Journal of Mathematical Imaging and Vision 20: 99–120, 2004


# A Variational Approach to Remove Outliers and Impulse Noise

MILA NIKOLOVA
*CMLA UMR 8536–ENS de Cachan, 61 av. du President Wilson, 94235 Cachan Cedex, France*
nikolova@cmla.ens-cachan.fr

**Abstract.** We consider signal and image restoration using convex cost-functions composed of a non-smooth data-fidelity term and a smooth regularization term. We provide a convergent method to minimize such cost-functions. In order to restore data corrupted with outliers and impulsive noise, we focus on cost-functions composed of an $\ell_1$ data-fidelity term and an edge-preserving regularization term. The analysis of the minimizers of these cost-functions provides a natural justification of the method. It is shown that, because of the $\ell_1$ data-fidelity, these minimizers involve an implicit detection of outliers. Uncorrupted (regular) data entries are fitted exactly while outliers are replaced by estimates determined by the regularization term, independently of the exact value of the outliers. The resultant method is accurate and stable, as demonstrated by the experiments. A crucial advantage over alternative filtering methods is the possibility to convey adequate priors about the restored signals and images, such as the presence of edges. Our variational method furnishes a new framework for the processing of data corrupted with outliers and different kinds of impulse noise.

**Keywords:** image denoising, impulse noise removal, non-smooth analysis, non-smooth optimization, outliers, restoration, regularization, signal denoising, total variation, variational methods

## 1. Introduction

We consider the problem where, given data $y \in \mathrm{R}^q$, the estimate $\hat{x} \in \mathrm{R}^p$ of an image or signal is defined as the minimizer of a convex cost-function $\mathcal{F}_y : \mathrm{R}^p \to \mathrm{R}$ which combines a data-fidelity term $\Psi_y$ and a regularization term $\mathcal{Q}$, weighted by a parameter $\beta > 0$:

$$\mathcal{F}_y(x) = \Psi_y(x) + \beta \mathcal{Q}(x), \quad \text{where}$$
$$\Psi_y(x) = \sum_{i=1}^{q} \psi_i\left(a_i^T x - y_i\right). \tag{1}$$

Such cost-functions are classical in regularization and Bayesian estimation [5, 12, 19, 28, 29]. If $x^*$ is the original (unknown) image or signal, every $y_i$ can be seen as a possibly noisy version of $a_i^T x^*$, where $a_i \in \mathrm{R}^p$ and $^T$ stands for transpose. Via the choice of $\{\psi_i\}$, the term $\Psi_y$ encourages $\hat{x}$ to be such that each $a_i^T \hat{x}$ is close to $y_i$, while $\mathcal{Q}$ pushes $\hat{x}$ to exhibit some a priori expected features. The trade-off between these two goals is controlled by $\beta$. Since [5, 15], a useful class of regularization functions are

$$\mathcal{Q}(x) = \sum_{i=1}^{r} \varphi\left(g_i^T x\right), \tag{2}$$

where $g_i \in \mathrm{R}^p$, for $i = 1, \ldots, r$, are difference operators and $\varphi : \mathrm{R} \to \mathrm{R}$ is called a potential function. It is frequently required that $\hat{x}$ contains smoothly varying regions and edges. The possibility that convex functions $\varphi$ give rise to minimizers $\hat{x}$ involving large differences $|g_i^T \hat{x}|$ at the locations of edges has been studied in [6, 7, 10, 22]. Examples of smooth and convex *edge-preserving* potential functions are

$$\varphi(t) = |t|^\alpha, \ 1 < \alpha \leq 2, \tag{3}$$
$$\varphi(t) = \sqrt{\alpha + t^2}, \tag{4}$$
$$\varphi(t) = 1 + |t|/\alpha - \log(1 + |t|/\alpha), \tag{5}$$
$$\varphi(t) = \log(\cosh(t/\alpha)), \tag{6}$$

where $\alpha > 0$ is a parameter.

We draw a particular attention to the choice of the data-fidelity term $\Psi_y$. In signal and image restoration and reconstruction, and in numerous inverse problems, the most usual choice is $\psi_i(t) = t^2$, for all $i = 1, \ldots, q$ [10, 12, 28, 29, 32]. In a statistical setting this choice corresponds with the assumption that data are corrupted with white Gaussian noise. Other forms for $\Psi_y$ arise e.g. in computed tomography and in astronomy [8, 27]. Let us emphasize that beyond a few exceptions [2, 3, 24], the data-fidelity terms $\Psi_y$ involved in cost-functions of the form (1) are always *smooth functions*. In contrast, this paper is devoted to non-smooth data-fidelity terms as specified below.

*H1.* For every $i = 1, \ldots, q$, the function $\psi_i : \mathrm{R} \to \mathrm{R}$ in (1) is convex, $\mathcal{C}^1$ on $\mathrm{R}\backslash\{0\}$ and its left-side and right-side derivatives at zero, denoted $\psi_i'(0^-)$ and $\psi_i'(0^+)$, respectively, satisfy[1] $\psi_i'(0^-) < 0 < \psi_i'(0^+)$.

In Section 2 we provide a convergent method to minimize the relevant non-smooth cost-functions. Incidentally, the same method can be applied to the total-variation cost-function for 1D signals.

Our interest in cost-functions having a non-smooth data-fidelity term is due to the property that typically, their minimizers $\hat{x}$ fit *exactly* a certain number of the data entries, i.e. that the set $\hat{h} = \{i : a_i^T \hat{x} = y_i\}$ is nonempty [24]. It easy to see that if *all* entries $y_i$, for $i = 1, \ldots, q$, are noticeably corrupted, $\hat{x}$ will be a poor estimate: such cost-functions are not adapted to deal with e.g. Gaussian, or Poisson, or Laplace measurement noise. Instead, we suppose that data $y$ are corrupted so that

*I1.* A certain number of the data entries $y_i$ are uncorrupted (regular);
*I2.* The corrupted data entries (outliers) are dissimilar with respect to their neighbors.

Typically, these are signals and images corrupted with outliers or impulse noise. Such perturbations are often due to bit errors in transmission, faulty memory locations, errors in analog-to-digital conversion, malfunctioning pixel elements in camera sensors. If $x^*$ denotes the original signal or image, corrupted entries usually take an arbitrary value in the interval $[\min_i x^*, \max_i x^*]$, or in the set $\{\min_i x^*, \max_i x^*\}$. Cleaning such noise needs to smooth *selectively* outliers while preserving the local features of the original signal or image and uncorrupted entries. Since [30], many different order-statistic filters have been derived in order to process outliers [4, 9, 13, 14, 20, 34, 35], e.g. recursive median, weighted median, hybrid median, center-weighted median, permutation weighted median. Being applied uniformly across the image, these filters tend to alter both corrupted and regular pixels. Decision-based filters operate over small windows where outliers are first detected and then selectively smoothed [1, 11, 21, 25, 33]. Their success is closely dependent on the reliability of the outlier decision rule and on the window size.

Our approach is different. In Section 3, data corrupted as mentioned in I1–I2 are denoised by minimizing cost-functions of the form (1) where $\Psi_y(x) = \sum_i |x_i - y_i|$ and $\mathcal{Q}$ is of the form (2) whith an edge-preserving $\varphi$. The minimizers of these cost-functions are analyzed in order to justify the method. We show that minimizers $\hat{x}$ involve an implicit detection of outliers which is reliable and stable. Regular data entries are kept unchanged ($\hat{x}_i = y_i$), whereas outliers incur edge-preserving smoothing via continuous (implicit) functions, based only on neighboring *regular* data entries, independently of the exact value of the outliers. The resultant method is robust and accurate. A critical advantage over order-statistic filtering methods is that outliers are detected and smoothed based on adequate priors on the sought-after image or signal. The experiments in Section 4 show that it gives rise to significant improvement in signal and image restoration over usual filtering methods. Our variational method provides a new framework for the processing of data contaminated with outliers and different kinds of impulse noise.

The proofs of all statements presented in this paper are outlined in the Appendix.

## 2. Minimization Method

Now we focus on the calculation of the minimizers $\hat{x}$ of functions $\mathcal{F}_y$ of the form (1) where $\Psi_y$ is non-smooth as specified in H1. We will make some additional assumptions which are presented below.

*H2.* The family $\{a_i : i = 1, \ldots, q\}$ is linearly independent.

This assumption is satisfied for denoising and many missing-data problems. It allows $\mathcal{F}_y$ to be put into the following form. If $q < p$, let us choose $a_{q+i} \in \mathrm{R}^p$, for $i = 1, \ldots, p - q$, so that $\{a_i : i = 1, \ldots, p\}$ spans $\mathrm{R}^p$. Let $A$ be the $p \times p$ matrix whose rows are $a_i^T$, for

$i = 1, \ldots, p$. For every $i = 1, \ldots, p$, let $e_i \in \mathbb{R}^p$ denote the $i$th vector of the canonical basis of $\mathbb{R}^p$ (i.e., $e_i[i] = 1$ and $e_i[j] = 0$ if $j \neq i$). Using the change of variables $z = Ax - \tilde{y}$, where $\tilde{y} = \sum_{i=1}^{q} y_i e_i + \sum_{i=q+1}^{p} 0\, e_i$, we will consider $F_y(z) = \mathcal{F}_y(A^{-1}(z + \tilde{y}))$,

$$F_y(z) = \sum_{i=1}^{q} \psi_i(z_i) + \beta Q_y(z), \tag{7}$$

where

$$Q_y(z) = \mathcal{Q}(A^{-1}(z + \tilde{y})). \tag{8}$$

Clearly, $\mathcal{F}_y$ reaches its minimum at $\hat{x} \in \mathbb{R}^p$ if, and only if, $F_y$ reaches its minimum at $\hat{z} = A\hat{x} - \tilde{y}$.

*H3.* For every $y \in \mathbb{R}^q$, the function $F_y : \mathbb{R}^p \to \mathbb{R}$ in (7) is 0-coercive, i.e. $F_y(z) \to \infty$ if $\|z\| \to \infty$.

*H4.* The function $Q_y$ in (8) is convex and $\mathcal{C}^1$-continuous. Moreover, for every $y \in \mathbb{R}^q$, and for every $\rho > 0$, there is $\eta > 0$ such that for every $i = 1, \ldots, p$,

$$Q_y(z + te_i) - Q_y(z) \geq t D_i Q_y(z) + \eta t^2, \quad \forall z \in \bar{B}(0, \rho), \ \ \forall t \in [-\rho, \rho], \tag{9}$$

where $D_i Q_y$ denotes the $i$th partial derivative of $Q_y$.

Although $Q_y$ is strongly convex [18] along each direction span$\{e_i\}$, for $i = 1, \ldots, p$, globally it can be non-strictly convex and non-coercive. Let us consider H4 when $\mathcal{Q}$ is of the form (2). According to (8),

$$Q_y(z) = \sum_{i=1}^{r} \varphi\big(b_i^T (z + \tilde{y})\big), \quad \text{where} \quad b_i^T = g_i^T A^{-1}, \quad \forall i = 1, \ldots, r. \tag{10}$$

Let $B$ be the $r \times p$ matrix with rows $b_i^T$, for $i = 1, \ldots, r$. Observe that all functions $\varphi$ in (3)–(6) are strongly convex on any bounded interval—for every $\delta > 0$, there is $\eta_\delta > 0$, such that[2]

$$\varphi(t) - \varphi(\tau) \geq \varphi'(\tau)\,(t - \tau) + \eta_\delta (t - \tau)^2, \quad \forall t \in [-\delta, \delta], \ \ \forall \tau \in [-\delta, \delta]. \tag{11}$$

*Remark 1.* Suppose that $B$ does not contain zero-valued columns. If $\varphi$ is $\mathcal{C}^1$ and satisfies (11), then $Q_y$ in (10) satisfies H4. The details are given in the Appendix.

### *2.1. Conditions for a Minimum*

Let us focus on the conditions for minimum of cost-functions of the form (7). The following simple lemma underlies many properties discussed in what follows.

**Lemma 1.** *Suppose that $\psi_i$ satisfies* H1. *Then*

$$t \geq 0 \Rightarrow \psi_i(t) - \psi_i(0) \geq \psi_i'(0^+)\, t,$$
$$t \leq 0 \Rightarrow \psi_i(t) - \psi_i(0) \geq \psi_i'(0^-)\, t.$$

Being convex by H1 and H4, and 0-coercive by H3, the function $F_y$ (and also $\mathcal{F}_y$), does admit a minimum for every $y \in \mathbb{R}^q$. This minimum is both local and global [18].

**Theorem 1.** *Let $F_y$ in* (7) *satisfy* H1 *and* H4. *Then $F_y$ reaches its minimum at $\hat{z} \in \mathbb{R}^p$ if, and only if,*

$$-\psi_i'(0^+) \leq \beta D_i Q_y(\hat{z}) \leq -\psi_i'(0^-) \quad \text{if } i \in \hat{h}, \tag{12}$$
$$\psi_i'(\hat{z}_i) + \beta D_i Q_y(\hat{z}) = 0 \quad \text{if } i \in \hat{h}^c, \tag{13}$$
$$D_i Q_y(\hat{z}) = 0 \quad \text{if } i \in \{1, \ldots, p\} \backslash (\hat{h}^c \cup \hat{h}), \tag{14}$$

*where $\hat{h}$ is defined by*

$$\hat{h} = \{i \in \{1, \ldots, q\} : \hat{z}_i = 0\}, \tag{15}$$

*and $\hat{h}^c$ is its complement in* $\{1, \ldots, q\}$. *Moreover, for any $i \in \hat{h}^c$, we have*

$$\beta D_i Q_y(\hat{z} - \hat{z}_i e_i) < -\psi_i'(0^+) \Rightarrow \hat{z}_i > 0, \tag{16}$$
$$\beta D_i Q_y(\hat{z} - \hat{z}_i e_i) > -\psi_i'(0^-) \Rightarrow \hat{z}_i < 0. \tag{17}$$

In Section 3 it is seen that $\hat{h}$ is usually non-empty. The example below illustrates that the minimum of $F_y$ may be non-strict. Since $F_y$ is 0-coercive, all minimizers are bounded.

*Example 1.* Consider the function $F_y : \mathbb{R}^2 \to \mathbb{R}$ given below

$$F_y(z) = |z_1| + |z_2| + \beta(z_1 - z_2 + y_1 - y_2)^2, \quad \text{subject to} \quad y_1 - y_2 < -1 - \frac{1}{2\beta}. \tag{18}$$

As shown in the Appendix, $F_y$ reaches its minimum at $\hat{z}$ if, and only if $\hat{z} \in Z$, where $Z$ reads

$$Z = \left\{ z \in \mathbb{R}^2 : 0 \le z_1 \le -(y_1 - y_2) - \frac{1}{2\beta} \text{ and } z_2 = z_1 + y_1 - y_2 + \frac{1}{2\beta} \right\}.$$

Notice that $Z$ is bounded.

