J Math Imaging Vis
DOI 10.1007/s10851-008-0125-y

Semi-explicit Solution and Fast Minimization Scheme for an Energy with ℓ_1-Fitting and Tikhonov-Like Regularization

Mila Nikolova

Abstract Regularized energies with ℓ_1-fitting have attracted a considerable interest in the recent years and numerous aspects of the problem have been studied, mainly to solve various problems arising in image processing. In this paper we focus on a rather simple form where the regularization term is a quadratic functional applied on the first-order differences between neighboring pixels. We derive a semi-explicit expression for the minimizers of this energy which shows that the solution is an affine function in the neighborhood of each data set. We then describe the volumes of data for which the same system of affine equations leads to the minimum of the relevant energy. Our analysis involves an intermediate result on random matrices constructed from truncated neighborhood sets. We also put in evidence some drawbacks due to the ℓ_1-fitting. A fast, simple and exact optimization method is proposed. By way of application, we separate impulse noise from Gaussian noise in a degraded image.

Keywords Non-smooth analysis · ℓ_1 data-fitting · Image denoising · Signal denoising · Random matrices · Tikhonov regularization · Impulsive noise · Nonsmooth optimization · Numerical methods

1 Introduction

For any positive integer $p > 0$ (e.g. the number of the pixels in an image) we denote

$$I \stackrel{\text{def}}{=} \{1, \ldots, p\}.$$

M. Nikolova (✉)
CMLA, ENS Cachan, CNRS, PRES UniverSud, 61 Av. President Wilson, 94230 Cachan, France
e-mail: nikolova@cmla.ens-cachan.fr

With every index $i \in I$ we associate a subset $N_i \subset I$ such that $\forall i \in I$ and $\forall j \in I$, we have

$$\begin{cases} i \notin N_i, \\ j \in N_i \ \Leftrightarrow \ i \in N_j. \end{cases} \tag{1}$$

Typically, N_i represents the set of the neighbors of i. Thus we call $(N_i)_{i\in I}$ the neighborhood system on I. We consider the following minimization problem: for any $y \in \mathbb{R}^p$,

$$\text{find} \quad \hat{x} \quad \text{such that} \quad \mathcal{F}(\hat{x}, y) = \min_{x\in\mathbb{R}^p} \mathcal{F}(x, y),$$

$$\text{where } \mathcal{F}(x, y) = \sum_{i=1}^{p} |x_i - y_i| + \frac{\alpha}{2}\sum_{i=1}^{p}\sum_{j\in N_i}(x_i - x_j)^2, \tag{2}$$

where $\alpha > 0$ is a parameter. In this paper we study how the solution $\hat{x}$ depends on y.

The practical context we have in mind is signal or image processing where y is the data and $\hat{x}$ is the sought-after signal or image. In the latter case we assume that all pixels of the image are arranged in a p-length vector and that $(N_i)_{i\in I}$ is an usual neighborhood system corresponding to the 4 or the 8 adjacent pixels. (E.g., for a $m \times n$-size picture whose columns are concatenated in a $p = mn$-length vector, the 4 adjacent neighbors of a pixel i in the interior of the image are $N_i = \{i - m, i - 1, i + m, i + 1\}$.) Any other configuration fitting (2) can also be considered. The continuous version of the energy in (2), where α corresponds to a Lagrange multiplier for an underlying constrained optimization problem, was considered in [19] to denoise smooth (H_0^1-regular) images. The authors propose and analyze an active-set method for solving the constrained non-smooth optimization problem.

Regularized energies with ℓ_1-fitting were considered by Alliney in [1, 2] in the context of one-dimensional filters.

Energies involving general non-smooth data-fitting terms were analyzed in [21] where it was shown that the corresponding minimizers $\hat{x}$ satisfy

$$\hat{x}_i = y_i \quad \text{for numerous indexes } i. \tag{3}$$

This result initially suggested applications for impulse noise processing. The geometric properties of images restored using ℓ_1-fitting and total-variation (TV) regularization were studied by Chan and Esedoglu in [9] and applied to the restoration of binary images in [10]. Deblurring under impulse noise using ℓ_1-fitting and different regularization terms was explored in [3–5]. Fast optimization methods for this kind of problems were proposed by [13, 14, 16]. The first two papers, written by Darbon and Sigelle, are based on graph-cuts and suppose that images are quantized. A lot of papers using ℓ_1-fitting with regularization were published in the last years and it is hard to evoke all of them.

In this paper we focus on ℓ_1-fitting with a simple Tikhonov-like regularization for two main reasons: (i) in this context it is possible to obtain a lot of explicit results on the properties of the solution and (ii) we can provide a fast and simple numerical scheme.

Our Contribution

Our main contribution is to show that $\mathcal{F}$, as defined in (2), is minimized by an $\hat{x}$ which is a *locally affine function of the data* y. We exhibit an affine formula giving the result and determine a subset of data where the same formula leads to the minimum of the relevant energy. Such a set is seen to be a polyhedron of $\mathbb{R}^p$ which is unbounded. Since many components of the solution satisfy $\hat{x}_i = y_i$, we decompose the remaining samples ($\hat{x}_i \neq y_i$) into connected components. We show that each connected component is estimated using a matrix with *non-negative entries* applied to the adjacent pixels satisfying $\hat{x}_i = y_i$, plus a fixed vector dependent only on the sign of all $y_i - \hat{x}_i$ in the connected component. An interesting intermediate result is to prove the invertibility and the positivity of this matrix (which is random, since it corresponds to randomly truncated neighborhoods). We also give a very fast and simple minimization scheme which easily recovers the most difficult points—those where the energy is non-differentiable, namely $\hat{x} = y_i$. It is fully explicit (there is no line-search) and involves only sums and comparisons to a fixed threshold. Numerical results on data contaminated with impulse noise are provided in order to illustrate the properties of the energy and the minimization scheme. Other applications can certainly be envisaged.

1.1 Notations and Definitions

We use the symbol $\|\cdot\|$ to denote the ℓ_2-norm of vectors,

$$\|x\| = \sqrt{\langle x, x\rangle}.$$

Next, we denote by $\mathbb{N}^*$ the positive integers. For a subset $\zeta \subset I$, the symbol $\#\zeta$ stands for its cardinality while ζ^c is the complement of ζ in I.

Definition 1 A subset $\zeta \subset I$ is said to be a *connected component with respect to* $(N_i)_{i\in I}$ if either ζ is a singleton, say $\zeta = \{i\}$ and $\zeta \cap N_i = \emptyset$, or if the following hold:

$$\forall (i,j) \in \zeta^2,\ \exists n \in \mathbb{N}^*,\ \exists (i_k)_{k=1}^n \quad \text{such that}$$
$$\begin{cases} i_1 = i \text{ and } i_n = j \\ i_k \in N_{i_{k+1}} \cap \zeta, \quad \forall k = 1,\dots,n-1, \end{cases} \tag{4}$$

$$N_\zeta \stackrel{\text{def}}{=} \left(\bigcup_{i\,\in\,\zeta} N_i\right) \setminus \zeta \subset \zeta^c. \tag{5}$$

By extending (1), we call N_ζ in (5) the neighborhood of ζ. The condition in (4) means that any i and j in ζ are connected by a sequence of neighbors (w.r.t. $(N_i)_{i\in I}$) that belong to ζ. The requirement in (5) means that ζ is maximal—there are no other elements in $I \setminus \zeta$ that can satisfy (4).

Guided by the property in (3), we systematically use the set-valued application J from $\mathbb{R}^p \times \mathbb{R}^p$ to the family of all subsets of I defined by

$$J(x,y) \stackrel{\text{def}}{=} \big\{ i \in I : x_i \neq y_i \big\}. \tag{6}$$

Note that $\mathcal{F}(.,y)$, as defined in (2), is nondifferentiable in the classical sense; it has continuous partial derivatives $\frac{\partial \mathcal{F}(x,y)}{\partial x_i}$ for $i \in J(x,y)$ but not for $i \in J^c(x,y)$. We recall some facts about nondifferentiable functions—see e.g. [18].

Definition 2 The right-side derivative of $\mathcal{F}(.,y)$ at x in the direction of u is defined by

$$\delta\mathcal{F}(x,y)(u) \stackrel{\text{def}}{=} \lim_{\varepsilon \searrow 0} \frac{\mathcal{F}(x+\varepsilon u, y) - \mathcal{F}(x,y)}{\varepsilon},$$

whenever this limit exists.

This one-sided derivative always exists for the energy in (2) since it is convex and continuous. Let us remind that the relevant left-side derivative is $-\delta\mathcal{F}(x,y)(-u)$. If $\mathcal{F}$ is nonsmooth at x along u, we have $\delta\mathcal{F}(x,y)(-u) \neq -\delta\mathcal{F}(x,y)(u)$ and by the convexity of $\mathcal{F}$ the following inequality holds: $-\delta\mathcal{F}(x,y)(-u) < \delta\mathcal{F}(x,y)(u)$. More details can be found e.g. in [18]. Otherwise, if $\mathcal{F}$ is smooth at x along u, we just have $-\delta\mathcal{F}(x,y)(-u) = \delta\mathcal{F}(x,y)(u)$. Finally, if $\mathcal{F}(.,y)$ is differentiable at x (in the classical sense), we have $\delta\mathcal{F}(x,y)(u) = \langle \nabla_x \mathcal{F}(x,y), u \rangle$ which is linear in u.

The components of a matrix A are denoted by $A(i,j)$. Given a vector $y \in \mathbb{R}^p$, its ith component is denoted by y_i or $y(i)$. We will write $y \geq 0$ to say that all components of

y are nonnegative. If $K \subset I$ is an ordered subset of indexes, say $K = (k_1, \ldots, k_{\#K})$, then y_K is the restriction of y to the indexes contained in K:

$$[y_K](i) = y_{k_i}, \quad \forall i = 1, \ldots, \#K. \tag{7}$$

For definiteness, we suppose in what follows that all subsequences are arranged in increasing order. For any $i = 1, \ldots, p$, the symbol e^i systematically denotes the ith vector of the canonical basis of $\mathbb{R}^p$, namely $e^i(i) = 1$ and $e^i(j) = 0$, $\forall j \neq i$, for all $i = 1, \ldots, p$.

2 Necessary and Sufficient Conditions for a Minimum

Since $\mathcal{F}(., y)$ is continuous, convex and coercive for every $y \in \mathbb{R}^p$, it always has a unique minimum and the latter is reached on a convex set. See [18, 24] for details. First we determine the necessary and sufficient conditions for $\mathcal{F}(., y)$ to have a minimum at $\hat{x}$.

Proposition 1 *For $y \in \mathbb{R}^p$, the function $\mathcal{F}(., y)$ given in (2) reaches its minimum at $\hat{x} \in \mathbb{R}^p$ if and only if*

$$\left| \hat{x}_i - \frac{1}{\#N_i} \sum_{j \in N_i} \hat{x}_j \right| \le \frac{1}{2\alpha \#N_i} \quad \forall i \in \hat{J}^c, \tag{8}$$

$$\begin{cases} \hat{x}_i - \dfrac{1}{\#N_i} \displaystyle\sum_{j \in N_i} \hat{x}_j = \dfrac{\sigma_i}{2\alpha \#N_i} & \forall i \in \hat{J}, \\ \text{for } \sigma_i = \operatorname{sign}(y_i - \hat{x}_i) \end{cases} \tag{9}$$

where $\hat{J} \stackrel{\text{def}}{=} J(\hat{x}, y)$ is defined according to (6).

