \bar{z}_{ijw}^{x-}, w = 1, 2, \dots, W$, are enough) to estimate $\delta^x f_{ij}$ in a linear Gaussian context. Denoting by $\widehat{\delta^x f_{ij}}$ the estimate and by $\sigma_{ij}^{2,x}$ its variance, $p(\delta^x f_{ij} | \bar{z}^x, \bar{z}^{x-})$ turns out to be the Gaussian $N(\widehat{\delta^x f_{ij}}, \sigma_{ij}^{2,x})$. Similarly, with an obvious use of notations, we have

$$p(\Delta^y f | \bar{z}^y, \bar{z}^{y-}) \simeq \prod_{\substack{i \in C \\ j \in F}} p(\partial^y f_{ij}) * p(\delta^y f_{ij} | \bar{z}^y, \bar{z}^{y-}),$$

with $p(\partial^y f_{ij})$ the flat density over the admissible set

$$|\partial^y f_{ij}| \leq K \Delta_F \tag{33}$$

and

$$p(\delta^y f_{ij} | \bar{z}^y, \bar{z}^{y-}) = N(\widehat{\delta^y f_{ij}}, \sigma_{ij}^{2,y}).$$

Computation of $\widehat{\delta^y f_{ij}}, \widehat{\delta^y f_{ij}}, \sigma_{ij}^{2,x}, \sigma_{ij}^{2,y}$ is carried out in Section 6, Appendix B, where it is shown also that the variances $\sigma_{ij}^{2,x}, \sigma_{ij}^{2,y}$ are independent of the pixel location.

Now, we are ready to compute the posterior probability density of f ,

$$\begin{aligned}
 p(f|z) &= p(z, f)/p(z) \\
 &= p(T'(z), f)/p(z) \\
 &= p(\bar{z}, \bar{z}^x, \bar{z}^{x-}, \bar{z}^y, \bar{z}^{y-}, f)/p(z) \\
 &= p(\bar{z}|\bar{z}^x, \bar{z}^{x-}, \bar{z}^y, \bar{z}^{y-}, f)p(\bar{z}^x, \bar{z}^{x-}, \bar{z}^y, \bar{z}^{y-}, f)/p(z) \\
 &\simeq p(\bar{z}|f)p(\bar{z}^x, \bar{z}^{x-}, \bar{z}^y, \bar{z}^{y-}, f)/p(z) \\
 &= p(\bar{z}|f)p(f|\bar{z}^x, \bar{z}^{x-}, \bar{z}^y, \bar{z}^{y-})p(\bar{z}^x, \bar{z}^{x-}, \bar{z}^y, \bar{z}^{y-})/p(z) \\
 &= p(\bar{z}|f)p(T''(f)|\bar{z}^x, \bar{z}^{x-}, \bar{z}^y, \bar{z}^{y-})p(\bar{z}^x, \bar{z}^{x-}, \bar{z}^y, \bar{z}^{y-})/p(z) \\
 &= p(\bar{z}|f)p(\bar{f}, \Delta^x f, \Delta^y f|\bar{z}^x, \bar{z}^{x-}, \bar{z}^y, \bar{z}^{y-})p(\bar{z}^x, \bar{z}^{x-}, \bar{z}^y, \bar{z}^{y-})/p(z) \\
 &= p(\bar{z}|f)p(\bar{f}|\Delta^x f, \Delta^y f, \bar{z}^x, \bar{z}^{x-}, \bar{z}^y, \bar{z}^{y-}) \\
 &\quad \times p(\Delta^x f, \Delta^y f|\bar{z}^x, \bar{z}^{x-}, \bar{z}^y, \bar{z}^{y-})p(\bar{z}^x, \bar{z}^{x-}, \bar{z}^y, \bar{z}^{y-})/p(z) \\
 &\simeq p(\bar{z}|f)p(\bar{f}|\Delta^x f, \Delta^y f)p(\Delta^x f|\bar{z}^x, \bar{z}^{x-}) \\
 &\quad \times p(\Delta^y f|\bar{z}^y, \bar{z}^{y-})p(\bar{z}^x, \bar{z}^{x-}, \bar{z}^y, \bar{z}^{y-})/p(z). \tag{34}
 \end{aligned}$$

In (34), we approximated the probability $p(\bar{z}|\bar{z}^x, \bar{z}^{x-}, \bar{z}^y, \bar{z}^{y-}, f)$ by $p(\bar{z}|f)$ and the probability $p(\bar{f}|\Delta^x f, \Delta^y f, \bar{z}^x, \bar{z}^{x-}, \bar{z}^y, \bar{z}^{y-})$ by $p(\bar{f}|\Delta^x f, \Delta^y f)$, since once we are given f the vectors $\bar{z}^x, \bar{z}^{x-}, \bar{z}^y, \bar{z}^{y-}$ carry no significant further information about the regular parts of f and z . Also, we approximated the probability $p(\Delta^x f, \Delta^y f|\bar{z}^x, \bar{z}^{x-}, \bar{z}^y, \bar{z}^{y-})$ by $p(\Delta^x f|\bar{z}^x, \bar{z}^{x-})p(\Delta^y f|\bar{z}^y, \bar{z}^{y-})$ in that, by the very way the relevant vectors were computed, $\Delta^x f$ and $\Delta^y f$ turn out to be independent and there is no significant information between $\Delta^x f$ and \bar{z}^y, \bar{z}^{y-} , as well as between $\Delta^y f$ and \bar{z}^x, \bar{z}^{x-} .

