# Denoising of frame coefficients using $\ell^1$ data-fidelity term and edge preserving regularization

Sylvain Durand* and Mila Nikolova†

* LAMFA UMR 6140–Université de Picardie 33 rue Saint-Leu, 80039 Amiens Cedex

* † CMLA UMR 8536–ENS de Cachan, 61 av. du Président Wilson, 94235 Cachan Cedex

e-mail: {sdurand,nikolova}@cmla.ens-cachan.fr

**Abstract**

We consider the denoising of a function (an image or a signal) containing smooth regions and edges. Classical ways to solve this problem are variational methods and shrinkage of a representation of the data in a basis or a frame. We propose a method which combines the advantages of both approaches. Following the wavelets shrinkage method of Donoho and Johnstone, we set to zero all frame coefficients with respect to a reasonable threshold. The shrunk frame representation involves both large coefficients corresponding to noise (outliers) and some coefficients, erroneously set to zero, leading to Gibbs-like oscillations in the estimate. We design a specialized (non-smooth) objective function allowing all these coefficients to be selectively restored, without modifying the other coefficients which are nearly faithful, using regularization in the domain of the restored function. We analyze the well-posedness and the main properties of this objective function. We also propose an approximation of this method which is accurate enough and very fast. We present numerical experiments with signals and images corrupted with white Gaussian noise, which are decomposed into a wavelets basis. The obtained results demonstrate the advantages of our approach over the main alternative methods.

## 1 Introduction

We consider the restoration of an original (unknown) function $u_o(s)$ defined on a (possibly finite) domain $\Omega$—an image or a signal containing smooth zones and edges—from noisy data

$$v = u_o + n,$$

where $n$ represents a perturbation. Restoration has to recover the features of $u_o$, lost because of the noise, according to prior smoothness constraints. In the literature, quite different approaches have been developed in order to deal with this classical but yet unsolved problem. We will discuss only variational methods and shrinkage estimators since they underly the method we propose in this paper. In variational methods, the restored function is defined as the minimizer of an objective function $\mathcal{F}_v$ which balances trade-off between closeness to data and smoothness constraints,

$$\mathcal{F}_v(u) = \mu \int_\Omega |u(s) - v(s)|^2 ds + \int_\Omega \varphi(|\nabla u(s)|)\, ds, \tag{1}$$

where $\nabla$ stands for gradient, $\varphi : \mathrm{R}_+ \to \mathrm{R}_+$ is called a potential function and $\mu > 0$ is a parameter. In their pioneering work, Tikhonov and Arsenin [46] considered $\varphi(t) = t^2$. However, this choice for $\varphi$ leads to smooth images with flattened edges. Under the usual assumptions that $\Omega$ is discrete and the noise $n$ is white and Gaussian, Bayesian maximum a posteriori estimators amount to minimize an objective function of the same form as (1), see e.g. [7, 21, 30]. Modeling $u$ as a Markov random field gave rise to many

different convex and nonconvex functions $\varphi$ [27, 8, 11, 31]. Even if nonconvex potential functions can yield minimizers involving sharp edges and smooth regions, there are no general methods to approximate a global minimizer of $\mathcal{F}_v$. Instead, there was an increasing interest to determine convex functions $\varphi$ which allow edges to be restored. Based on a fine analysis of the minimizers of $\mathcal{F}_v$ as solutions of PDE's on a continuous domain $\Omega$, Rudin, Osher and Fatemi [42] exhibited that $\varphi(t) = |t|$ leads to images involving edges. Their method is at the origin of a large amount of papers dedicated to edge-preserving convex potential functions, see e.g. [1, 16, 48]. A recent overview of variational methods can be found in [5]. A systematic default of the images and signals restored using edge-preserving convex functions $\varphi$ is that the amplitude of edges is underestimated. This is particularly annoying if the sought-after function has spiky areas since the later are subjected to erosion; see for instance Fig. 4 in Section 7.

Shrinkage estimators operate on a decomposition of data $v$ into a frame of $L^2(\Omega)$, say $\{w_i : i \in J\}$ where $J$ is a set of indexes. Let $W$ be the corresponding frame operator, i.e. $(Wv)[i] = \langle v, w_i \rangle$, $\forall i \in J$, and $\widetilde{W}$ be a left inverse of $W$, giving rise to the dual frame $\{\widetilde{w}_i : i \in J\}$. The frame coefficients of $v$ read $y = Wv$ and are contaminated with noise $Wn$. The idea is to denoise them by shrinkage using a symmetric function $\tau : \mathrm{R} \to \mathrm{R}$ satisfying $0 \leq \tau(t) \leq t$ for all $t \geq 0$ and to generate a denoised function, denoted $v_\tau$, according to

$$v_\tau = \sum_{i \in J} \tau((Wv)[i])\, \widetilde{w}_i = \sum_{i \in J} \tau(y[i])\, \widetilde{w}_i. \tag{2}$$

Since the inaugural work of Donoho and Johnstone in [22], shrinkage estimators are a popular and fast tool to denoise images and signals. The latter paper addresses orthogonal wavelets transforms for $W$ and discrete domains $\Omega$ of finite cardinality $\#\Omega$, and considers two different choices for $\tau$: given $T > 0$, hard thresholding corresponds to

$$\tau(t) = \begin{cases} 0 & \text{if } |t| \leq T, \\ t & \text{otherwise,} \end{cases} \tag{3}$$

while soft-thresholding corresponds to

$$\tau(t) = \begin{cases} 0 & \text{if } |t| \leq T, \\ t - T\mathrm{sign}(t) & \text{otherwise.} \end{cases} \tag{4}$$

Both soft and hard thresholding are asymptotically optimal in the minimax sense if $n$ is white Gaussian noise of standard deviation $\sigma$ and

$$T = \sigma \sqrt{2 \log_e \#\Omega}. \tag{5}$$

This threshold is difficult to use in practice because it increases with the size of $\#\Omega$. Other limitations of these methods are discussed later on. Refinements of these methods have been proposed where an appropriate threshold $T$ is used for each scale of the coefficients [23]. Denoising of coefficients has also been considered using maximum a posteriori estimation [50, 44, 35, 6, 3]. The restored coefficients are set to minimize an objective function similar to (1),

$$F(x) = \|x - y\|^2 + \sum_i \mu_i \varphi(|x_i|),$$

where $\varphi : \mathrm{R} \to \mathrm{R}$ is a potential function and $\{\mu_i\}$ are weights related to the scale so that the second term in the above expression conveys a multiscale prior on $x$. The restored coefficients can be put into the form $\hat{x}_i = \tau(y_i)$ where $\tau : \mathrm{R} \to \mathrm{R}$ is defined by

$$\tau(y_i) = \arg\min_{t \in \mathrm{R}} \left\{ (t - y_i)^2 + \mu_i \varphi(|t|) \right\}, \tag{6}$$

for every $y_i \in \mathrm{R}$. This approach hence generalizes hard and soft thresholding, defined in (3)-(4), and gives rise to many other shrinkage functions $\tau$; a review can be found in [3]. A particular interest has been carried on priors defined by $\varphi(t) = |t|^\alpha$ for $0 < \alpha \leq 2$, see e.g. [6, 4].

Typically, the functions denoised by variational methods (1) or by shrinkage estimators as given in (2) and (6), exhibit quite a different appearance. For illustration, on can compare Figures 3 and 4 in Section 7. Nevertheless, striking equivalences between these methods in the case of total-variation or Besov regularization, and soft thresholding (4), has been exhibited in [14, 17] and investigated further in [45, 36].

The major problems with shrinkage methods is that shrinking large coefficients can entail oversmoothing of edges, while shrinking small coefficients towards zero yields Gibbs-type oscillations in the vicinity of edges. On the other hand, if shrinkage is not sufficiently strong, some coefficients bearing mainly the noise will remain almost unchanged—we call such coefficients *outliers*—and (2) suggests they generate artifacts with the shape of the functions $\widetilde{w}_i$ of the frame. This effect can be observed in Fig. 1 in Section 4. A lot of studies have been carried out in order to determine functions $\varphi$ in (6) which faithfully account for the statistical distribution of the coefficients $x_i$. An inherent difficulty comes from the fact that coefficients between different scales are not independent, as assumed in (6). Another limitation comes from the necessity to use in practical methods only a limited number of scales and coefficients. It turns out that priors on the coefficients $x$ cannot adequately address important features of the restored function such as the presence of edges and smooth regions.

In order to introduce such information in the restoration, several authors [10, 19, 15, 25, 33, 32, 12, 24] investigated the idea to combine the information contained in the large coefficients $y[i]$ with pertinent priors directly on the sought-after function $u$. Although based on different motives, these "hybrid" methods amount to define the restored function $\hat{u}$ as

$$\text{minimize } \Phi(u) = \int_\Omega \varphi(|\nabla u(s)|)\, ds \text{ subject to } \hat{u} \in \{u : |(W(u-v))[i]| \leq \mu_i, \forall i \in J\}, \tag{7}$$

where in a general way, $\{\mu_i\}$ are determined based on $y$. In the first such method, introduced in [10], $\varphi(t) = t^2$. General functions $\varphi$ are considered in [19]. In order to remove pseudo-Gibbs oscillations, the authors of [25, 33, 12, 24] focused on $\varphi(t) = |t|$ and did various choices for the operator $W$. In [10, 19, 25, 24], orthogonal bases have been used for $W$, [12] have focused on curvelets transform, while [33] have considered unions of wavelet bases. These methods differ also in the choice of parameters $\{\mu_i\}_{i\in J}$. In [33, 32], all $\mu_i$ are equal and determined by the level of the noise. In other methods, the choice of $\mu_i$ takes into account the magnitude of data coefficients $y[i]$ as well. In [19, 25], $\mu_i$ relevant to large data coefficients $y[i]$ are set to 0, while those corresponding to small coefficients are set to $\infty$ in [25]. If the use of an edge-preserving function for $\varphi$ is clearly a pertinent choice, the strategy for the selection of parameters $\{\mu_i\}_{i\in J}$ remains an open question.

In section 2 we provide a critical analysis of the strategies adopted by the authors cited above. Our first conclusion is that the choice for each $\mu_i$ must take into account the magnitude of the relevant data coefficient $y[i]$, and it corroborates the approach followed by [19, 25]. However, deciding on the value of $\mu_i$ based solely on $y[i]$, as done in these papers, is too rigid since there are either correct data coefficients that incur smoothing ($\mu_i > 0$), or noisy coefficients that are left unchanged ($\mu_i = 0$). A way to alleviate to this situation is to determine $\{\mu_i\}_{i\in J}$ based both on the data and on a prior regularization term. This is precisely the project of this paper. It is carried out by defining restored coefficients $\hat{x}$ to minimize an

objective function of the form

$$F_y(x) = \Psi(x, y) + \Phi(\widetilde{W}x), \tag{8}$$

where $\Psi$ is a specially designed data-fitting term and $\Phi$ is of the form (7). More precisely, $F_y$ is designed in such a way that its minimizer $\hat{x}$ involves a classification of the restored coefficients as faithful ($\hat{x}[i] = y[i]$), essentially noise ($\hat{x}[i] = 0$) and coefficients restored by fitting to the regularization term. Following [39, 40], we focus on a new family of objective functions where $\Psi$ is non-smooth. The design of $F_y$ is presented in section 2. Section 3 is dedicated to the existence of a minimizer $\hat{x}$ of $F_y$, and to its uniqueness. To do this analysis, we essentially follow [47]. Our choice for $F_y$ is justified in section 4. In section 5 we study some properties of the minimizer $\hat{x}$ of $F_y$ which give rise to practical bounds for the parameters. Experiments on denoising a signal and an image, presented in section 7, demonstrate the effectiveness of our method over existing denoising schemes.