Classically, the minimum of $F_y$ is strict if one of the conditions hold: (i) $p = q$ and $\{\psi_i\}_{i=1}^q$ are strictly convex; (ii) the function $Q_y$ is strictly convex. Let $\mathbb{1} \in \mathbb{R}^r$ read $\mathbb{1}_i = 1$ for all $i = 1, \ldots, r$. If $\mathcal{Q}$ is as in (2) and $\{g_i^T\}_{i=1}^r$ are difference operators, then $g_i^T \mathbb{1} = 0$ for all $i$, so (ii) fails. Moreover, in Section 3 we use $\psi_i(t) = |t|$, in which case (i) fails as well.

**Proposition 1.** *Let $F_y$ be of the form* (7) *where $Q_y$ is as given in* (10). *Let* H1 *and* H4 *be satisfied. Let $\hat{z} \in \mathbb{R}^p$ satisfy* (12), (13) *and* (14). *Suppose also the following*:

(i) *$\varphi$ is strictly convex on* R;
(ii) *the set $\hat{h}^0 = \{i \in \hat{h}$ such that* (12) *is strict$\}$ is nonempty*;
(iii) *if $u \in \ker B$, there is $i \in \hat{h}^0$ such that $u_i \neq 0$.*

*Then $F_y$ reaches its minimum at $\hat{z}$ and the latter is strict.*

Assumption (i) holds for all functions $\varphi$ in (3)–(6). Although (ii) fails in Example 1, we see in Section 3 that it usually holds. When $A$ is the identity and $\{g_i^T\}_{i=1}^r$ are first-order difference operators, $\ker B = \text{span}\{\mathbb{1}\}$, hence (iii) holds. *Usually, the minimum of $F_y$, and so the minimum of $\mathcal{F}_y$, is strict.*

### 2.2. Relaxation-Based Minimization

Since $F_y$ is non-smooth, the calculation of $\hat{z}$ needs a special care. Similar optimization problems arise along with non-smooth regularization where $\mathcal{F}_y$ is of the form (1) with $\Psi_y$ smooth and $\mathcal{Q}$ as given in (2) with $\varphi$ non-smooth at zero. A popular choice is $\varphi(t) = |t|$, which corresponds with total-variation methods, pioneered in [26]. The non-convex function $\varphi(t) = |t|/(\alpha + |t|)$ has been proposed in [16]. Gradient descent minimization is well-known to fail when $\mathcal{F}_y$ is non-smooth [17, 18]. Since [31, 32], $\varphi(t) = |t|$ is usually replaced by $\varphi_\nu(t) = \sqrt{t^2 + \nu}$, for $\nu > 0$ "small enough", and the obtained solution is an approximation of the desired one. In [16], the function which is minimized is a half-quadratic smooth approximation of $\mathcal{F}_y$. In [23], a continuation method is used for various non-smooth functions $\varphi$: given $\varphi_\nu$—a smooth approximation of $\varphi$, with $\lim_{\nu \searrow 0} \varphi_\nu \to \varphi$—a sequence of minimizers $\hat{x}_\nu$ for $\nu$ decreasing from 1 to 0 is tracked by local minimization.

We focus on relaxation-based minimization. Let $z^{(0)} \in \mathbb{R}^p$ be a starting point. At every iteration $k = 1, 2, \ldots$, the new iterate $z^{(k)}$ is obtained from $z^{(k-1)}$ by calculating successively each one of its entries $z_i^{(k)}$ using one-dimensional minimization:
for any $i = 1, \ldots, p$, find $z_i^{(k)}$ such that

$$\forall t \in \mathbb{R},\ F_y\big(z_1^{(k)}, \ldots, z_{i-1}^{(k)}, z_i^{(k)}, z_{i+1}^{(k-1)}, \ldots, z_p^{(k-1)}\big) \le F_y\big(z_1^{(k)}, \ldots, z_{i-1}^{(k)},\ t,\ z_{i+1}^{(k-1)}, \ldots, z_p^{(k-1)}\big). \quad (19)$$

Notice that the order of updating the entries $z_i^{(k)}$, for $i \in \{1, \ldots, p\}$, can be any. If $F_y$ is strictly convex, coercive and $\mathcal{C}^1$, the sequence $z^{(k)}$ defined by (19) converges to the unique minimizer of $F_y$ [17]. If $F_y$ is non-smooth, $z^{(k)}$ may not converge to a minimizer of $F_y$. Nevertheless, it is shown in [17] that if $F_y$ is of the form (7) where $\psi_i(t) = \alpha_i |t|$, with $\alpha_i \ge 0$, for all $i$, and $Q_y$ is strictly convex and 0-coercive, then $z^{(k)}$ reaches the minimizer of $F_y$ as $k \to \infty$. We cannot apply this result because in our context the functions $\psi_i$ have a more general form, and $Q_y$ may be non-strictly convex and non-coercive.

According to (19), the solution obtained at step $(i-1)$ of iteration $k$ is

$$\big(z_1^{(k)}, \ldots, z_{i-1}^{(k)}, z_i^{(k-1)}, z_{i+1}^{(k-1)}, \ldots, z_p^{(k-1)}\big). \quad (20)$$

Let H1, H3 and H4 hold. Then the function $t \to F_y(z_1^{(k)}, z_2^{(k)}, \ldots, z_{i-1}^{(k)}, t, z_{i+1}^{(k-1)}, \ldots, z_p^{(k-1)})$ is strictly convex and 0-coercive. Hence, $z_i^{(k)}$ in (19) is well defined and unique [18]. Given the solution at step $(i-1)$ of iteration $k$, the entry $z_i^{(k)}$ is determined using Theorem 1:

- if $i \in \{1, \ldots, q\}$, calculate $\xi_i^{(k)} = \beta D_i Q_y(z_1^{(k)}, \ldots, z_{i-1}^{(k)}, 0, z_{i+1}^{(k-1)}, \ldots, z_p^{(k-1)})$;

$$\text{if } -\psi_i'(0^+) \le \xi_i^{(k)} \le -\psi_i'(0^-), \quad \text{then } z_i^{(k)} = 0; \quad (21)$$

otherwise $z_i^{(k)}$ is the (unique) solution of

$$\psi_i'\big(z_i^{(k)}\big) + \beta D_i Q_y \big(z_1^{(k)}, \dots, z_{i-1}^{(k)}, z_i^{(k)}, z_{i+1}^{(k-1)}, \dots, z_p^{(k-1)}\big) = 0, \tag{22}$$

where

$$\begin{cases} z_i^{(k)} > 0 & \text{if } \xi_i^{(k)} < -\psi_i'(0^+), \\ z_i^{(k)} < 0 & \text{if } \xi_i^{(k)} > -\psi_i'(0^-); \end{cases}$$

- if $q < p$ and $i \in \{q+1, \dots, p\}$, then $z_i^{(k)}$ is the (unique) solution of

$$D_i Q_y\big(z_1^{(k)}, \dots, z_{i-1}^{(k)}, z_i^{(k)}, z_{i+1}^{(k-1)}, \dots, z_p^{(k-1)}\big) = 0. \tag{23}$$

Knowing the sign of $z_i^{(k)}$ in (22) is crucial because $\psi_i'$ is discontinuous at zero. Notice that if the minimum in (19) occurs at a point where the function is non-smooth, the minimizer is given exactly by (21).

**Theorem 2.** *Let $F_y : \mathrm{R}^p \to \mathrm{R}$ be of the form* (7) *and suppose that* H1, H2, H3 *and* H4 *are satisfied. For $k \to \infty$, the sequence $z^{(k)}$ defined by* (19) *converges to a point $\hat{z}$ such that $F_y(\hat{z}) \le F_y(z)$, for all $z \in \mathrm{R}^p$.*

An important argument in the proof of this theorem is the lemma given next.

**Lemma 2.** *There is a constant $\eta > 0$ such that for every $k \in \mathbf{N}$,*

$$\begin{aligned} &F_y\big(z_1^{(k)}, \dots, z_{i-1}^{(k)}, z_i^{(k-1)}, z_{i+1}^{(k-1)}, \dots, z_p^{(k-1)}\big) \\ &- F_y\big(z_1^{(k)}, \dots, z_{i-1}^{(k)}, z_i^{(k)}, z_{i+1}^{(k-1)}, \dots, z_p^{(k-1)}\big) \\ &\ge \beta\eta\big(z_i^{(k-1)} - z_i^{(k)}\big)^2, \quad \forall i \in \{1, \dots, p\}. \end{aligned}$$

### *2.3. Total Variation Cost-Function for 1D Signals*

When $x$ is an one-dimensional signal, the discrete form of the total-variation cost-function reads

$$\mathcal{F}_y(x) = \|Ax - y\|^2 + \beta \sum_{i=1}^{p-1} |x_i - x_{i+1}|,$$

where $A \in \mathrm{R}^{r\times p}$ is given. Let the entries of $G \in \mathrm{R}^{p\times p}$ read $G_{i,i} = 1$, $G_{i,i+1} = -1$ and $G_{i,j} = 0$ if $j < i$ or $j > i+1$, for all $i = 1, \dots, p-1$, and $G_{p,i} = 1/p$, for all $i = 1, \dots, p$. Using the change of variables $z = Gx$,

$$F_y(z) = \mathcal{F}_y(G^{-1}z) = \beta \sum_{i=1}^{p-1} |z_i| + Q_y(z),$$

where $Q_y(z) = \|Bz - y\|^2$ and $B = AG^{-1}$.

For every $i = 1, \dots, p$, the $i$th column of $B$ is denoted $\mathrm{b}_i$ and is supposed non-zero. Assumptions H1, H2, H3 and H4 clearly hold. The minimization scheme follows from (21) and (22). For every $k \in \mathbf{N}$,

- if $i \in \{1, \dots, p-1\}$, calculate $\xi_i^{(k)} = -2\mathrm{b}_i^T y$ $2\mathrm{b}_i^T B(z_1^{(k)}, z_2^{(k)}, \dots, z_{i-1}^{(k)}, 0, z_{i+1}^{(k-1)}, \dots, z_p^{(k-1)})$;

  if $\big|\xi_i^{(k)}\big| \le \beta$, then $z_i^{(k)} = 0$,

  if $\xi_i^{(k)} < -\beta$, then $z_i^{(k)} = -\dfrac{\xi_i^{(k)} + \beta}{2\|\mathrm{b}_i\|^2} > 0$,

  if $\xi_i^{(k)} > \beta$, then $z_i^{(k)} = -\dfrac{\xi_i^{(k)} - \beta}{2\|\mathrm{b}_i\|^2} < 0$;

- $z_p^{(k)} = -\dfrac{\xi_p^{(k)}}{2\|\mathrm{b}_p\|^2}$.

Notice that for images H2 is not satisfied, so this method cannot be applied.

## 3. Detection and Smoothing of Outliers

### *3.1. Proposed Method*

We will process data $y \in \mathrm{R}^p$ corrupted as mentioned in I1–I2 in Section 1 by minimizing

$$\mathcal{F}_y(x) = \sum_{i=1}^{p} |x_i - y_i| + \beta \mathcal{Q}(x), \quad \text{where}$$

$$\mathcal{Q}(x) = \frac{1}{2}\sum_{i=1}^{p}\sum_{j\in\mathcal{N}_i} \varphi(x_i - x_j). \tag{24}$$

Here $\mathcal{N}_i$ denotes the set of the neighbors of $i$, for every $i = 1, \dots, p$. As usually [5, 15], we have $i \notin \mathcal{N}_i$ and $j \in \mathcal{N}_i \Leftrightarrow i \in \mathcal{N}_j$, for all $i$ and $j$. If $x$ is an 1D signal, $\mathcal{N}_i = \{i, i+1\}$, if $i = 2, \dots, p-1$; for a 2D image, $\mathcal{N}_i$ is the set of the 4, or the 8 pixels adjacent to $i$. An illustration is given in Fig. 1(a). Put

$$N = \max_{i=1}^{p} \#\mathcal{N}_i, \tag{25}$$

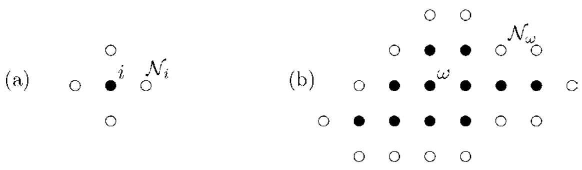



*Figure 1.* Neighborhoods. (a) The 4 nearest neighbors $\mathcal{N}_i$ (circles) of a pixel $i$ (bullet). (b) A connected subset $\omega$ (bullets) and its neighborhood $\mathcal{N}_\omega$ (circles) with respect to the neighborhood $\{\mathcal{N}_i\}$ in (a).

where # stands for cardinality. We consider that $\#\mathcal{N}_i = N$ is constant for any $i$ which is not on the boundaries of $x$. The function $\varphi$ in (24) is $\mathcal{C}^1$, convex, symmetric, and edge-preserving, as those given in (3)–(6).

Suppose that $\mathcal{F}_y$ reaches its minimum at $\hat{x}$ and put $\hat{h} = \{i : \hat{x}_i = y_i\}$. The idea of our method is that *every $y_i$, for $i \in \hat{h}$, is uncorrupted (i.e. regular), since $\hat{x}_i = y_i$; in contrast, every $y_i$, for $i \in \hat{h}^c$, can be an outlier since $\hat{x}_i \neq y_i$, in which case $\hat{x}_i$ is an estimate of the original entry*. In particular, we define the *outlier detector* function

$$y \to \hat{h}^c = \{i \in \{1, \ldots, p\} : \hat{x}_i \neq y_i\}, \qquad (26)$$

where $\hat{x}$ is such that $\mathcal{F}_y(\hat{x}) \leq \mathcal{F}_y(x), \ \forall x \in \mathrm{R}^p$.

The rationale of this method relies on the properties of the minimizers $\hat{x}$ of $\mathcal{F}_y$ when $y$ involves outliers.

### 3.2. Important Properties of $\hat{\mathbf{x}}$

Theorem 1 is straightforward to adapt to (24).

**Corollary 1.** *For $y \in \mathrm{R}^p$ given, consider $\mathcal{F}_y$ of the form* (24), *where $\varphi$ is $\mathcal{C}^1$ and convex. The function $\mathcal{F}_y$ reaches its minimum at $\hat{x} \in \mathrm{R}^p$ if, and only if,*

$$i \in \hat{h} \Rightarrow \left| \sum_{j \in \mathcal{N}_i} \varphi'(y_i - \hat{x}_j) \right| \leq \frac{1}{\beta}, \qquad (27)$$

$$i \in \hat{h}^c \Rightarrow \sum_{j \in \mathcal{N}_i} \varphi'(\hat{x}_i - \hat{x}_j) = \frac{\sigma_i}{\beta}, \quad \textit{for} \quad \sigma_i = \mathrm{sign}\left( \sum_{j \in \mathcal{N}_i} \varphi'(y_i - \hat{x}_j) \right), \qquad (28)$$

*where*

$$\hat{h} = \{i \in \{1, \ldots, p\} : \hat{x}_i = y_i\}. \qquad (29)$$

By (27), the subset $Y_* \subset \mathrm{R}^p$ given below is the set of all outlier-free data,

$$Y_* = \left\{ y \in \mathrm{R}^p : \left| \sum_{j \in \mathcal{N}_i} \varphi'(y_i - y_j) \right| \leq \frac{1}{\beta}, \quad \forall\, i = 1, \ldots, p \right\}, \qquad (30)$$

since $\mathcal{F}_y$ reaches its minimum at $\hat{x} = y$ for every $y \in Y_*$. Observe that $Y_*$ is of dimension $p$ and contains signals or images with smoothly varying and textured areas, and edges.