Finally, we note that, if $p(\partial^x f_{ij}), p(\partial^y f_{ij})$ are approximated by Gaussian densities with zero mean and variance $(K\Delta_F)^2 = c^2$, $p(f|z)$ is zero out of the admissible set defined in (28)–(30), while inside the same set is proportional (via a factor independent of f) to the product

$$\begin{aligned}
 &\prod_{\substack{i \in F \setminus C \\ j \in F \setminus C}} \exp\{-[z(i, j) - (H(f))_{ij}]^2/\sigma_n^2\} \\
 &\times \prod_{\substack{i \in F \setminus C \\ j \in F \setminus C}} \exp\{-[(\Delta^x f_{ij})^2 + (\Delta^y f_{ij})^2]/c^2\} \\
 &\times \prod_{\substack{i \in F \\ j \in C}} \exp\{-[\Delta^x f_{ij} - \widehat{\delta^x f_{ij}}]^2/(c^{-2} + \sigma^{-2})^{-1}\} \\
 &\times \prod_{\substack{i \in C \\ j \in F}} \exp\{-[\Delta^y f_{ij} - \widehat{\delta^y f_{ij}}]^2/(c^{-2} + \sigma^{-2})^{-1}\}.
 \end{aligned}$$

Thus the MAP estimate for f is achieved by minimizing on the admissible set (28)–(30) the cost function

$$\begin{aligned}
 J(f) = & \sum_{i,j \in F \setminus C} \{ [z(i,j) - (H(f))_{ij}]^2 / \sigma_n^2 + [(\Delta^x f_{ij})^2 + (\Delta^y f_{ij})^2] / c^2 \} \\
 & + \sum_{\substack{i \in F \\ j \in C}} [\Delta^x f_{ij} - \widehat{\delta^x f_{ij}}]^2 / (c^{-2} + \sigma^{-2})^{-1} \\
 & + \sum_{\substack{i \in C \\ j \in F}} [\Delta^y f_{ij} - \widehat{\delta^y f_{ij}}]^2 / (c^{-2} + \sigma^{-2})^{-1}. \tag{35}
 \end{aligned}$$

This minimization problem enjoys the following properties:

- (i) it is a quadratic programming problem (the cost functional is a quadratic form to be minimized under linear constraints);
- (ii) standard routines are available for its numerical solution, even in the case of large size problems;
- (iii) the existence of solutions is guaranteed by the compactness of the admissible set (28)–(30) and the continuity of J ;
- (iv) the uniqueness of the solution holds, by the convexity of the admissible set and the strict convexity of J .

4. Testing the Procedure

In this section, we present some preliminary results obtained by applying the proposed procedure for image reconstruction to some simulated data and real data.

As mentioned already in Section 3.2, the data have been preprocessed via convolution by $W = 3$ one-dimensional kernels $N_w^{(2)}$, with zero mean and variances σ_w^2 , $w = 1, 2, 3$. The number of kernels as well as their variances have been chosen in order to comply with different requirements: noise filtering, smoothing of regular component, discriminating among adjacent edges. Quantitative choice criteria are provided in Refs. 2–9, which of course account for the variances σ_d^2 , σ_n^2 , the ratio $r = \Delta_C / \Delta_F$ and the modulus of continuity K .

The optimization phase, discussed in Section 3.3, follows such pre-processing phase. The optimization problem has the convex quadratic cost function (35) and the admissible set of constraints (28). Such a problem is solved iteratively by means of a standard implementation of the barrier method (Ref. 12). A Cholewsky factorization of the matrix of the quadratic part of the cost function is performed. The constraints are then replaced

by penalty terms (barrier function) added to the cost function. Then, an unconstrained problem is solved. This is iterated by varying the barrier terms, until the optimality conditions are satisfied numerically. Numerical precision is of order 10^{-6} .

To test the procedure, we generated simulated data corresponding to simple geometrical forms. Those images (underlying data f) are reported in Figs. 1, 13, 19; in all cases, the regular component was taken as

$$\gamma(x, y) = 0.1 \sin(0.2x + 0.08y);$$

the piecewise constant singular component $f(x, y) - \gamma(x, y)$ assumes the values indicated in the same figures.

The above underlying data f have been degraded according to the model (2)–(3) for different values of σ_n^2 , σ_d^2 , thus generating the measured data z represented in Figs. 2, 7, 10, 14, 20. The results obtained for suitable values of Δ_C and σ_w^2 , $w = 1, 2, 3$, are shown in Figs. 3, 4, 8, 9, 11, 12, 15, 16, 21. Note that the results given in Figs. 4 and 21 are obtained by images featuring edges not belonging to the set L_C : this in order to check the robustness of the procedure.

The usefulness of the various steps in the preprocessing phase was evaluated by waiving the convolution via the kernels $N_w^{(2)}$ (and therefore the

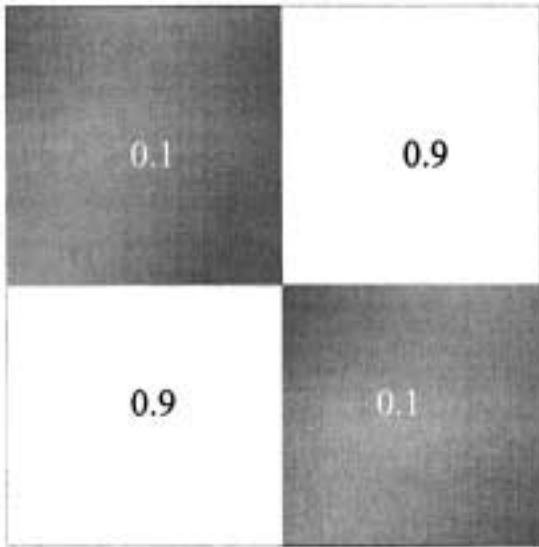


Fig. 1. First simulated uncorrupted image.

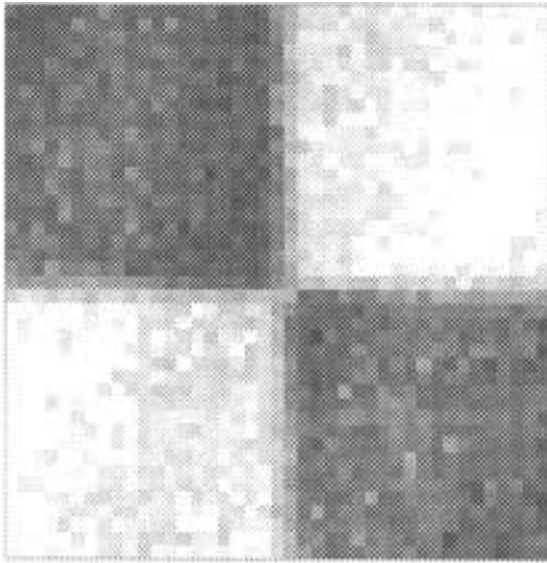


Fig. 2. First simulated corrupted image, $\sigma_n^2 = 0.01$, $\sigma_d^2 = 1$.