## 2 Design of an objective function

We start this section with an analysis of the information borne by the data coefficients

$$y[i] = \langle w_i, u \rangle, \quad \forall i \in J.$$

For normalized frame transforms, we can suppose that the noise on each coefficient

$$\eta[i] = \langle w_i, n \rangle, \quad \forall i \in J,$$

is zero mean and has constant variance, let it be denoted $\sigma^2$. However, visual degradation induced by noisy coefficients in different frequency bands is ill-assorted. Noise corresponding to a function $\widetilde{w}_i$ is of the form $\eta[i]\widetilde{w}_i$ where $\eta[i]$ is random, zero-mean and has a fixed variance. When $\widetilde{w}_i$ is low frequency, then it is nearly zero at each point since the frame is normalized. Hence the noise component $\eta[i]\widetilde{w}_i$ is nearly invisible. This suggests we can take $\hat{x}[i] = y[i]$. An additional argument for such a choice is the following. In many frame transforms (e.g. wavelets) all functions $w_i$ whose mean is non-zero are also low-frequency. Taking $\hat{x}[i] = y[i]$ then allows the mean of the original $u_o$ to be preserved. Let $I_* \subset J$ denote the subset of all such coefficients. If one restores all other coefficients, namely $\hat{x}[i]$ for $i \in I = J \setminus I_*$, the sought-after function $\hat{u}$ reads

$$\hat{u} = \sum_{i \in I} \hat{x}[i]\ \widetilde{w}_i + \sum_{i \in I_*} y[i]\ \widetilde{w}_i = \widetilde{W}\hat{x} + \widetilde{W}_*\ y. \tag{9}$$

In order to simplify the notations, we will write $\widetilde{W}$ for the restriction of the frame operator to $I$, and $\widetilde{W}_*$ for its restriction to $I_*$. Another important observation is that for the classes of functions $u_o$ we consider—composed of homogeneous regions and edges—the coefficients $\langle w_i, u_o \rangle$ corresponding to a low-pass function $w_i$ have a large magnitude. In contrast, for the high-frequency functions $w_i$ we have

$$\langle w_i, u_o \rangle \approx 0, \quad \text{for many indexes } i, \tag{10}$$

whereas $\langle w_i, u_o \rangle$ has a significative magnitude only in connection with to edges. In particular, the data coefficients $y_i$ for all indexes $i$ mentioned in (10) contain essentially noise, $y_i \approx \eta[i]$. E.g., the values of more than 95% of them are contained in $[-2\sigma, 2\sigma]$ if $\eta$ is Gaussian noise. If there are other less noisy data coefficients with values in the essential range of the noise, they cannot be distinguished from the

former coefficients. No reliable information on $u_o$ can be extracted from data coefficients whose values are in the essential range of the noise. This suggest we split the set $I = J \setminus I_*$ into $I = I_0 \cup I_1$,

$$I_0 = \{i \in I : |y[i]| \leq T\}, \tag{11}$$

$$I_1 = I \setminus I_0, \tag{12}$$

where $T > 0$ (e.g. of the order of $2\sigma$), and consider the restoration of the coefficients $\hat{x}[i]$ corresponding to each subset separately. Our discussion allows us to characterize the goals of the restoration for each subset as presented below.

$(G_0)$ The coefficients $y[i]$ for $i \in I_0$ are usually high-frequency components which can be

- Noise data coefficients, if they correspond to (10). The best restoration for these coefficients is certainly $\hat{x}[i] = 0$.
- Coefficients $y[i]$ which correspond to edges and other details in $u_o$. Since $y[i]$ is difficult to distinguish from the noise, the relevant $\hat{x}[i]$ will be restored based on the edge-preserving priors conveyed by $\Phi$ in (8). Notice that a careful restoration certainly leads to a nonzero $\hat{x}[i]$, since otherwise $\hat{x}[i] = 0$ would generate Gibbs-like oscillations in $\hat{u}$.

$(G_1)$ $I_1$ addresses two types of coefficients $y[i]$:

- Large coefficients which bear the main features of the sought-after function. They verify $y[i] \approx \langle w_i, u_o \rangle$ and must be kept intact, i.e. $\hat{x}[i] = y[i]$.
- Coefficients which are highly contaminated by noise, characterized by $|y[i]| \gg |\langle w_i, u_o \rangle|$. We call them *outliers* because if we had $\hat{x}[i] = y[i]$, then $\hat{u}$ would contain an artifact with the shape of $\widetilde{w}_i$ since $\sum_j \hat{x}[j]\widetilde{w}_j + (y[i] - \langle w_i, u_o \rangle)\,\widetilde{w}_i$. Instead, $\hat{x}[i]$ will be restored according to the prior carried by $\Phi$.

Observe that dropping all data coefficients $y[i]$ for $i \in I_0$ amounts to consider

$$y_T[i] = \begin{cases} 0 & \text{if } |y[i]| \leq T, \\ y[i] & \text{otherwise.} \end{cases} \tag{13}$$

In fact, $y_T$ is obtained from $y$ by hard thresholding, $y_T[i] = \tau(y[i])$, where $\tau$ is the function given in (3). We require hence that the minimizer $\hat{x}$ of $F_y$ achieves all the goals stated in $(G_0)$-$(G_1)$. In particular, $\hat{x}$ must involve an implicit classification between coefficients that must fit to $y_T$ exactly and coefficients that must be restored. For the latter, $\hat{x}$ have to provide a pertinent restoration. Qualitatively speaking, restored coefficients have to fit $y_T$ exactly if they are in accordance with the regularization term and have to be restored else. In order to design an objective function whose minimizer satisfies $(G_0)$-$(G_1)$ we follow [39, 40] where objective functions with $\Psi$ non-smooth at the origin are considered. These objective functions were shown to have the property that a certain number of the restored coefficients satisfy $\hat{x}[i] = y_T[i]$ (in fact, if they are in accordance with the prior), whereas the other coefficients are restored by fitting to the regularization term. Our attention being restricted to convex objective functions, the most natural choice is $\Psi(x, y) = \|x - y_T\|_1$ where $\|.\|_1$ denotes the $\ell_1$ norm. So, $F_y$ is of the form

$$F_y(x) = \sum_{i \in I_1} \lambda_i \, |(x - y)[i]| + \sum_{i \in I_0} \lambda_i \, |x[i]| + \Phi(x), \tag{14}$$

$$\Phi(x) = \int_\Omega \varphi(|\nabla \widetilde{W} x|)\, ds, \tag{15}$$

and the sought-after solution $\hat{x}$ minimizes $F_y$ over the subset $\{x \in J : x[i] = y[i],\ \forall i \in I_*\}$. The regularization $\Phi$ brings the priors about the local features of the restored function. Its role is critical on the regions corresponding to wavelet coefficients which are either outliers or are erroneously set to zero. The images and signals we wish to restore are supposed to involve smooth regions and edges. To this end, we focus on *edge-preserving* convex potential functions $\varphi$ which have been studied by many authors [42, 11, 31, 9, 16]. An essential distinction between these potential functions is the differentiability of $t \to \varphi(|t|)$ at the origin. Since [38], it is known that if $t \to \varphi(|t|)$ is non-smooth at zero, restored images and signals $\widetilde{W}\hat{x}$ involve constant regions. Such a property does not correspond to real-world images and signals. In contrast, if $\Phi$ is smooth, they contain smoothly varying regions and possibly edges. We hence focus on potential functions $\varphi$ of the latter kind, which means that $\varphi'(0^+) = 0$. Examples of such functions are [28, 11, 9, 16, 49]

$$\varphi(t) = t^\alpha, \quad 1 < \alpha \leq 2, \tag{16}$$

$$\varphi(t) = \sqrt{\alpha + t^2}, \tag{17}$$

$$\varphi(t) = \log(\cosh(\alpha t)), \tag{18}$$

$$\varphi(t) = |t| - \alpha \log\left(1 + \frac{|t|}{\alpha}\right), \tag{19}$$

$$\varphi(t) = \begin{cases} t^2/2 & \text{if} \quad |t| \leq \alpha, \\ \alpha|t| - \alpha^2/2 & \text{if} \quad |t| > \alpha, \end{cases} \tag{20}$$

where $\alpha > 0$ is a parameter. Notice that $\alpha = 1$ in (16) or $\alpha = 0$ in (17) leads to $\varphi(t) = |t|$ in which case $\varphi'(0^+) = 1$.

**Remark 1** In (13), we would not recommend the use of another shrinkage function $\tau$ to construct $y_T$ since it will alter *all* the data coefficients, without restoring them faithfully. In contrast, we base our restoration on data preserving all the initial information on the sought-after image or signal. To this end, we choose a threshold $T$ considerably smaller than (5).

The considerations on the information content of noisy coefficients $y$ provide a tool to make a critical analysis of the hybrid methods mentioned in the Introduction, namely [10, 19, 25, 33, 12, 24]. Let us come back to the formulation given in (8). If the use of an edge-preserving function for $\varphi$ is clearly a pertinent choice, the strategy for the selection of parameters $\{\mu_i\}_{i \in J}$ remains an open question. In order to enable noise removal, all $\mu_i$ corresponding to coefficients degraded by the noise must be large enough to allow the noise to be smoothed by the regularization term. Then the residual $|(Wv)[i] - (W\hat{u})[i]|$ between data and restored coefficients is non-zero and increases with the magnitude of the coefficients. Hence, large $\mu_i$ for large data coefficients $(Wv)[i]$ entail oversmoothing of important features in the restored image or signal, as it can be observed in Fig. 11. This can be seen as a weakness of the method of [33]. A further observation, exploited in [19, 25], is that for each data coefficient, its signal-to-noise ratio is better (i.e. higher) if its magnitude is larger, and vice-versa. The $\mu_i$ for data coefficients which are likely to have a large SNR can be small, or null. The question of how to decide which coefficient is large and less degraded, and which is small and highly degraded, is difficult to solve by thresholding. An optimistic (i.e. low) thresholding rule gives $\mu_i = 0$ for a large number of data coefficients; it is very likely that some of them—typically those of medium magnitude—are highly corrupted, thus leading to "outliers" in the restored coefficients. A pessimistic (i.e. high) thresholding rule gives $\mu_i > 0$ for a large amount of data coefficients, thus leading to oversmoothing in the restored $\hat{u}$. More flexibility is hence necessary in the

restoration of these coefficients. Furthermore, the restoration of small coefficients is important to avoid Gibbs-type oscillations in the vicinity of edges and to obtain smooth regions between the edges; the latter may need that some of these coefficients remain nearly null, or null.

# 3 Well-posedness of the minimization problem

In this section we focus on the existence and the uniqueness of the minimization problem formulated in (14)-(15).

## 3.1 Existence

It is clear that $F_y$ does have a minimizer when $\#\Omega$ is finite. This question requires more investigation when $\Omega$ is an open subset of $\mathrm{R}^d$. For all $u \in \mathrm{BV}(\Omega)$, denote by $Du$ its weak derivative and recall the Lebesgue decomposition

$$Du = \nabla u \, \mathcal{L}_d + D_s u,$$

where $\mathcal{L}_d$ is the Lebesgue measure on $\mathrm{R}^d$, $\nabla u \in (L^1(\Omega))^d$ is the Radon-Nikodym derivative of $Du$ and $D_s u$ is singular with respect to $\mathcal{L}_d$. In order to ensure the existence of a minimizer we must change the objective function in its relaxation on $BV - w^*$,

$$\begin{aligned} F_y(x) &= \sum_{i \in I_1} \lambda_i \, |(x-y)[i]| + \sum_{i \in I_0} \lambda_i \, |x[i]| \\ &+ \varphi(|D\widetilde{W}x|)(\Omega), \end{aligned} \tag{21}$$

where $\varphi(|D\widetilde{W}x|)(\Omega) = \int_\Omega \varphi(|\nabla \widetilde{W}x|)\, ds + |D_s \widetilde{W}x|(\Omega)$. We consider $F_y$ as a function on $\ell^2(J)$ although $\varphi(|D\widetilde{W}x|)(\Omega)$ is not finite for all $x \in \ell^2(J)$. Hence the so defined $F_y$ is $\overline{\mathrm{R}}$ valued.