In (28), $\sigma_i \in \{-1, 1\}$, since (27) is false for every $i \in \hat{h}^c$. Suppose there is an outlier at $i$, then $\hat{x}_i$ satisfies (28). Two situations can occur:

- $y_i$ is quite larger than its neighbors, $\sum_{j \in \mathcal{N}_i} \varphi'(y_i - \hat{x}_j) > \frac{1}{\beta}$, which entails that $\sigma_i = 1$. We can then write $\sum_{j \in \mathcal{N}_i} \varphi'(y_i - \hat{x}_j) > \frac{\sigma_i}{\beta} = \sum_{j \in \mathcal{N}_i} \varphi'(\hat{x}_i - \hat{x}_j)$. Since $\varphi'$ is increasing on R, we have $\hat{x}_i < y_i$.
- $y_i$ is quite smaller than its neighbors, $\sum_{j \in \mathcal{N}_i} \varphi'(y_i - \hat{x}_j) < -\frac{1}{\beta}$, then $\sigma_i = -1$. In a similar way it is found that now $\hat{x}_i > y_i$.

In both cases $y_i$ incurs smoothing since $\hat{x}_i$ is closer to its neighbors $\hat{x}_j$ for $j \in \mathcal{N}_i$ than $y_i$.

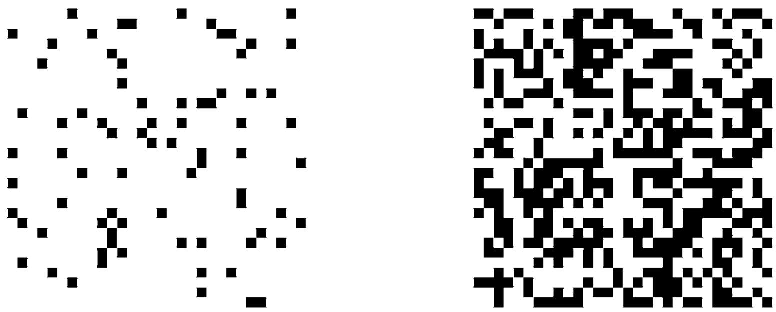
**Lemma 3.** *Let $\mathcal{F}_y$ be as in Corollary* 1. *Let $\hat{x}$, $\hat{h}$ and $\sigma_i$, for all $i \in \hat{h}^c$, read as there. Define $Y_{\hat{x}}$ by*

$$Y_{\hat{x}} = \left\{ \gamma \in \mathrm{R}^p : \begin{array}{|ll} \gamma_i = y_i & \textit{if } i \in \hat{h}, \\ \sigma_i \gamma_i \geq \sigma_i y_i & \textit{if } i \in \hat{h}^c. \end{array} \right\} \qquad (31)$$

*Then for every $\gamma \in Y_{\hat{x}}$, the function $\mathcal{F}_\gamma$ reaches its minimum at $\hat{x}$.*

All functions $\mathcal{F}_\gamma$ for $\gamma \in Y_{\hat{x}}$ reach their minimum at the same $\hat{x}$, even if $\gamma_i \to \sigma_i \infty$, for any $i \in \hat{h}^c$. Thus, any $\hat{x}_i$ for $i \in \hat{h}^c$ (the estimate of an outlier) is *independent* of the exact value of $\gamma_i$ (the outlier).

In the following, given a vector $v \in \mathrm{R}^p$ and a subset $h = \{h_1, \ldots, h_{\#h}\} \subset \{1, \ldots, p\}$, we write $v_h \in \mathrm{R}^{\#h}$ for the restriction of $v$ to those entries of $v$ whose indexes are in $h$, i.e. $v_h[i] = v[h_i]$, for $i = 1, \ldots, \#h$. If data are slightly corrupted (e.g., Fig. 2(a)), many outliers are isolated—if $y_i$ is an outlier, its neighbors are not outliers. When data are highly corrupted (e.g., Fig 2(b)), many outliers are neighbors and form patches. Then



(a) 10% impulse noise. (b) 45% impulse noise.

*Figure 2.* Locations of outliers in black, locations of regular data entries in white.

we expect that $\hat{h}^c$ involves many indexes which are neighbors. Let us represent $\hat{h}^c$ as a union of connected components,[3] say $\omega_k$ for $k = 1, \ldots, m$, with respect to $\{\mathcal{N}_i\}_{i=1}^p$. Given $\omega$, a connected component of $\hat{h}^c$, let $f_y$ be the function which for every $x_\omega \in \mathrm{R}^{\#\omega}$ yields

$$f_y(x_\omega) = \sum_{i\in\omega}\left(\sum_{j\in\mathcal{N}_i\cap\hat{h}}\varphi(x_i - y_j) + \frac{1}{2}\sum_{j\in\mathcal{N}_i\cap\omega}\varphi(x_i - x_j) - \frac{\sigma_i x_i}{\beta}\right). \quad (32)$$

Let $\hat{x}$ minimizes $\mathcal{F}_y$ as given in (24) and $\hat{h}$ be defined as in (29). Let $\omega$ be a connected component of $\hat{h}^c$. By (28), all entries $\hat{x}_i$, for $i \in \omega$, are determined by the following system of $\#\omega$ equations:[4]

$$D_i f_y(\hat{x}_\omega) = 0, \quad \forall\, i \in \omega, \quad (33)$$

where

$$\sigma_i = \mathrm{sign}\left(\sum_{j\in\mathcal{N}_i\cap\hat{h}}\varphi'(y_i - y_j) + \sum_{j\in\mathcal{N}_i\cap\omega}\varphi'(y_i - \hat{x}_j)\right).$$

We see that all the influence of an $y_i$ for $i \in \omega$ (an outlier) on the minimizer $\hat{x}_\omega$ of $f_y$ is reduced to the binary value $\sigma_i$. Notice that $\hat{x}_\omega$ is a function of the entries $y_j$ for $j \in N_\omega$, where $N_\omega$ is the neighborhood of $\omega$ with respect to $\{\mathcal{N}_i\}$ (see Fig. 1(b)):

$$N_\omega = \left(\bigcup_{i\in\omega}\mathcal{N}_i\right)\Big\backslash\, \omega \subset \hat{h}, \quad (34)$$

and that $\hat{x}_\omega$ is independent of all $y_j$ for $j \notin \mathcal{N}_\omega \cup \omega$.

**Lemma 4.** *Suppose that $\varphi$ is $\mathcal{C}^1$ and strictly convex. Let $\hat{h} \subset \{1, \ldots, p\}$ and $\hat{h}^c$ be nonempty. Consider $\omega$ a connected component of $\hat{h}^c$. We have*:

(a) *For every $\gamma \in \mathrm{R}^p$, the function $f_\gamma$ defined according to* (32) *is strictly convex.*
(b) *Given $y \in \mathrm{R}^p$, let $\hat{x}_\omega$ be such that* (33) *is satisfied. Then there exists $\xi > 0$ and a continuous function $\mathcal{X}_\omega$ such that $\gamma \in B(y, \xi)$ leads to $D_i f_\gamma(\mathcal{X}_\omega(\gamma)) = 0$, for all $i \in \omega$.*

Hence, $\hat{x}_\omega$ results from a continuous minimizer function, namely $x_\omega$ , hence it is stable under perturbations of $y$.

**Theorem 3.** *For $y \in \mathrm{R}^p$ given, let $\mathcal{F}_y$ be of the form* (24), *where $\varphi$ is $\mathcal{C}^1$ and strictly convex. Let $\mathcal{F}_y$ reach its minimum at $\hat{x}$. Define $\hat{h}$ and $\sigma_i$, for every $i \in \hat{h}^c$, as in Corollary* 1. *Suppose that $\hat{h}$ is nonempty and that for every $i \in \hat{h}$, the inequality in* (27) *is strict. Then there is $\rho > 0$ such that if $Y_{\hat{h}}$ reads*

$$Y_{\hat{h}} = \left\{\gamma \in \mathrm{R}^p : \left|\begin{array}{ll} |\gamma_i - y_i| \le \rho & \forall\, i \in \hat{h} \\ \sigma_i\gamma_i \ge \sigma_i y_i - \rho & \forall\, i \in \hat{h}^c \end{array}\right.\right\},$$

*then for every $\gamma \in Y_{\hat{h}}$, the relevant $\mathcal{F}_\gamma$ reaches its minimum at an $\hat{\chi} \in \mathrm{R}^p$ such that*

$$\{i \in \{1, \ldots, p\} : \hat{\chi}_i = \gamma_i\} = \hat{h} \quad (35)$$

*and* $\mathrm{sign}(\sum_{j\in\mathcal{N}_i}\varphi'(\gamma_i - \hat{\chi})) = \sigma_i$, *for every $i \in \hat{h}^c$.*

By this theorem, $\mathrm{R}^p$ contains *open domains $Y_{\hat{h}}$* corresponding to different $\hat{h}$. Real data do belong to such domains and give rise to minimizers $\hat{x}$ for which $\hat{h} = \{i : \hat{x}_i = y_i\}$ is nonempty.

### 3.3. Meaning of the Mathematical Results

Let us consider how outliers are processed in some typical configurations.

*Example 2* (Constant area). Let $i \in \{1, \ldots, p\}$ be such that no $j \in \mathcal{N}_i$ belongs to the boundaries of $x$. Without loss of generality, consider that $y_k = 0$ for all $k \in (\bigcup_{j \in \mathcal{N}_i} \mathcal{N}_j) \cup \mathcal{N}_i$. Let $\theta > 0$ be such that

$$\varphi'(\theta) = \frac{1}{\beta N}. \tag{36}$$

Based on Corollary 1, it is easy to see that

$$|y_i| \leq \theta \Rightarrow \hat{x}_j = y_j, \quad \forall j \in \mathcal{N}_i \cup \{i\},$$
$$|y_i| > \theta \Rightarrow \begin{cases} \hat{x}_i = \theta \ \mathrm{sign}(y_i), \\ \hat{x}_j = y_j, \quad \forall j \in \mathcal{N}_i. \end{cases}$$

Supposing that $\varphi$ is $\mathcal{C}^2$ and admits 3 derivatives, and noticing that $\varphi'(0) = 0$, we have $\varphi'(\hat{x}_i) = \varphi''(0)\ \hat{x}_i + \hat{x}_i^2 \varepsilon(\hat{x}_i)$, where $\varepsilon(\hat{x}_i)$ goes to zero as $\hat{x}_i \to 0$. Then

$$\hat{x}_i = \frac{\varphi'(\hat{x}_i) - \hat{x}_i^2 \varepsilon(\hat{x}_i)}{\varphi''(0)}. \tag{37}$$

Grossly speaking, the more $\varphi''(0)$ is large, the more smoothing is improved, i.e. $\hat{x}_i$ is closer to zero.

*Example 3* (Breakpoint in an 1D signal). Let $y \in \mathrm{R}^p$ be such that for some $i \in \{3, \ldots, p-2\}$,

$$y_{i-2} = y_{i-1} = 0 \quad \text{and} \quad y_{i+1} = y_{i+2} = \delta > 0.$$

We suppose that $\beta < 1/\varphi'(\delta)$. Let $\theta > 0$ be the constant such that

$$\varphi'(\theta) + \varphi'(\theta - \delta) = \frac{1}{\beta}. \tag{38}$$

Notice that $\theta > \delta$. It is not difficult to check that

$$y_i \in [0, \theta] \Rightarrow \begin{cases} \hat{x}_{i-1} = y_{i-1} = 0, \\ \hat{x}_i = y_i, \\ \hat{x}_{i+1} = y_{i+1} = \delta, \end{cases}$$

whereas

$$y_i > \theta \Rightarrow \begin{cases} \hat{x}_{i-1} = y_{i-1} = 0, \\ \hat{x}_i = \theta \in (\delta, y_i), \\ \hat{x}_{i+1} = y_{i+1} = \delta. \end{cases}$$

The conditions for an edge between two constant areas in a 2D image are basically the same.

Several consequences of the theory presented in Section 3.2 are worth to emphasize.

*Detection of outliers.* The inequality in (27) provides the rule to decide whether an entry $y_i$ is an outlier or not. It is important to notice that $y_i$ is compared only with *faithful neighbors*—regular entries $y_j$ for $j \in \mathcal{N}_i \cap \hat{h}$ and estimates of outliers $\hat{x}_j$ for $j \in \mathcal{N}_i \cap \hat{h}^c$. This is crucial for the reliability of the detection of outliers and allows very restricted neighborhoods $\{\mathcal{N}_i\}$ to be used (e.g. the four nearest neighbors in the case of images). Since $\varphi'$ is increasing on R and $\varphi'(0) = 0$, we see that if $y_i$ is too dissimilar with respect to its neighbors, (27) fails and $y_i$ is replaced by an $\hat{x}_i$ satisfying (28). The constant $\sigma_i \in \{-1, 1\}$ gives the direction of the deviation of $y_i$. By Lemma 3, $\hat{x}_i$ remains the same, independently of the amplitude of this deviation. Hence the robustness of $\hat{x}$. Theorem 3 shows that the detection of outliers is *stable*: the set $\hat{h}^c$ is constant under small perturbations of regular data entries $y_i$ for $i \in \hat{h}$ and under arbitrary deviations of outliers $y_i$ for $i \in \hat{h}^c$ in their directions $\sigma_i$.

*Restoration of sets of neighboring outliers.* Let $\hat{x}$ minimize $\mathcal{F}_y$ and $\hat{h}$ read as in (29). If $\omega$ is a connected component of $\hat{h}^c$, then $\hat{x}_\omega$ is the minimizer of a function $f_y$ of the form (32). Observe that $f_y$ is a *regularized cost-function* similar to (1). Its first term encourages every boundary entry for $\omega$, namely $\hat{x}_i$ for an $i \in \omega$ such that $\mathcal{N}_i \cap \hat{h} \neq \emptyset$, to fit neighboring regular data entries $y_j$ for $j \in \mathcal{N}_i \cap \hat{h}$. Its second term is a smoothness constraint on $\hat{x}_\omega$ since it favors neighboring entries for $\omega$, say $\hat{x}_i$ and $\hat{x}_j$, with $i, j \in \omega$ and $j \in \mathcal{N}_i$, to have similar values. This term is absent if $\#\omega = 1$. The last term introduces a small bias. Based on the theory of edge-preserving regularization, we can expect that *edges in $\hat{x}_\omega$ are well restored if $\varphi$ is a good edge-preserving function*. By Lemma 4, $\hat{x}_\omega$ results from a continuous minimizer function $\mathcal{X}_\omega$ which depends only on the neighboring regular data entries $y_i$ for $i \in \mathcal{N}_\omega$ and is independent of the value of the outliers $y_i$ for $i \in \omega$.