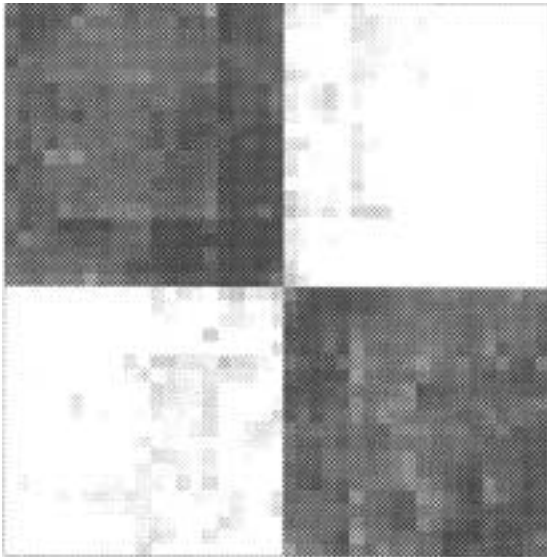


Fig. 3. Reconstruction obtained from the data of Fig. 2 assuming $\Delta_C = 5$, $\sigma_w = 1, 1.3, 1.5$. It leads to $m = 0.6$.

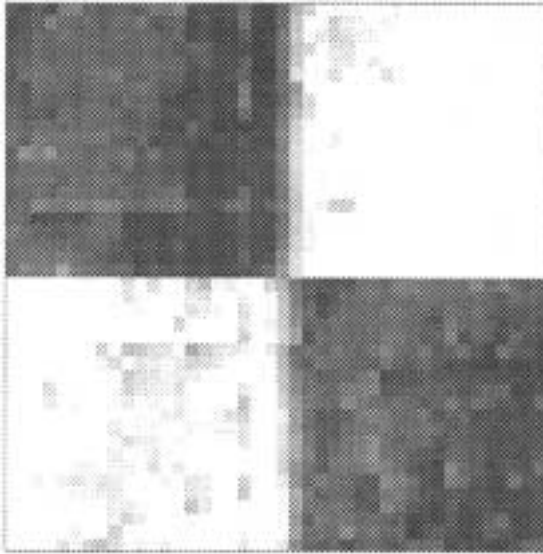


Fig. 4. Reconstruction obtained from the data of Fig. 2 assuming $\Delta_c = 5$, $\sigma_w = 1, 1.3, 1.5$, and vertical edge out of the grid. It leads to $m = 0.8$.

estimation of $(\delta^x f_{ij}, \delta^y f_{ij})$ or even allowing edges to belong to the whole of L_F . The related results are reported in Figs. 5, 17, 6, 18.

An overall figure of the achieved improvement, in the case of simulated data, has been drawn from the relative mean-square error

$$m = E_f / E_i,$$

where E_f is the mean-square error between the reconstructed image and the true underlying image and E_i is the mean-square error between the initial data and the true image. Obviously, $m \geq 0$; the case $m = 0$ corresponds to the ideal unrealistic case of perfect reconstruction of the true image; the case $m \in (0, 1)$ corresponds to an improvement of the reconstructed image with respect to the data; while the case $m > 1$ corresponds to a worsening (this latter case is indeed possible in particular when the available data are of very high quality).

The proposed procedure exhibits good capability of recovering the edges which are actually included in L_C , even in the case of severe data corruption. The edges out of L_C are also identified, although with an obvious increase in uncertainty. In any case, a priori knowledge of L_C and estimation of the jump size on L_C via data preprocessing contribute positively to the accuracy of the reconstructed image.

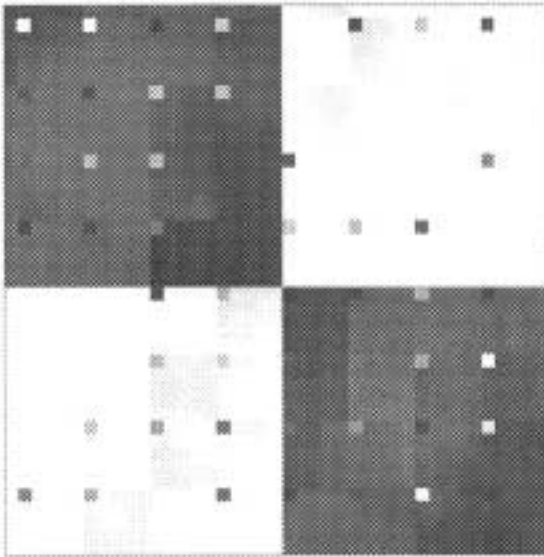


Fig. 5. Reconstruction obtained from the data of Fig. 2 assuming $\Delta_c = 5$ and without preprocessing. It leads to $m = 0.8$.

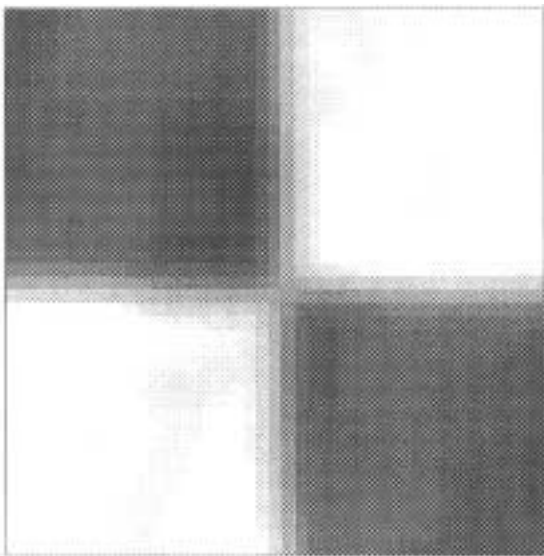


Fig. 6. Reconstruction obtained from the data of Fig. 2 without preprocessing and allowing the edges to belong to the whole of L_f . It leads to $m = 0.98$.

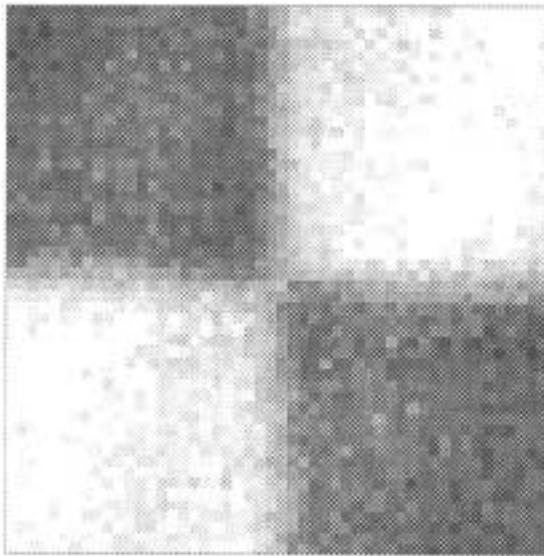


Fig. 7. First simulated corrupted image, $\sigma_n^2 = 0.01$, $\sigma_d^2 = 4$.