**Theorem 1** *For $y \in \ell^2(J)$ and $T > 0$ given, consider $F_y$ as defined in (21), where $\Omega \in \mathrm{R}^d$ is open and bounded, its boundary $\partial\Omega$ is Lipschitz, $\varphi$ is convex and there is a constant $a > 0$ such that*

$$t - a \leq \varphi(t) \leq t + a, \qquad \forall t \in \mathrm{R}_+. \tag{22}$$

*Furthermore, we suppose that*

$$w_i \in L^1(\Omega) \quad \textit{and} \quad \int_\Omega w_i(s) ds = 0, \quad \forall i \in I. \tag{23}$$

*Suppose that either the assumptions in (B), or those in (F), are satisfied:*

*(B)*
1. *$\{w_i\}_{i \in J}$ is a Riesz basis of $L^2(\Omega)$ ;*
2. *we have $d = 1$ or $d = 2$ ;*

*(F)*
1. *$\{w_i\}_{i \in J}$ is a frame of $L^2(\Omega)$ and the operator $\widetilde{W}$ is the pseudo-inverse of $W$ ;*
2. *if $d \geq 2$, $w_i \in L^d(\Omega)$, for all $i \in J$ ;*
3. *$\lambda_{\min} = \min_{i \in I} \lambda_i > 0$ .*

*Then $F_y$ has a minimizer in $\{x \in \ell^2(J) : x[i] = y[i], \forall i \in I_*\}$.*

Recall that the pseudo-inverse of $W$ reads $\widetilde{W} = (W^*W)^{-1}W^*$ where $W^*$ is the adjoint operator. More details can be found in [34]. The assumption (22) is general enough for edge-preserving regularization. It holds for all functions given in (17)-(20), as well as for (16) if $\alpha = 1$.

**Proof.** Our proof is essentially inspired by [47]. For simplicity, we denote by $K$ a positive constant whose value can change from an equation to another.

Let $\{x_n\}_{n\geq 1}$ be a minimizing sequence for (21). Then

$$|F_y(x_n)| \leq K, \quad \forall n \geq 1, \tag{24}$$

and using (22),

$$K \geq \int_\Omega \varphi(|\nabla \widetilde{W} x_n|\, ds) \geq \int_\Omega |\nabla \widetilde{W} x_n|\, ds - a|\Omega|, \quad \forall n \geq 1.$$

Combining this with the fact that $\Omega$ is bounded shows that

$$\int_\Omega |\nabla \widetilde{W} x_n|\, ds \leq K, \quad \forall n \geq 1. \tag{25}$$

From (24) yet again, $\left|D_s \widetilde{W} x_n\right|(\Omega) \leq K$, for all $n \geq 1$, which combined with (25) shows that

$$\left|D \widetilde{W} x_n\right|(\Omega) \leq K, \quad \forall n \geq 1. \tag{26}$$

Let $\chi_\Omega$ denote the characteristic function of $\Omega$. Put

$$c = \frac{1}{|\Omega|}\left(\int_\Omega \widetilde{W} x_n\right)\chi_\Omega.$$

Notice that $c$ is independent of $n$ since the mean of $\widetilde{W} x_n$ is fixed by the model ($w_i$ has zero mean for all $i \in I_*$). Using the Poincaré-Wirtinger inequality,

$$\|\widetilde{W} x_n - c\|_{L^p(\Omega)} \leq K, \quad \text{where } p = 2 \text{ if } d = 1 \text{ and } p = \frac{d}{d-1} \text{ if } d \geq 2.$$

Since $\Omega$ is bounded, we have $\|\widetilde{W} x_n\|_{L^p(\Omega)} \leq K$, and hence,

$$\|\widetilde{W} x_n\|_{L^1(\Omega)} \leq K.$$

It follows from (26) and the latter equation that $\widetilde{W} x_n \in BV$ and that

$$\|\widetilde{W} x_n\|_{BV} \leq K, \quad \forall n \geq 1. \tag{27}$$

Then there exist $u \in BV$ and a subsequence, denoted $\{x_n\}$ again, such that

$$\widetilde{W} x_n \rightharpoonup u \text{ in } L^p, \tag{28}$$

$$D\widetilde{W} x_n \rightharpoonup Du \text{ in } w*\mathcal{M}(\Omega), \tag{29}$$

where $\mathcal{M}(\Omega)$ is the set of signed measures on $\Omega$ with bounded total variation. Since $\varphi(|\,.\,|)(\Omega)$ is weak$*$ lower semicontinuous on $\mathcal{M}(\Omega)$, we have

$$\varphi(|Du|)(\Omega) \leq \liminf_{n\to\infty} \varphi(|D\widetilde{W} x_n|)(\Omega). \tag{30}$$

Suppose that there exists $\hat{x} \in \{x \in \ell^2(J) :\ x[i] = y[i],\ \forall i \in I_*\}$ such that

$$\begin{cases} u = \widetilde{W}\hat{x}, \\ x_n[i] \to \hat{x}[i], \quad \forall i \in I. \end{cases} \tag{31}$$

By Fatou's Lemma,

$$\sum_{i\in I} \lambda_i \left|\hat{x}[i] - y_T[i]\right| \leq \lim_{n\to\infty} \inf \sum_{i\in I} \lambda_i \left|x_n[i] - y_T[i]\right|.$$

Combining this result with (30) shows that

$$F_y(\hat{x}) \leq \lim_{n\to\infty} \inf F_y(x_n).$$

In the following we will show that if either (B) or (F) holds, there is an $\hat{x}$ and a subsequence of $\{x_n\}$ such that (31) is satisfied.

- Consider that (B) holds. Using (28) for $p = 2$, we can write that

  $$\langle \widetilde{W}x_n, w_i\rangle \to \langle u, w_i\rangle, \quad \forall i \in I.$$

  From (B)-1, $\langle \widetilde{W}x_n, w_i\rangle = x_n[i]$, for all $i \in J$. If we put $\hat{x}[i] = \langle u, w_i\rangle$, for every $i \in J$, the above expression yields that

  $$x_n[i] \to \hat{x}[i], \quad \forall i \in I. \tag{32}$$

- Consider that the assumptions (F) hold. Based on (24),

  $$\begin{aligned} K &\geq \sum_{i\in I} \lambda_i \left|x_n[i] - y_T[i]\right| \\ &\geq \lambda_{\min} \sum_i \left|x_n[i] - y_T[i]\right| \\ &= \lambda_{\min} \left\|x_n - y_T\right\|_{\ell^1(I)} \\ &\geq \lambda_{\min} \left\|x_n - y_T\right\|_{\ell^2(I)} \\ &\geq \lambda_{\min} \left\|x_n\right\|_{\ell^2(I)} - \lambda_{\min} \left\|y_T\right\|_{\ell^2(I)}. \end{aligned}$$

  Using (F)-3, we deduce that $\|x_n\|_{\ell^2(I)} \leq K$ for all $n \geq 1$. It follows that there are $\hat{x} \in \{x \in \ell^2(J) : x[i] = y[i],\ \forall i \in I_*\}$ and a subsequence, denoted $\{x_n\}$ again, so that

  $$x_n \rightharpoonup \hat{x} \quad \text{in} \quad \ell^2(J). \tag{33}$$

  In particular, we have

  $$x_n[i] \to \hat{x}[i], \quad \forall i \in J, \tag{34}$$

  $$(W\widetilde{W}x_n)[i] \to (W\widetilde{W}\hat{x})[i], \quad \forall i \in J, \tag{35}$$

  since $W\widetilde{W}$ is the orthogonal projection onto $\mathrm{Im}W$. Using that by (F)-2, we have $w_i \in (L^p)'$ for every $i$, (28) leads to

  $$\langle \widetilde{W}x_n, w_i\rangle \to \langle u, w_i\rangle \quad \forall i \in J, \tag{36}$$

  which is equivalent to

  $$(W\widetilde{W}x_n)[i] \to (Wu)[i], \quad \forall i \in J.$$

From the last two results,

$$W\widetilde{W}\hat{x} = Wu.$$

Then

$$\widetilde{W}\hat{x} = \widetilde{W}W\widetilde{W}\hat{x} = \widetilde{W}Wu = u.$$

$\diamond$

**Remark 2** In the case of a signal or an image ($d = 1$ or 2, respectively), assumptions (B)-(F) boil down to write that either $\{w_i\}_{i\in J}$ is a Riesz basis of $L^2(\Omega)$ or $\lambda_{\min} > 0$.

**Remark 3** In the case of frames, assumption (F)-2 saying that $w_i \in L^d(\Omega)$ is usually satisfied since in practice $w_i \in L^\infty(\Omega) \cup BV(\Omega)$.

## 3.2 Uniqueness

The following theorem applies eather when $\Omega$ is finite or when $\Omega$ and $F_y$ are as in Theorem 1.

**Theorem 2** *Suppose $\varphi$ is convex. If $\hat{x}_1$ and $\hat{x}_2$ are two minimizers of $F_y$, then*

$$\nabla\widetilde{W}\hat{x}_1 \propto \nabla\widetilde{W}\hat{x}_2, \qquad \text{a.e. on } \Omega. \tag{37}$$

*Moreover, if $\varphi$ is strictly convex, then*

$$\nabla\widetilde{W}\hat{x}_1 = \nabla\widetilde{W}\hat{x}_2, \qquad \text{a.e. on } \Omega. \tag{38}$$

Before to prove the theorem, we comment its meaning in the situations where any solution $\hat{u} = \widetilde{W}\hat{x}$ is smooth (e.g. $J$ is finite and $\widetilde{W}$ is smooth). In these cases $D\hat{u} = \nabla\hat{u}$, so (37) or (38) hold everywhere on $\Omega$. As the gradient of an image at any point is orthogonal to the level line passing through this point, Theorem 2 means that $\widetilde{W}\hat{x}_1$ and $\widetilde{W}\hat{x}_2$ have the same level lines. In other words, these two images are obtained one from another by a local change of contrast. Noticing also that in practice $\int_\Omega \widetilde{w}_i\, ds = 0$, for all $i \in I$, choosing $\varphi$ strictly convex ensures that there is a unique smooth $\hat{u} = \widetilde{W}\hat{x}$.