Remark 1 It is worth emphasizing that from the definition of $\hat{J}$, (8) corresponds to

$$\hat{x}_i = y_i, \quad \forall i \in \hat{J}^c.$$

Remark 2 Equations (8) and (9) show that the restored $\hat{x}$ (e.g. an image or a signal) involves a firm bound on the difference between each sample $\hat{x}_i$ and the mean of its neighbors:

$$\left| \hat{x}_i - \frac{1}{\#N_i} \sum_{j \in N_i} \hat{x}_j \right| \le \frac{1}{2\alpha \#N_i}, \quad \forall i \in I.$$

According to the value of α, textures or non-spiky noises can be preserved in the solution $\hat{x}$.

Proof Being convex, $\mathcal{F}(., y)$ reaches its minimum at $\hat{x}$ if and only if (see e.g. [18], vol. I, Theorem 2.2.1 on p. 253)

$$\delta\mathcal{F}(\hat{x}, y)(u) \ge 0, \quad \forall u \in \mathbb{R}^p, \tag{10}$$

where $\delta\mathcal{F}(\hat{x}, y)(u)$ is the right-side derivative of $\mathcal{F}(., y)$ at $\hat{x}$ in the direction of u, see Definition 2. Let us decompose $\mathcal{F}(\hat{x}, y)$ in the following way:

$$\mathcal{F}(\hat{x}, y) = \sum_{i \in \hat{J}^c} f_{y_i}(\hat{x}_i) + \Phi_{\hat{J}^c}(\hat{x}) + \tilde{\mathcal{F}}(\hat{x}),$$

where

$$f_{y_i}(x_i) = |x_i - y_i|,$$
$$\Phi_{\hat{J}^c}(x) = \frac{\alpha}{2} \sum_{i \in \hat{J}^c} \sum_{j \in N_i} (x_i - x_j)^2,$$
$$\tilde{\mathcal{F}}(x) = \sum_{i \in \hat{J}} \left(|x_i - y_i| + \frac{\alpha}{2} \sum_{j \in N_i} (x_i - x_j)^2 \right).$$

Clearly, f_{y_i} is nondifferentiable at $\hat{x}_i$ for every $i \in \hat{J}^c$ (since $\hat{x}_i = y_i$) whereas $\Phi_{\hat{J}^c}$ and $\tilde{\mathcal{F}}$ are differentiable at $\hat{x}$. Using Definition 2, for any $u \in \mathbb{R}^p$,

$$\delta\mathcal{F}(\hat{x}, y)(u) = \sum_{i \in \hat{J}^c} \left(\delta f_{y_i}(\hat{x}_i)(u_i) + u_i \left. \frac{\partial \Phi_{\hat{J}^c}(x)}{\partial x_i} \right|_{x=\hat{x}} \right) + \sum_{i \in \hat{J}} u_i \left. \frac{\partial \tilde{\mathcal{F}}(x)}{\partial x_i} \right|_{x=\hat{x}}. \tag{11}$$

Next we calculate the terms involved in these sums. For every $i \in \hat{J}^c$ we have

$$\delta f_{y_i}(\hat{x}_i)(u_i) = \lim_{\varepsilon \searrow 0} \frac{|\hat{x}_i + \varepsilon u_i - y_i| - |\hat{x}_i - y_i|}{\varepsilon} = |u_i| = u_i \operatorname{sign}(u_i).$$

It is easy to find that

$$\frac{\partial \Phi_{\hat{J}^c}(x)}{\partial x_i} = 2\alpha \#N_i \left(x_i - \frac{1}{\#N_i} \sum_{j \in N_i} x_j \right).$$

Note that the constant 2 comes from the regularization term in (2) which involves both $(x_i - x_j)^2$ and $(x_j - x_i)^2$.

For every $i \in \hat{J}$, we have $\hat{x}_i \neq y_i$ and the partial derivative below is well defined:

$$\left. \frac{\partial \tilde{\mathcal{F}}(x)}{\partial x_i} \right|_{x=\hat{x}} = \operatorname{sign}(\hat{x}_i - y_i) + 2\alpha \#N_i \left(\hat{x}_i - \frac{1}{\#N_i} \sum_{j \in N_i} \hat{x}_j \right).$$

Introducing these results in the expression for $\delta\mathcal{F}(\hat{x}, y)$ in (11) yields:

$$\delta\mathcal{F}(\hat{x}, y)(u) = \sum_{i \in \hat{J}^c} u_i \left(\operatorname{sign}(u_i) + 2\alpha \#N_i \left(\hat{x}_i - \frac{1}{\#N_i} \sum_{j \in N_i} \hat{x}_j \right) \right) + \sum_{i \in \hat{J}} u_i \left(\operatorname{sign}(\hat{x}_i - y_i) + 2\alpha \#N_i \left(\hat{x}_i - \frac{1}{\#N_i} \sum_{j \in N_i} \hat{x}_j \right) \right). \tag{12}$$

If $\hat{J}^c$ is empty, the first sum in (12) is null and (8) is void. Consider next that $\hat{J}^c$ is nonempty. Let us apply (10) to the expression of $\delta\mathcal{F}(\hat{x}, y)$ in (12) with e^i and $-e^i$ (as defined at the end of Sect. 1.1) for $i \in \hat{J}^c$:

$$\delta\mathcal{F}(\hat{x}, y)(e^i) \geq 0, i \in \hat{J}^c$$
$$\Rightarrow 1 + 2\alpha\#N_i\Big(\hat{x}_i - \frac{1}{\#N_i}\sum_{j\in N_i}\hat{x}_j\Big) \geq 0,$$
$$\delta\mathcal{F}(\hat{x}, y)(-e^i) \geq 0,\ \ i \in \hat{J}^c$$
$$\Rightarrow 1 - 2\alpha\#N_i\Big(\hat{x}_i - \frac{1}{\#N_i}\sum_{j\in N_i}\hat{x}_j\Big) \geq 0,$$

since $e^i\operatorname{sign}(e^i) = -e^i\operatorname{sign}(-e^i) = 1$. Combining these results leads to

$$-1 \leq 2\alpha\#N_i\Big(\hat{x}_i - \frac{1}{\#N_i}\sum_{j\in N_i}\hat{x}_j\Big) \leq 1.$$

Hence (8).

If $\hat{J}$ is empty, the second sum in (12) is null and (9) is void. Consider next that $\hat{J}$ is nonempty. Let us apply (10) to the expression of $\delta\mathcal{F}(\hat{x}, y)$ in (12) with e^i and $-e^i$ for $i \in \hat{J}$:

$$\delta\mathcal{F}(\hat{x}, y)(e^i) \geq 0,\ \ i \in \hat{J}$$
$$\Rightarrow \operatorname{sign}(\hat{x}_i - y_i) + 2\alpha\#N_i\Big(\hat{x}_i - \frac{1}{\#N_i}\sum_{j\in N_i}\hat{x}_j\Big) \geq 0,$$
$$\delta\mathcal{F}(\hat{x}, y)(-e^i) \geq 0,\ \ i \in \hat{J}$$
$$\Rightarrow -\operatorname{sign}(\hat{x}_i - y_i) - 2\alpha\#N_i\Big(\hat{x}_i - \frac{1}{\#N_i}\sum_{j\in N_i}\hat{x}_j\Big) \geq 0.$$

Combining the last two results shows that

$$\operatorname{sign}(\hat{x}_i - y_i) + 2\alpha\#N_i\Big(\hat{x}_i - \frac{1}{\#N_i}\sum_{j\in N_i}\hat{x}_j\Big) = 0,$$

which is equivalent to

$$\hat{x}_i - \frac{1}{\#N_i}\sum_{j\in N_i}\hat{x}_j - \frac{\operatorname{sign}(y_i - \hat{x}_i)}{2\alpha\#N_i} = 0, \quad \forall i \in \hat{J}. \tag{13}$$

This amounts to (9). □

Lemma 1 *We assume the conditions of Proposition 1. The constant $\sigma_i \in \{-1, +1\}$ in (9) satisfies*

$$\sigma_i = \operatorname{sign}(y_i - \hat{x}_i) = \operatorname{sign}\Big(y_i - \frac{1}{\#N_i}\sum_{j\in N_i}\hat{x}_j\Big), \quad \forall i \in \hat{J}. \tag{14}$$

Proof We know that $\forall i \in \hat{J}$ we have

$$\text{either} \quad y_i - \frac{1}{\#N_i}\sum_{j\in N_i}\hat{x}_j < -\frac{1}{2\alpha\#N_i} < 0$$
$$\text{or} \quad y_i - \frac{1}{\#N_i}\sum_{j\in N_i}\hat{x}_j > \frac{1}{2\alpha\#N_i} > 0,$$

since otherwise we would find $\hat{x}_i = y_i$ and $i \in \hat{J}^c$ according to (8). These two cases are considered below.

- Consider that $y_i - \frac{1}{\#N_i}\sum_{j\in N_i}\hat{x}_j < 0$. Subtracting (13) from the latter inequality yields

$$y_i - \frac{1}{\#N_i}\sum_{j\in N_i}\hat{x}_j - \Big(\hat{x}_i - \frac{1}{\#N_i}\sum_{j\in N_i}\hat{x}_j - \frac{\operatorname{sign}(y_i - \hat{x}_i)}{2\alpha\#N_i}\Big) < 0.$$

  Using that $u = |u|\operatorname{sign}(u)$ for any $u \in \mathbb{R}$, this is equivalent to

$$\operatorname{sign}(y_i - \hat{x}_i)\Big(|\hat{x}_i - y_i| + \frac{1}{2\alpha\#N_i}\Big) < 0.$$

  Since the expression between the big parentheses is positive, we find that

$$\sigma_i = \operatorname{sign}(y_i - \hat{x}_i) = \operatorname{sign}\Big(y_i - \frac{1}{\#N_i}\sum_{j\in N_i}\hat{x}_j\Big) = -1.$$

- Consider that $y_i - \frac{1}{\#N_i}\sum_{j\in N_i}\hat{x}_j > 0$. Subtracting (13) from the latter inequality yields

$$y_i - \frac{1}{\#N_i}\sum_{j\in N_i}\hat{x}_j - \Big(\hat{x}_i - \frac{1}{\#N_i}\sum_{j\in N_i}\hat{x}_j - \frac{\operatorname{sign}(y_i - \hat{x}_i)}{2\alpha\#N_i}\Big) > 0.$$

  This is equivalent to

$$\operatorname{sign}(y_i - \hat{x}_i)\Big(|\hat{x}_i - y_i| + \frac{1}{2\alpha\#N_i}\Big) > 0.$$

  It follows that

$$\sigma_i = \operatorname{sign}(y_i - \hat{x}_i) = \operatorname{sign}\Big(y_i - \frac{1}{\#N_i}\sum_{j\in N_i}\hat{x}_j\Big) = 1.$$

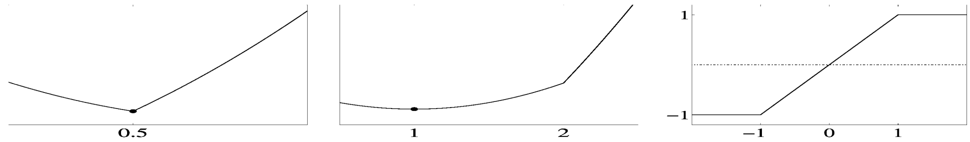
The proof is complete. □

Now we can state a more handy formulation of the minimality condition given in Proposition 1.