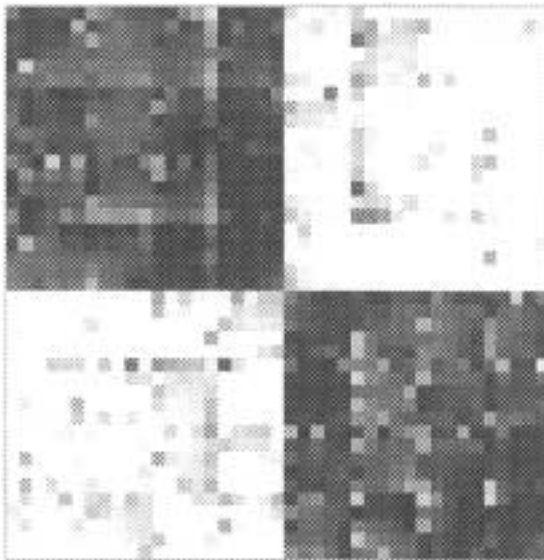


Fig. 8. Reconstruction obtained from the data of Fig. 7 assuming $\Delta_C = 5$, $\sigma_w = 1, 1.3, 1.5$. It leads to $m = 0.7$.

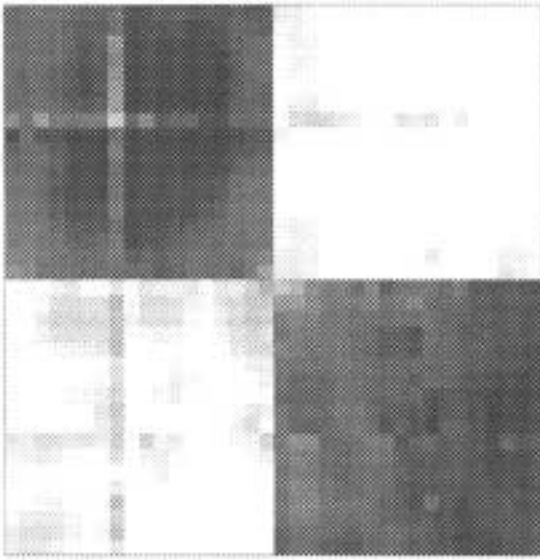


Fig. 9. Reconstruction obtained from the data of Fig. 7 assuming $\Delta_C = 10$, $\sigma_w = 2, 2.6, 3$. It leads to $m = 0.6$.

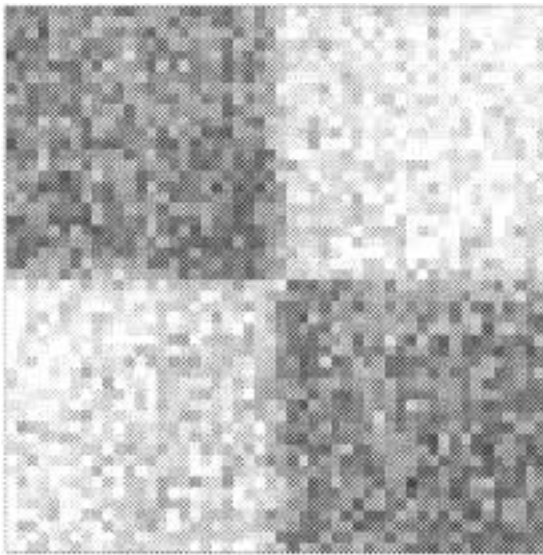


Fig. 10. First simulated corrupted image, $\sigma_n^2 = 0.1$, $\sigma_d^2 = 1$.

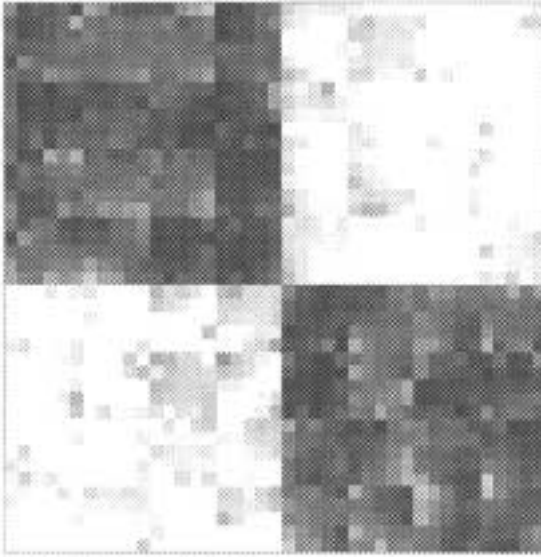


Fig. 11. Reconstruction obtained from the data of Fig. 10 assuming $\Delta_C = 5$, $\sigma_w = 1, 1.3, 1.5$. It leads to $m = 0.5$.

The last application refers to a real image corrupted by unknown blurring and noise (Fig. 22). Now, the issue arises of preliminarily estimating σ_n^2, σ_d^2 : we did not deal with that issue here, and we simply used a few guessed values for the above variances. The results shown in Figs. 23–25 were obtained by processing a detail of the whole image. Note that the quality level of the three results is similar: the procedure features do not seem critical with respect to the choice of the above parameters.

5. Appendix A

For any fixed $i \in F, w \in \{1, 2, \dots, W\}$, let us define the vectors

$$\zeta^x = \begin{bmatrix} z(i, 0) \\ z(i, r) \\ \vdots \\ z(i, r[N/r]) \end{bmatrix}, \quad \tilde{\zeta}^x = \begin{bmatrix} \tilde{z}_{i0w} - \sum_{k \in F \setminus C} z(i, k) N_w^{(2)}(0 - k) \\ \tilde{z}_{irw} - \sum_{k \in F \setminus C} z(i, k) N_w^{(2)}(r - k) \\ \vdots \\ \tilde{z}_{i, r[N/r], w} - \sum_{k \in F \setminus C} z(i, k) N_w^{(2)}(r[N/r] - k) \end{bmatrix}.$$

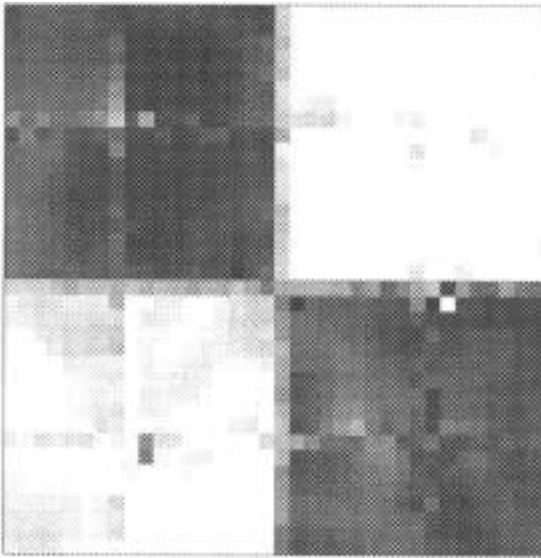


Fig. 12. Reconstruction obtained from the data of Fig. 10 assuming $\Delta_C = 10$, $\sigma_w = 2, 2.6, 3$. It leads to $m = 0.44$.