*Stability of the minimizers $\hat{x}$ of $\mathcal{F}_y$.* Theorem 3 shows that the minimizer $\hat{x}$ of $\mathcal{F}_y$ fits all regular data entries $y_i$ for $i \in \hat{h}$ when these incur weak perturbations, while $\hat{x}$ remains unchanged under arbitrary perturbations of outliers $y_i$ for $i \in \hat{h}^c$ in the directions determined by

$\sigma_i$. Thus, the minimizers $\hat{x}$ of $\mathcal{F}_y$ are stable and are not contaminated by the amplitude of the outliers.

*Bounds on $\beta$.* Suppose that $x^*$ is composed of the most typical configurations—smoothly varying and textured areas, edges—for the signals and images which are restored. By Corollary 1, the condition ensuring that $\mathcal{F}_{x^*}$ reaches its minimum at $x^*$ is $\beta \leq (\max_{i=1}^{p} |\sum_{j\in\mathcal{N}_i} \varphi'(x_i^* - x_j^*)|)^{-1}$. This bound for $\beta$ can be quite small—e.g., consider that $x^*$ is an image with edges of high amplitude and contours with sharp creases. Observe that by (36) and (38), decreasing $\beta$ increases the threshold $\theta$ over which a data entry is detected as an outlier, whereas by (38), for $\beta$ fixed, $\theta$ is higher at edges than in homogeneous regions. Then outliers which are not dissimilar enough with respect to their neighbors, and which are near to edges, may be omitted. If $\beta$ is larger than the bound given above, some regular data entries $y_i$ which are near to edges can falsely be detected as outliers and yield $i \in \hat{h}^c$. Any such $y_i$ will be replaced by an $\hat{x}_i$ which minimizes an $f_y$ of the form (32). Nevertheless, if $\varphi$ has good edge-preserving properties, this $\hat{x}_i$ will preserve the edges and the resultant error will be small.

On the other hand, if data $y \in \mathbb{R}^p$ contain outliers, we must have $\beta > (\max_{i=1}^{p} |\sum_{j\in\mathcal{N}_i} \varphi'(y_i - y_j)|)^{-1}$, since otherwise Corollary 1 shows that $\mathcal{F}_y$ reaches its minimum at $\hat{x} = y$. Suppose that $K$ among all $p$ entries of $y$ are corrupted with outliers. In an ideal case when outliers are isolated and located in completely homogeneous areas, we should take $\beta = (|\sum_{j\in\mathcal{N}_{i_K}} \varphi'(y_{i_K} - y_j)|)^{-1}$, where $\{i_k\}_{k=1}^{p}$ is the rearrangement of the indexes $\{1, \ldots, p\}$ defined by $|\sum_{j\in\mathcal{N}_{i_1}} \varphi'(y_{i_1} - y_j)| \geq \cdots \geq |\sum_{j\in\mathcal{N}_{i_p}} \varphi'(y_{i_p} - y_j)|$. However, in order to have a good $\theta$, even if outliers occur near to edges and form patches, we need $\beta > (|\sum_{j\in\mathcal{N}_{i_K}} \varphi'(y_{i_K} - y_j)|)^{-1}$.

*Why choosing $\psi(t) = |t|$ in* (24)*?* Let us consider $\mathcal{F}_y(x) = \sum_{i=1}^{p} \psi(x_i - y_i) + \beta \mathcal{Q}(x)$, where $\mathcal{Q}$ is as in (24). If $\psi$ is smooth at zero, the set $h$ is empty for almost all $y \in \mathbb{R}^p$ [24]. So consider that and $\psi$ satisfies H1 along with $\psi(t) = \psi(-t)$ and $\psi'(0^+) = 1$. The necessary and sufficient condition for minimum given in Corollary 1 remains the same except that (28) now reads

$$\psi'(\hat{x}_i - y_i) + \beta \sum_{j\in\mathcal{N}_i} \varphi'(\hat{x}_i - \hat{x}_j) = 0, \quad \text{if } i \in \hat{h}^c.$$

If $\psi'$ is not constant on $(-\infty, 0)$ and $(0, \infty)$, the estimate $\hat{x}_i$ of an outlier $y_i$ depends on the exact value of $y_i$. In such a case, the important property stated in Lemma 3 does not hold. The only convex and symmetric $\psi$ such that $\psi'$ is constant on $(-\infty, 0)$ and $(0, \infty)$, is $\psi(t) = \lambda|t|$, for $\lambda \in \mathbb{R}$.

### 3.4. Minimization Algorithm

We calculate $\hat{x}$ as $\hat{x} = \hat{z} + y$ where $\hat{z}$ minimizes $F_y$, the equivalent form for $\mathcal{F}_y$ introduced in (7):

$$F_y(z) = \sum_{i=1}^{p} |z_i| + \beta Q_y(z), \quad \text{where}$$
$$Q_y(z) = \frac{1}{2} \sum_{i=1}^{p} \sum_{j\in\mathcal{N}_i} \varphi(z_i + y_i - z_j - y_j).$$

Based on Theorem 3, we can expect that $\hat{z}_i = 0$ for a certain number of indexes $i$. This suggests we initialize with $z^{(0)} = 0$. At each iteration $k$, for every $i = 1, \ldots, p$, we calculate

$$\xi_i^{(k)} = \beta \sum_{j\in\mathcal{N}_i} \varphi'(y_i - z_j - y_j), \quad \text{where}$$
$$Z = \left(z_1^{(k)}, \ldots, z_{i-1}^{(k)}, 0, z_{i+1}^{(k-1)}, \ldots, z_p^{(k-1)}\right); \quad (39)$$

if $|\xi_i^{(k)}| \leq 1$, then $z_i^{(k)} = 0$,
if $|\xi_i^{(k)}| > 1$, then find $z_i^{(k)}$ by solving
$\beta \sum_{j\in\mathcal{N}_i} \varphi'(z_i^{(k)} + y_i - z_j - y_j) = \operatorname{sign}(\xi_i^{(k)})$,
knowing that $\operatorname{sign}(z_i^{(k)}) = -\operatorname{sign}(\xi_i^{(k)})$.

*Remark 2.* The calculation of each $z_i^{(k)}$ involves only the entries whose indexes are in $\mathcal{N}_i$. So, at each step we can update simultaneously any subset of entries $\{i_1, \ldots, i_K\} \subset \{1, \ldots, p\}$ provided that $i_j \cap N_{i_k} = \emptyset$ for all $j, k \in \{1, \ldots, K\}$ with $j \neq k$. The minimization can be implemented in a parallel way.

(a) Original signal. (b) Data $y$ "—", original $x^*$ "...".

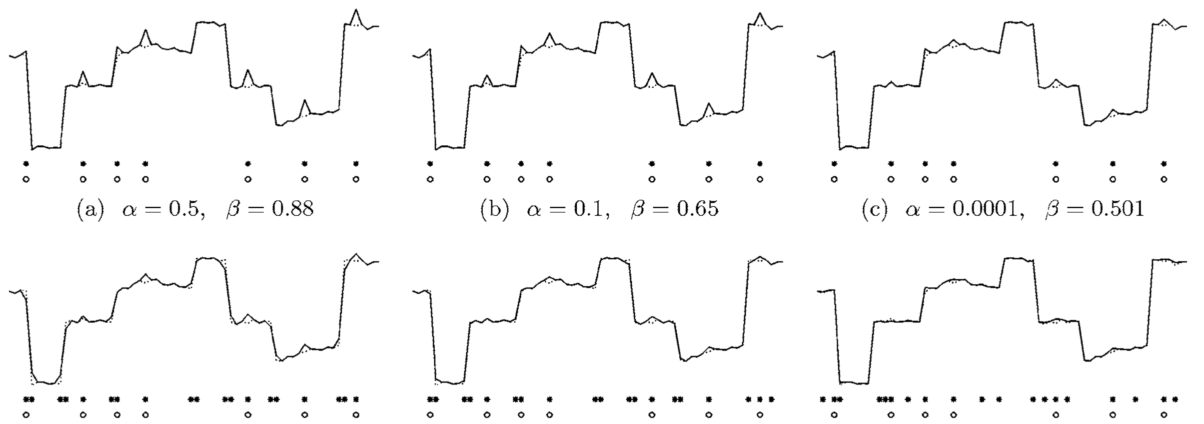
*Figure 3.* A simple example of signal contaminated with outliers.

If $\varphi$ is $\mathcal{C}^1$, symmetric and satisfies (11), $F_y$ satisfies all assumptions H1–H4. By Theorem 2, the sequence $\hat{z}^{(k)}$ generated by (39) converges to the sought $\hat{z}$.

## 4. Experiments

### *4.1. Explanatory Experiment*

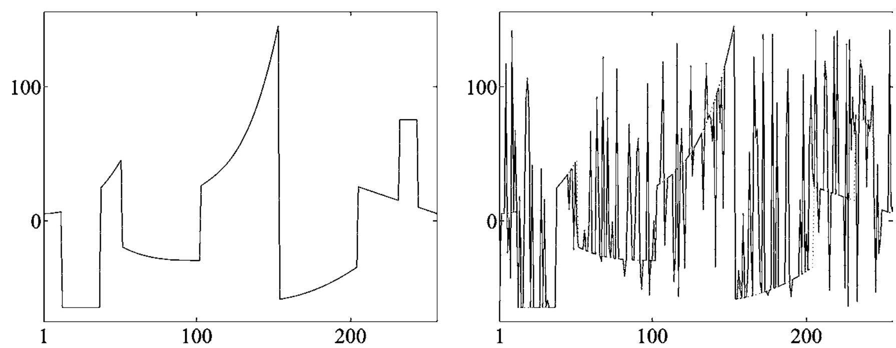
This simple experiment illustrates the main features of the minimizers of cost-functions of the form (24). We consider an $\mathcal{F}_y$ where $\mathcal{N}_i = \{i, i+1\}$ and $\varphi(t) = \sqrt{\alpha + t^2}$. The original signal $x^*$, shown in Fig. 3(a), contains sharp edges and slightly textured zones. The data $y$, plotted in Fig. 3(b), contain 7 outliers. The minimizers $\hat{x}$ of $\mathcal{F}_y$ are presented in Fig. 4, where each column corresponds to a different value of $\alpha$. According to (27), $\beta \leq 0.5$ yields $\hat{x} = y$, since $|\sum_{j\in\mathcal{N}_i} \varphi'(y_i - y_j)| < 2\sup_{t\in\mathbb{R}} |\varphi'(t)| = 2 \leq 1/\beta$, for all $i$. The solutions displayed in (a), (b) and (c) correspond to the "smallest" $\beta$ leading to a minimizer $\hat{x}$ such that $\hat{x}_i \neq y_i$ if, and only if, $y_i$ is an outlier, i.e. $\hat{h}^c = \{i : \hat{x}_i \neq y_i\} = \{i : y_i \neq x_i^*\}$. Outliers are better smoothed when $\alpha$ is small—as in (c)—which corresponds to a large value for $\varphi''(0)$, as suggested by (37). Comparing the first and the second row of Fig. 4 shows that for $\alpha$ fixed, smoothing of outliers is improved by increasing $\beta$. The set $\hat{h}^c$, corresponding to (d), (e) and (f), contains all outliers in $y$ plus 12 regular entries of $y$ which are "erroneously" smoothed. When $\alpha$ is large—as in (d)—this entails blurring of edges. When $\alpha$ is small, as in (f), edges are well preserved and the error $|\hat{x}_i - x_i^*|$ is small.



(a) $\alpha = 0.5$, $\beta = 0.88$ (b) $\alpha = 0.1$, $\beta = 0.65$ (c) $\alpha = 0.0001$, $\beta = 0.501$

(d) $\alpha = 0.5$, $\beta = 1.8$ (e) $\alpha = 0.1$, $\beta = 1.44$ (f) $\alpha = 0.0001$, $\beta = 0.54$

*Figure 4.* Minimizers $\hat{x}$ of $\mathcal{F}_y$ where $\varphi(t) = \sqrt{\alpha + t^2}$ and $\mathcal{N}_i = \{i-1, i+1\}$: restored signal $\hat{x}$ "—", original signal $x^*$ "...", locations of outliers in $y$ "o", estimated locations of outliers $\hat{h}^c = \{i : \hat{z}_i \neq 0\}$ "$*$".




(a) Original signal $x^*$. (b) Data $y$ with 45% random-valued noise.

*Figure 5.* Piece-wise polynomial signal in 45% random-valued noise.