**Proof.** Let us decompose $F_y$ as follows:

$$F_y = G(x) + H_y(x),$$

where

$$G(x) = \int_\Omega \varphi(|\nabla(\widetilde{W}x)(s)|)\, ds, \tag{39}$$

$$H_y(x) = \sum_{i\in I_1} \lambda_i\, |(x-y)[i]| + \sum_{i\in I_0} \lambda_i\, |x[i]| + \left|D_s(\widetilde{W}x)\right|(\Omega). \tag{40}$$

Since $F_y$ is convex, $(\hat{x}_1 + \hat{x}_2)/2$ is a minimizer of $F_y$ as well. We can write that

$$F_y\left(\frac{\hat{x}_1 + \hat{x}_2}{2}\right) = \frac{F_y(\hat{x}_1) + F_y(\hat{x}_2)}{2}. \tag{41}$$

From the convexity of $G$ and $H_y$,

$$G\left(\frac{\hat{x}_1 + \hat{x}_2}{2}\right) \leq \frac{G(\hat{x}_1) + G(\hat{x}_2)}{2},$$

$$H_y\left(\frac{\hat{x}_1 + \hat{x}_2}{2}\right) \leq \frac{H_y(\hat{x}_1) + H_y(\hat{x}_2)}{2}.$$

If one of these inequalities is strict, (41) shows that the other inequality cannot be satisfied. It follows that both inequalities are in fact equalities. In particular,

$$\int_\Omega \varphi\left(\left|\nabla\left(\widetilde{W}\frac{\hat{x}_1+\hat{x}_2}{2}\right)(s)\right|\right)\,ds = \int_\Omega \frac{\varphi(|\nabla(\widetilde{W}\hat{x}_1)(s)|)+\varphi(|\nabla(\widetilde{W}\hat{x}_2)(s)|)}{2}\,ds. \tag{42}$$

Using that $\varphi$ is convex, the inequality below holds for almost every $s\in\Omega$:

$$\varphi\left(\left|\nabla\left(\widetilde{W}\frac{\hat{x}_1+\hat{x}_2}{2}\right)(s)\right|\right) \le \frac{\varphi(|\nabla(\widetilde{W}\hat{x}_1)(s)|)+\varphi(|\nabla(\widetilde{W}\hat{x}_2)(s)|)}{2}.$$

If the last inequality was strict on a subset of $\Omega$ of positive measure, we would find that the left-hand side of (42) is strictly smaller than its right-hand side. We deduce that for almost every $s\in\Omega$,

$$\varphi\left(\left|\nabla\left(\widetilde{W}\frac{\hat{x}_1+\hat{x}_2}{2}\right)(s)\right|\right) = \frac{1}{2}\varphi(|\nabla(\widetilde{W}\hat{x}_1)(s)|)+\frac{1}{2}\varphi(|\nabla(\widetilde{W}\hat{x}_2)(s)|).$$

Using a triangular inequality,

$$\left|\nabla\left(\widetilde{W}\frac{\hat{x}_1+\hat{x}_2}{2}\right)(s)\right| \le \frac{\left|\nabla(\widetilde{W}\hat{x}_1)(s)\right|+\left|\nabla(\widetilde{W}\hat{x}_2)(s)\right|}{2}.$$

Using that $\varphi$ is increasing, this entails that

$$\begin{aligned}\varphi\left(\left|\nabla\left(\widetilde{W}\frac{\hat{x}_1+\hat{x}_2}{2}\right)(s)\right|\right) &\le \varphi\left(\frac{1}{2}\left|\nabla(\widetilde{W}\hat{x}_1)(s)\right|+\frac{1}{2}\left|\nabla(\widetilde{W}\hat{x}_2)(s)\right|\right)\\ &\le \frac{1}{2}\varphi(|\nabla(\widetilde{W}\hat{x}_1)(s)|)+\frac{1}{2}\varphi(|\nabla(\widetilde{W}\hat{x}_2)(s)|)\\ &= \varphi\left(\left|\nabla\left(\widetilde{W}\frac{\hat{x}_1+\hat{x}_2}{2}\right)(s)\right|\right).\end{aligned} \tag{43}$$

The above inequalities are therefore equalities. Using that $\varphi$ is strictly increasing, it follows that for almost every $s\in\Omega$,

$$\left|\nabla(\widetilde{W}\hat{x}_1)(s)+\nabla(\widetilde{W}\hat{x}_2)(s)\right| = \left|\nabla(\widetilde{W}\hat{x}_1)(s)\right|+\left|\nabla(\widetilde{W}\hat{x}_2)(s)\right|.$$

We conclude that, for almost every $s\in\Omega$,

$$\nabla(\widetilde{W}\hat{x}_1)(s) \propto \nabla(\widetilde{W}\hat{x}_2)(s).$$

If $\varphi$ is strictly convex, (43) leads to the result in (38). The proof is complete. ◇

# 4 Rationale of the objective function

The function $F_y$ in (21) belongs to the family

$$\begin{aligned}F_y(x) &= \sum_{i\in I_1}\psi_i(|(x-y)[i]|)+\sum_{i\in I_0}\psi_i(|x[i]|)\\ &+ \varphi(|D\widetilde{W}x|)(\Omega),\end{aligned} \tag{44}$$

where $\psi_i:\mathrm{R}_+\to\mathrm{R}_+$, $i\in I_0\cup I_1$ are $\mathcal{C}^1$, convex and increasing functions, and $\varphi$ is convex as well, as specified above. To simplify the presentation, we take $\psi_i(0)=0$ for all $i$. The justification of the choice we made in (21) relies on an analysis of the necessary and sufficient conditions for a minimum of $F_y$ as given in (44). Following [29], $F_y$ reaches its minimum at $\hat{x}$ if and only if $0\in\partial F_y(\hat{x})$ where the set $\partial F_y(\hat{x})$ is the subdifferential of $F_y$ at $\hat{x}$. In our case, this condition yields

- $\forall i \in I_1$,

$$\hat{x}[i] = y[i] \quad \Rightarrow \quad \exists g \in \partial_i \Phi(\hat{x}) \ : \ |g| \leq \psi_i'(0), \tag{45}$$

$$\hat{x}[i] \neq y[i] \quad \Rightarrow \quad -\psi_i'\big(|(\hat{x}-y)[i]|\big)\mathrm{sign}\big((\hat{x}-y)[i]\big) \in \partial_i \Phi(\hat{x}), \tag{46}$$

- $\forall i \in I_0$,

$$\hat{x}[i] = 0 \quad \Rightarrow \quad \exists g \in \partial_i \Phi(\hat{x}) \ : \ |g| \leq \psi_i'(0), \tag{47}$$

$$\hat{x}[i] \neq 0 \quad \Rightarrow \quad -\psi_i'(|\hat{x}[i]|)\mathrm{sign}\big(\hat{x}[i]\big) \in \partial_i \Phi(\hat{x}), \tag{48}$$

where $\partial_i \Phi(x)$ is the subdifferential of $\Phi$ at $\hat{x}$ on the subspace spanned by $e_i$ defined by

$$e_i[i] = 1 \quad \text{and} \quad e_i[j] = 0 \text{ if } j \neq i.$$

Notice that these conditions for a minimum hold both for $F_y$ smooth and nonsmooth. If $\Phi$ is differentiable on a neighborhood of $x$ in the direction of $e_i$, then $g = \partial_i \Phi(x)$ is the $i$th partial derivative, namely

$$\partial_i \Phi(x) = \partial \Phi / \partial x[i]\,(x) = \int_\Omega \varphi'(|\nabla \widetilde{W} x|)\,(\nabla \widetilde{w}_i)^T \,\frac{\nabla \widetilde{W} x}{\left|\nabla \widetilde{W} x\right|}\, ds. \tag{49}$$

For simplicity, in the expression above we consider that $\widetilde{w}_i$ is differentiable. If $t \to \psi_i(|t|)$ are smooth functions we have $\psi_i'(0) = 0$ for all $i$, in which case (45) and (47) become equalities. Let us emphasize that the classical choice for $\psi_i$ is $\psi_i(t) = t^2$, for all $i$. The clue of our method concerns the choice of $\psi_i$ in (44). Below we derive a set of necessary conditions for $\psi_i$ enabling the basic desiderata $(G_0 - G_1)$ in section 2 to be satisfied.

## 4.1 Restoration of large noisy coefficients ($I_1$).

First we consider all coefficients $y[i]$ for $i \in I_1$ where $I_1$ is defined by (11)-(12). Let us recall that

$$|y[i]| > T \quad \Leftrightarrow \quad i \in I_1.$$

- **Preservation of significant coefficients.** Consider that for some $i \in I_1$ we have

$$y[i] \approx (\widetilde{W} u_o)[i].$$

In order to prevent $\hat{u}$ from erosion, a good choice for $\hat{x}[i]$ is certainly $\hat{x}[i] = y[i]$. In words, $\hat{x}[i]$ is required to satisfy (45) where we have $\psi_i'(0) \geq 0$ because $\psi_i$ is increasing on $\mathbf{R}_+$. If it happens that $y[i]$ fits the prior in such a way that $0 \in \partial_i \Phi(\hat{x})$, then (45) holds for any $\psi_i$. However, having $0 \in \partial_i \Phi(\hat{x})$ along with $\hat{x}[i] = y[i]$ is a highly special case. It is enough to see this in the scalar case when $F_y : \mathbf{R} \to \mathbf{R}_+$ reads $F_y(x) = \psi(|x-y|) + \varphi(|wx|)$. If $y \neq 0$, the function $x \to \varphi(|wx|)$ is differentiable on a neighborhood of $y$ and the above requirement, namely $0 = \varphi'(|wy|)$, cannot be satisfied when $\varphi$ is convex and increasing on $\mathbf{R}_+$. We can have $\hat{x} = y$ only for $y = 0$. The general situation corresponds to $0 \notin \partial_i \Phi(\hat{x})$ if $\hat{x}[i] = y[i]$, hence a necessary condition enabling (45) to hold is that the constant $\lambda_i$ below,

$$\lambda_i = \psi_i'(0) \tag{50}$$

satisfies $\lambda_i > 0$. Since all $\psi_i$ are convex and $\psi_i(0) = 0$, this implies that

$$\psi_i(t) \geq \lambda_i t \quad \forall t > 0. \tag{51}$$

- **Suppression of outliers.** The noise $n$ being unbounded, for any choice of $T$ in (13) we can have highly corrupted coefficients $y[i]$ for $i \in I_1$ corresponding to

$$\big|(Wu_o)[i]\big| \ll T.$$

Any such $y[i]$ bears no information on the true $(Wu_o)[i]$, so the best choice is that $\hat{x}[i]$ fits the prior as well as possible. By (44) this will occur if the contribution of $\psi_i$ to $F_y$ is as small as possible. Given that $\psi_i$ satisfies (51), we are induced to choose

$$\psi_i(t) = \lambda_i t \quad \text{for all} \quad t \geq 0, \tag{52}$$

where $\lambda_i > 0$.

## 4.2 Restoration of coefficients in $I_0$.

Now we focus on all coefficients $y[i]$ for $i \in I_0$ where $I_0$ is defined by (11). For reminder,

$$\big|y[i]\big| \leq T \quad \Leftrightarrow \quad i \in I_0.$$

These coefficients usually correspond to high-frequency components in the restored signal or image.

- **Suppression of noise coefficients in $I_0$.** A pertinent choice of a frame means that we have a sparse representation, that is

$$|(Wu_o)[i]| \approx 0,$$

for many coefficients $i$. Most of the relevant coefficients $y[i]$ are likely to satisfy $\big|y[i]\big| \leq T$. For all these coefficients, we claim that the most reasonable choice is $\hat{x}[i] = 0$. If $\psi_i = 0$ on $\mathrm{R}_+$ for all $i \in I_0$, these coefficients will be restored according to the prior $\Phi$, but then there is no guarantee they are close to zero. The argument is that $\Phi$ promotes signals and arguments that are locally homogeneous (or even locally constant in the case of TV regularization), separated by edges. Unconstrained coefficients can hence yield spurious edges in regions where the gradient is important, in order to break them into regions with a smaller gradient.