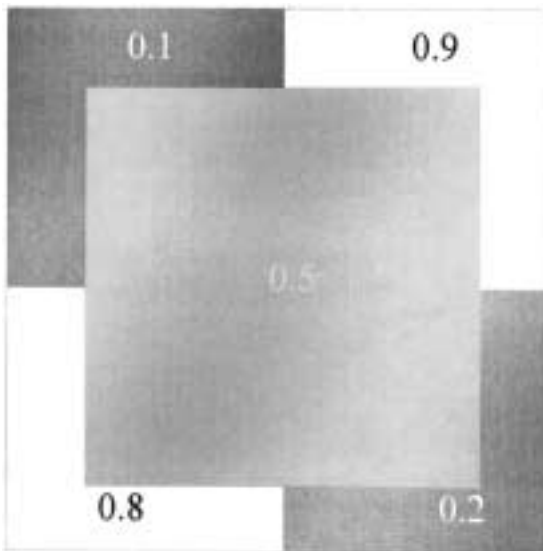


Fig. 13. Second simulated uncorrupted image.

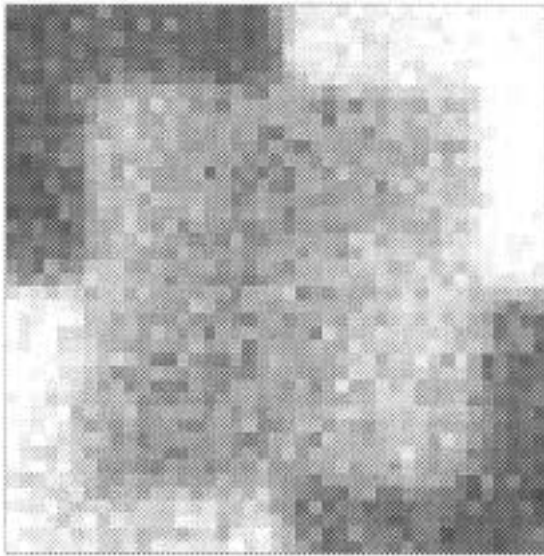


Fig. 14. Second simulated corrupted image, $\sigma_n^2 = 0.01$, $\sigma_d^2 = 1$.

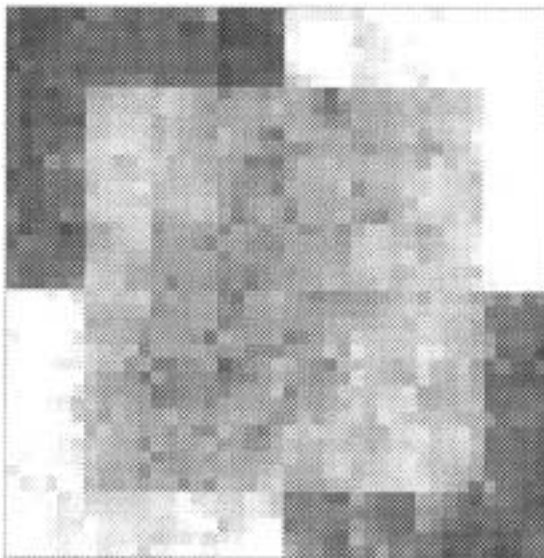


Fig. 15. Reconstruction obtained from the data of Fig. 14 assuming $\Delta_c = 5$, $\sigma_w = 1, 1.3, 1.5$. It leads to $m = 0.7$.

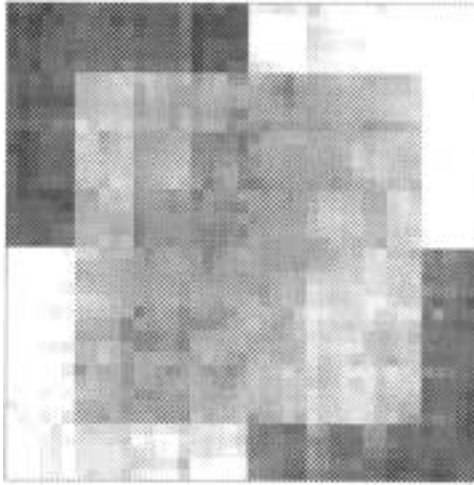


Fig. 16. Reconstruction obtained from the data of Fig. 14 assuming $\Delta_c = 5$, $\sigma_w = 2, 2.6, 3$. It leads to $m = 0.64$.

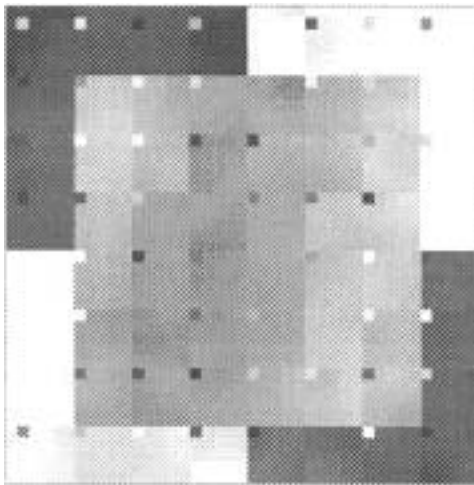


Fig. 17. Reconstruction obtained from the data of Fig. 14 assuming $\Delta_c = 5$ and without preprocessing. It leads to $m = 1.05$.