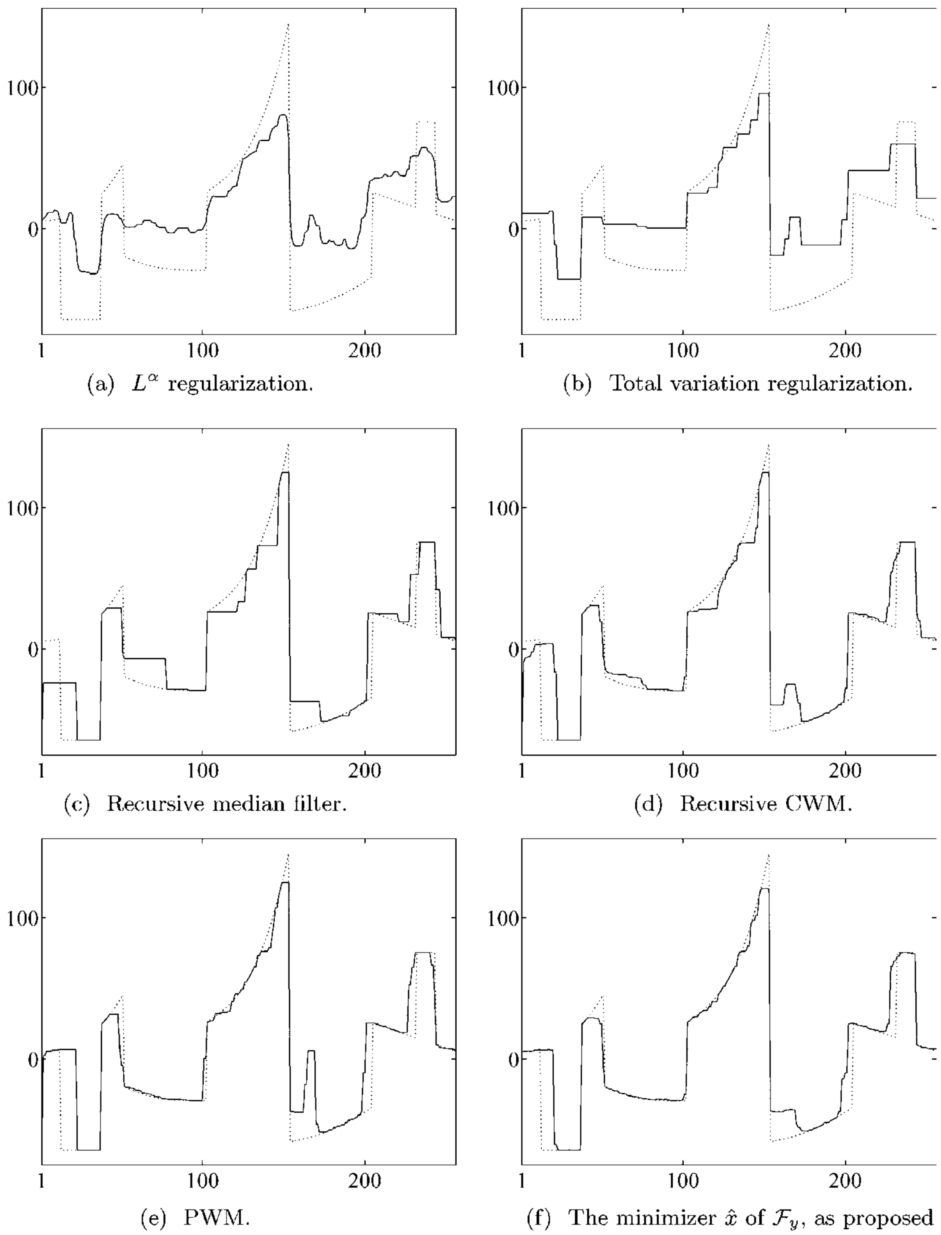
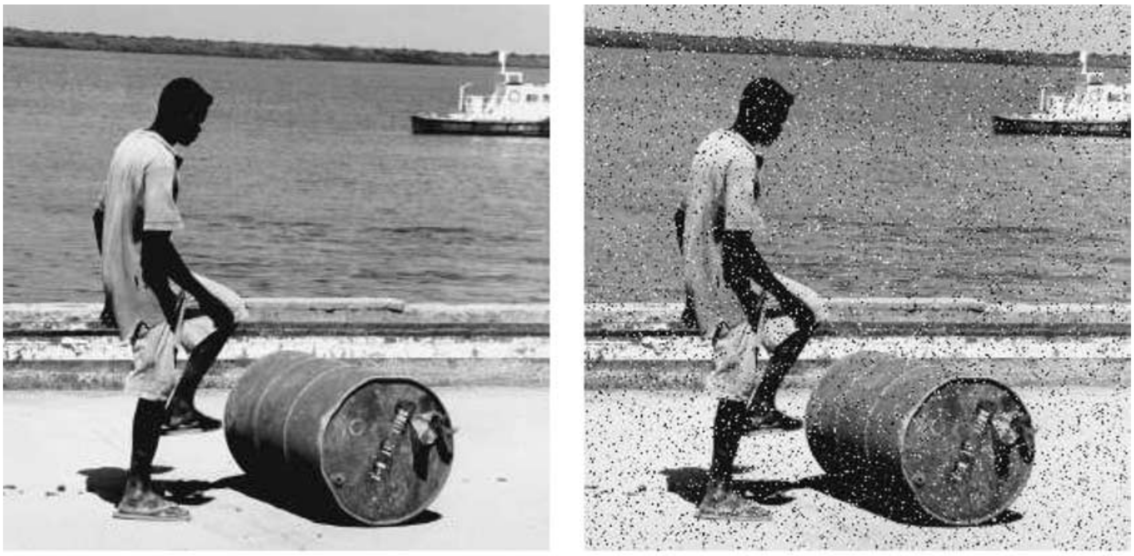



*Figure 6.* Restoration using different methods: restored signal "—", original signal "· · ·".

### *4.2. Signal with 45% Random-Valued Impulse Noise*

The original signal $x^*$ is shown in Fig. 5(a). Data $y$ in Fig. 5(b) are corrupted with 45% impulse noise with values uniformly distributed on $[\min_{1\le i\le p} x_i^*, \max_{1\le i\le p} x_i^*]$. The results in Fig. 6(a) and (b) are obtained by minimizing classical cost-functions $\mathcal{F}_y(x) = \|x-y\|^2 + \beta \sum_{i=1}^{p-1} |x_i - x_{i+1}|^\alpha$. The minimizer in (a) corresponds to $\alpha = 1.3$ and $\beta = 100$, the one in (b) to $\alpha = 1$ (total-variation regularization) and $\beta = 120$. Cost-functions of this kind cannot deal with outliers. We present experiments with several order-statistic (OS) filters: median filter, center-weighted median (CWM) filter[5] and permutation-weighted (PWM) filter [9].[6] For each method, the parameters (window size, number of iterations,



(a) Original image $x^*$ (size $300 \times 300$). (b) Data $y$ with 10% random-valued noise.

*Figure 7.* Original picture and data with random-valued noise.

recursivity, other parameters) are tuned to yield the best result. The solution in Fig. 6(c) is calculated with a 9-window recursive median filter. The result in (d) is obtained with a 9-window recursive CWM for $\alpha = 2$. The one in (e) correspond to a 9-window PWM filter for $\alpha = 4$. All solutions obtained using OS filters—(c), (d) and (e)—exhibit important defects. The solution in (f) is the minimizer of $\mathcal{F}_y(x) = \sum_{i=1}^{p} |x_i - y_i| + 0.65 \sum_i (\sum_{j \in \mathcal{N}_i} \varphi(x_i - x_j) + 0.6 \sum_{j \in \mathcal{N}_i'} \varphi(x_i - x_j))$, where $\mathcal{N}_i = \{i-1, i+1\}$, $\mathcal{N}_i' = \{i-2, i+2\}$ and $\varphi(t) = \sqrt{0.2 + t^2}$. Our method provides significant improvement over the other methods.

### 4.3. Picture with 10% Random-Valued Impulse Noise

The original image $x^*$ is shown in Fig. 7(a). In Fig. 7(b), 10% of the pixels have random values uniformly distributed on $[\min_i x^*, \ \max_i x^*]$. Denoising results using different methods are displayed in Fig. 8, where all parameters are finely tuned. The restoration in Fig. 8(a) corresponds to one iteration of a $3 \times 3$ window median filter. The image in (b) is calculated using a $3 \times 3$ window recursive CWM for $\alpha = 3$. The result in (c) corresponds to a $3 \times 3$ window PWM filter for $\alpha = 4$. These images are slightly blurred, the texture of the sea is deformed, and several outliers still remain.

The image $\hat{x}$ in (d) is the minimizer $\hat{x}$ of $\mathcal{F}_y$ as given in (24), with $\varphi(t) = |t|^{1.1}$, $\mathcal{N}_i$ the set of the 4 adjacent neighbors and $\beta = 0.3$. All details are well preserved and the image is difficult to distinguish from the original $x^*$. Indeed, for 85% of the pixels, $\left|\hat{x}_i - x_i^*\right| / \Delta \leq 2\%$, where $\Delta = \max_i x^* - \min_i x^*$. Based on the experiments in Section 4.1, one can expect that a smaller $\beta$ can reduce the number of regular data entries erroneously detected as outliers, but that detected outliers are not smoothed enough. The minimizer $\hat{x}$ of $\mathcal{F}_y$ for $\beta = 0.24$ is shown in Fig. 9(a). Now, for 90% of the pixels, $|\hat{x}_i - x_i^*| / \Delta \leq 2\%$. The image in (b), denoted $\tilde{x}$, is obtained using *conditional median* on $\hat{h}^c$. More precisely, $\tilde{x}_i = y_i$ for all $i \in \hat{h}$. For every $i \in \hat{h}^c$ with $\mathcal{N}_i \subset \hat{h}$, we take $\tilde{x}_i = \text{median}\{y_j : j \in \mathcal{N}_i\}$. For every $i$ belonging to a larger connected component $\omega \subset \hat{h}^c$, we take $\tilde{x}_i$ equal to the median of the 2 nearest $y_j \in N_\omega \subset \hat{h}$, where $N_\omega$ is as in (34). This solution is almost undistinguishable from Fig. 8(d).

### 4.4. Picture with 45% Salt-and-Pepper Noise

45% of the entries in Fig. 10(a), with locations uniformly distributed over the grid of the image, are equal

(a) Median filtering ($\|\hat{x}-x^*\|_2 = 4155$).

(b) Recursive CWM ($\|\hat{x}-x^*\|_2 = 3566$).

(c) PWM ($\|\hat{x}-x^*\|_2 = 3984$).

(d) The proposed method ($\|\hat{x}-x^*\|_2 = 2934$).

*Figure 8.* Restoration from a picture with 10% random-valued noise.

either to $\min_i x^*$, of to $\max_i x^*$, with probability $1/2$. The image in Fig. 10(b) is obtained after 2 iterations of a $3 \times 3$ window recursive median filter. This image has a poor resolution and exhibits a stair-case effect. The result in (c) is calculated using a $5 \times 5$ window recursive CWM filter for $\alpha = 7$. The image in (d) results from a $7 \times 7$ window PWM filter for $\alpha = 14$. Although the images in (c) and (d) are better restored, the resolution is poor, there are artifacts along the edges and the texture of the sea is destroyed. The images in Fig. 11 are the minimizers of $\mathcal{F}_y$ as given in (24) where $\varphi(t) = |t|^{1.3}$ and $\mathcal{N}_i$ is the set of the 4 adjacent neighbors. The image in (a) corresponds to $\beta = 0.18$ and the one in (b) to $\beta = 0.2$. The quality of these restorations is clearly improved: the contours are neater, the texture of the sea in better preserved and some details on the boat can be distinguished.

(a) The minimizer $\hat{x}$ of $\mathcal{F}_y$ for $\beta = 0.24$ ($\|\hat{x}-x^*\|_2=3355$).

(b) Conditional median on $\hat{h}^c$ ($\|\hat{x}-x^*\|_2 = 3325$).

*Figure 9*. A variation of the proposed method.

(a) Data $y$ with 45% salt-and-pepper noise.

(b) Recursive median filter ($\|\hat{x}-x^*\|_2 = 7825$).

(c) Recursive CWM ($\|\hat{x}-x^*\|_2 = 7497$).

(d) PWM ($\|\hat{x}-x^*\|_2 = 6265$).

*Figure 10*. Picture with 45% salt-and-pepper noise.

(a) $\alpha = 1.3$ and $\beta = 0.18$ ($\|\hat{x} - x^*\|_2 = 6064$). (b) $\alpha = 1.3$ and $\beta = 0.2$ ($\|\hat{x} - x^*\|_2 = 6126$).

*Figure 11.* The minimizer $\hat{x}$ of $\mathcal{F}_y$.

## Appendix

For $x^* \in \mathrm{R}^p$ and $\rho > 0$, we denote $B(x^*, \rho) = \{x \in \mathrm{R}^p : \|x^* - x\| < \rho\}$ where $\|\cdot\|$ is the Euclidian norm on $\mathrm{R}^p$. Furthermore, $\overline{B}(x^*, \rho)$ is the relevant closed ball.

*Calculations relevant to Remark 1.* Put $b_\infty = \max_{i=1}^p \|b_j\|$. For any $\rho > 0$, we have

$$z \in \bar{B}(0, \rho),\ t \in [-\rho, \rho] \ \Rightarrow\ \left|b_j^T(z + te_i + \tilde{y})\right| \\ \leq\ b_\infty(2\rho + \|\tilde{y}\|),\quad \forall i, j \in \{1, \ldots, p\}.$$

Let $\eta_\delta > 0$ be the constant in (11) corresponding to $\delta = b_\infty(2\rho + \|\tilde{y}\|)$. For every $z \in \bar{B}(0, \rho)$ and $t \in [-\rho, \rho]$, and for all $i = 1, \ldots, p$,

$$\begin{aligned} &Q_y(z + te_i) - Q_y(z) \\ &\quad = \sum_{j=1}^r \left(\varphi\big(b_j^T(z + te_i + \tilde{y})\big) - \varphi\big(b_j^T(z + \tilde{y})\big)\right) \\ &\quad \geq t \sum_{j=1}^r \varphi'\big(b_j^T(z + \tilde{y})\big)\big(b_j^T e_i\big) + t^2 \eta_\delta \sum_{j=1}^r \big(b_j^T e_i\big)^2 \\ &\quad = t D_i Q_y(z) + \eta t^2, \end{aligned}$$

where $\eta = \min_{i=1}^p \eta_\delta \sum_{j=1}^r (b_j^T e_i)^2 > 0$ since $B$ does not involve zero-valued columns.

**Proof of Lemma 1:** If $t = 0$, the result is trivial. So consider that $t > 0$. Let $\tau \in (0, t)$, then $t - \tau/t \in (0, 1)$. Since $\psi_i$ is convex,

$$\begin{aligned} \psi_i(\tau) &= \psi_i\left(\frac{t - \tau}{t}0 + \frac{\tau}{t}t\right) \leq \frac{t - \tau}{t}\psi_i(0) + \frac{\tau}{t}\psi_i(t) \\ &= \psi_i(0) + \frac{\psi_i(t) - \psi_i(0)}{t}\tau. \end{aligned}$$

It follows that

$$\frac{\psi_i(\tau) - \psi_i(0)}{\tau} \leq \frac{\psi_i(t) - \psi_i(0)}{t}. \tag{40}$$

Since $\psi_i'(0^+)$ is well defined and finite,

$$\psi_i'(0^+) = \lim_{\tau \searrow 0} \frac{\psi_i(\tau) - \psi_i(0)}{\tau} \leq \frac{\psi_i(t) - \psi_i(0)}{t}.$$

The result for $t < 0$ is obtained in a symmetric way.

**Proof of Theorem 1:** Let $F_y$ reach its minimum at $\hat{z}$. Then for every $t > 0$ and $\tau > 0$,

$$\frac{F_y(\hat{z} - te_i) - F_y(\hat{z})}{-t} \leq 0 \leq \frac{F_y(\hat{z} + \tau e_i) - F_y(\hat{z})}{\tau}, \\ \forall i = 1, \ldots, p. \tag{41}$$

Using (7), for every $i = 1, \ldots, q$, and for every $t > 0$ and $\tau > 0$, we have

$$\frac{\psi_i(\hat{z}_i - t) - \psi_i(\hat{z}_i)}{-t} + \beta \frac{Q_y(\hat{z} - te_i) - Q_y(\hat{z})}{-t} \leq 0$$
$$\leq \frac{\psi_i(\hat{z}_i + \tau) - \psi_i(\hat{z}_i)}{\tau} + \beta \frac{Q_y(\hat{z} + \tau e_i) - Q_y(\hat{z})}{\tau}.$$

Let us consider this expression at the limit when $t \searrow 0$ and $\tau \searrow 0$. If $\hat{z}_i = 0$, we find

$$\psi_i'(0^-) + \beta D_i Q_y(\hat{z}) \leq 0 \leq \psi_i'(0^+) + \beta D_i Q_y(\hat{z}),$$

and hence (12). It $\hat{z}_i \neq 0$, we get

$$\psi_i'(\hat{z}_i) + \beta D_i Q_y(\hat{z}) \leq 0 \leq \psi_i'(\hat{z}_i) + \beta D_i Q_y(\hat{z}),$$

which leads to (13). If $q < p$, for every $i \in \{q + 1, \ldots, p\}$, the expression in (41) reads

$$\beta \frac{Q_y(\hat{z} - te_i) - Q_y(\hat{z})}{-t} \leq 0 \leq \beta \frac{Q_y(\hat{z} + \tau e_i) - Q_y(\hat{z})}{\tau}.$$

For $t \searrow 0$ and $\tau \searrow 0$, we find (14).