Theorem 1 *For $y \in \mathbb{R}^p$, the function $\mathcal{F}(., y)$ given in (2) reaches its minimum at $\hat{x} \in \mathbb{R}^p$ if and only if*

$$\Big|y_i - \frac{1}{\#N_i}\sum_{j\in N_i}\hat{x}_j\Big| \leq \frac{1}{2\alpha\#N_i} \quad \forall i \in \hat{J}^c, \tag{15}$$

Fig. 1 Illustration of Example 1



$\mathcal{F}(.,y)$ in (18) for $\alpha = 1$, $y = 0.5$. By (19), $\hat{x} = y = 0.5$.
(a)

$\mathcal{F}(.,y)$ in (18) for $\alpha = 1$, $y = 2$. By (19), $\hat{x} = \frac{\text{sign}(y)}{\alpha} = 1$.
(b)

$y \to \hat{x}(y)$ according to (19) for $\alpha = 1$.
(c)

$$\begin{cases} \hat{x}_i - \dfrac{1}{\#N_i} \displaystyle\sum_{j \in N_i \cap \hat{J}} \hat{x}_j = \dfrac{1}{\#N_i} \displaystyle\sum_{j \in N_i \cap \hat{J}^c} y_j + \dfrac{\sigma_i}{2\alpha \#N_i} & \forall i \in \hat{J}, \\ \textit{for } \sigma_i = \text{sign}\Big(y_i - \frac{1}{\#N_i} \sum_{j \in N_i} \hat{x}_j \Big) \end{cases} \tag{16}$$

where $\hat{J} \overset{\text{def}}{=} J(\hat{x}, y)$ *is defined according to* (6).

Proof Using Remark 1, we obtain (15) directly from (8) and split the first equation in (9) to obtain the first equation in (16). Then we use the expression for σ_i derived in Lemma 1. □

Remark 3 Using (8) in Proposition 1, it is easy to see that

$$\big[\hat{J} = \emptyset\big] \iff \big[\hat{J}^c = I\big] \iff \big[\hat{x}_i = y_i,\ \forall i \in I\big] \iff \big[y \in W_I\big],$$

where

$$W_I \overset{\text{def}}{=} \left\{ y \in \mathbb{R}^p : \left| y_i - \frac{1}{\#N_i} \sum_{j \in N_i} y_j \right| \le \frac{1}{2\alpha \#N_i}, \quad \forall i \in I \right\}. \tag{17}$$

Obviously, W_I is a polyhedron enclosed between $p = \#I$ pairs of affine hyperplanes in $\mathbb{R}^p$. Its Lebesgue measure in $\mathbb{R}^p$ is clearly positive. Obtaining a solution $\hat{x} = y$ is useless, so α should be such that $y \notin W_I$, hence we need

$$\alpha > \alpha_{\min},$$

where

$$\alpha_{\min} \overset{\text{def}}{=} \min_{i \in I} \left(2\#N_i \left| y_i - \frac{1}{\#N_i} \sum_{j \in N_i} y_j \right| \right)^{-1}.$$

Example 1 Consider the cost-function $\mathcal{F} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$,

$$\mathcal{F}(x, y) = |x - y| + \frac{\alpha}{2} x^2, \tag{18}$$

illustrated in Fig. 1(a) and (b). For any $y \in \mathbb{R}$, the energy $\mathcal{F}(., y)$ above is minimized by

$$\hat{x} = \begin{cases} y & \text{if } |y| \le \dfrac{1}{\alpha}, \\ \dfrac{\text{sign}(y)}{\alpha} & \text{if } |y| > \dfrac{1}{\alpha}. \end{cases} \tag{19}$$

The solution $\hat{x}$ as a function of y is plotted in Fig. 1(c).

- observe that $\hat{x}$ fits the data y whenever $|y| \le 1/\alpha$;
- otherwise, for $|y| \ge \frac{1}{\alpha}$, there is a threshold effect, so the value of $\hat{x}$ is independent of the exact value of y, it depends only on its sign.

The function $y \to \hat{x}$ is linear on each one of the subsets $(-\infty, -\frac{1}{\alpha})$, $[-\frac{1}{\alpha}, \frac{1}{\alpha}]$ and $(\frac{1}{\alpha}, +\infty)$.

Example 2 is *quite pathological* since under some conditions, the minimizer is not unique.

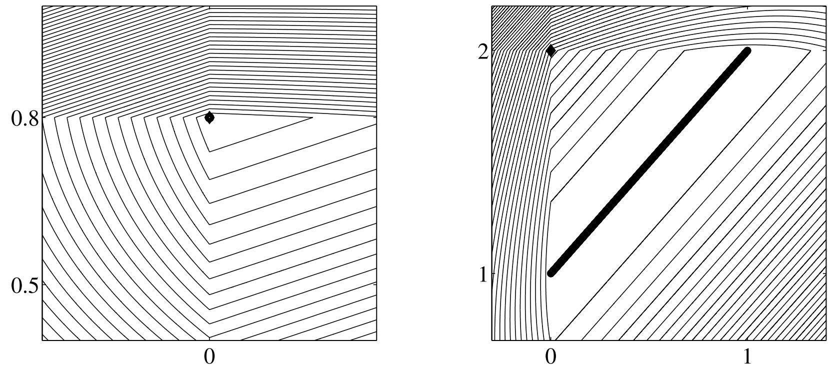
Example 2 Consider the cost-function $\mathcal{F} : \mathbb{R}^2 \times \mathbb{R}^2_+ \to \mathbb{R}$ of the form (2),

$$\begin{aligned} \mathcal{F}(x, y) &= \sum_{i=1}^{2} |x_i - y_i| + \frac{\alpha}{2} \sum_{i=1}^{2} \sum_{j \in N_i} (x_i - x_j)^2 \\ &= |x_1 - y_1| + |x_2 - y_2| + \alpha (x_1 - x_2)^2, \end{aligned} \tag{20}$$

where the simplification comes from the facts that $N_1 = \{2\}$ and $N_2 = \{1\}$, and hence $\#N_i = 1$ for $i = 1, 2$. Several cases arise according to the values of y_1 and y_2.

- If $|y_1 - y_2| \le \frac{1}{2\alpha}$, Proposition 1 shows that $\hat{x}_1 = y_1$ and $\hat{x}_2 = y_2$, so $\hat{J} = \emptyset$. Such a case is illustrated on Fig. 2(a).
- For definiteness, let $y_2 - y_1 > \frac{1}{2\alpha}$. Using Proposition 1, we find that the solution is given by the segment $[(y_1, y_1 + \frac{1}{2\alpha}), (y_2 - \frac{1}{2\alpha}, y_2)]$ (see Fig. 2(b)). Its extreme points correspond to $\hat{J}^c = \{1\}$ and $\hat{J}^c = \{2\}$. This pathological behavior is due to the fact that in the interior of the segment, $\hat{x}_1$ and $\hat{x}_2$ are calculated using (9) which is simplified to a unique equation, $1 + 2\alpha(\hat{x}_1 - \hat{x}_2) = 0$. Notice also that the connected component ζ (see Definition 1) reads $\zeta = \hat{J} = I = \{1, 2\}$.

More generally, if $\hat{J} = I$, then $\hat{x}$ is calculated using (9) only and the minimizer may not be unique. As seen in Sect. 7



$(\hat{x}_1, \hat{x}_2) = (y_1, y_2) = (0, 0.8).$
(a)

$(y_1, y_2) = (0, 2),\ \{\hat{x}\} = [(0,1),(1,2)]$
(b)

Fig. 2 Both figures present the level lines of $\mathcal{F}$ in (20) for $\alpha = 0.5$ near the data and the minimum. In both figures, the data point y is plotted with a "$\diamond$". On the left, the minimizer $\hat{x}$ match the data y. *On the right*, the minimizer is any element of the segment $[(0,1),(1,2)]$ (plotted with a *thick line*)

(Experiments), the case when $\hat{J}^c$ is empty (i.e. $\hat{J} = I$) is exceptional; otherwise minimizers are well defined, and hopefully, interesting for practical applications.

Remark 4 We know from Proposition 1 in [22] that if $\hat{J}^c$ is nonempty and if the condition in (8) involves at least one strict inequality, then the minimizer $\hat{x}$ of $\mathcal{F}$ is *unique*. We can conjecture that the cases when $\hat{J}^c$ is empty or (8) involves only equalities are quite exceptional in practice. If for some $y \in \mathbb{R}^p$ the set $\hat{J}^c$ is empty, then the minimality conditions are given by (9) for all $i \in I$. Since $\mathcal{F}(., y)$ admits a minimizer for every $y \in \mathbb{R}^p$, this system admits at least one solution; the latter it is not necessarily unique. In any case, such a solution has no practical interest since it is quasi-independent of the data; it also indicates that the parameter α is too large.

3 Random Matrices from Restricted Neighborhoods

Lemma 2 *Let $\zeta = \{k_1, \dots, k_q\} \subset I$, where $q \overset{\text{def}}{=} \#\zeta < p$, be any connected component of I w.r.t $(N_i)_{i \in I}$ (see (1) and Definition 1). Then the $q \times q$ matrix L given below*:

$$L(i,j) = \begin{cases} 1 & \text{if } i = j, \\ -\dfrac{1}{\#N_{k_i}} & \text{if } k_j \in N_{k_i} \cap \zeta, \\ 0 & \text{otherwise}, \end{cases} \tag{21}$$

is invertible and its inverse has non-negative entries, i.e.

$$(L)^{-1}(i,j) \geq 0, \quad \forall i, j. \tag{22}$$

Remark 5 Observe that $L(i,i) = 1$ for all i while for $i \neq j$ we have

$$L_{i,j} < 0 \quad \Longleftrightarrow \quad L_{j,i} < 0.$$

Indeed, if $k_i \in \zeta$ and $j \in N_{k_i} \cap \zeta$, we find $L(i,j) = -(\#N_{k_i})^{-1}$. By (1), we also have $k_i \in N_{k_j} \cap \zeta$, hence $L(j,i) = -(\#N_{k_j})^{-1}$. Notice that we can have $L_{i,j} \neq L_{j,i}$ since different neighborhoods can have different sizes.

A matrix of the form (21) is presented in the example below.

Example 3 Consider an image where for every pixel i which is not at the boundary of the image, N_i is composed of its 8 adjacent neighbors. However, the pixels at the boundaries have less neighbors. Consider a connected component $\zeta = \{k_1, k_2, k_3, k_4, k_5\}$ as presented below in (a):

$$\begin{array}{cccccc}
 & & & \cdots & & \\
 & \circ & - & \boxed{k_2} & - & \circ \\
\cdots & | & \times & | & \times & | \\
 & \boxed{k_1} & - & \boxed{k_3} & - & \circ \\
 & | & \times & | & \times & | \\
 & \circ & - & \boxed{k_4} & - & \boxed{k_5}
\end{array}$$

(a) Connected component ζ at the lower right end of the image.

$$L = \begin{bmatrix} 1 & -1/8 & -1/8 & -1/8 & 0 \\ -1/8 & 1 & -1/8 & 0 & 0 \\ -1/8 & -1/8 & 1 & -1/8 & -1/8 \\ \mathbf{-1/5} & 0 & \mathbf{-1/5} & 1 & \mathbf{-1/5} \\ 0 & 0 & \mathbf{-1/3} & \mathbf{-1/3} & 1 \end{bmatrix}$$

(b) The matrix L according to (21).