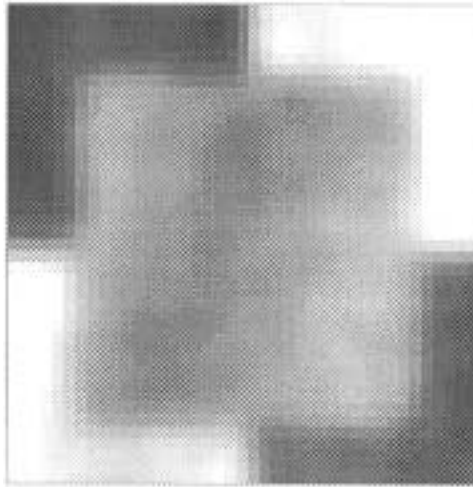


Fig. 18. Reconstruction obtained from the data of Fig. 14 without preprocessing and allowing the edges to belong to the whole of L_F . It leads to $m = 0.73$.

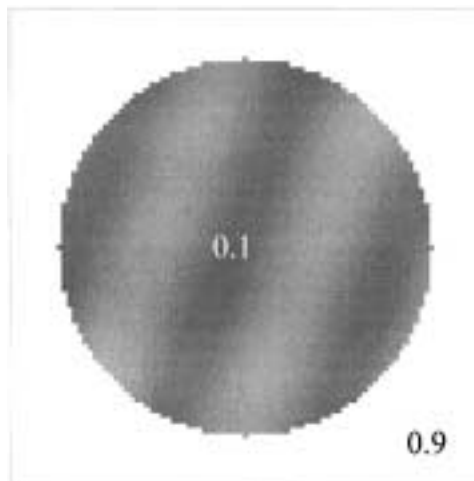


Fig. 19. Third simulated uncorrupted image.

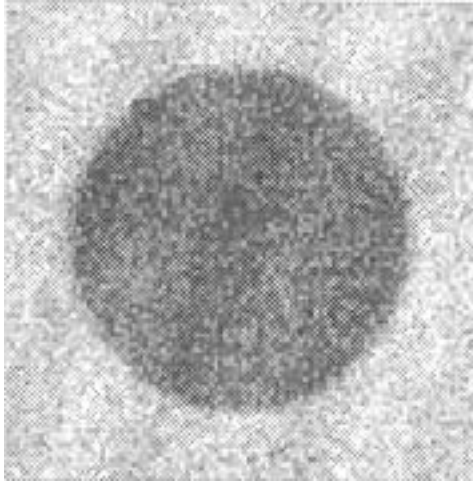


Fig. 20. Third simulated corrupted image, $\sigma_n^2 = 0.1$, $\sigma_d^2 = 1$.

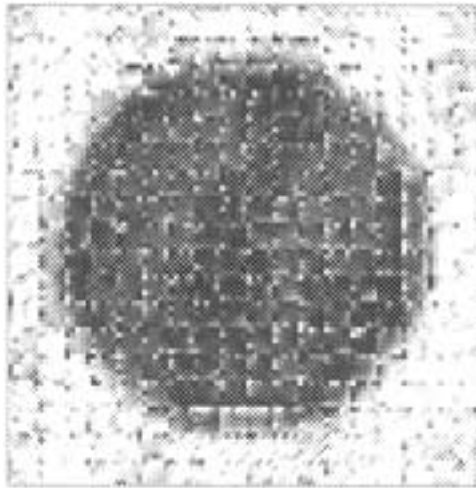


Fig. 21. Reconstruction obtained from the data of Fig. 20 assuming $\Delta_C = 5$, $\sigma_w = 1, 1.3, 1.5$. It leads to $m = 0.9$.

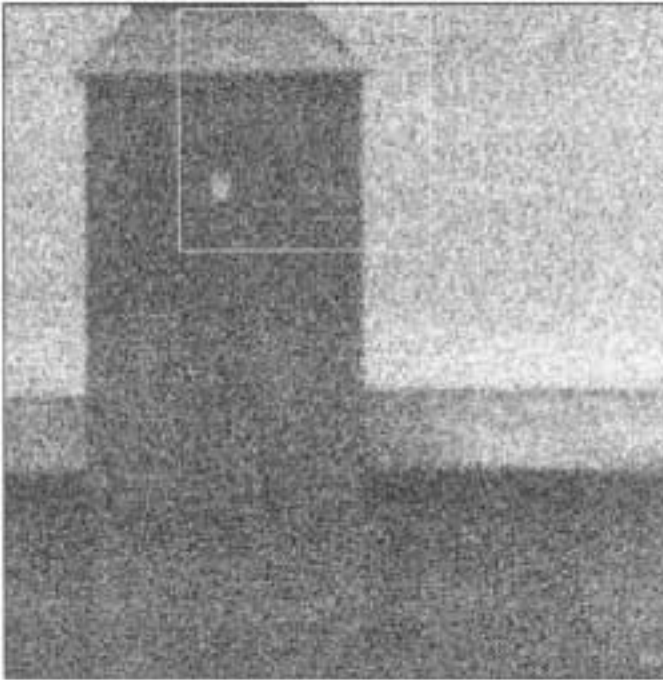


Fig. 22. Real image corrupted by unknown blurring and noise (the detail which will be processed is evidentiated).

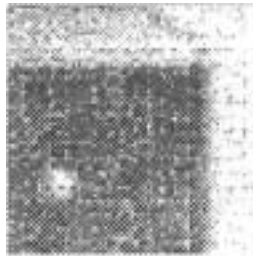


Fig. 23. Reconstruction obtained from the data of Fig. 22 guessing $\sigma_n^2 = 0.1$, $\sigma_d^2 = 1$, and assuming $\Delta_C = 5$, $\sigma_w = 1, 1.3, 1.5$.



Fig. 24. Reconstruction obtained from the data of Fig. 22 guessing $\sigma_n^2 = 0.2$, $\sigma_d^2 = 1$, and assuming $\Delta_c = 5$, $\sigma_w = 1, 1.3, 1.5$.

Note that $\tilde{\zeta}^x$ is known, once \bar{z} , \bar{z}^x are so, and that, from (20), we draw

$$\tilde{\zeta}^x = H\zeta^x,$$

where

$$H = \begin{bmatrix} N_w^{(2)}(0) & N_w^{(2)}(r) & \cdots & N_w^{(2)}(r \lfloor N/r \rfloor) \\ N_w^{(2)}(r) & N_w^{(2)}(0) & \cdots & N_w^{(2)}(r \lfloor N/r \rfloor - 1) \\ \dots & \dots & \dots & \dots \\ N_w^{(2)}(r \lfloor N/r \rfloor) & N_w^{(2)}(r \lfloor N/r \rfloor - 1) & \cdots & N_w^{(2)}(0) \end{bmatrix}.$$

In a similar way, we may define the vectors ζ^y , $\tilde{\zeta}^y$ and again we have

$$\tilde{\zeta}^y = H\zeta^y.$$

The possibility of recovering uniquely ζ^x from $\tilde{\zeta}^x$ and ζ^y from $\tilde{\zeta}^y$ corresponds clearly to the nonsingularity of the matrix H .