Reciprocally, let $\hat{z}$ satisfy (12), (13) and (14). For any $u \in \mathrm{R}^p$, let us consider $\Delta(u)$ given by

$$\begin{aligned}\Delta(u) &= F_y(\hat{z} + u) - F_y(\hat{z}) \\ &= \sum_{i \in \hat{h}} (\psi_i(u_i) - \psi_i(0)) + \sum_{i \in \hat{h}^c} (\psi_i(\hat{z}_i + u_i) - \psi_i(\hat{z}_i)) \\ &\quad + \beta(Q_y(\hat{z} + u) - Q_y(\hat{z})).\end{aligned} \tag{42}$$

When $\hat{h}^c$ is nonempty, put

$$\zeta = \frac{1}{2} \min_{i \in \hat{h}^c} |\hat{z}_i|;$$

then $\zeta > 0$ by (15). Since for every $i \in \hat{h}^c$, the function $\psi_i$ is convex and $\mathcal{C}^1$ on $B(\hat{z}_i, \zeta)$,

$$\psi_i(\hat{z}_i + u_i) - \psi_i(\hat{z}_i) \geq u_i \psi_i'(\hat{z}_i), \quad \forall u \in B(0, \zeta), \forall i \in \hat{h}^c. \tag{43}$$

Similarly, since $Q_y$ is convex and $\mathcal{C}^1$,

$$Q_y(\hat{z} + u) - Q_y(\hat{z}) \geq \sum_{i=1}^{p} u_i D_i Q_y(\hat{z}), \quad \forall u \in \mathrm{R}^p. \tag{44}$$

Introducing (43) and (44) into (42) shows that for all $u \in B(0, \zeta)$,

$$\Delta(u) \geq \sum_{i \in \hat{h}} (\psi_i(u_i) - \psi_i(0) + \beta u_i D_i Q_y(\hat{z})) \tag{45}$$
$$+ \sum_{i \in \hat{h}^c} u_i (\psi_i'(\hat{z}_i) + \beta D_i Q_y(\hat{z})) \tag{46}$$
$$+ \beta \sum_{i=q+1}^{p} u_i D_i Q_y(\hat{z}). \tag{47}$$

Consider first that $\hat{h}$ is non-empty. Applying Lemma 1 and (12) for every $i \in \hat{h}$ yields

$$\begin{aligned} u_i \geq 0 &\Rightarrow \psi_i(u_i) - \psi_i(0) + \beta u_i D_i Q_y(\hat{z}) \\ &\quad \geq \psi_i'(0^+)\, u_i + \beta u_i D_i Q_y(\hat{z}) \geq 0, \\ u_i \leq 0 &\Rightarrow \psi_i(u_i) - \psi_i(0) + \beta u_i D_i Q_y(\hat{z}) \\ &\quad \geq \psi_i'(0^-)\, u_i + \beta u_i D_i Q_y(\hat{z}) \geq 0. \end{aligned} \tag{48}$$

Then the term on the right side of (45) is non-negative. If $\hat{h}$ is empty, this term is absent. If $\hat{h}^c$ is nonempty, (13) implies that (46) is null; otherwise (46) is absent. If $p > q$, (47) is null by (14), otherwise it is absent. In all cases, $\Delta(u) \geq 0$, for all $u \in B(0, \zeta)$. It follows that $F_y$ reaches its minimum at $\hat{z}$.

- Since $t \to \beta D_i Q_y(z + (t - z_i)e_i)$ is increasing on R by H2 and $\psi_i$ is convex, if $\tau < 0 < t$, we have

$$\begin{aligned} &\psi_i'(\tau) + \beta D_i Q_y(\hat{z} + (\tau - \hat{z}_i)e_i) \\ &\quad \leq \psi_i'(0^-) + \beta D_i Q_y(\hat{z} - \hat{z}_i e_i) \\ &\quad < \psi_i'(0^+) + \beta D_i Q_y(\hat{z} - \hat{z}_i e_i) \\ &\quad \leq \psi_i'(t) + \beta D_i Q_y(\hat{z} + (t - \hat{z}_i)e_i). \end{aligned}$$

If (16) holds, i.e. if $\psi_i'(0^+) + \beta D_i Q_y(\hat{z} - \hat{z}_i e_i) < 0$, we see that (13) cannot be reached unless $\hat{z}_i \geq 0$. Since $\hat{z}_i \neq 0$, it follows that $\hat{z}_i > 0$. In the same way we find (17).

*Example 1* (*details*). We have $p = q$. We first check that the conditions of Theorem 1 hold for $\hat{z} = (0,\ y_1 - y_2 + \frac{1}{2\beta} \in Z$. In this case, $\hat{h} = \{1\}$. The condition on $y$ in (18) shows that $\hat{z}_2 < 0$. Then

$$\begin{aligned} \beta D_1 Q_y(\hat{z}) &= 2\beta(\hat{z}_1 - \hat{z}_2 + y_1 - y_2) = -1, \\ \psi'(\hat{z}_2) + \beta D_2 Q_y(\hat{z}) &= -1 + 2\beta(-\hat{z}_1 + \hat{z}_2 - y_1 + y_2) \\ &= 0. \end{aligned}$$

So, (12) and (13) are satisfied. Let us now consider $\hat{z} = (-(y_1 - y_2) - \frac{1}{2\beta},\ 0) \in Z$, which corresponds to

$\hat{h} = \{2\}$. Since $\hat{z}_1 > 0$,

$$\psi'(\hat{z}_1) + \beta D_1 Q_y(\hat{z}) = 1 + 2\beta(\hat{z}_1 - \hat{z}_2 + y_1 - y_2) = 0,$$
$$\beta D_2 Q_y(\hat{z}) = 2\beta(-\hat{z}_1 + \hat{z}_2 - y_1 + y_2) = 1.$$

Now again, (12) and (13) are satisfied. Notice that in both cases, (12) is non-strict.

Next we consider all $\hat{z} \in Z$ such that $0 < \hat{z}_1 < -(y_1 - y_2) - \frac{1}{2\beta}$, in which case $\hat{z}_1 > 0$ and $\hat{z}_2 < 0$. We have $\hat{h} = \{\emptyset\}$ and (12) is empty. We see that (13) is satisfied since

$$\psi'(\hat{z}_1) + \beta D_1 Q_y(\hat{z}) = 1 + 2\beta(\hat{z}_1 - \hat{z}_2 + y_1 - y_2) = 0,$$
$$\psi'(\hat{z}_2) + \beta D_2 Q_y(\hat{z}) = -1 + 2\beta(-\hat{z}_1 + \hat{z}_2 - y_1 + y_2) = 0.$$

It is not difficult to verify that (12) and (13) cannot be satisfied if $\hat{z} \notin Z$.

**Proof of Proposition 1:** Let us consider $\Delta(u)$ as given in (42) for any $u \in \mathrm{R}^p$.

- Let $j \in \{1, \ldots, r\}$ be such that $b_j^T u \neq 0$. By (i) we have $\varphi(b_j^T(\hat{z}+u+y)) - \varphi(b_j^T(\hat{z}+y)) > (b_j^T u)\varphi'(b_j^T(\hat{z}+y))$. Using (i) again,

$$Q_y(\hat{z}+u) - Q_y(\hat{z}) = \sum_{i=1}^{r} \left(\varphi\left(b_i^T(\hat{z}+u+y)\right) - \varphi\left(b_i^T(\hat{z}+y)\right)\right) > \sum_{i=1}^{r} \left(b_i^T u\right)\varphi'\left(b_i^T(\hat{z}+y)\right).$$

Inserting this result into (45)–(47) shows that $\Delta(u) > 0$.

- Let $u \in \ker B$. Using (ii), let $j \in \hat{h}^0$ be such that $u_j \neq 0$. Since (12) is strict for this $j$, (48) becomes $\psi_j(u_j) - \psi_j(0) + \beta u_j D_j Q_y(\hat{z}) > 0$. Hence $\Delta(u) \geq \psi_j(u_j) - \psi_j(0) + \beta u_j D_j Q_y(\hat{z}) > 0$.

**Proof of Theorem 2:** For any $i = 1, \ldots, p$, let $z_{[i]}^{(k)}$ denote the intermediate solution at step $i$ of iteration $k$,

$$z_{[i]}^{(k)} = \left(z_1^{(k)}, z_2^{(k)}, \ldots, z_{i-1}^{(k)}, z_i^{(k)}, z_{i+1}^{(k-1)}, \ldots, z_p^{(k-1)}\right). \tag{49}$$

For $i = 0$, put $z_{[0]}^{(k)} = z^{(k-1)}$. Notice that $z_{[p]}^{(k)} = z^{(k)}$. For every $k \in \mathbf{N}$, (19) shows that $F_y(z_{[i]}^{(k)}) \leq F_y(z_{[i-1]}^{(k)})$, for every $i = 1, \ldots, p$. Hence,

$$F_y\left(z^{(k)}\right) \leq F_y\left(z^{(k-1)}\right), \quad \forall k \in \mathbf{N}. \tag{50}$$

The sequence $F_y(z^{(k)})$ is monotonically decreasing and bounded below by $F_y(\hat{z})$, hence it converges. In particular, there is a radius $\delta > 0$ such that

$$z^{(k)} \in B(0, \delta), \quad \forall k \in \mathbf{N}. \tag{51}$$

For every $k \in \mathbf{N}$ we can write

$$F_y\left(z^{(k-1)}\right) - F_y\left(z^{(k)}\right) = \sum_{i=1}^{p} \left(F_y\left(z_{[i-1]}^{(k)}\right) - F_y\left(z_{[i]}^{(k)}\right)\right) \tag{52}$$
$$\geq \beta\eta \sum_{i=1}^{p} \left(z_i^{(k-1)} - z_i^{(k)}\right)^2, \tag{53}$$

where the inequality in (53) is obtained by applying Lemma 2 to every term on the right side of (52). It follows that the sequence $z^{(k)}$ is convergent. Put

$$\hat{z} = \lim_{k\to\infty} z^{(k)} \quad \text{and} \quad \hat{h} = \{i \in \{1, \ldots, q\} : \hat{z}_i = 0\}.$$

We will show that $\hat{z}$ satisfies the conditions for minimum given in Theorem 1.

- If $\hat{h}$ is nonempty, for every $i \in \hat{h}$, the convergence of $z_i^{(k)}$ to 0 can be produced in two different ways.
  - Consider that there is an integer $n_i$ such that $z_i^{(k)} = 0$, for all $k \geq n_i$. Using (21),

$$-\psi_i'(0^+) \leq \beta D_i Q_y\left(z_1^{(k)}, \ldots, z_{i-1}^{(k)}, 0, z_{i+1}^{(k-1)}, \ldots, z_p^{(k-1)}\right) \leq -\psi_i(0^-), \quad \forall k \geq n_i.$$

    Since $D_i Q_y$ is continuous, we get (12) when $k \to \infty$.
  - Otherwise, there is a subsequence, for simplicity denoted $z_i^{(k)}$, such that $z_i^{(k)} \neq 0$, for all $k \in \mathbf{N}$. Any such $z_i^{(k)}$ satisfies (22), so we have

$$\beta D_i Q_y\left(z_1^{(k)}, \ldots, z_{i-1}^{(k)}, z_i^{(k)}, z_{i+1}^{(k-1)}, \ldots, z_p^{(k-1)}\right) + \psi_i\left(z_i^{(k)}\right) = 0, \quad \forall k \in \mathbf{N}. \tag{54}$$

    We will show that there is an integer $n_i$ and a constant $\sigma_i \in \{-1, 1\}$ such that

$$k \geq n_i \;\Rightarrow\; \operatorname{sign}\left(z_i^{(k)}\right) = \sigma_i. \tag{55}$$

Suppose the contrary: for every $k$ there is $j_k > k$ so that $\mathrm{sign}(z_i^{(j_k)}) = -\mathrm{sign}(z_i^{(k)})$. Then there is a subsequence, denoted $z_i^{(k)}$ again, such that $\mathrm{sign}(z_i^{(k)}) = (-1)^k$, for all $k \in \mathbf{N}$. Using (22) and H1,

$$k \quad \text{odd} \Rightarrow$$
$$\beta D_i Q_y\big(z_1^{(k)}, \ldots, z_{i-1}^{(k)},\ z_i^{(k)},\ z_{i+1}^{(k-1)}, \ldots, z_p^{(k-1)}\big) = -\psi_i'\big(z_i^{(k)}\big) \geq -\psi_i'(0^-),$$
$$k \quad \text{even} \Rightarrow$$
$$\beta D_i Q_y\big(z_1^{(k)}, \ldots, z_{i-1}^{(k)},\ z_i^{(k)},\ z_{i+1}^{(k-1)}, \ldots, z_p^{(k-1)}\big) = -\psi_i'\big(z_i^{(k)}\big) \leq -\psi_i'(0^+).$$

This result contradicts the fact that since $Q_y$ is $\mathcal{C}^1$,

$$\lim_{k\to\infty} D_i Q_y\big(z_1^{(k)}, \ldots, z_{i-1}^{(k)},\ z_i^{(k)},\ z_{i+1}^{(k-1)}, \ldots, z_p^{(k-1)}\big) = D_i Q_y(\hat{z}).$$

If $\sigma_i = 1$, we have $z_i^{(k)} > 0$, for all $k \geq n_i$. For $k \to \infty$, we have $z_i^{(k)} \searrow 0$ and thus $\psi_i'(z_i^{(k)}) \searrow \psi_i'(0^+)$. Combining this with (54) shows that for $k \to \infty$ we get $\psi_i'(0^+) + \beta D_i Q_y(\hat{z}) = 0$. If $\sigma_i = -1$, in a similar way we obtain $\psi_i'(0^-) + \beta D_i Q_y(\hat{z}) = 0$. In both cases, (12) is satisfied.

- If $\hat{h}^c$ is nonempty, (22) shows that for all $k \in \mathbf{N}$,

$$\beta D_i Q_y\big(z_1^{(k)}, \ldots, z_{i-1}^{(k)}, z_i^{(k)}, z_{i+1}^{(k-1)}, \ldots, z_p^{(k-1)}\big) + \psi_i'\big(z_i^{(k)}\big) = 0, \quad \forall i \in \hat{h}^c.$$

Put

$$\zeta := \frac{1}{2} \min_{i \in \hat{h}^c} |\hat{z}_i|.$$

There is $n \in \mathbf{N}$ such that $z^{(k)} \in B(\hat{z}, \zeta)$, for all $k \geq n$. Since for every $i \in \hat{h}^c$, the function $z \to \psi_i'(z_i) + \beta D_i Q(z)$ is continuous on $B(\hat{z}, \zeta)$, at the limit $k \to \infty$ we find (13).