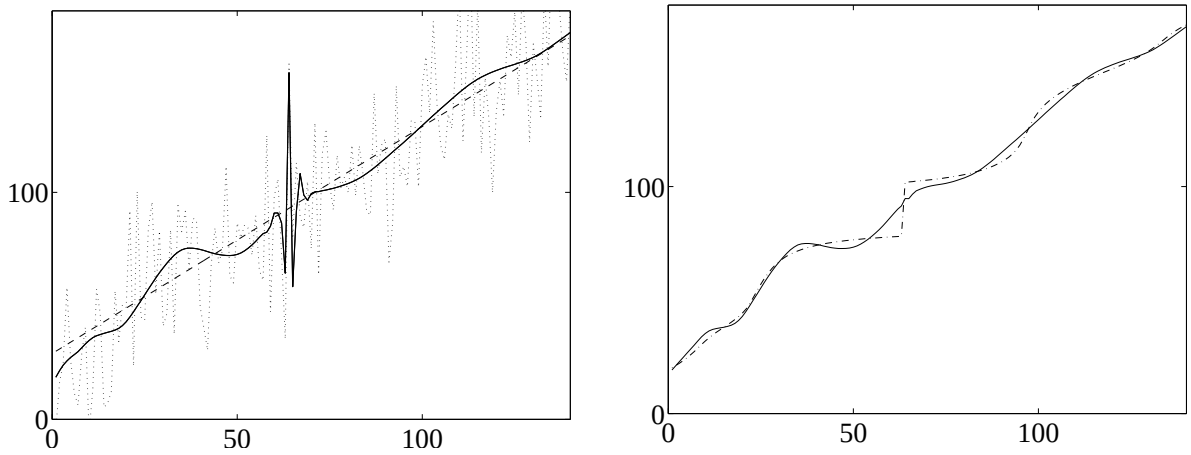
The above reasoning is illustrated in Fig. 1. The ramp-shaped data in (a) contain many coefficients indexed by $I_0$ (i.e. smaller than $T$) as well as one outlier that gives rise to a wavelet-shaped artifact in $Wy_T$. Both solutions in (b) correspond to the minimization of an $F$ as given in (44) where $\forall i \in I_1$ we take $\psi_i(t) = \lambda_i t$ with $\lambda_i > 0$, as proposed in § 4.1. Furthermore, the solution plotted with a dashed line corresponds to $\psi_i(t) = 0$ on $\mathrm{R}$ $\forall i \in I_0$ while the solution plotted with a solid line corresponds to $\psi_i(t) = \lambda_i t$ with $\lambda_i =$const$> 0$ $\forall i \in I_0 \cup I_1$. As explained in § 4.1, this outlier is restored in both solutions. However, the solution corresponding to $\psi_i = 0$, $\forall i \in I_0$ exhibits a spurious edge near the restored outlier and several coefficients indexed by $I_0$ have large values. In contrast, the solution corresponding to $\lambda_i > 0$, $\forall i \in I_0$ pushes these coefficients to zero and thus the continuity of the restored ramp is preserved.

Trivial necessary conditions enabling to keep null the noise coefficients in $I_0$ are that for every $i \in I_0$ we have

$$\psi_i(0) < \psi_i(t), \quad \forall t > 0.$$

This, combined with the convexity of $\psi_i$ yields

$$\psi_i'(t) > 0 \quad \forall t > 0, \quad \forall i \in I_0. \tag{53}$$




(a) Noisy data (....) corresponding to an original slope (- - -) and the solution $Wy_T$ (—) obtained by hard thresholding.

(b) Solution for $\lambda_i = 0$ for all $i \in I_0$ (- - -); Solution for $\lambda_i$ =const, $\forall i \in I_0 \cup I_1$ (—-).

Figure 1: Data in (a) and restorations using (14) in (b).

- **Suppression of Gibbs-like effect.** Even if $T$ in (13) is small compared to the optimal (5), we can have coefficients $y[i]$ for $i \in I_0$ corresponding to large $(Wu_o)[i]$ whose magnitude is just below $T$, i.e.

$$\Big|(Wu_o)[i]\Big| \lessapprox T.$$

  As mentioned in section 2, keeping $\hat{x}[i] = 0$ would entail Gibbs-like oscillations in the restored $\hat{u}$. A reasonable choice for $\hat{x}[i]$ is certainly non-zero, in which case $\hat{x}[i]$ should be restored with the aid of equation (48) and realize a good fit to the prior, in spite of the fact that $\psi_i'(|\hat{x}[i]|) > 0$ according to (53). Hence the requirements that the equation (48) holds, namely $-\psi_i'(|\hat{x}[i]|)\text{sign}(\hat{x}[i]) \in \partial_i\Phi(\hat{x})$, where $\psi_i'(|\hat{x}[i]|) > 0$ remains close to zero even if $\hat{x}[i]$ has a large value. The latter requirement suggests we fix $\psi_i'(t) = \lambda_i > 0$ for all $t > 0$ and finally

$$\psi_i(t) = \lambda_i t \quad \forall t \geq 0, \quad \forall i \in I_0.$$

### 4.3 On the choice of $\varphi$

We have already explained in section 2 why $\varphi$ must be a convex edge-preserving function as those given in (16)-(20). For all of them, the function $t \to \varphi(|t|)$ is smooth (i.e. $\varphi'(0) = 0$), except for $\varphi(t) = t$ which corresponds to the total-variation regularization. The smoothness of $t \to \varphi(|t|)$ at the origin has important consequences on the solution. Since [37, 38, 41], it is well known why the minimizers corresponding to a non-smooth at zero $t \to \varphi(|t|)$ tend to be constant on some regions. Such a stair-casing effect is undesirable in our context. This effect is weakened by our $\ell_1$ data-fidelity term since it constrains many coefficients to fit exactly the data rather than the regularization term. Nevertheless, each time we have "neighboring" outliers, corresponding to functions $\widetilde{w}_i$ that address the same region of $u$ (i.e. whose supports overlap), the solution is again prone to exhibit stair-casing, especially in the case of signals. This argument suggests a preference for functions $\varphi$ satisfying $\varphi'(0) = 0$—for instance (17), in which case $t \to \varphi(|t|)$ can also be seen as a smooth approximation of $|t|$.

# 5 Some practical properties of the solution

Let us focus on $F_y$ as given in (14) where $t \to \varphi(|t|)$ can be either smooth or non-smooth at zero. The conditions for a minimum presented in section 4 now have a simpler form:

- $\forall i \in I_1$

$$\hat{x}[i] = y[i] \quad \Rightarrow \quad \exists g \in \partial_i \Phi(\hat{x}) \ : \ |g| \leq \lambda_i, \tag{54}$$

$$\hat{x}[i] \neq y[i] \quad \Rightarrow \quad -\lambda_i \text{sign}\big((\hat{x} - y)[i]\big) \in \partial_i \Phi(\hat{x}). \tag{55}$$

- $\forall i \in I_0$

$$\hat{x}[i] = 0 \quad \Rightarrow \quad \exists g \in \partial_i \Phi(\hat{x}) \ : \ |g| \leq \lambda_i, \tag{56}$$

$$\hat{x}[i] \neq 0 \quad \Rightarrow \quad -\lambda_i \text{sign}\big(\hat{x}[i]\big) \in \partial_i \Phi(\hat{x}), \tag{57}$$

where $\partial_i \Phi(x)$—the subdifferential of $\Phi$ at $\hat{x}$ on the subspace spanned by $e_i$—reads

$$\partial_i \Phi(x) = \int_\Omega \varphi'(|\nabla \widetilde{W} x|) \, (\nabla \widetilde{w}_i)^T \, \frac{\nabla \widetilde{W} x}{\left|\nabla \widetilde{W} x\right|} \, ds + \varphi'(0) \int_{\Omega_x} |\nabla \widetilde{w}_i| \, ds \times [-1, 1], \tag{58}$$

where $\Omega_x$ is the complement of the support of $\nabla \widetilde{W} x$ and we systematically define

$$\frac{\nabla u}{|\nabla u|} = 0 \quad \text{if} \quad \nabla u = 0.$$

A detailed analysis of these formulas is contained in [2]. Notice that the second term in the expression for $\partial_i \Phi(x)$ is null if $t \to \varphi(|t|)$ is smooth since then $\varphi'(0) = 0$.

## 5.1 Bounds for $\boldsymbol{\lambda_i}$

Some orientations for the choice of $\lambda_i$ can be derived from the conditions for minimum (54)-(57). The next proposition gives a bound for $\lambda_i$ over which the method boils down to a hard thresholding of the coefficients.

**Proposition 1** *Let $\varphi : \mathrm{R}_+ \to \mathrm{R}_+$ be convex, differentiable and satisfy (22). If $\lambda_i > \int_\Omega |\nabla \widetilde{w}_i| ds$ then for every $y$ we have*

$$\hat{x}[i] = \begin{cases} y[i] & \text{if} \quad i \in I_1 \\ 0 & \text{if} \quad i \in I_0 \end{cases}$$

**Proof.** From the assumptions on $\varphi$,

$$\varphi'(t) \leq 1, \quad \forall t \geq 0. \tag{59}$$

Hence the first term in (58) satisfies

$$\left| \int_\Omega \varphi'(|\nabla \widetilde{W} x|) \, (\nabla \widetilde{w}_i)^T \, \frac{\nabla \widetilde{W} x}{\left|\nabla \widetilde{W} x\right|} \, ds \right| \leq \int_\Omega \big|\nabla \widetilde{w}_i\big| ds, \quad \forall x.$$

Using (58) yet again,

$$\sup \big\{ |g| : g \in \partial_i \Phi(x), \forall x \big\} \leq \int_\Omega \big|\nabla \widetilde{w}_i\big| ds.$$

It then follows that for every $x$,

$$g \in \partial_i \Phi(x) \quad \Rightarrow \quad |g| < \lambda_i.$$

The conclusion is obtained with the aid of (54)-(57). ◇

Conversely, if we wish that the coefficient $\hat{x}[i]$ can be restored at least for some noisy data, it is necessary that

$$\lambda_i \leq \int_\Omega |\nabla \widetilde{w}_i| ds, \quad \forall i \in I_0 \cup I_1. \tag{60}$$

A qualitative argument explained next gives rise to a lower bound for $\lambda_i$. The minimizer $\hat{x}$ of $F_y$ can also be expressed as

$$\hat{x} = \lim_{t \to \infty} x_t$$

where for every $t > 0$,

$$\frac{dx_t}{dt} \in -\partial F_y(x_t) \tag{61}$$

and $x_0 = \sum_{i \in I_1} y[i] e_i$. We focus on a restricted regions of the signal or the image $\sum_{i \in I_1} y[i] \widetilde{w}_i$ such that there is an isolated outlier, say $y[k] = \delta_0 > T$. Locally, we can then assume that $\nabla \widetilde{W} x_0 \approx \delta_0 \nabla \widetilde{w}_k$. By the continuity of the evolution scheme (61), $x_0[k] = \delta_0$ and $x_t[k]$ will be progressively smoothed as far as $t$ increases. It is reasonable to require that this evolution does not affect the neighboring coefficients $x_t[i]$ (where $i$ is such that $\mathrm{supp}\widetilde{w}_i \cap \mathrm{supp}\widetilde{w}_k$ is significant) since they are not outliers: we simply wish to avoid the situation seen in Fig. 1(a). This means that $0 \in \partial_i F_y(x_t)$, so that

$$\begin{aligned} \lambda_i + \varphi'(0) \int_{\Omega_x} |\nabla \widetilde{w}_i| \, ds &\geq \left| \int_\Omega \varphi'(\nabla \widetilde{W} x_t) \left(\nabla \widetilde{w}_i\right)^T \frac{\nabla \widetilde{W} x_t}{\left|\nabla \widetilde{W} x_t\right|} \, ds \right| \\ &\approx \left| \int_\Omega \varphi'(\delta_t |\nabla \widetilde{w}_k|) \left(\nabla \widetilde{w}_i\right)^T \frac{\nabla \widetilde{w}_k}{|\nabla \widetilde{w}_k|} \, ds \right|. \end{aligned}$$

While $\delta_t$ is large enough, we can say that $\varphi'(\delta_t |\nabla \widetilde{w}_k|) \approx 1$. This suggests it is reasonable to require that

$$\left| \int_\Omega \left(\nabla \widetilde{w}_i\right)^T \frac{\nabla \widetilde{w}_k}{|\nabla \widetilde{w}_k|} \, ds \right| \leq \lambda_i. \tag{62}$$

Based on (60) and (62), parameters $\lambda_i$ will be chosen in such a way that

$$\left| \int_\Omega \nabla \widetilde{w}_i^T \frac{\nabla \widetilde{w}_k}{|\nabla \widetilde{w}_k|} \, ds \right| \leq \lambda_i \leq \int_\Omega |\nabla \widetilde{w}_i| \, ds, \quad \forall k \in I \setminus \{i\}. \tag{63}$$

**Parameters when $\{w_k\}$ is a wavelet basis.** We consider henceforth a wavelet basis generated by $2^d - 1$ mother wavelets $w^m$, for $m \in \{1, \ldots, 2^d - 1\}$, with dual wavelets $\widetilde{w}^m$ and defined on $\Omega \subset \mathrm{R}^d$. Let $j$ and $\kappa$ denote the scale and the space (or time) parameters, respectively. In such a case, $I$ is an arrangement of all indexes $(j, \kappa, m)$ and $\widetilde{w}_k$ is of the form

$$\widetilde{w}^m_{j,\kappa}(s) = 2^{-\frac{dj}{2}} \widetilde{w}^m(2^{-j} s - \kappa).$$

Using a change of variables, the upper bound in (63) is

$$\int_\Omega |\nabla(\widetilde{w}^m_{j,\kappa})| \, ds = 2^{\left(\frac{d}{2}-1\right)j} \int_\Omega |\nabla \widetilde{w}^m| \, ds.$$

This suggests we take, for $i \in \{0, 1\}$,

$$\lambda^m_{j,\kappa} = 2^{\left(\frac{d}{2}-1\right)j} \lambda^m_i, \quad \forall (j, \kappa, m) \in I_i,$$

where

$$\left|\int_\Omega (\nabla(\widetilde{w}_{k,j}^{m'}))^T \frac{\nabla \widetilde{w}^m}{|\nabla \widetilde{w}^m|}\, ds\right| \leq \lambda_i^m \leq \int_\Omega |\nabla \widetilde{w}^m(s)|\, ds, \quad \forall (j,\kappa,m') \neq (0,0,m).$$

More generally, for any frame $\{\widetilde{w}_i\}_{i\in J}$, the parameters $\{\lambda_i\}$ can be chosen proportionally to $\int_\Omega |\nabla \widetilde{w}_i(s)|\, ds$.