Theorem 5.1. For any fixed w, N, r , the matrix H is negative definite.

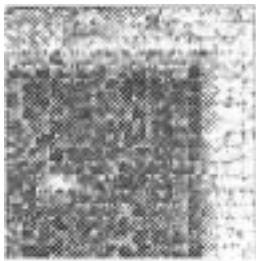


Fig. 25. Reconstruction obtained from the data of Fig. 22 guessing $\sigma_n^2 = 0.2$, $\sigma_d^2 = 2$, and assuming $\Delta_c = 5$, $\sigma_w = 1, 1.3, 1.5$.

Proof. Let us consider the function Φ defined as

$$\Phi(\omega) = \exp[-(1/2)\omega^2\sigma_w^2], \quad \omega \in \mathbb{R}.$$

Since Φ is a nonnegative, even, and integrable function, there exists a wide sense stationary (real) process v , continuous in the mean-square sense, such that $\Phi = \Phi_v$, where Φ_v denotes the power spectral density of v (Ref. 13, Theorem I.5.1). The correlation function R_v of v is then given by

$$\begin{aligned} R_v(\tau) &= \int_{-\infty}^{\infty} e^{j\omega\tau}\Phi_v(\omega) d\omega \\ &= (1/\sqrt{2\pi}\sigma_w) \exp[-(1/2)\tau^2/\sigma_w^2] \\ &= N_w(\tau). \end{aligned}$$

Since R_v is twice differentiable in $\tau=0$, and therefore everywhere, there exists a process u such that $u(t) = dv(t)/dt$, in the mean square sense, with correlation function (Ref. 13, Corollary V.2.2)

$$\begin{aligned} R_u(\tau) &= -d^2R_v(\tau)/d\tau^2 \\ &= -N_w^{(2)}(\tau). \end{aligned}$$

This suffices for H to be negative semidefinite. Next, we show that H is actually negative definite. To that purpose, we observe that the power spectral density of u is given by

$$\Phi_u(\omega) = \omega^2\Phi_v(\omega) = \omega^2 \exp[-(1/2)\omega^2\sigma_w^2],$$

which clearly is nonzero in at least $r[N/r]+1$ values of ω [indeed, $\Phi_u(\omega) \neq 0, \forall \omega \neq 0$]. Then, the conclusion follows by an argument similar to the one developed in Ref. 14, Property 1, page 122. \square

As a matter of fact, Theorem 5.1 allows us to establish a discrete version of the result proved in Ref. 2, Theorem 10, namely, the identifiability of an atomic measure from the knowledge of its mixture by a $N^{(2)}$ -like kernel over a number of points equal to the number of its atoms.

6. Appendix B

From (18)–(19), we have

$$\begin{aligned} &z(i, j) - z(i, j - 1) \\ &= \sum_{h, k \in F} N_{\Sigma_d}(i - h, j - k)[\partial^x f_{hk} + \delta^x f_{hk}] + \eta(i, j), \quad i, j \in F, \end{aligned} \tag{36}$$

where

$$\eta(i, j) = n(i, j) - n(i, j - 1).$$

Now, let us denote by $N_{d_2+w}^{(2)}$ the second-order derivative of the 1D Gaussian kernel of zero mean and variance $\sigma_{d_2}^2 + \sigma_w^2$. For $i \in F, j \in C$, from (36) it follows that

$$\tilde{z}_{ijw}^x - \tilde{z}_{ijw}^{x-} = \sum_{h,k \in F} N_{d_1}(i-h)N_{d_2+w}^{(2)}(j-k)[\partial^x f_{hk} + \delta^x f_{hk}] + \tilde{\eta}_{ijw}^x,$$

where

$$\tilde{\eta}_{ijw}^x = \sum_{k \in F} N_w^{(2)}(j-k)\eta(i, k).$$

Now, we exploit the fact that $\partial^x f_{hk} + \delta^x f_{hk}$ does not vary significantly with respect to h , for h in the range $(i - 3\sigma_{d_1}, i + 3\sigma_{d_1})$, where $N_{d_1}(i-h)$ is different significantly from zero. Furthermore, $\partial^x f_{hk}$ does not vary significantly with respect to

$$k \in (j - 3\sqrt{\sigma_{d_2}^2 + \sigma_w^2}, j + 3\sqrt{\sigma_{d_2}^2 + \sigma_w^2}),$$

while $\delta^x f_{hk}$ is zero for $k \in F \setminus C$ and it is equal to $\delta^x f_{hj}$ on the only value of $k \in C$ where $N_{d_2+w}^{(2)}(j-k)$ is significantly nonzero, that is, for $k = j$. Therefore, for $i \in F, j \in C$, we approximate

$$\begin{aligned} \tilde{z}_{ijw}^x - \tilde{z}_{ijw}^{x-} &\simeq \sum_{k \in F} N_{d_2+w}^{(2)}(j-k)[\partial^x f_{hk} + \delta^x f_{hk}] + \tilde{\eta}_{ijw}^x \\ &\simeq \sum_{k \in F} N_{d_2+w}^{(2)}(j-k)\delta^x f_{ik} + \tilde{\eta}_{ijw}^x \\ &\simeq N_{d_2+w}^{(2)}(0)\delta^x f_{ij} + \tilde{\eta}_{ijw}^x, \quad w = 1, 2, \dots, W. \end{aligned} \tag{37}$$

Let us define the following vectors of size W :

$$q^{(ij,x)} = \begin{bmatrix} \tilde{z}_{ij,1}^x - \tilde{z}_{ij,1}^{x-} \\ \vdots \\ \tilde{z}_{ij,W}^x - \tilde{z}_{ij,W}^{x-} \end{bmatrix}, \tag{38}$$

$$u^x = \begin{bmatrix} N_{d_2+1}^{(2)}(0) \\ \vdots \\ N_{d_2+W}^{(2)}(0) \end{bmatrix}, \tag{39}$$

$$\tilde{\eta}^{(ij,x)} = \begin{bmatrix} \tilde{\eta}_{ij,1}^x \\ \vdots \\ \tilde{\eta}_{ij,W}^x \end{bmatrix}. \tag{40}$$