Observe that k_5 is on the last column and on the last row of the image, and that k_4 is on its last row. Then

$$\#N_{k_4} = 5 \quad \text{and} \quad \#N_{k_5} = 3, \quad \text{while } \#N_{k_i} = 8,\ i = 1, 2, 3.$$

We have: $N_{k_1} \cap \zeta = \{k_2, k_3, k_4\}$, $N_{k_2} \cap \zeta = \{k_1, k_3\}$, $N_{k_3} \cap \zeta = \{k_1, k_2, k_4, k_5\}$, $N_{k_4} \cap \zeta = \{k_1, k_3, k_5\}$ and $N_{k_5} \cap \zeta = \{k_3, k_4\}$. The matrix L corresponding to (21) is presented in (b).

Proof The proof of the lemma is based on the following equivalence result[1] [11]: a matrix $L \in \mathbb{R}^{q\times q}$ is invertible and satisfies (22) *if and only if* the implication below holds true:

$$[Lv \geq 0] \quad \Rightarrow \quad [v \geq 0], \tag{23}$$

where the expression $v \geq 0$ means that $v_i \geq 0$, for every $i = 1, \dots, q$. In what follows, we will prove that the matrix L defined by (21) satisfies (23). Notice that

$$[Lv]_i = v_i - \frac{1}{\#N_{k_i}} \sum_{j \in N_{k_i} \cap \zeta} v_j. \tag{24}$$

Below we show several preliminary implications that will help us to prove (23).

- Let $b < 0$ and $u \in \mathbb{R}^r$ for $r \in \mathbb{N}^*$, we have the implication

$$\left[b - \sum_{i=1}^{r} u_i \geq 0 \right] \Rightarrow \left[\exists j \in \{1, \dots, r\} : u_j \leq \frac{1}{r} b < 0 \right]. \tag{25}$$

  Indeed, if the statement on the right is false, i.e. $u_i > \frac{b}{r}$ for all i, then

$$\sum_{i=1}^{r} u_i > \sum_{i=1}^{r} \frac{b}{r} = b.$$

  Then the inequality on the left hand side cannot be satisfied.
- Let us prove the following implication:

$$\begin{bmatrix} a \stackrel{\text{def}}{=} v_i < 0 \text{ and } N_{k_i} \subset \zeta \\ \text{AND} \\ [Lv]_i \geq 0 \end{bmatrix} \Rightarrow \begin{bmatrix} v_j = a < 0, \quad \forall j \in N_{k_i} \cup \{k_i\} & (i) \\ \text{OR} & \\ \exists j \in N_{k_i} \text{ such that } v_j < v_i = a < 0 & (ii) \end{bmatrix}. \tag{26}$$

  By (25), $\exists j_1 \in N_{k_i}$ such that

$$\frac{1}{\#N_{k_i}} v_{j_1} \leq \frac{1}{\#N_{k_i}} v_i.$$

  If $v_{j_1} < v_i$, then (26)(ii) holds.
  Else $v_{j_1} = v_i = a < 0$. We can write down that

$$\left(1 - \frac{1}{\#N_{k_i}}\right) v_i - \frac{1}{\#N_{k_i}} \sum_{j \in N_{k_i} \setminus \{j_1\}} v_j \geq 0.$$

  Using (25) yet again, $\exists j_2 \in N_{k_i} \setminus \{j_1\}$ such that

$$\frac{1}{\#N_{k_i}} v_{j_2} \leq \frac{1}{\#N_{k_i} - 1} \left(1 - \frac{1}{\#N_{k_i}}\right) v_i.$$

  Hence

$$v_{j_2} \leq v_{k_i} = v_{j_1} = a < 0.$$

  If $v_{j_2} < v_i$, then (26)(ii) holds. Otherwise, $v_{j_2} = v_i = v_{j_1} = a < 0$.
  Iterating this reasoning $\#N_{k_i}$ times shows the result.
- We have a reciprocal of (i) in (26):

$$\begin{bmatrix} v_j = a < 0, \quad \forall j \in N_{k_i} \cup \{k_i\} \\ \text{AND} \\ [Lv]_i \geq 0 \end{bmatrix} \Rightarrow \left[N_{k_i} \subset \zeta \right]. \tag{27}$$

  Suppose that the right side of (27) is false, i.e. $\#\big(N_{k_i} \cap \zeta\big) \leq \#N_{k_i} - 1$ then

$$[Lv]_i = a - \frac{\#(N_{k_i} \cap \zeta)}{\#N_{k_i}} a = a \frac{\#N_{k_i} - \#(N_{k_i} \cap \zeta)}{\#N_{k_i}} < 0$$

  where the last inequality comes from the fact that the fraction is strictly positive.
- We have the following implication:

$$\begin{aligned} &\left[a \stackrel{\text{def}}{=} v_i < 0 \text{ and } N_{k_i} \cap \zeta \neq N_{k_i} \right] \\ &\quad \Rightarrow \quad \left[\exists j \in N_{k_i} : v_j < v_i = a < 0 \right]. \end{aligned} \tag{28}$$

  Notice that $\#(N_{k_i} \cap \zeta) < \#N_{k_i}$. Then using (25), $\exists j \in N_{k_i}$ such that

$$\frac{1}{\#N_{k_i}} v_j \leq \frac{1}{\#(N_{k_i} \cap \zeta)} v_i$$

  which shows (28).

With the help of these implications, we will prove (23) by contradiction. So, suppose that $Lv \geq 0$ but that

$$\exists j_1 \in \{1, \dots, q\} \quad \text{such that} \quad v_{j_1} \stackrel{\text{def}}{=} a < 0.$$

[1] For completeness, we remind the proof given in [11]. If L is invertible and (22) is true, and $Lv \geq 0$, then clearly $v = L^{-1}(Lv) \geq 0$. Conversely, let (23) holds. If $Lv = 0$ then $L(-v) = 0$, hence $v \geq 0$ and $-v \geq 0$, i.e. $v = 0$ which shows that L is invertible. Since the ith column of L^{-1} is $v_i \stackrel{\text{def}}{=} L^{-1}e^i$, we have $Lv_i = e^i \geq 0$ and hence $v_i \geq 0$. The same holds true for all columns v_i, $1 \leq i \leq q$, hence $(L)^{-1}(i, j) \geq 0$, $\forall (i, j)$.

According to N_{j_1} (either $N_{j_1} \subset \zeta$ or $N_{j_1} \cap \zeta \neq N_{j_1}$) we apply (26) or (28) and find that

$$\exists j_2 \in N_{j_1} \quad \text{such that} \quad v_{j_2} \leq v_{j_1} \stackrel{\text{def}}{=} a < 0.$$

Since ζ is connected, we can in the same way visit all N_{j_i} for $i \in \{1, \ldots, q\}$ and thus find a decreasing sequence, say of length $n \leq q$, denoted $(v_{j_i})_{i=1}^n$ whose elements are $\leq a < 0$. Since $\zeta \neq I$, there exists $k \leq n$ and $j_k \in \zeta$ such that $N_{j_k} \cap \zeta \neq N_{j_k}$ in which case (27) implies that $v_{j_k} < v_{j_{k-1}}$. We have thus found that

$$b \stackrel{\text{def}}{=} v_{j_n} \leq \cdots \leq v_{j_k} < v_{j_{k-1}} \leq \cdots \leq j_1 \leq v_{j_1} = a < 0.$$

Since we have visited all N_{j_i} for $i \in \{1, \ldots, q\}$, we can write that

$$b = \min_{1 \leq i \leq q} v_i. \tag{29}$$

Consider now N_{j_n}. By (29),

$$v_j \geq v_{j_n}, \quad \forall j \in N_{j_n}. \tag{30}$$

If we had

$$v_{j_n} = v_j = b, \quad \forall j \in N_{j_n},$$

then (27) shows that $N_{j_n} \subset \zeta$ in which case we can again iterate (26) and (28), hence we have not visited all N_{j_i} for $i \in \{1, \ldots, q\}$. If follows that

$$\exists j \in N_{k_{j_n}} \quad \text{such that} \quad v_j > b. \tag{31}$$

Using (30) and (31), we can write down the following:

$$\begin{aligned}
\left[Lv\right]_{j_n} &= v_{j_n} - \frac{1}{\#N_{k_{j_n}}} \sum_{j \in N_{k_{j_n}} \cap \zeta} v_j \\
&= b - \frac{1}{\#N_{k_{j_n}}} \sum_{j \in N_{k_{j_n}} \cap \zeta} v_j \\
&< b - \frac{\#(N_{k_{j_n}} \cap \zeta)}{\#N_{k_{j_n}}} b \\
&= b \frac{\#N_{k_{j_n}} - \#(N_{k_{j_n}} \cap \zeta)}{\#N_{k_{j_n}}} \\
&\leq 0.
\end{aligned}$$

Thus $\left[Lv\right]_{j_n} < 0$ which contradicts the assumption that $Lv \geq 0$. It follows that (23) is true. The proof is complete. □

4 Semi-explicit Expression for Minimizers

Given $\hat{x}$—a minimizer of $\mathcal{F}(., y)$—let $\zeta = \{i\} \subset \hat{J}$ be a connected component with respect to $(N_i)_i$, in the sense of Definition 1, which in particular implies that $N_i \subset \hat{J}^c$. Then Theorem 1 tells us that

$$\hat{x}_i = \frac{1}{\#N_i} \sum_{j \in N_i} y_j + \frac{\sigma_i}{2\alpha \#N_i}, \tag{32}$$

where $\sigma_i \in \{-1, +1\}$ is given by Lemma 1 and now reads $\sigma_i = \operatorname{sign}(y_i - \frac{1}{\#N_i} \sum_{j \in N_i} y_j)$.

A more general result, considering arbitrary connected components ζ is presented next. Since Lemma 2 holds for connected components ζ such that $\#\zeta < p$, we will exclude the case when $\hat{J}^c$ is empty (i.e. $\hat{J} = I$ and the unique connected component is $\zeta = I$, hence $\#\zeta = p$).

Proposition 2 *For $y \in \mathbb{R}^p$, let $\mathcal{F}(., y)$ reach its minimum at $\hat{x}$ with $\hat{J} = J(\hat{x}, y) \neq \emptyset$, $\hat{J} \neq I$, where J is defined in (6). Let*

$$\zeta = \{k_1, \ldots, k_q\} \subset \hat{J} \tag{33}$$

be any connected component w.r.t. $(N_i)_{i \in I}$ (see Definition 1) and its neighborhood read $N_\zeta = \{n_1, \ldots, n_{\#N_\zeta}\}$. Let us define $L^\zeta \in \mathbb{R}^{q \times q}$, $Q^\zeta \in \mathbb{R}^{q \times \#N_\zeta}$ and $d^\zeta \in \mathbb{R}^q$ as it follows:

$$L^\zeta(i, j) = \begin{cases} 1 & \text{if } i = j \\ -\dfrac{1}{\#N_{k_i}} & \text{if } k_j \in N_{k_i} \cap \zeta, \\ 0 & \text{otherwise}, \end{cases} \quad 1 \leq i, j \leq q; \tag{34}$$

$$Q^\zeta(i, j) = \begin{cases} \dfrac{1}{\#N_{k_i}} & \text{if } n_j \in N_{k_i} \cap N_\zeta, \\ 0 & \text{otherwise}, \end{cases}$$
$$1 \leq i \leq q, \ 1 \leq j \leq \#N_\zeta; \tag{35}$$

$$d^\zeta(i) = \frac{\sigma_{k_i}}{2\alpha \#N_{k_i}}, \quad \sigma_{k_i} = \operatorname{sign}(y_{k_i} - \hat{x}_{k_i}), \ 1 \leq i \leq q. \tag{36}$$

Then $\hat{x}_\zeta$ reads

$$\hat{x}_\zeta = A^\zeta y_{N_\zeta} + b^\zeta,$$

where

(a) *the matrix $A^\zeta \in \mathbb{R}^{q \times \#N_\zeta}$ satisfies $A^\zeta(i, j) \geq 0$, $\forall i, j$ and reads $A^\zeta = (L^\zeta)^{-1} Q^\zeta$;*
(b) *the vector $b^\zeta \in \mathbb{R}^q$ reads $b^\zeta = (L^\zeta)^{-1} d^\zeta$.*

It is easy to see that A^ζ depends only on $\{N_i : i \in \zeta \cup N_\zeta\}$, and that b^ζ depends only on $\{N_i : i \in \zeta \cup N_\zeta\}$ and $\{\sigma_i : i \in \zeta\}$.