- If $q < p$, for all $k \in \mathbf{N}$ we have

$$D_i Q_y\big(z_1^{(k)}, \ldots, z_{i-1}^{(k)},\ z_i^{(k)},\ z_{i+1}^{(k-1)}, \ldots, z_p^{(k-1)}\big) = 0, \quad \forall i \in \{q+1, \ldots, p\}.$$

When $k \to \infty$, we find (14).

**Proof of Lemma 2.2:** Let $\delta > 0$ be the radius mentioned in (51). Then for all $k \in \mathbf{N}$ and $i = 1, \ldots, p$,

$$z_{[i]}^{(k)} \in \bar{B}(0, \delta) \quad \text{and } \big|z_i^{(k-1)} - z_i^{(k)}\big| \leq 2\delta.$$

Let $\eta > 0$ be the constant mentioned in H2 relevant to $\rho = 2\delta$. Noticing that by (49)

$$z_{[i-1]}^{(k)} = z_{[i]}^{(k)} + \big(z_i^{(k-1)} - z_i^{(k)}\big)e_i,$$

the inequality in (9) shows that for all $k \in \mathbf{N}$ and $i \in \{1, \ldots, p\}$,

$$Q_y\big(z_{[i-1]}^{(k)}\big) - Q_y\big(z_{[i]}^{(k)}\big) \geq \big(z_i^{(k-1)} - z_i^{(k)}\big)\, D_i Q_y\big(z_{[i]}^{(k)}\big) + \eta\big(z_i^{(k-1)} - z_i^{(k)}\big)^2. \quad (56)$$

If $q < p$ and $i \in \{q+1, \ldots, p\}$,

$$F_y\big(z_{[i-1]}^{(k)}\big) - F_y\big(z_{[i]}^{(k)}\big) = \beta\big(Q_y\big(z_{[i-1]}^{(k)}\big) - Q_y\big(z_{[i]}^{(k)}\big)\big).$$

In this case, $D_i Q_y(z_{[i]}^{(k)}) = 0$. Introducing this and (56) in the expression above shows the statement.

Using (56), for all $k \in \mathbf{N}$ and $i \in \{1, \ldots, q\}$,

$$F_y\big(z_{[i-1]}^{(k)}\big) - F_y\big(z_{[i]}^{(k)}\big) = \psi_i\big(z_i^{(k-1)}\big) - \psi_i\big(z_i^{(k)}\big) + \beta\big(Q_y\big(z_{[i-1]}^{(k)}\big) - Q_y\big(z_{[i]}^{(k)}\big)\big) \geq \Delta + \beta\eta\big(z_i^{(k-1)} - z_i^{(k)}\big)^2, \quad (57)$$

where

$$\Delta = \psi_i\big(z_i^{(k-1)}\big) - \psi_i\big(z_i^{(k)}\big) + \beta\big(z_i^{(k-1)} - z_i^{(k)}\big) D_i Q_y\big(z_{[i]}^{(k)}\big). \quad (58)$$

Observe that for any $i \in \{1, \ldots, q\}$,

$$t \neq 0 \text{ and } \tau\ \mathrm{sign}(t) \geq 0 \Rightarrow \psi_i(\tau) - \psi_i(t) \geq \psi_i'(t)\,(\tau - t). \quad (59)$$

Indeed, the inequality is evident if $\tau\ \mathrm{sign}(t) > 0$ and it remains true when $\tau \to 0$.

Several cases arise according to the position of $z_i^{(k)}$ and $z_i^{(k-1)}$ with respect to zero.

- $z_i^{(k)} = 0$ and $z_i^{(k-1)} \geq 0$. Using Lemma 1, $\psi_i(z_i^{(k-1)}) - \psi_i(z_i^{(k)}) \geq \psi_i'(0^+)(z_i^{(k-1)} - z_i^{(k)})$. Introducing this into (58) yields

$$\Delta \geq \big(z_i^{(k-1)} - z_i^{(k)}\big)\big(\psi_i'(0^+) + \beta D_i Q_y\big(z_{[i]}^{(k)}\big)\big).$$

Using (21), the term among the large parentheses is non-negative, hence $\Delta \geq 0$.

- $z_i^{(k)} = 0$ and $z_i^{(k-1)} \leq 0$. We just replace $\psi_i'(0^+)$ by $\psi_i'(0^-)$ in the expressions above and find that $\Delta \geq 0$.

- $z_i^{(k)} \neq 0$ and $\text{sign}(z_i^{(k-1)}\ z_i^{(k)}) \geq 0$. Using (59), $\psi_i(z_i^{(k-1)}) - \psi_i(z_i^{(k)}) \geq \psi_i'(z_i^{(k)})(z_i^{(k-1)} - z_i^{(k)})$. Then

$$\Delta \geq \big(z_i^{(k-1)} - z_i^{(k)}\big)\big(\psi_i'\big(z_i^{(k)}\big) + \beta D_i Q_y\big(z_{[i]}^{(k)}\big)\big).$$

Using (22) we see that that $\psi_i'(z_i^{(k)}) + \beta D_i Q_y(z_{[i]}^{(k)}) = 0$, hence $\Delta = 0$.

- $z_i^{(k)} < 0$ and $z_i^{(k-1)} > 0$. We consider $\psi_i$ on the intervals $(z_i^{(k)}, 0)$ and on $(0, z_i^{(k-1)})$ separately. Applying Lemma 1 to $\psi_i(z_i^{(k-1)}) - \psi_i(0)$, and (59) to $\psi_i(0) - \psi_i(z_i^{(k)})$, yields

$$\begin{aligned}
&\psi_i\big(z_i^{(k-1)}\big) - \psi_i\big(z_i^{(k)}\big)\\
&\quad = \big(\psi_i\big(z_i^{(k-1)}\big) - \psi_i(0)\big) + \big(\psi_i(0) - \psi_i\big(z_i^{(k)}\big)\big)\\
&\quad \geq \psi_i'(0^+) z_i^{(k-1)} - \psi_i'\big(z_i^{(k)}\big) z_i^{(k)}\\
&\quad = \psi_i'\big(z_i^{(k)}\big)\big(z_i^{(k-1)} - z_i^{(k)}\big)\\
&\quad\quad + \big(\psi_i'(0^+) - \psi_i'\big(z_i^{(k)}\big)\big) z_i^{(k-1)}.
\end{aligned}$$

Since $z_i^{(k)} < 0$ we have $\psi_i'(z_i^{(k)}) < 0$; then $(\psi_i'(0^+) - \psi_i'(z_i^{(k)}))z_i^{(k-1)} > 0$. Hence,

$$\psi_i\big(z_i^{(k-1)}\big) - \psi_i\big(z_i^{(k)}\big) > \psi_i'\big(z_i^{(k)}\big)\big(z_i^{(k-1)} - z_i^{(k)}\big).$$

Inserting the last inequality into (58) shows that

$$\Delta > \big(z_i^{(k-1)} - z_i^{(k)}\big)\big(\psi_i'\big(z_i^{(k)}\big) + \beta D_i Q_y\big(z_{[i]}^{(k)}\big)\big). \quad (60)$$

According to (22), the term between the large parentheses is null, hence $\Delta > 0$.

- $z_i^{(k)} > 0$ and $z_i^{(k-1)} < 0$. Similarly, we consider $(z_i^{(k-1)}, 0)$ and $(0, z_i^{(k)})$ separately. We only have to replace $\psi'(0^+)$ with $\psi'(0^-)$ in the expressions above and find that $\Delta > 0$.

In all cases, $\Delta \geq 0$. Introducing this into (57) shows the lemma.

**Proof of Lemma 3:** Consider an arbitrary $i \in \hat{h}^c$. Since $y_i$ does not satisfy (27), we have $\sigma_i \sum_{j\in\mathcal{N}_i} \varphi'(y_i - \hat{x}_j) > \frac{1}{\beta} > 0$. Since $\varphi'$ is increasing on R,

$$\sigma_i \gamma_i \geq \sigma_i y_i \Rightarrow \sigma_i \sum_{j\in\mathcal{N}_i} \varphi'(\gamma_i - \hat{x}_j) > \frac{1}{\beta} > 0.$$

Hence $\text{sign}(\sum_{j\in\mathcal{N}_i} \varphi'(\gamma_i - \hat{x}_j)) = \sigma_i$, for all $i \in \hat{h}^c$. Combining this with the fact that $\gamma_i = y_i$, for all $i \in \hat{h}$, shows that $\hat{x}$ satisfies both conditions (27) and (28) with respect to $\mathcal{F}_\gamma$.

**Proof of Lemma 4:** To prove (a), we show that for all $\gamma \in \mathrm{R}^p$, the function $x_\omega \to (D_i f_\gamma(x_\omega) : i \in \omega)$ is strictly monotone [18]. For arbitrary $x_\omega$ and $v_\omega \neq 0$, consider $\Delta$ as given below

$$\begin{aligned}
\Delta &= \sum_{i\in\omega} (D_i f_\gamma(x_\omega + v_\omega) - D_i f_\gamma(x_\omega)) v_i\\
&= \sum_{i\in\omega} v_i \Bigg( \sum_{j\in\mathcal{N}_i\cap\hat{h}} (\varphi'(x_i + v_i - \gamma_j) - \varphi'(x_i - \gamma_j))\\
&\quad + \sum_{j\in\mathcal{N}_i\cap\omega} (\varphi'(x_i - x_j + v_i - v_j) - \varphi'(x_i - x_j)) \Bigg)\\
&= \sum_{i\in\omega} \Bigg( \sum_{j\in\mathcal{N}_i\cap\hat{h}} v_i(\varphi'(x_i + v_i - \gamma_j) - \varphi'(x_i - \gamma_j))\\
&\quad + \frac{1}{2} \sum_{j\in\mathcal{N}_i\cap\omega} (v_i - v_j)(\varphi'(x_i - x_j + v_i - v_j)\\
&\quad - \varphi'(x_i - x_j)) \Bigg).
\end{aligned}$$

Since $\varphi$ is strictly convex,

$$(\varphi'(t + \varepsilon) - \varphi'(t))\, \varepsilon > 0, \quad \forall t \in \mathrm{R}, \ \forall \varepsilon \neq 0. \quad (61)$$

Then $\Delta \geq 0$ since all the terms in the last expression for $\Delta$ are non-negative. If $\#\omega = 1$, let $\omega = \{i\}$. Then $\mathcal{N}_i \subset \hat{h}$ and, using (61),

$$\Delta = \sum_{j\in\mathcal{N}_i} v_i(\varphi'(x_i + v_i - \gamma_j) - \varphi'(x_i - \gamma_j)) > 0,$$

since $v_i \neq 0$. Consider now that $\#\omega > 1$. If there are $i \in \omega$ and $j \in \omega$, such that $v_i \neq v_j$,

$$\begin{aligned}
\Delta \geq \frac{1}{2}(v_i - v_j)(&\varphi'(x_i - x_j + v_i - v_j)\\
&- \varphi'(x_i - x_j)) > 0.
\end{aligned}$$

Otherwise, there is $c \neq 0$ such that $v_i = c$, for all $i \in \omega$. Since $\hat{h} \neq \emptyset$, there is $i \in \omega$ such that $\mathcal{N}_i \cap \hat{h} \neq \emptyset$.

Then

$$\Delta \geq \sum_{j\in\mathcal{N}_i\cap\hat{h}} c(\varphi'(x_i + c - \gamma_j) - \varphi'(x_i - \gamma_j)) > 0.$$

Hence (a).

Combining (a) with (33) shows that $f_y$ reaches its minimum at $\hat{x}_\omega$ and that the latter is strict. For an arbitrary $\delta > 0$, define

$$\mu = \min\{f_y(x_\omega) - f_y(\hat{x}_\omega) : x_\omega \in S(\hat{x}_\omega, \delta)\}, \quad \text{where} \\ S(\hat{x}_\omega, \delta) = \{x_\omega : \|x_\omega - \hat{x}_\omega\| = \delta\}. \tag{62}$$

Notice that $(x_\omega, \gamma) \to f_\gamma(x_\omega)$ is $\mathcal{C}^1$. Since $S(\hat{x}_\omega, \delta)$ is compact, $\mu > 0$. Moreover, there is $\xi > 0$ such that

$$\gamma \in B(y, \xi) \Rightarrow |f_\gamma(x_\omega) - f_y(x_\omega)| \leq \frac{\mu}{2}, \\ \forall x_\omega \in \bar{B}(\hat{x}_\omega, \delta). \tag{63}$$

Then we find that if $\gamma \in B(y, \xi)$,

$$x_\omega \in S(\hat{x}_\omega, \delta) \Rightarrow f_\gamma(x_\omega) - f_\gamma(\hat{x}_\omega) = (f_\gamma(x_\omega) \\ - f_y(x_\omega)) + (f_y(x_\omega) - f_y(\hat{x}_\omega)) \\ \geq -\frac{\mu}{2} + \mu,$$

where the inequality comes from (63) and (62). Hence, for every $\gamma \in B(y, \xi)$, the function $f_\gamma$ reaches its minimum at a point $\hat{\chi}_\omega$ which lies in the interior of $B(\hat{x}_\omega, \delta)$; this minimum is strict according to (a). Using [7], there is a unique, continuous minimizer function $\mathcal{X}_\omega : B(y, \xi) \to \mathrm{R}^{\#\omega}$, such that for every $\gamma \in B(y, \xi)$, the function $f_\gamma$ reaches its minimum at $\mathcal{X}_\omega(\gamma)$, and such that $\hat{x}_\omega = \mathcal{X}_\omega(y)$. Hence,

$$\gamma \in B(y, \xi) \Rightarrow D_i f_\gamma(\mathcal{X}_\omega(\gamma)) = 0, \quad \forall i \in \omega.$$

**Proof of Theorem 3:** For every $i \in \{1, \ldots, p\}$, let $\Phi_i : \mathrm{R}^p \times \mathrm{R}^p \to \mathrm{R}$ be the function

$$\Phi_i(x, \gamma) = \sum_{j\in\mathcal{N}_i\cap\hat{h}} \varphi'(\gamma_i - \gamma_j) + \sum_{j\in\mathcal{N}_i\cap\hat{h}^c} \varphi'(\gamma_i - x_j).$$

Consider first that $\hat{h} = \{1, \ldots, p\}$. Since $\hat{h}^c = \emptyset$, for every $i = 1, \ldots, p$, we have

$$\Phi_i(x, \gamma) = \Phi_i(0, \gamma) = \sum_{j\in\mathcal{N}_i} \varphi'(\gamma_i - \gamma_j)$$

and $|\Phi_i(0, \gamma)| < 1/\beta$ because (27) is strict by assumption. Put

$$\mu_1 = \max_{i=1}^{p} |\Phi_i(0, y)|,$$

then $\mu_1 < \dfrac{1}{\beta}$. Since $\{\Phi_i\}_{i=1}^p$ are continuous functions, there is $\rho > 0$ such that

$$\gamma \in B(y, \rho) \Rightarrow |\Phi_i(0, \gamma) - \Phi_i(0, y)| \\ < \frac{1}{\beta} - \mu_1, \quad \forall i \in \{1, \ldots, p\}.$$

Then for every $i \in \{1, \ldots, p\}$,

$$\gamma \in B(y, \rho) \Rightarrow |\Phi_i(0, \gamma)| \leq |\Phi_i(0, y)| \\ + |\Phi_i(0, \gamma) - \Phi_i(0, y)| < \mu_1 + \frac{1}{\beta} - \mu_1.$$

Equivalently, for every $\gamma \in B(y, \rho)$, (27) is satisfied by $\hat{\chi} = \gamma$. Hence $\mathcal{F}_\gamma$ reaches its minimum at $\hat{\chi} = \gamma$.