Then, (37) becomes

$$q^{(ij,x)} = u^x \delta^x f_{ij} + \tilde{\eta}^{(ij,x)}. \quad (41)$$

We note that $\tilde{\eta}^{(ij,x)}$ is a Gaussian vector, with zero mean and covariance matrix

$$\Psi = E[\tilde{\eta}^{(ij,x)} \tilde{\eta}^{(ij,x)T}] = \{\psi_{st}\},$$

with

$$\begin{aligned} \psi_{st} &= E[\tilde{\eta}_{ij,s}^x \tilde{\eta}_{ij,t}^x] \\ &= E\left[\sum_{k \in F} N_s^{(2)}(j-k) \eta(i, k) \sum_{h \in F} N_t^{(2)}(j-h) \eta(i, h) \right] \\ &= \sum_{h, k \in F} N_s^{(2)}(j-k) N_t^{(2)}(j-h) E\{[n(i, k) - n(i, k-1)][n(i, h) - n(i, h-1)]\} \\ &= \sigma_n^2 \sum_{k \in F} [2N_s^{(2)}(j-k) N_t^{(2)}(j-k) - N_s^{(2)}(j-k) N_t^{(2)}(j-k-1) \\ &\quad - N_s^{(2)}(j-k) N_t^{(2)}(j-k+1)] \\ &= 2\sigma_n^2 [N_{s+t}^{(4)}(0) - N_{s+t}^{(4)}(1)], \quad s, t = 1, 2, \dots, W, \end{aligned}$$

where $N_{s+t}^{(4)}$ is the fourth-order derivative of a 1D Gaussian kernel with zero mean and variance $\sigma_s^2 + \sigma_t^2$.

From (41), we get the Markov estimate of $\delta^x f_{ij}$,

$$\begin{aligned} \widehat{\delta^x f_{ij}} &= u^{xT} \Psi^{-1} q^{(ij,x)} / u^{xT} \Psi^{-1} u^x \\ &= \delta^x f_{ij} + u^{xT} \Psi^{-1} \tilde{\eta}^{(ij,x)} / u^{xT} \Psi^{-1} u^x. \end{aligned} \quad (42)$$

Then, given $\tilde{z}^x, \tilde{z}^{x-}$, we see that $\delta_{f_{ij}}^x$ is Gaussian kernel with mean $\widehat{\delta^x f_{ij}}$ and variance

$$\sigma_{ij}^{2,x} = (u^{xT} \Psi^{-1} u^x)^{-1} \quad (43)$$

independent of i, j . In a similar way, given $\tilde{z}^y, \tilde{z}^{y-}$, one shows that $\delta_{f_{ij}}^y$ is Gaussian with mean

$$\widehat{\delta^y f_{ij}} = u^{yT} \Psi^{-1} q^{(ij,y)} / u^{yT} \Psi^{-1} u^y$$

and variance

$$\sigma_{ij}^{2,y} = (u^{yT} \Psi^{-1} u^y)^{-1},$$

where $q^{(ij,y)}$ and u^y are defined similarly to (38) and (39),

$$q^{(ij,y)} = \begin{bmatrix} \tilde{z}_{ij,1}^y - \tilde{z}_{ij,1}^{y-} \\ \vdots \\ \tilde{z}_{ij,W}^y - \tilde{z}_{ij,W}^{y-} \end{bmatrix},$$

$$u^y = \begin{bmatrix} N_{d_1+1}^{(2)}(0) \\ \vdots \\ N_{d_1+W}^{(2)}(0) \end{bmatrix}.$$

References

1. BRUNI, C., DE SANTIS, A., and KOCH, G., *Optimization over Spaces of Discontinuous 2D Functions*, Journal of Optimization Theory and Applications, Vol. 106, pp. 277–283, 2000.
2. BRUNI, C., DE SANTIS, A., IACOVIELLO, D., and KOCH, G., *Modeling for Edge Detection Problems in Blurred Noisy Images*, IEEE Transactions on Image Processing, Vol. 10, pp. 1447–1453, 2001.
3. MUMFORD, D., and SHAH, J., *Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems*, Communications on Pure and Applied Mathematics, Vol. 42, pp. 577–685, 1989.
4. MUMFORD, D., and SHAH, J., *Boundary Detections by Minimizing Functionals*, Image Understanding, Edited by S. Ulman and W. Richards, Ablex Publishing Company, New Jersey, Vol. 1, pp. 19–43, 1989.
5. DAL MASO, G., MOREL, J. M., and SOLIMINI, S., *A Variational Method in Image Segmentation: Existence and Approximation Results*, Acta Mathematica, Vol. 168, pp. 89–151, 1992.
6. MOREL, J. M., and SOLIMINI, S., *Variational Methods in Image Segmentation*, Birkhäuser, Basel, Switzerland, 1995.
7. BRUNI, C., DE SANTIS, A., and KOCH, G., *Convergence and Compactness Properties for a Family of 2D Discontinuous Signals*, Acta Applicandae Mathematicae, Vol. 62, pp. 131–153, 2000.
8. BRUNI, C., DE SANTIS, A., and IACOVIELLO, D., *Multiscale Analysis and Segmentation in Blurred and Noisy Images*, Proceedings of COMCON 8, Crete, Greece, 2001 (to appear).
9. BRUNI, C., DE SANTIS, A., and IACOVIELLO, D., *A Multiscale Procedure for the Segmentation of Degraded Images*, IEEE Transactions on Image Processing (to appear).
10. DAUBECHIES, I., *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, Pennsylvania, Vol. 61, 1992.
11. MALLAT, S., and ZHONG, S., *Characterization of Signals from Multiscale Edges*, IEEE Transactions on Automatic Control, Vol. 34, pp. 316–321, 1989.

12. LUSTIG, I. J., MARSTEN, R., and SHANNO, D. F., *Interior-Point Methods for Linear Programming: Computational State of the Art*, ORSA Journal of Computing, Vol. 6, pp. 1–14, 1994.
13. GIKHMAN, I. V., and SKOROKHOD, A. V., *Introduction to the Theory of Random Processes*, W. B. Saunders Company, Philadelphia, Pennsylvania, 1965.
14. SODERSTROM, T., and STOICA, P., *System Identification*, Prentice Hall International, New York, NY, 1989.