Proof Since $\zeta \subset \hat{J}$, (9) in Proposition 1 shows that the entries of $\hat{x}_\zeta$ satisfy

$$\hat{x}_{k_i} - \frac{1}{\#N_{k_i}} \sum_{j \in N_{k_i}} \hat{x}_j = \frac{\sigma_{k_i}}{2\alpha \#N_{k_i}}, \quad \forall i = 1, \ldots, q. \tag{37}$$

For every $k_i \in \zeta$, we can decompose N_{k_i} as

$$N_{k_i} = (N_{k_i} \cap \zeta) \cup (N_{k_i} \cap N_\zeta).$$

Since $N_\zeta \subset \hat{J}^c$ (see Definition 1), we have $\hat{x}_j = y_j$ for every $j \in N_\zeta$. Introducing this into (37) yields

$$\hat{x}_{k_i} - \frac{1}{\#N_{k_i}} \sum_{k_j \in N_{k_i} \cap \zeta} \hat{x}_{k_j} = \frac{1}{\#N_{k_i}} \sum_{n_j \in N_{k_i} \cap N_\zeta} y_j + \frac{\sigma_{k_i}}{2\alpha \# N_{k_i}},$$
$$\forall i = 1, \dots, q. \tag{38}$$

This is a system of q affine equations with q unknowns, which can be expressed in matrix form:

$$L^\zeta \hat{x}_\zeta = Q^\zeta y_{N_\zeta} + d^\zeta, \tag{39}$$

where L^ζ, Q^ζ and d^ζ are given in (34), (35) and (36), respectively.

The matrix L^ζ is clearly of the form (21) and $q < p$ since by assumption, $\hat{J}$ is strictly included in I. According to Lemma 2, L^ζ is invertible and the components of $(L^\zeta)^{-1}$ are ≥ 0. Hence A^ζ in statement (a) and b^ζ in statement (b) of the proposition are well defined. Moreover, it is clear that $Q^\zeta(i, j) \geq 0$ for all i, j which entails that $A^\zeta(i, j) \geq 0$, for all i, j. The proof is complete. □

Using Proposition 2, we can formulate a semi-explicit expression for $\hat{x}$, the minimizer of $\mathcal{F}(., y)$ over $\mathbb{R}^p$. The theorem below furnishes another formulation of the minimality conditions stated in Theorem 1.

Theorem 2 *For $y \in \mathbb{R}^p$, let $\mathcal{F}(., y)$ reach its minimum at $\hat{x}$ with $\hat{J} = J(\hat{x}, y)$, $\hat{J} \neq I$, for J as defined in (6). Let m be the number of all connected components of $\hat{J}$ $(N_i)_{i \in I}$ (see Definition 1), say ζ_ℓ for $\ell = 1, \dots, m$:*

$$\hat{J} = \bigcup_{\ell=1}^{m} \zeta_\ell. \tag{40}$$

Then $\hat{x}$ reads

$$\hat{x}_{\zeta_\ell} = A^{\zeta_\ell} y_{N_{\zeta_\ell}} + b^{\zeta_\ell}, \quad \ell = 1, \dots, m, \tag{41}$$

$$\hat{x}_i = y_i, \quad \forall i \in \hat{J}^c, \tag{42}$$

where A^{ζ_ℓ} and b^{ζ_ℓ}, for every $\ell = 1, \dots, m$, are as exhibited in Proposition 2.

Furthermore, for any $i \in \hat{J}$ there exist $a_i \in \mathbb{R}^{\#(\hat{J}^c)}$ with $a_i \geq 0$ and $\beta_i \in \mathbb{R}$ such that the system in (41) is equivalent to

$$\hat{x}_i = \langle a_i, y_{\hat{J}^c} \rangle + \beta_i, \quad \forall i \in \hat{J}. \tag{43}$$

Notice that (42)–(43) can be derived directly from Theorem 1 and Remark 1; however, this is not enough to say that all entries of a_i are nonnegative, for all $i \in \hat{J}$.

Proof If $\hat{J}$ is empty, then (41) is void and (42) holds for every $i \in I = \hat{J}^c$. Consider next that $\hat{J}$ is nonempty. Equations (41) and (42) are a straightforward consequence of Theorem 1 and Proposition 2.

The statement in (43) exploits the observation that $N_{\zeta_\ell} \subset \hat{J}^c$, $\forall \ell \in \{1, \dots, m\}$. Via a reordering of the components of each ζ_ℓ and each N_{ζ_ℓ}, the minimizer $\hat{x}$ can be put into the form (43)–(42) where every a_i contains a row of the matrix A^{ζ_ℓ} such that $i \in \zeta_\ell$, the remaining terms, being null, and every β_i is an element of b^{ζ_ℓ} such that $i \in \zeta_\ell$. Since the components of every A^{ζ_ℓ} are ≥ 0, it follows that all components of a_i are ≥ 0, for every $i \in \hat{J}$. The proof is complete. □

Observe that for every $i \in \hat{J}$, the linear operator a_i depends only on $(N_i)_{i \in I}$ and that the constant β_i depend only on $(N_i)_{i \in I}$ and $\{\sigma_i : i \in \hat{J}\}$.

Remark 6 The pixels belonging to a connected component $\zeta_\ell \subset \hat{J}$ are calculated only based on the data points y_i which are neighbors of ζ_ℓ, namely y_i for $i \in N_{\zeta_\ell}$. All data samples $y_i \in I \setminus \{\zeta_\ell \cup N_{\zeta_\ell}\}$ have no contribution. In this sense, the restoration of each $\hat{x}_{\zeta_\ell}$ is local.

5 Minimizer $\hat{x}$ Is a Locally Affine Function of the Data y

In this section we exhibit subsets of data in $\mathbb{R}^p$ leading either to the same minimizer point $\hat{x}$, or that satisfy the same system of equations, as exhibited in (41)–(42).

Proposition 3 *Given $y \in \mathbb{R}^p$, let $\mathcal{F}(., y)$ reach its minimum at $\hat{x}$ with $\hat{J} \stackrel{\text{def}}{=} J(\hat{x}, y)$, for J defined according to (6) and put*

$$\sigma_i = \operatorname{sign}(y_i - \hat{x}_i), \quad \forall i \in \hat{J}. \tag{44}$$

Consider the subset given below:

$$V_{\hat{J}} = \left\{ y' \in \mathbb{R}^p : \ y'_i = y_i, \forall i \in \hat{J}^c, \right.$$
$$\left. \begin{bmatrix} y'_i > \frac{1}{2\alpha \# N_i} + \frac{1}{\# N_i} \sum_{j \in N_i} \hat{x}_j & \text{if } i \in \hat{J}, \sigma_i = +1 \\ y'_i < -\frac{1}{2\alpha \# N_i} + \frac{1}{\# N_i} \sum_{j \in N_i} \hat{x}_j & \text{if } i \in \hat{J}, \sigma_i = -1 \end{bmatrix} \right\}.$$

Then for every $y' \in V_{\hat{J}}$, the function $\mathcal{F}(., y')$ reaches its minimum at $\hat{x}$.

Proof Let us consider $\hat{x}$ that minimizes $\mathcal{F}(., y)$. We will show that $\hat{x}$ satisfies the conditions for a minimum of $\mathcal{F}(., y')$, for any $y' \in V_{\hat{J}}$.

Consider an arbitrary $y'_i \in V_{\hat{J}}$ for $i \in \hat{J}$. From the definition of $V_{\hat{J}}$, it is seen that for every $i \in \hat{J}$ we have

$$y'_i - \frac{1}{\#N_i} \sum_{j \in N_i} \hat{x}_j > \frac{1}{2\alpha \# N_i} > 0 \quad \text{if } \sigma_i = +1;$$

$$y'_i - \frac{1}{\#N_i} \sum_{j \in N_i} \hat{x}_j < -\frac{1}{2\alpha \# N_i} < 0 \quad \text{if } \sigma_i = -1.$$

Hence $\forall y' \in V_{\hat{J}}$,

$$\operatorname{sign}\left(y'_i - \frac{1}{\#N_i} \sum_{j \in N_i} \hat{x}_j \right) = \sigma_i, \quad \forall i \in \hat{J}, \tag{45}$$

where σ_i is given in (44). Taking into account that for every $i \in \hat{J}^c$ we have $y'_i = y_i = \hat{x}_i$, (16) in Theorem 1 reads

$$\hat{x}_i - \frac{1}{\#N_i} \sum_{j \in N_i \cap \hat{J}} \hat{x}_j = \frac{1}{\#N_i} \sum_{j \in N_i \cap \hat{J}^c} \hat{x}_j + \frac{\operatorname{sign}(y_i - \hat{x}_i)}{2\alpha \# N_i},$$
$$\forall i \in \hat{J},$$

where σ_i satisfies (45). By Theorem 1, $\mathcal{F}(., y')$ reaches its minimum at $\hat{x}$.

If $\hat{J}^c$ is empty, the system above amounts to say that (9) in Proposition 1 holds for all $i \in I$. So the conclusion is the same. □

Remark 7 Let us emphasize that $V_{\hat{J}}$ is a subset whose Lebesgue measure in $\mathbb{R}^{\#(\hat{J})}$ is infinite and that it can be seen as a cone whose origin is translated.

Remark 8 If for $y \in \mathbb{R}^p$, $\mathcal{F}(., y)$ reaches its minimum at $\hat{x}$ such that $\hat{J} = I$ (i.e. $\hat{J}^c$ is empty), Proposition 3 tells us that for every $y' \in W_{I,\sigma}$, where

$$W_{I,\sigma} \equiv W_{\hat{J},\sigma} = \left\{ y' \in \mathbb{R}^p : \forall i \in I, \left[\begin{array}{ll} y'_i > \frac{1}{2\alpha\#N_i} + \frac{1}{\#N_i}\sum_{j\in N_i} \hat{x}_j & \text{if } \sigma_i = +1 \\ y'_i < -\frac{1}{2\alpha\#N_i} + \frac{1}{\#N_i}\sum_{j\in N_i} \hat{x}_j & \text{if } \sigma_i = -1 \end{array} \right] \right\}$$

and

$$\sigma_i = \operatorname{sign}(y_i - \hat{x}_i), \quad \forall i \in I, \mathcal{F}(., y')$$

reaches its minimum at $\hat{x}$, that is

$$\hat{x}' = \hat{x}, \quad \forall y' \in W_{I,\sigma}.$$

Next we determine a set $W_{\hat{J},\sigma} \subset \mathbb{R}^p$ (as large as possible), such that for every $y \in W_{\hat{J},\sigma}$, the energy $\mathcal{F}(., y)$ reaches its minimum at an $\hat{x}$ calculated by (41)–(42) using the same matrices A^{ζ_ℓ} and vectors b^{ζ_ℓ}. For arbitrary sets $\hat{J}^c \subset I$ and $\sigma \in \{-1, +1\}^{\#(\hat{J})}$, $\hat{J} \neq I$ it is possible that there is no $\hat{x}$ satisfying

$$\hat{x}_i = y_i \quad \text{for all } i \in \hat{J}^c$$

and

$$y_i \neq \hat{x}_i \quad \text{with } \operatorname{sign}(y_i - \hat{x}_i) = \sigma_i \text{ for all } i \in \hat{J}.$$

For this reason, we start with the minimizer $\hat{x}$ relevant to a given y and then we determine a set of data y' such that the relevant solution $\hat{x}'$ is calculated using exactly the same affine equation applied to every y' in this set.