We have observed that the minimizers $\hat{x}$ of $F_y$ are very stable with respect to the choice of the parameters $\{\lambda_i\}$. This can be explained by the fact that since $F_y$ is nonsmooth, minimizers $\hat{x}$ are located at "kinks" which are stable with respect to parameters and data.

## 5.2 An analytical example

Let $\mathbb{1}$ be a constant vector. Suppose that on a neighborhood of the index $k$, our input data $y_T$, obtained by (13), are of the form

$$y_T = W\mathbb{1} + \delta e_k, \tag{64}$$

where $y[k] = \delta > T$ is an outlier. The function denoised by hard-thresholding is

$$\widetilde{W} y_T = \widetilde{W} W \mathbb{1} + \delta \widetilde{W} e_k = \mathbb{1} + \delta \widetilde{w}_k. \tag{65}$$

Clearly, it contains an artifact with the shape of $\widetilde{w}_k$. This artifact can be suppressed if we choose the total-variation regularization, as seen below.

**Proposition 2** *Let $F_y$ be as in (14) where $\varphi(t) = t$ and for every $i \in I$, (60) is satisfied. Consider that $y_T$ reads as in (64). Then $\mathcal{F}_y$ reaches its minimum at*

$$\hat{x} = W\mathbb{1}. \tag{66}$$

**Proof.** For every $i \in I$ we have

$$\partial_i \Phi(\hat{x}) = \int_\Omega |\nabla \widetilde{w}_i| ds \times [-1, 1].$$

For every $i \neq k$ we can choose $g = 0$, then $g \in \partial_i \Phi(\hat{x})$ and $|g| \leq \lambda_i$, which show that (54) and (56) hold. Applying (60) for $\lambda_k$ shows that (55) holds as well. Hence $0 \in \partial_i F_y(\hat{x})$, for all $i \in I$. Then $F_y$ reaches its minimum at $\hat{x}$. ◇

When $\varphi'(0) = 0$ it is easy to see that $\partial_i \Phi(W\mathbb{1}) = 0$, for all $i \in I$ which shows that $\partial_k \Phi(W\mathbb{1}) \neq \lambda_k$ and hence $\hat{x} \neq W\mathbb{1}$. Nevertheless, the artifact in (65) can be smoothed arbitrarily well if $\varphi$ is a good edge-preserving function. We will show that under mild conditions, the minimizer $\hat{x}$ of $F_y$ reads

$$\hat{x} = W\mathbb{1} + \varepsilon \widetilde{w}_k, \tag{67}$$

where $\varepsilon$ satisfies

$$\int_\Omega \varphi'(|\varepsilon\, \nabla \widetilde{w}_k|)\, |\nabla \widetilde{w}_k|\, ds = \lambda_k. \tag{68}$$

It is important to notice that we have $\varepsilon \approx 0$ if $\varphi$ has a steep increase near to zero which is the case for edge-preserving functions $\varphi$. E.g., when $\varphi$ is of the form (17), we have $\varepsilon = \sqrt{\alpha/C}$ where $C$ is the unique solution of the equation

$$\int_\Omega \frac{|\nabla \widetilde{w}_k|^2}{\sqrt{|\nabla \widetilde{w}_k|^2 + C}}\, ds = \lambda_k.$$

Clearly, $\varepsilon$ decreases to zero when $\alpha \searrow 0$. For $\varphi$ of the form (16), the equation in (68) yields

$$\alpha \varepsilon^{\alpha-1} \int_\Omega |\nabla \widetilde{w}_k|^\alpha = \lambda_k.$$

Then

$$\varepsilon = \left( \frac{\lambda_k}{\alpha \int_\Omega |\nabla \widetilde{w}_k|^\alpha} \right)^{\frac{1}{\alpha-1}}.$$

If we choose that the inequality in (60) is strict, then the term between the parentheses is strictly smaller than 1. Then $\varepsilon \searrow 0$ if $\alpha \searrow 1$.

The assumption mentioned above is that

$$\left| \int_\Omega \varphi'(|\varepsilon \nabla \widetilde{w}_k|) (\nabla \widetilde{w}_i)^T \frac{\nabla \widetilde{w}_k}{|\nabla \widetilde{w}_k|} \, ds \right| \lessapprox \left| \int_\Omega (\nabla \widetilde{w}_i)^T \frac{\nabla \widetilde{w}_k}{|\nabla \widetilde{w}_k|} \, ds \right|, \quad \forall i \neq k, \tag{69}$$

which is realist since $\varphi'(t) \lessapprox 1$ when $t$ is beyond a restricted neighborhood of zero. Let us now verify that $\hat{x}$ as given in (67) does indeed minimize $F_y$. We have

$$\partial_i \Phi(\hat{x}) = \int_\Omega \varphi'(|\varepsilon \nabla \widetilde{w}_k|) (\nabla \widetilde{w}_i)^T \frac{\nabla \widetilde{w}_k}{|\nabla \widetilde{w}_k|} \, ds.$$

Combining (69) and (63) shows that

$$|\partial_i \Phi(\hat{x})| \leq \lambda_i, \quad \forall i \neq k,$$

so that (54) and (56) hold. The $k$th entry of $\hat{x}$ is $\hat{x}[k] = \varepsilon$ since (68) is in fact the condition given in (55). Hence $F_y$ reaches its minimum at $\hat{x}$ as given in (67).

# 6 Minimization schemes

For practical calculation, the signal or image $u$ is defined on a discrete finite grid, so we can consider that $u \in \mathrm{R}^p$. The discrete equivalent of the regularization term $\varphi$ in (15) reads

$$\Phi(x) = \sum_{i=1}^{p} \varphi\big(|\nabla_i \widetilde{W} x|\big), \tag{70}$$

where for every $u \in \mathrm{R}^p$, $\nabla_i u \in \mathrm{R}^2$ is a discrete approximation of the gradient of $u$ at $i$ if $\big(\nabla_i u\big)[1]$ is the difference between the pixel $u[i]$ and its north adjacent neighbor and $\big(\nabla_i u\big)[2]$ is the difference between $u[i]$ and its left adjacent neighbor. When $u$ is a signal, we can formally write that $\big(\nabla_i u\big)[2] = 0$, for all $i$. Let us also mention that the norm $|.|$ in (70) is defined as

$$|z| = \sqrt{z[1]^2 + z[2]^2}, \quad \forall z \in \mathrm{R}^2. \tag{71}$$

Then discrete equivalent of (58) is

$$\partial_i \Phi(x) = \sum_j \varphi'(|\nabla_j \widetilde{W} x|) \big(\nabla_j \widetilde{w}_i\big)^T \frac{\nabla_j \widetilde{W} x}{|\nabla_j \widetilde{W} x|}.$$

The function $F_y$ in (14) is non-smooth and several approaches can be envisaged for its minimization. The subgradient descent scheme is quite easy to implement. Put $x_0 = y_T$ and, for all $k \in \mathbf{N}$, compute

$$x_{k+1} = x_k - t_k g_k,$$

where $g_k$ is a subgradient of $F_y$ at $x_k$ and $t_k > 0$. Using classical results on minimization methods (see [43]), we can prove that, if $\lim_{k\to\infty} t_k = 0$ and $\sum_{k=0}^{\infty} t_k = \infty$, then

$$\lim_{k\to\infty} x_k = \hat{x}.$$

A better alternative is the adaptive level-set method proposed in [20].

Instead, we focus on a relaxation-based method, proposed in [40], which is properly adapted to $\ell_1$ data-fidelity terms as those involved in (14), even if it requires that the regularization $\Phi$ is smooth. The main interest of this method is the facility to recover the components $i$ of $I_1$ such that $\hat{x}_i = y_i$ and the components $i$ of $I_0$ satisfying $\hat{x}[i] = 0$, as well as the relative easiness of implementation.

Let $x^0 = y_T$ be the starting point. As each iteration $k = 1, 2, ...$, the new iterate $x^k$ is obtained from $x^{k-1}$ by updating successively each one of its components $x^k[i]$ using one-dimensional minimization. Let the solution obtained at step $i-1$ of iteration $k$ read

$$(x^k[1], \ldots, x^k[i-1], x^{k-1}[i], \ldots, x^{k-1}[p]).$$

The new entry $x^k[i]$ is determined according to the following rule:

- if $i \in I_1$, compute

$$K = \partial_i \Phi\big(x^k[1], \ldots, x^k[i-1],\ y[i], x^{k-1}[i+1],\ \ldots, x^{k-1}[p]\big) \tag{72}$$

  - if $|K| < \lambda_i$ then $x^k[i] = y[i]$
  - else $x^k[i]$ is the unique solution of the equation in $t$

$$\partial_i \Phi\big(x^k[1], \ldots, x^k[i-1],\ t, x^{k-1}[i+1],\ \ldots, x^{k-1}[p]\big) = \lambda_i \text{sign}(K) \tag{73}$$

    where we know that $\text{sign}(x^k[i] - y[i]) = -\text{sign}(K)$

- if $i \in I_0$, compute

$$K = \partial_i \Phi\big(x^k[1], \ldots, x^k[i-1],\ 0, x^{k-1}[i+1],\ \ldots, x^{k-1}[p]\big) \tag{74}$$

  - if $|K| < \lambda_i$ then $x^k[i] = 0$
  - else $x^k[i]$ is the unique solution of the equation

$$\partial_i \Phi\big(x^k[1], \ldots, x^k[i-1],\ t, x^{k-1}[i+1],\ \ldots, x^{k-1}[p]\big) = \lambda_i \text{sign}(K) \tag{75}$$

    where it is known that $\text{sign}(x^k[i]) = -\text{sign}(K)$

Observe that the components $x[i]^k$ that fit exactly the data (i.e. $x[i]^k = y[i]$ for $i \in I_1$ and $x[i]^k = 0$ for $i \in I_0$) are easily found by checking only the sign of an inequality. Their computation is hence very accurate. Moreover, in practice most of the components of the solution fit exactly the data-fidelity term. Conversely, at each iteration, most of the components are easily and accurately found. The other components, corresponding to (73) and (75), are obtained using one-dimensional line-search. Here, the knowing of the sign of $\text{sign}(x^k[i] - y[i])$ and $\text{sign}(x^k[i])$ constitutes an important simplification. The finding of these signs uses the fact that since $\Phi$ is convex, the function $S_i$ given below

$$S_i(t) = \partial_i \Phi\big(x^k[1], \ldots, x^k[i-1],\ t, x^{k-1}[i+1],\ \ldots, x^{k-1}[p]\big), \quad \forall t \in \mathrm{R},$$

is monotonous increasing on R. Consider that $x^k[i]$, $i \in I_1$, has to satisfy (46). Then $\hat{x}[i] \neq y[i]$. Consider first that

$$x^k[i] > y[i],$$

in which case $S_i(x^k[i]) \geq S_i(y[i])$. The equation in (46) now reads $\lambda_i + S_i(x^k[i]) = 0$, hence we can write

$$S_i(y[i]) \leq S_i(x^k[i]) < 0.$$

The result follows from the observation that $K$ in (72) satisfies $K = S_i(y[i])$. The case when $x^k[i] < y[i]$ in established in a symmetric way. The reasoning in the case when $x^k[i]$, $i \in I_0$, has to satisfy (75) is basically the same, so we skip it. The convergence of $x^k$ towards $\hat{x}$ has been established in [40].