Consider now that $\hat{h}^c$ is nonempty. Let the connected components of $\hat{h}^c$ read $\omega_k$ for $k = 1, \ldots, m$. For every $k \in \{1, \ldots, m\}$, let $\xi_k > 0$ and $\mathcal{X}_{\omega_k} : B(y, \xi_k) \to \mathrm{R}^{\#\omega_k}$ be as in Lemma 4. Put $\rho_1 = \min_{k=1}^m \xi_k$. For every $i \in \hat{h}$, put $\mathcal{X}_i(\gamma) = \gamma_i$. The resultant $\mathcal{X}$ is a continuous function from $B(y, \rho_1)$ to $\mathrm{R}^p$ which satisfies $\mathcal{X}(y) = \hat{x}$ and

$$\sum_{j\in\mathcal{N}_i\cap\hat{h}} \varphi'(\mathcal{X}_i(\gamma) - \gamma_j) + \sum_{j\in\mathcal{N}_i\cap\hat{h}^c} \varphi'(\mathcal{X}_i(\gamma) - \mathcal{X}_j(\gamma)) \\ = \frac{\sigma_i}{\beta}, \forall i \in \hat{h}^c. \tag{64}$$

Define the constants $\mu_1$ and $\mu_2$ by

$$\mu_1 = \max_{i\in\hat{h}} |\Phi_i(\hat{x}, y)|, \quad \mu_2 = \min_{i\in\hat{h}^c} |\Phi_i(\hat{x}, y)|. \tag{65}$$

The inequality (27) is strict for every $i \in \hat{h}$, so $\mu_1 < 1/\beta$, and it is false for every $i \in \hat{h}^c$, so $\mu_2 > 1/\beta$. By the continuity of $\mathcal{X}$ and $\{\Phi_i\}_{i=1}^p$, there is $\rho_2 \in (0, \rho_1]$ such that for every $\gamma \in B(y, \rho_2)$, we have

$$|\Phi_i(\mathcal{X}(\gamma), \gamma) - \Phi_i(\hat{x}, y)| < \frac{1}{\beta} - \mu_1, \quad \forall i \in \hat{h},$$

$$|\Phi_i(\mathcal{X}(\gamma), \gamma) - \Phi_i(\hat{x}, y)| < \mu_2 - \frac{1}{\beta}, \quad \forall i \in \hat{h}^c.$$

For every $i \in \hat{h}$ we have $\mathcal{X}_i(\gamma) = \gamma_i$ and

$$\gamma \in B(y, \rho_2) \Rightarrow \left| \sum_{j \in \mathcal{N}_i} \varphi'(\gamma_i - \mathcal{X}_j(\gamma)) \right| = |\Phi_i(\mathcal{X}(\gamma),\ \gamma)| \leq |\Phi_i(\mathcal{X}(\gamma),\ \gamma) - \Phi_i(\hat{x},\ y)| + |\Phi_i(\hat{x},\ y)| < \frac{1}{\beta},$$

where the last inequality comes from (66) and (65). Hence, $\mathcal{X}(\gamma)$ satisfies (27) for every $i \in \hat{h}$. Furthermore, for every $i \in \hat{h}^c$ we have

$$\gamma \in B(y, \rho_2) \Rightarrow |\Phi_i(\mathcal{X}(\gamma),\ \gamma)| \geq |\Phi_i(\hat{x},\ y)| - |\Phi_i(\mathcal{X}(\gamma),\ \gamma) - \Phi_i(\hat{x},\ y)| > \frac{1}{\beta},$$

where the last inequality is due to (66) and (65). By the continuity of $\{\Phi_i\}_{i=1}^{p}$ and $\mathcal{X}$, we deduce that

$$\gamma \in B(y, \rho_2) \Rightarrow \text{sign}(\Phi_i(\mathcal{X}(\gamma),\ \gamma)) = \sigma_i, \quad \forall\, i \in \hat{h}^c. \tag{66}$$

Combining this with (64) shows that $\mathcal{X}(\gamma)$ satisfies (28), for every $i \in \hat{h}^c$. By Corollary 1, $\mathcal{F}_\gamma$ reaches its minimum at $\hat{\chi} = \mathcal{X}(\gamma)$ and the latter satisfies (35). Hence $B(y, \rho_2) \in Y_{\hat{h}}$.

If $\rho > 0$ is such that $|\gamma_i - y_i| \leq \rho$, for all $i \in \{1, \ldots, p\}$, then $\|\gamma - y\| < \rho_2$. Let $\gamma \in Y_{\hat{h}}$ satisfy

$$\begin{aligned} |\gamma_i - y_i| &\leq \rho &&\text{if } i \in \hat{h}, \\ \gamma_i &= y_i - \sigma_i \rho &&\text{if } i \in \hat{h}^c. \end{aligned}$$

Then $\mathcal{F}_\gamma$ reaches its minimum at an $\hat{\chi} = \mathcal{X}(\gamma)$ which satisfies (35). By Lemma 3.2, for any $y'$ such that $y'_i = \gamma_i$ if $i \in \hat{h}$, and $\sigma_i y'_i \geq \sigma_i y_i - \rho$ if $i \in \hat{h}^c$, the relevant $\mathcal{F}_{y'}$ reaches its minimum at the same $\hat{\chi}$. It follows that the equality on the right side of (66) and (35) are satisfied for all $\gamma \in Y_{\hat{h}}$.

## Notes

1. Since $\psi_i$ is convex, $\psi_i(0^+) = \lim_{\varepsilon \searrow 0}(\psi_i(\varepsilon) - \psi_i(0))/\varepsilon$ and $\psi_i(0^-) = \lim_{\varepsilon \searrow 0}(\psi_i(-\varepsilon) - \psi_i(0))/(-\varepsilon)$ are well defined and finite [18]. We write $\varepsilon \searrow 0$ to specify that $\varepsilon$ goes to zero by positive values.
2. If $\varphi$ is $\mathcal{C}^2$ and $\varphi''(t) > 0$, for all $t \in \mathrm{R}$, put $\eta_\delta = (1/2)\min_{|t| \leq \delta} \varphi''(t)$, then $\eta_\delta > 0$. Using a Taylor expansion, $\varphi(t) - \varphi(\tau) = \varphi'(\tau)(t - \tau) + (t-\tau)^2 \int_0^1 (1-s)\varphi''(\tau + s(t-\tau))ds \geq \varphi'(\tau)(t-\tau) + \eta_\delta(t-\tau)^2$, for all $t \in [-\delta, \delta]$ and $\tau \in [-\delta, \delta]$.
3. $\omega$ is a singleton, say $\omega = \{i\}$, if $\mathcal{N}_i \subset \hat{h}$. Otherwise, for every $i, j \in \omega$ there are $i_j \in \omega$, for $j = 1, \ldots, n$, such that $i_1 \in \mathcal{N}_i, i_2 \in \mathcal{N}_{i_1}, \ldots, j \in \mathcal{N}_{i_n}$, and for every $i \in \omega$; moreover, if $\mathcal{N}_i \cap \hat{h}^c$ is nonempty, then $\mathcal{N}_i \cap \hat{h}^c \subset \omega$.
4. Let $\omega = \{i_1, \ldots, i_n\}$, then $x_\omega = \{x_{i_1}, \ldots, x_{i_n}\}$. To avoid ambiguity, let us precise that $D_i f_y(x_\omega) = \frac{\partial f_y}{\partial x_{i_j}}(\hat{x}_\omega)$ if $i = i_j$.
5. CWM filter is defined using a window $\mathcal{N}_i \cup \{i\}$ and a replication parameter $\alpha \in \mathbf{N}$. The output at $i$ is the median of the set of all $y_j$, for $j \in \mathcal{N}_i$, and the current entry $y_i$ replicated $\alpha$ times. Thus, $\alpha$ is the weight of the central entry.
6. PWM filter is defined using a window $\mathcal{N}_i \cup \{i\}$ and a rank threshold $\alpha$. The output at $i$ is $y_i$ if the rank of $y_i$ among all entries in the window is rated between $\alpha$ and $\#\mathcal{N}_i - \alpha$; otherwise, it is the median of all $y_j$ for $j \in \mathcal{N}_i \cup \{i\}$.

## References


1. E. Abreu, M. Lightstone, S.K. Mitra, and K. Arakawa, "A new efficient approach for the removal of impulse noise from highly corrupted images," *IEEE Transactions on Image Processing*, Vol. 5, pp. 1012–1025, 1996.
2. S. Alliney, "Digital filters as absolute norm regularizers," *IEEE Transactions on Medical Imaging*, Vol. MI-12, pp. 173–181, 1993.
3. S. Alliney, "A property of the minimum vectors of a regularizing functional defined by means of absolute norm," *IEEE Transactions on Signal Processing*, Vol. 45, pp. 913–917, 1997.
4. G.R. Arce, T.A. Hall, and K.E. Barner, "Permutation weighted order statistic filters," *IEEE Transactions on Image Processing*, Vol. 4, pp. 1070–1083, 1995.
5. J.E. Besag, "Digital image processing: Towards Bayesian image analysis," *Journal of Applied Statistics*, Vol. 16, pp. 395–407, 1989.
6. M. Black and A. Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications to early vision," *International Journal of Computer Vision*, Vol. 19, pp. 57–91, 1996.
7. C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving map estimation," *IEEE Transactions on Image Processing*, Vol. 2, pp. 296–310, 1993.
8. C. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Transactions on Image Processing*, Vol. 5, pp. 480–492, 1996.
9. A.C. Bovik, *Handbook of Image and Video Processing*, Academic Press: New York, 2000.
10. P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Transactions on Image Processing*, Vol. 6, pp. 298–311, 1997.
11. T. Chen and H.R. Wu, "Adaptive impulse detection using center-weighted median filters," *IEEE Transactions on Image Processing*, Vol. 8, pp. 1–3, 2001.
12. G. Demoment, "Image reconstruction and restoration: Overview of common estimation structure and problems," *IEEE Transactions on Acoustics Speech and Signal Processing*, Vol. ASSP-37, pp. 2024–2036, 1989.

13. M. Fisher and J. Paredes, "Weighted median image sharpeners for world wide web," *IEEE Transactions on Image Processing*, Vol. 11, pp. 717–727, 2002.
14. A. Flaig, G. R. Arce, and K. Barner, "Affine order statistics filters: Medianization of linear FIR filters," *IEEE Transactions on Signal Processing*, 1998.
15. D. Geman, "Random fields and inverse problems in imaging," Vol. 1427, École d'Été de Probabilités de Saint-Flour XVIII—1988, Springer-Verlag, Lecture Notes in Mathematics ed., 1990, pp. 117–193.
16. D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Transactions on Image Processing*, Vol. IP-4, pp. 932–946, 1995.
17. R. Glowinski, J. Lions, and R. Trémolières, *Analyse numérique des inéquations variationnelles*, Vol. 1, 1st edition Dunod: Paris, 1976.
18. J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, Vols. I and II. Springer-Verlag: Berlin, 1996.
19. D. Keren and M. Werman, "Probabilistic analysis of regularization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-15, pp. 982–995, 1993.
20. S.J. Ko and Y.H. Lee, "Center weighted median filters and their applications to image enhancement," *IEEE Transactions on Circuits and Systems*, Vol. 38, pp. 984–993, 1991.
21. Y.-H. Lee and S. Tantaratana, "Decision-based order statistic filters," *IEEE Transactions on Acoustics Speech and Signal Processing*, Vol. 38, pp. 406–420, 1990.
22. S. Li, "Markov random field modeling in computer vision," 1st edition. Springer-Verlag: New York, 1995.
23. M. Nikolova, "Markovian reconstruction using a GNC approach," *IEEE Transactions on Image Processing*, Vol. 8, pp. 1204–1220, 1999.
24. M. Nikolova, "Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers," *SIAM Journal of Numerical Analysis*, Vol. 40, pp. 965–994, 2001.
25. G. Pok and A.S. Nair, "Selective removal of impulse noise based on homogeneity level information," *IEEE Transactions on Image Processing*, Vol. 12, pp. 85–92, 2003.
26. L. Rudin, S. Osher, and C. Fatemi, "Nonlinear total variation based noise removal algorithm," *Physica*, Vol. 60 D, pp. 259–268, 1992.
27. J.-L. Starck, E.P., and F. Murtagh, "Deconvolution in astronomy: A review," Publications of the Astronomical Society of the Pacific, 114 2002.
28. A. Tarantola, *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier Science Publishers: Amsterdam, 1987.
29. A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems*, Winston: Washington DC, 1977.
30. J.W. Tukey, "Nonlinear (nonsuperimposable) methods for smoothing data," in *Conf. Rec.*, Eacson, 1974.
31. C.R. Vogel and M.E. Oman, "Iterative method for total variation denoising," *SIAM Journal of Scientific Computing*, Vol. 17, pp. 227–238, 1996.
32. C.R. Vogel and M.E. Oman, "Fast, robust total variation-based reconstruction of noisy, blurred images," *IEEE Transactions on Image Processing*, Vol. 7, pp. 813–824, 1998.
33. Z. Wang and D. Zhang, "Restoration of impulse noise corrupted images using long-range correlation," *IEEE Signal Processing Letters*, Vol. 5, pp. 4–7, 1998.
34. Windyga, "Fast impulsive noise removal," *IEEE Transactions on Image Processing*, Vol. 10, pp. 173–179, 2001.
35. L. Yin, R. Yang, M. Gabbouj, and Y. Neuvo, "Weighted median filters: A tutorial," *IEEE Transactions on Circuit Theory*, Vol. 41, pp. 157–192, 1996.

**Mila Nikolova** is Scientific researcher with the CNRS (National Center for Scientific Research). In 2003, she joined the CMLA (Center for Mathematics and their Applications), ENS de Cachan. Her research interests include inverse problems and reconstruction of images and signals.