Theorem 3 *Given $y \in \mathbb{R}^p$, let $\mathcal{F}(., y)$ reach its minimum at $\hat{x}$ with $\hat{J} \stackrel{\text{def}}{=} J(\hat{x}, y)$, $\hat{J} \neq I$, where J is defined by (6). Put*

$$\sigma_i = \operatorname{sign}(y_i - \hat{x}_i), \quad \forall i \in \hat{J}. \tag{46}$$

Let $a_i \in \mathbb{R}^{\#\hat{J}}$ and $\beta_i \in \mathbb{R}$ be such that

$$\hat{x}_i = \langle a_i, y_{\hat{J}^c} \rangle + \beta_i, \quad \forall i \in \hat{J}, \tag{47}$$

according to (43) in Theorem 2.

For all $i \in I$, define the constants $\overline{\beta_i}$ and the affine applications $h_i : \operatorname{span}\{e^i, i \in \hat{J}^c\} \to \mathbb{R}$ by

$$\overline{\beta_i} = \frac{1}{\#N_i} \sum_{j \in N_i \cap \hat{J}} \beta_j, \tag{48}$$

$$\langle h_i, y'_{\hat{J}^c} \rangle = \frac{1}{\#N_i} \sum_{j \in N_i \cap \hat{J}^c} y'_j + \frac{1}{\#N_i} \sum_{j \in N_i \cap \hat{J}} \langle a_j, y'_{\hat{J}^c} \rangle, \quad \forall y' \in \mathbb{R}^p. \tag{49}$$

Consider the subset given below:

$$W_{\hat{J},\sigma} = \left\{ y' \in \mathbb{R}^p : \left[\begin{array}{ll} \text{(a)} \left| y'_i - \langle h_i, y'_{\hat{J}^c} \rangle - \overline{\beta_i} \right| \le \dfrac{1}{2\alpha\#N_i} & \forall i \in \hat{J}^c \\ \text{(b)} \left\{ \begin{array}{ll} y'_i > \frac{1}{2\alpha\#N_i} + \langle h_i, y'_{\hat{J}^c} \rangle + \overline{\beta_i} & \text{if } \sigma_i = +1 \\ y'_i < -\frac{1}{2\alpha\#N_i} + \langle h_i, y'_{\hat{J}^c} \rangle + \overline{\beta_i} & \text{if } \sigma_i = -1 \end{array} \right\} & \forall i \in \hat{J} \end{array} \right] \right\}. \tag{50}$$

Then for every $y' \in W_{\hat{J},\sigma}$, the function $\mathcal{F}(.,y')$ reaches its minimum at $\hat{x}'$ given by

$$\hat{x}'_i = y'_i, \quad \forall i \in \hat{J}^c, \tag{51}$$

$$\hat{x}'_i = \langle a_i, y'_{\hat{J}^c}\rangle + \beta_i, \quad \forall i \in \hat{J}, \tag{52}$$

where a_i and β_i, $\forall i \in \hat{J}$, are the same as those given in (47).

Observe that the first sum in (49) involves only y'_j for $j \in \hat{J}^c$, so h_i is defined on $\operatorname{span}\{e^i, i \in \hat{J}^c\}$, indeed. Remind that from the definition of $\hat{J}$, we have $\hat{x}_i = y_i$ for every $i \in \hat{J}^c$.

Proof For an arbitrary $y' \in W_{\hat{J},\sigma}$, let $\hat{x}'$ read as in (51)–(52). We will show that $\hat{x}'$ satisfies the minimality conditions for $\mathcal{F}(.,y')$ stated in Theorem 1.

From the construction of $\overline{\beta_i}$ and h_i, see (48) and (49), respectively, we derive

$$\begin{aligned}
&\langle h_i, y'_{\hat{J}^c}\rangle + \overline{\beta_i} \\
&\quad = \frac{1}{\#N_i}\sum_{j\in N_i\cap\hat{J}^c} y'_j + \frac{1}{\#N_i}\sum_{j\in N_i\cap\hat{J}} (\langle a_j, y'_{\hat{J}^c}\rangle + \beta_j) \\
&\quad = \frac{1}{\#N_i}\sum_{j\in N_i} \hat{x}'_j, \quad \forall i \in I,
\end{aligned} \tag{53}$$

where the last expression comes from (51)–(52). Combining (53) with (a) in the definition of $W_{\hat{J},\sigma}$ in (50) shows that

$$\begin{aligned}
&\left|y'_i - \left(\langle h_i, y'_{\hat{J}^c}\rangle + \overline{\beta_i}\right)\right| = \left|y'_i - \frac{1}{\#N_i}\sum_{j\in N_i}\hat{x}'_j\right| \le \frac{1}{2\alpha\#N_i}, \\
&\quad \forall i \in \hat{J}^c.
\end{aligned} \tag{54}$$

If $\hat{J}$ is empty, then $\hat{x} = y$. Using (51), the inequality in (54) reads $|y'_i - \frac{1}{\#N_i}\sum_{j\in N_i} y'_j| \le \frac{1}{2\alpha\#N_i}$, $\forall i \in \hat{J}^c = I$. By Proposition 1, $\mathcal{F}(.,y')$ reaches its minimum at $\hat{x}' = \hat{y}'$. Notice that in this case (47), (50)(b) and (52) are void.

Consider next that $\hat{J}$ is nonempty. Combining (53) with (b) in (50) shows that for every $i \in \hat{J}$ we have

$$\begin{aligned}
&y'_i - \left(\langle h_i, y'_{\hat{J}^c}\rangle + \overline{\beta_i}\right) \\
&\quad = y'_i - \frac{1}{\#N_i}\sum_{j\in N_i}\hat{x}'_j \begin{cases} > \dfrac{1}{2\alpha\#N_i} > 0 & \text{if } \sigma_i = +1, \\ < -\dfrac{1}{2\alpha\#N_i} < 0 & \text{if } \sigma_i = -1. \end{cases}
\end{aligned}$$

It follows that for all $i \in \hat{J}$ we have $|y'_i - \frac{1}{\#N_i}\sum_{j\in N_i}\hat{x}'_j| > \frac{1}{2\alpha\#N_i}$ and

$$\operatorname{sign}\left(y'_i - \frac{1}{\#N_i}\sum_{j\in N_i}\hat{x}'_j\right) = \operatorname{sign}(y_i - \hat{x}_i) = \sigma_i. \tag{55}$$

By Theorem 2 we know that (47) is equivalent to (16) in Theorem 1. Hence $\hat{x}'_i$, $\forall i \in \hat{J}$, as defined in (52), is equivalent to

$$\hat{x}'_i - \frac{1}{\#N_i}\sum_{j\in N_i\cap\hat{J}}\hat{x}'_j = \frac{1}{\#N_i}\sum_{j\in N_i\cap\hat{J}^c} y'_j + \frac{\sigma_i}{2\alpha\#N_i}, \quad \forall i \in \hat{J}, \tag{56}$$

for σ_i as given in (46), for all $i \in \hat{J}$. Combining (54), (55) and (56) shows that $\hat{x}'$ satisfies all conditions for a minimizer of $\mathcal{F}(.,y')$, according to Theorem 1. □

Remark 9 Observe that $W_{\hat{J},\sigma}$ is a nonempty polyhedron in $\mathbb{R}^p$ which is in addition unbounded (hence its Lebesgue measure is infinite). By Remark 8, this holds true for $\hat{J} = I$ as well. For every $y \in \mathbb{R}^p$ we can exhibit a $W_{\hat{J},\sigma}$ such that the relevant $\mathcal{F}(.,y')$ are minimized by solutions satisfying the same system of affine equations (and in particular, $\hat{J}^c$ remains the same). The set of all feasible sets $(\hat{J},\sigma)$ is finite. The union of all possible $W_{\hat{J},\sigma}$ corresponds to a partition of $\mathbb{R}^p$—the set of all possible data.

6 A Fast Exact Numerical Method

Since $\mathcal{F}$ is non-smooth, the calculation of $\hat{x}$ needs a specific optimization scheme. Similar optimization problems are encountered along with total-variation methods where $\mathcal{F}(x,y) = \|Ax - y\|^2 + \alpha\sum_i \|G_i x\|$, where for every $i = 1,\dots,p$, G_i is a discrete approximation of the gradient operator at i, or $\mathcal{F}(x,y) = \|Ax - y\|_1 + \alpha\Phi(x)$ where Φ is an edge-preserving regularization term. Some authors use continuation methods [9, 20, 25]. In these cases, the non-smooth $|.|$ is approximated by a family of smooth functions φ_ν, $\nu > 0$, *e.g.* $\varphi_\nu(t) = \sqrt{t^2 + \nu}$. For every $\nu > 0$, the minimizer $\hat{x}_\nu$ of the relevant $\mathcal{F}_\nu(x,y)$ can be calculated using classical optimization tools. It can be shown that $\hat{x}_\nu$ converges to the sought-after $\hat{x}$ as ν decreases to zero. However, the convergence is quite slow, especially when ν approaches zero. Most of the authors just fix $\nu > 0$ and minimize a smooth approximation of the original energy, let us cite among many others [3–5, 8, 27]. Notice that in such a case, the solution *cannot* exhibit the properties relevant to the nonsmooth term ($\hat{x}_i = y_i$ in our case) as proven in [21, 23].

Subgradient methods see e.g. [12, 26], are slow and find the features of $\hat{x}$ where $\mathcal{F}(.,y)$ is nondifferentiable after a huge amount of iterations.

A method for a case slightly similar to the our was derived in [19] based on active sets and Lagrange multipliers. An adaptation to our context might be possible; but our optimization problem is much simpler so we prefer to take a benefit from this. The methods in [13, 14] suppose that the image is quantized which is not our case. The minimization method in [16] is exact, but its implementation using an interior point method introduces an approximation towards the points where the energy is nondifferentiable. We focus on our method presented in [22]—an extension of a method exposed in a textbook [17] published in 1976—since it finds in priority the points of $\hat{x}$ where $\mathcal{F}(., y)$ is non-differentiable. Its specialization to the energy in (2) has an appreciable simplicity since all steps are fully explicit. It is of the type *one coordinate at a time*. Each iteration consists of p steps where all components are updated successively.

Let the current solution read $\overline{x}$. At step i, we minimize the scalar function

$$t \to \mathcal{F}(\overline{x} + (t - \overline{x}_i)e^i,\ y)$$

with respect to t. The updating equations come from Proposition 1. The numerical scheme reads:

- Start iterations with initial guess

  $$x^{(0)} = y.$$

- For every iteration $k = 1, 2, \ldots,$ consider $i = k$ modulo p and compute

  $$\phi_i^{(k)} = y_i - \frac{1}{\#N_i}\sum_{j\in N_i} x_j^{(k-1)}. \tag{57}$$

  Then update $x_i^{(k)}$ according to the following rule:

  $$\text{if} \quad \left|\phi_i^{(k)}\right| \le \frac{1}{2\alpha\#N_i} \quad \text{then} \quad x_i^{(k)} = y_i, \tag{58}$$

  $$\text{else} \quad x_i^{(k)} = y_i - \phi_i^{(k)} + \frac{\operatorname{sign}(\phi_i^{(k)})}{2\alpha\#N_i}. \tag{59}$$

- Stop when $|x_i^{(k)} - x_i^{(k-1)}|$ is small enough for all $i = 1, \ldots, p$.