If $\Phi$ corresponds to a TV regularization, we have to take a smooth approximation of it—for instance $\varphi(t) = \sqrt{t^2 + \alpha}$ for $\alpha \gtrapprox 0$—in order to apply the method above. An alternative that allows to use TV regularization without smooth approximation is the method proposed in [26, 13]. Let us notice that the implementation of the last method is tricky.

# 7 Experiments

## 7.1 Denoising of a signal

We consider the restoration of the 512-length original signal in Fig. 2 from the data shown there, contaminated with white Gaussian noise with standard deviation $\sigma = 10$. The restoration in Fig. 3 is obtained using the sure-shrink method [23] and the toolbox WaveLab. The result displayed in Fig. 4 is the minimizer of a function $\mathcal{F}_v$ of the form (1) where $\varphi$ is as given in (17), for $\alpha = 0.1$ and $\lambda = 0.01$. Smooth zones are rough, edges are slightly smoothed and spikes are eroded, while some diffused noise is still visible on the signal.

The restorations presented next are based on wavelet coefficients where $W$ is an orthogonal basis of Daubechies wavelets with 8 vanishing moments and thresholded data $y_T$ are obtained according to (13). The optimal $T$, as given in (5), reads $T = 35$. The wavelet-thresholding estimate $\widetilde{W}y_T$ is shown in Fig. 5. It involves important Gibbs artifacts, as well as wavelet-shaped oscillations due to aberrant coefficients. Using the same coefficients $y_T$, we calculated the minimizer $\hat{x}$ of $F_y$ as given in (14) where $\varphi$ is as given in (17), $\alpha = 0.05$, $\lambda_{j,\kappa} = 0.5 \times 2^{-j/2}$ if $(j, \kappa) \in I_0$ and $\lambda_{j,\kappa} = 1.5 \times 2^{-j/2}$ if $(j, \kappa) \in I_1$. The resultant restoration $\hat{u} = \widetilde{W}\hat{x}$, shown in Fig. 6, involves sharp edges and well denoised smooth pieces.

The noisy signal $v$ is also restored by translation invariant wavelet thresholding with optimal threshold $T = 35$. The obtained result is displayed in Fig. 7. Comparing to the signal restored with decimated wavelets shown in Fig. 5, we observe that wavelet shaped artifacts and Gibbs oscillations are reduced, but they remain still well visible.

We show in Fig. 8 the minimizer obtained with the model given in (7) for $\mu_i = 30$, $\forall i \in J$. Recall that for large $\mu_i$, spikes are oversmoothed, while if $\mu_i$ is too small, aberrant coefficients are not properly restored. Both artifacts are visible on the displayed signal which corresponds to an intermediate choice for $\mu_i$.

Next we consider $y_T$, obtained by (13) for $T = 23$. These coefficients have a richer information content, but the relevant estimate $\widetilde{W}y_T$, seen in Fig. 9, manifests Gibbs artifacts and many wavelet-shaped artifacts. Below we compare restorations where $F_y$ is of the form (44) for different choices of $\psi_i$. In spite of the considerations developed in section 4, it seems intuitive to take $\psi_{j,\kappa}(t) = \lambda_{j,\kappa}t^2$ in (44). Such

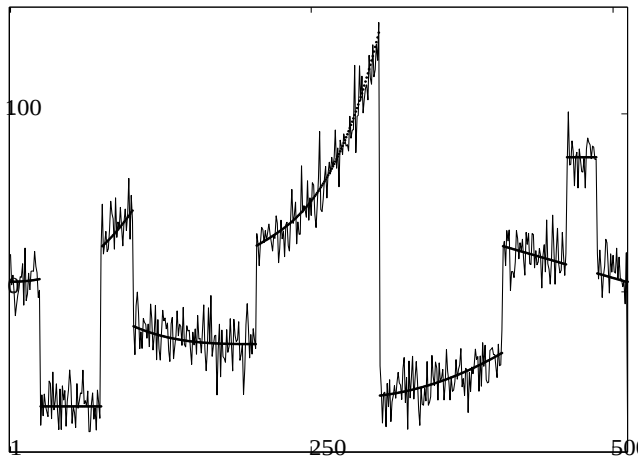



Figure 2: Original signal (dotted line) and noisy data (solid line).

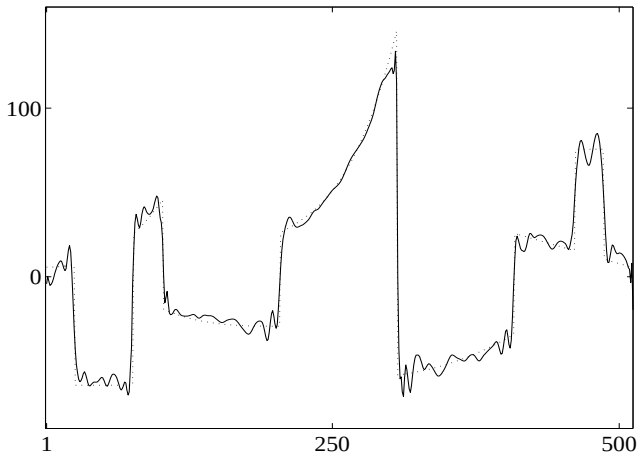



Figure 3: Denoising using the Donoho-Johnstone's *Sure-shrink* method.

a restoration is displayed in Fig. 10 where $\alpha = 0.05$ $\lambda_{j,\kappa} = 0.1$ if $(j,\kappa) \in I_0$, and $\lambda_{j,\kappa} = 0.2$ if $(j,\kappa) \in I_1$. The Gibbs oscillations are well removed but, because of the quadratic form of $\psi_{j,\kappa}$ for $(j,\kappa) \in I_1$, outliers overcontribute to $F_y$ and biases the estimate. Another possibility which may seem reasonable is to cancel the term indexed by $I_0$, i.e. to consider $\psi_{j,\kappa}(t) = 0$ for $(j,\kappa) \in I_0$. The result can be seen in Fig. 11 where $\psi_{j,\kappa}(t) = 0.2\,t$ for all $(j,\kappa) \in I_1$ and $\alpha = 0.05$. Once again, the thresholded coefficients are well restored but we observe that leaving too much freedom to these coefficients prevents the method from removing the outliers efficiently. Fig. 12 illustrates the proposed method: $F_y$ is of the form (14) with $\varphi$ as given in (17), and the same parameters as in Fig. 6, namely $\alpha = 0.05$, $\lambda_{j,\kappa} = 0.5 \times 2^{-j/2}$ if $(j,\kappa) \in I_0$ and $\lambda_{j,\kappa} = 1.5 \times 2^{-j/2}$ if $(j,\kappa) \in I_1$. In this restoration, edges are neat and polynomial parts are well recovered. Fig. 10 illustrates how restored coefficients $\hat{x}$ are placed with respect to $y_T$ and the coefficients of the original signal $Wu_o$. In particular, we observe how erroneously thresholded coefficients are restored and how outliers are smoothed.

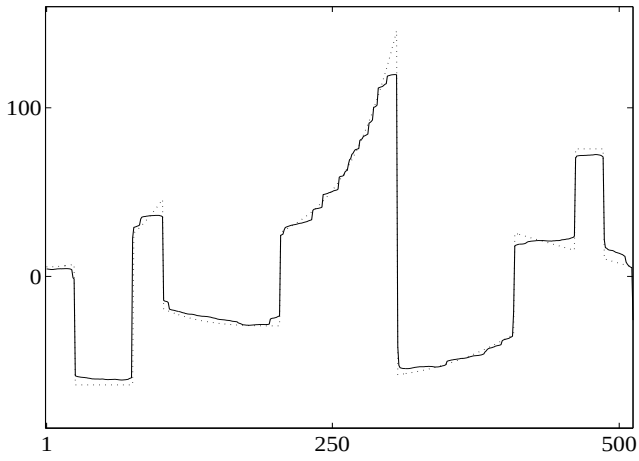



Figure 4: Denoising by minimizing $\mathcal{F}_v$ as given in (1) where $\varphi(t) = \sqrt{0.05 + t^2}$ and $\lambda = 0.01$.

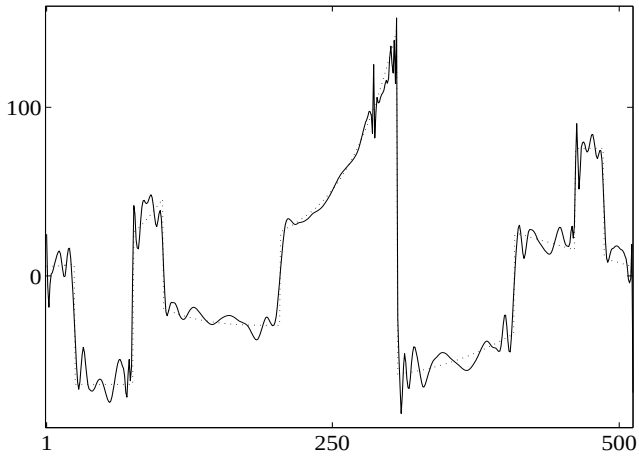



Figure 5: Denoising using wavelets thresholding with Donoho-Johnstone's optimal threshold $T = 35$.

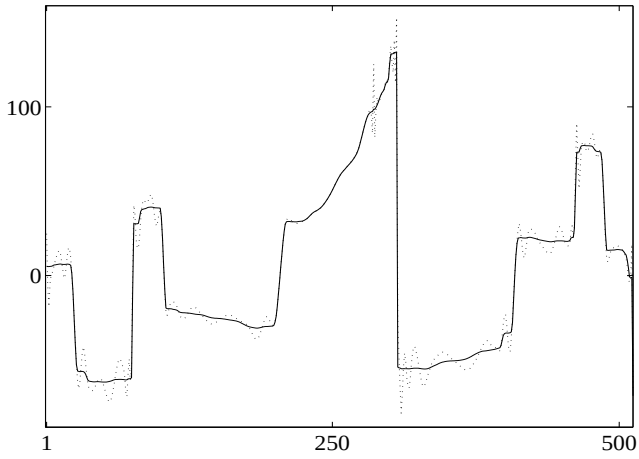



Figure 6: Denoising by restoration of the wavelet coefficients relevant to Fig. 5 using $F_y$ in (14) with $\varphi(t) = \sqrt{0.05 + t^2}$, $\lambda_{j,\kappa} = 0.5 \times 2^{-j/2}$ if $(j,\kappa) \in I_0$, $\lambda_{j,\kappa} = 1.5 \times 2^{-j/2}$ if $(j,\kappa) \in I_1$.