Theorem 4 *The sequence $x^{(k)}$ defined by (57) and (58)–(59) satisfies $\lim_{k\to\infty} x^{(k)} = \hat{x}$ where $\hat{x}$ is such that $\mathcal{F}(\hat{x}, y) \le \mathcal{F}(x, y)$, for every $x \in \mathbb{R}^p$.*

Proof The proof is based on Theorem 2 in [22]. The assumptions H1, H2 and H3 required there are now trivially satisfied. The last one, H4, amounts to require that the regularization term is locally strongly convex with respect to each component x_i independently, for all $i \in I$; this is also true in the present case. The rest of the proof is to calculate the intermediate steps. These come directly from Proposition 1. □

Next we give some comment on the minimization algorithm. The convergence of the algorithm presented above is guaranteed for any initial $x^{(0)}$. However, after the developments in Sect. 2, we can expect that $\hat{x}$ satisfies $\hat{x}_i = y_i$ for numerous indexes i. So initializing with $x^{(0)} = y$ should speed up the convergence. Notice that the algorithm can easily be extended to constraints of the form

$$d_i^- \le x_i \le d_i^+, \quad \text{where } d_i^- < d_i^+,\ 1 \le i \le p.$$

Let us emphasize the extreme simplicity of the numerical scheme—it involves only summations and comparisons to a fixed threshold, and there is no line search. This explains the speed of the method. For all data points which match the solution—the relevant pixels are updated according to (58), which is just a comparison to a threshold. Hence the minimization with respect to the pixels such that $\hat{x}_i = y_i$ is *exact*. This is crucial since in practice numerous samples of the solution fit exactly the relevant data samples, as evoked in (3).

Remark 10 The algorithm presented above is straightforward to adapt to energies form

$$\mathcal{F}(x, y) = \|x\|_1 + \beta\|Ax - y\|^2,$$

where A is a wavelet basis, or a frame, or a linear operator. Such problems are often encountered in approximation, compression, coding and compressive sensing.

Running Time The computation time depends clearly on the size of the image and the neighborhoods, the stopping criterion and the value of α; it is higher for larger values of α. We worked on a PC (Dell Latitude, D620, Genuine Intel(R) CPU T2500, 2.00 GHz and 1.00 RAM) running on Windows XP Professional and used Matlab 7.2. We did some comparisons on a 512×512 image (Fig. 4(g)), considered the four adjacent neighbors for $(N_i)_{i\in I}$, $\alpha = 0.01$ and the stopping rule was based on the value of $\|x^{(k)} - x^{(k-1)}\|_\infty$. A precision of 0.1 was reached after 24 iterations and needed 3 seconds CPU time. For a precision of 0.001, we had 76 iterations and 9 seconds.

Let us notice that the algorithm can be implemented in a parallel way which can speed up its convergence using Matlab.

7 Experiments

The results on nonsmooth data-fitting along with a smooth regularization shown in this paper, as well in [21, 22] clearly indicate that processing data by minimizing such an energy cannot be successful unless there are some nearly noise-free data samples. The main reason comes from the property sketched in (3). Satisfying results with such energies were obtained for denoising and deblurring under impulse noise

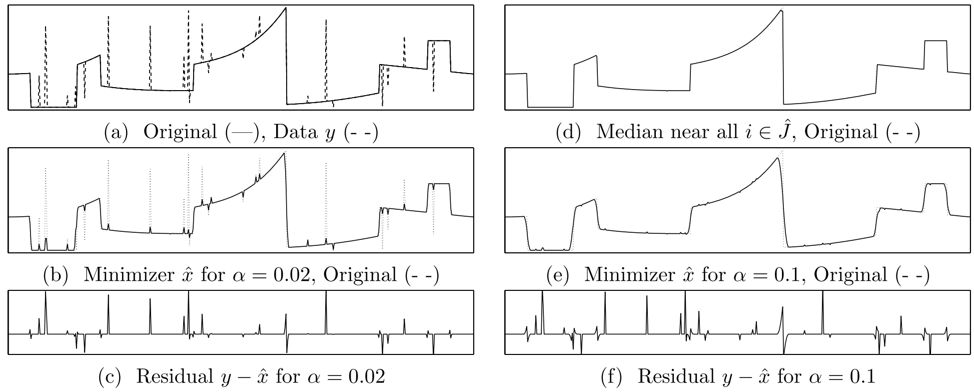


Fig. 3 Piecewise polynomial under random-valued impulse noise

[3–5, 19, 22], as well as for the denoising of frame coefficients [15], among others. For this reason, in our experiments we focus on signals and images contaminated with spiky noises. New applications will certainly appear, especially since strong results on the properties of the solutions are already available.

7.1 A Toy Illustration

The original piecewise polynomial signal is plotted in Fig. 3(a) with a solid line. Data y, plotted in Fig. 3(a) with a dashed line, corresponds to 5% random-valued impulse noise. The minimizer $\hat{x}$ for $\alpha = 0.02$, shown in Fig. 3(b), still contains some outliers; however the residual $y - \hat{x}$ in Fig. 3(c) shows that the set $\hat{J} = \{i : \hat{x}_i \neq y_i\}$ matches the locations of the outliers contained in y. In the next Fig. 3(d) we apply a simple median filter locally, only in the neighborhood of the samples $i \in \hat{J}$. The resultant restoration fits the original signal. Figure 3(e) displays the minimizer $\hat{x}$ for $\alpha = 0.1$: the outliers are sufficiently smoothed but the edges in the original signal are oversmoothed—this is not surprising since the regularization term is quadratic, hence it is not edge-preserving. The residuals in Fig. 3(f) show that $\hat{J}$ in this case is larger than the set of the outliers in data y. This effect is not surprising at all since the regularization is not edge-preserving. This example suggest that an undersmoothed $\hat{x}$ can be a good indicator to locate the outliers in the data. At a second step, the samples belonging to $\hat{J}$ need a pertinent postprocessing.

7.2 Image Denoising under Random-Valued Impulse Noise

The image in Fig. 4(b) is contaminated with 40% random-valued impulse noise. All results presented here correspond to the best choice of the parameters for each method. The solution in (c) is obtained using 5 iteration center-weighted median[2] (CWM) filter with a 5×5 window and multiplicity parameter 8. The restoration in (d) corresponds to permutation-weight[3] (PWM) filter on 7×7 window with rank parameter 22. A detailed description of the CWM and the PWM filters can be found e.g. in the textbook [6]. The image in (e) is obtained using the two-phase method described in [7]. The restoration in (f) is the minimizer $\mathcal{F}(., y)$ for $\alpha = 0.08$—outliers are removed but the edges are slightly oversmoothed. The image in (g) is the minimizer of $\mathcal{F}(., y)$ for $\alpha = 0.01$. As expected from the previous example, the outliers are not cleaned but $\hat{J}$ approximates well the locations of the outliers in the data. The final restoration in (h) is obtained using a local median smoother near each component in $\hat{J}$. The zoom presented below shows that the latter result has better preserved edges and a more natural appearance, compared with the other methods.

7.3 Cleaning Noisy Data from Outliers

In different applications, data y result from outlier-free degradations (e.g., distortion, blurring, quantization errors, electronic noise), plus impulsive noise. It is well known that classical reconstruction methods fail in the presence of outliers. A preliminary processing, aimed at removing the outliers, is usually needed before to apply standard reconstruction methods. Preprocessing is often realized using median-based filters. It is critical that preprocessing keeps intact all the information contained in the outlier-free data.

In our experiment, the sought image x^*, shown in Fig. 5(b), is related to x_o in (a), by $x^* = x_o + n$, where n is white Gaussian noise with 20 dB SNR. The histogram of

[2]**CWM filter.** Fix a number of iterations T, an integer multiplicity parameter $\mu \geq 1$ and a window N_i, $i \notin N_i$ of neighbors for each $i \in I$. At iteration t, pixel x_i is replaced by $\mathrm{Median}\{\underbrace{\{x_i, \dots, x_i\}}_{\mu \text{ times}} \cup \{x_j, j \in N_i\}\}$.

[3]**PWM filter.** Choose a number of iterations T, a window N_i, $i \notin N_i$ of neighbors for each $i \in I$ and a rank parameter ρ such that $1 \leq \rho \leq 0.5\#N_i$. At iteration t, all pixels $\{x_j, j \in N_i \cup \{i\}\}$ are sorted in increasing order. If the ranking r of x_i satisfies either $r < \rho$ or $r > \#N_i - \rho$, then x_i is replaced by $\mathrm{Median}\{x_j, j \in N_i \cup \{i\}\}$; otherwise x_i remains unchanged.

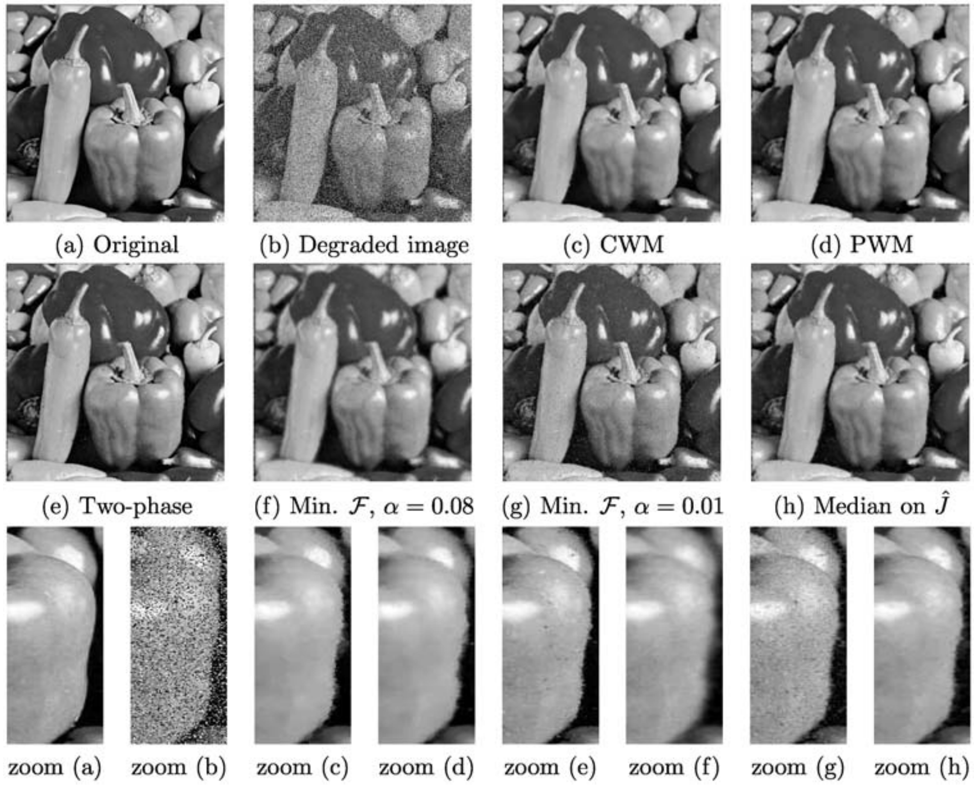


Fig. 4 Random-valued impulse noise cleaning using different methods

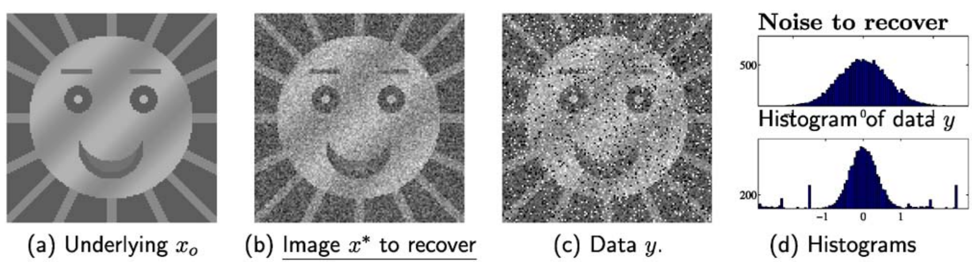


Fig. 5 Data with two-stage degradation

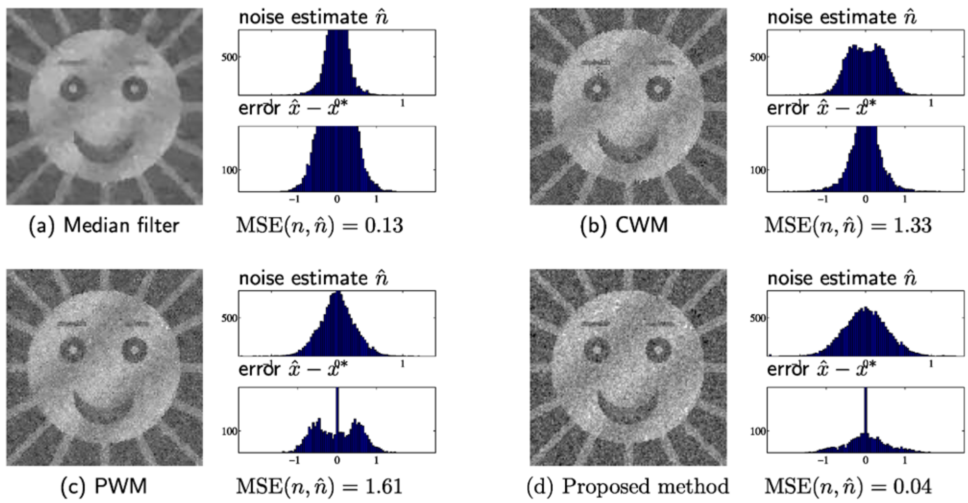
n is plotted in (d), top. Our goal is to restore x^* (the image contaminated with Gaussian noise only) based on the data y, shown in (c), which contain 10 % salt-and-pepper noise. Restoring x^* is a challenge since the white Gaussian noise must be preserved. The relevance of an estimate $\hat{x}$ is evaluated by both, the error $\hat{x} - x^*$ and the closeness of the estimated noise $\hat{n} = \hat{x} - x_o$ to the initial noise n. To this end, we use the Mean-square error (MSE) defined by $\mathrm{MSE}(n, \hat{n}) = \frac{1}{p}\|n - \hat{n}\|_2$. All images in this subsection are displayed using the same gray value scaling.

All results in Fig. 6 correspond to parameters leading to the best removal of the outliers. The image in (a) corresponds to one iteration of median filter over a 3×3 window. Almost all data samples are altered and the estimated noise, $\hat{n} = \hat{x} - x_o$, is quite concentrated near zero. The image in (b) is calculated using a center weighted median (CWM) filter with a 5×5 window and multiplicity parameter for the central pixel equal to 14. Even though the error $\mathrm{MSE}(\hat{x}, x^*)$ is reduced, the histogram of the noise estimate $\hat{n}$, plotted on the right, top, deviates considerably from the initial distribution, shown in Fig. 5. The image in (c) corresponds to one iteration of permutation weighted median (PWM) filter, for a 5×5 window and rank parameter 4. In all these estimates, the distribution of the noise estimate $\hat{n}$ is quite different from the distribution of n. Figure 6(d) displays the issue of the minimization of $\mathcal{F}$ as given in (2) for $\alpha = 1.3$. It achieves a good preservation of the statistics of the noise in x^*, as seen from the histogram of the estimated noise $\hat{n}$—Fig. 6(d), right, top. Moreover, the error $\hat{x} - x^*$ is essentially concentrated around zero, as seen in Fig. 6(d), right, down. Notice that this experiment takes a benefit from our Remark 2.

8 Conclusions and Perspectives

In this paper we show that the minimizers $\hat{x}$ of energies $\mathcal{F}(., y)$ of the form (2) are locally affine functions of the

Fig. 6 Various restorations



data, as exhibited in Theorem 3. Energies involving a nonsmooth data-fitting are known to produce solutions that partially fit the data, as evoked in (3). An important property that we found is that the pixels $\hat{x}_\zeta$ of any connected subset ζ that do not fit (3) is restored using a simple function of the form $\hat{x}_\zeta = A^\zeta y_{N_\zeta} + \beta^\zeta$ where all entries of A^ζ are ≥ 0 and β^ζ is a fixed vector that depends only on the sign of $(y_i - \hat{x}_i)$ for $i \in \zeta$. If we have some knowledge that the data y follow a simple distribution on a bounded domain of $\mathbb{R}^p$, it should be possible to evaluate the probability to obtain a solution given by the same set of affine equations. Even though we propose a fast minimization scheme, we hope that the obtained theoretical results can help to conceive faster minimization schemes.

Semi-explicit solutions of this kind are hard to exhibit for general regularization terms. The results of this paper suggest how are restored the pixels satisfying $\hat{x}_i \neq y_i$ under more general regularization terms. We consider this study as a starting point for the analysis of more elaborated energies involving ℓ_1 data-fitting.

The approach adopted in this paper can be used to analyze, as well as to derive new minimization schemes, for energies of the form $\mathcal{F}(x, y) = \|x\|_1 + \beta\|Ax - y\|^2$, or minimization problems such as: minimize $\|x\|_1$ subject to $\|Ax - y\|^2 \leq \tau$ for $\tau > 0$, or minimize $\|x\|_1$ subject to $Ax = y$, where A is a wavelet basis or a frame, or any linear operator. Such problems arise customarily in approximation, in coding and compression, and in compressive sensing. In these cases, $\#\hat{J}^c/p$ corresponds to the level of sparsity of the obtained solution.

Acknowledgement The author would like to thank the Reviewers for their pertinent remarks encouragement. The author would like to thank Jianfeng CAI (*National University of Singapore*) for providing the experiment with the two-phase method in Fig. 4.

References

1. Alliney, S.: Digital filters as absolute norm regularizers. IEEE Trans. Signal Process. **SP-40**, 1548–1562 (1992)
2. Alliney, S.: A property of the minimum vectors of a regularizing functional defined by means of absolute norm. IEEE Trans. Signal Process. **45**, 913–917 (1997)
3. Bar, L., Brook, A., Sochen, N., Kiryati, N.: Deblurring of color images corrupted by impulsive noise. IEEE Trans. Image Process. **16**, 1101–1111 (2007)
4. Bar, L., Sochen, N., Kiryati, N.: Image deblurring in the presence of salt-and-pepper noise. In: Proceeding of 5th International Conference on Scale Space and PDE methods in Computer Vision. LNCS, vol. 3439, pp. 107–118. Springer, Berlin (2005)
5. Bar, L., Sochen, N., Kiryati, N.: Semi-blind image restoration via Mumford-Shah regularization. IEEE Trans. Image Process. **15**, 483–493 (2006)
6. Bovik, A.C.: Handbook of Image and Video Processing. Academic, New York (2000)
7. Chan, R., Hu, C., Nikolova, M.: An iterative procedure for removing random-valued impulse noise. IEEE Signal Process. Lett. **11**, 921–924 (2004)
8. Chan, T., Chen, K.: An optimization-based multilevel algorithm for total variation image denoising. SIAM J. Multiscale Model. Simul. **5**, 615–645 (2006)
9. Chan, T., Esedoglu, S.: Aspects of total variation regularized l^1 function approximation. SIAM J. Appl. Math. **65**, 1817–1837 (2005)
10. Chan, T., Esedoglu, S., Nikolova, M.: Algorithms for finding global minimizers of image segmentation and denoising models. SIAM J. Appl. Math. **66**, 1632–1648 (2006)
11. Ciarlet, P.G.: Introduction à l'analyse numérique matricielle et à l'optimisation, 5th edn. Collection mathématiques appliquées pour la maîtrise. Dunod, Paris (2000)
12. Combettes, P., Luo, J.: An adaptive level set method for nondifferentiable constrained image recovery. IEEE Trans. Image Process. **11**, 1295–1304 (2002)
13. Darbon, J., Sigelle, M.: Image restoration with discrete constrained total variation part I: Fast and exact optimization. J. Math. Imaging Vis. **26**, 261–276 (2006)
14. Darbon, J., Sigelle, M.: Image restoration with discrete constrained total variation part II: Levelable functions, convex and non-convex cases. J. Math. Imaging Vis. **26**, 277–291 (2006)
15. Durand, S., Nikolova, M.: Denoising of frame coefficients using l1 data-fidelity term and edge-preserving regularization. SIAM J. Multiscale Model. Simul. **6**, 547–576 (2007)

16. Fu, H., Ng, M., Michael, K., Nikolova, M., Barlow, J.L.: Efficient minimization methods of mixed $\ell_1 - \ell_l$ and $\ell_2 - \ell_1$ norms for image restoration. SIAM J. Sci. Comput. **27**, 1881–1902 (2005)
17. Glowinski, R., Lions, J., Trémolières, R.: Analyse numérique des inéquations variationnelles, vol. 1, 1st edn. Dunod, Paris (1976)
18. Hiriart-Urruty, J.-B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms, vol. I and II. Springer, Berlin (1996)
19. Kärkkäinen, T., Kunisch, K., Majava, K.: Denoising of smooth images using l1-fitting. Computing **74**, 353–376 (2005)
20. Nikolova, M.: Markovian reconstruction using a GNC approach. IEEE Trans. Image Process. **8**, 1204–1220 (1999)
21. Nikolova, M.: Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers. SIAM J. Numer. Anal. **40**, 965–994 (2002)
22. Nikolova, M.: A variational approach to remove outliers and impulse noise, J. Math. Imaging Vis. 20 (2004)
23. Nikolova, M.: Weakly constrained minimization. Application to the estimation of images and signals involving constant regions. J. Math. Imaging Vis. **21**, 155–175 (2004)
24. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
25. Rudin, L., Osher, S., Fatemi, C.: Nonlinear total variation based noise removal algorithm. Physica D **60**, 259–268 (1992)
26. Shor, N.Z.: Minimization Methods for Non-Differentiable Functions, vol. 3. Springer, New York (1985)
27. Vogel, C.R., Oman, M.E.: Fast, robust total variation-based reconstruction of noisy, blurred images. IEEE Trans. Image Process. **7**, 813–824 (1998)

Mila Nikolova received the Ph.D. degree from the Université de Paris-Sud, France, in 1995. She got a Habilitation to direct research in 2006. Currently, she is Senior Research Fellow, 1st class, with the National Center for Scientific Research (CNRS), France, and performs her research at the Centre de Mathématiques et de Leurs Applications (CMLA), ENS de Cachan, France. Her research interests are in mathematical Image and signal reconstruction, Inverse problems, Regularization and variational methods and the properties of their solutions, Scientific computing.