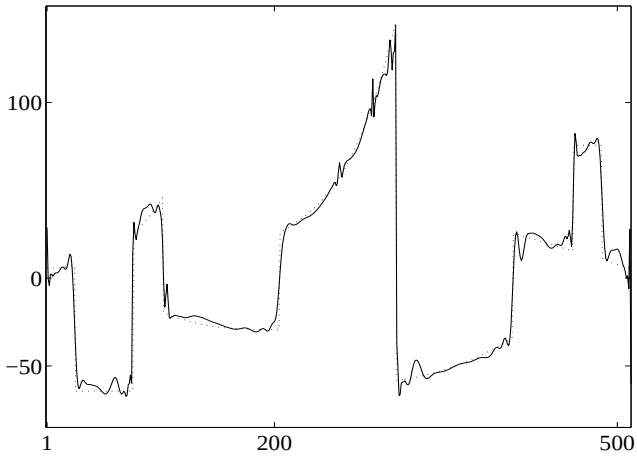



Figure 7: Denoising by translation invariant wavelets thresholding with Donoho-Johnstone's optimal threshold $T = 35$.

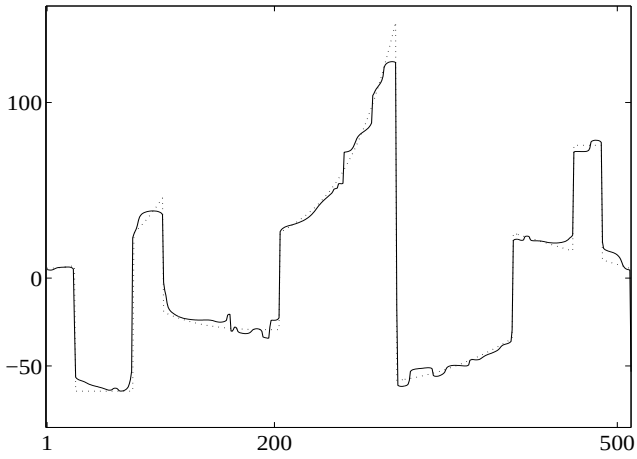



Figure 8: Denoising using (7) with $\varphi(t) = \sqrt{0.05 + t^2}$ and $\mu_i = 30$.

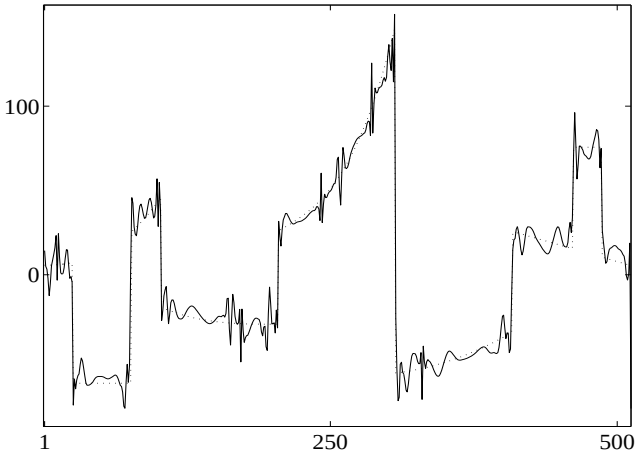



Figure 9: Denoising using wavelets thresholding with an under-optimal threshold $T = 23$.

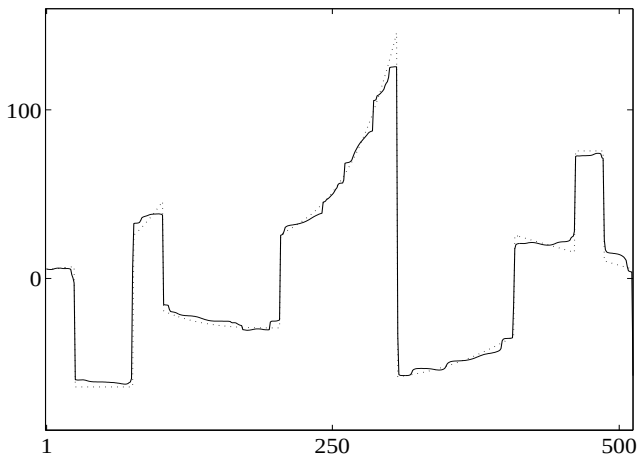



Figure 10: Restoration of the wavelet coefficients relevant to Fig. 9 by minimizing $F_y$ in (44) with $\varphi(t) = \sqrt{0.05 + t^2}$, $\psi_i(t) = 0.1t^2$ if $i \in I_0$ and $\psi_i(t) = 0.2t^2$ if $i \in I_1$.

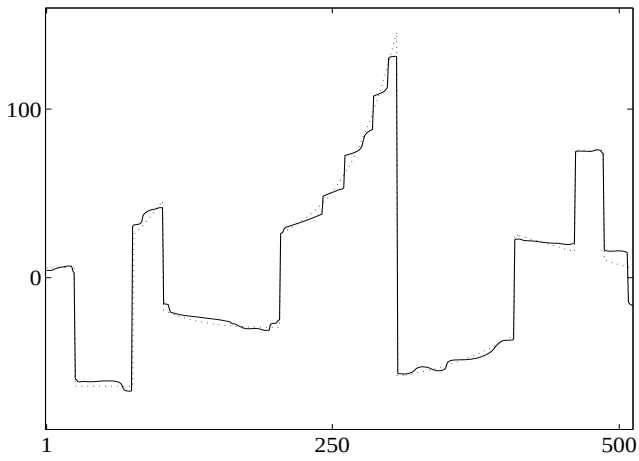



Figure 11: Restoration of Fig. 9 using $F_y$ in (44) where $\varphi(t) = \sqrt{0.05 + t^2}$, $\psi_i(t) = 0$ if $i \in I_0$ and $\psi_i(t) = 0.2t$ if $i \in I_1$.

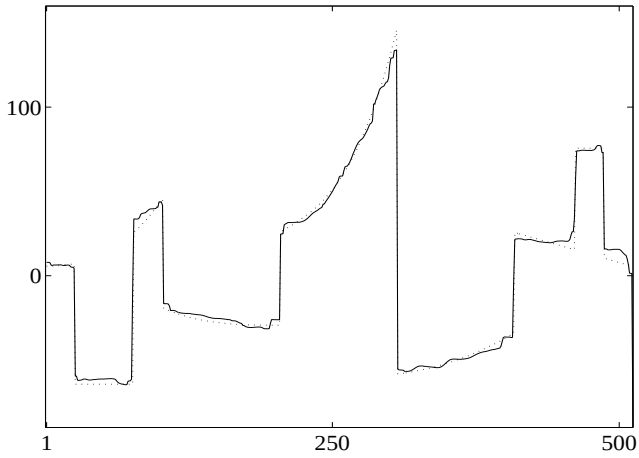



Figure 12: The proposed method: restoration of Fig. 9 using $F_y$ in (14) with $\varphi(t) = \sqrt{0.05 + t^2}$, $\lambda_{j,\kappa} = 0.5 \times 2^{-j/2}$ if $(j,\kappa) \in I_0$ and $\lambda_{j,\kappa} = 1.5 \times 2^{-j/2}$ if $(j,\kappa) \in I_1$.

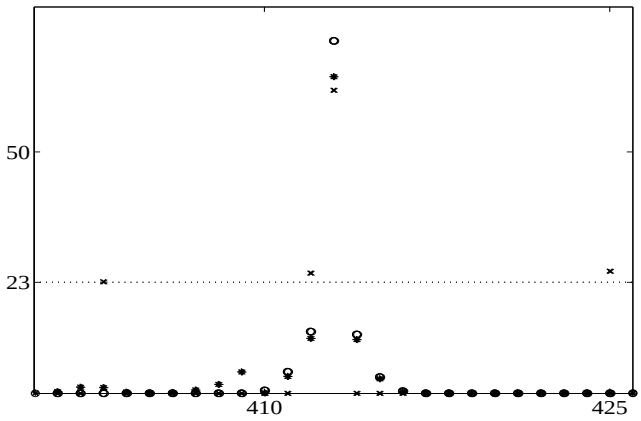



Figure 13: Magnitude of wavelet coefficients: $*$ signal restored by the proposed method (Fig. 9), $\circ$ original signal, $\times$ thresholded noisy signal (Fig. 6).

### 7.2 Denoising of an image

In this experiment we consider the denoising of the $256 \times 256$ picture of Lena, Fig. 14 (a), from noisy data obtained by adding white Gaussian noise with standard deviation 20. The restoration in Fig. 15 (a) is obtained by thresholding the wavelet coefficients, see (13), with respect to Donoho-Johnstone's threshold, given in (5), which now reads $T = 100$. This image is very smooth, a lot of details are lost, and Gibbs oscillations are visible near the edges. In Fig. 15 (b) we show the result from total-variation restoration which corresponds to $\mathcal{F}_v$ of the form (1) with $\varphi(t) = t$ and $\lambda = 0.03$. As expected, this restoration exhibits a stair-casing effect since it is constant on many regions. The image in Fig. 16 (a) is obtained by thresholding the wavelet coefficients with respect to $T = 50$. This $T$ is smaller than Donoho-Johnstone's threshold and the image presents many wavelet-shaped oscillations due to aberrant wavelet coefficients, as well as some Gibbs oscillations. It is used as input data for the specialized objective function $F_y$ given in (14), where $\varphi$ is as given in (17). The restoration in Fig. 16 (b) is obtained for $\lambda_i = 0.5$ if $i \in I_0$, $\lambda_i = 1.5$ if $i \in I_0$. This image has a quite natural appearance, and edges and texture are better preserved.

The numerical cost of variational methods become a real burden when images have a large size. In order to circumvent this problem, we have tested an approximation of the proposed method. Let $y_T$ be the wavelet transform of the thresholded image. According to (45), the minimizer $\hat{x}$ of $F_y$ satisfy

$$|\partial_i \Phi(\hat{x})| \leq \lambda_i, \quad \forall i \in I_1.$$

The idea of this approximation is to test for every $i \in I_1$ whether or not $|\partial_i \Phi(y_T)| > \lambda_i$. If $|\partial_i \Phi(y_T)| \leq \lambda_i$, we take simply $\hat{x}[i] = y_T[i]$. Otherwise, if $|\partial_i \Phi(y_T)| > \lambda_i$, we consider that $y_T[i]$ is an outlier. To restore such an outlier, we can take for the relevant $\hat{x}[i]$ either the median or the mean of the neighboring coefficients at the same scale. When outliers arise in homogeneous regions, we can just set $\hat{x}[i] = 0$. The Gibbs oscillations are not considered in this approximated method, so we have $\hat{x}[i] = y_T[i] = 0$ for all $i \in I_0$. The image obtained by this method for $T = 50$ and $\lambda_i = 5$ for all $i \in I_1$, is displayed on Fig. 17 (a). Let us emphasize that the image of the error $v_\tau - \hat{u}$, presented in Fig. 17 (b), exhibits the oscillations due to aberrant wavelet coefficients and that it does not present any structural information. This approximated method being computationally fast, it can be extended to translation invariant wavelets [18]. In Fig. 18 (a) we show the restoration obtained by the standard translation invariant wavelets thresholding, corresponding to $T = 50$ again. Although its quality is improved with respect to the image in Fig. 16 (a), it involves a lot of wavelet-shaped artifacts. This image is used as input data to our fast approximated method. The obtained restoration, shown in Fig. 18 (b), is of high quality, since edges and details are nicely recovered.

## 8 Conclusion

We proposed a method to denoise images and signals by restoring the thresholded frame coefficients of the noisy data. The restored coefficients minimize a specially designed objective function which allows the erroneously thresholded coefficients to be restored and the outliers to be removed, without substantially modifying the remaining coefficients. Our method is not sensitive to the probability distribution of the noise. We present numerical experiments with orthogonal bases of Daubechies wavelets. These experiments demonstrate the effectiveness of our method over alternative denoising methods.

(a) Original image. (b) Noisy image.

Figure 14: Original and noisy images.

(a) Wavelets thresholding with the optimal thereshold $T = 100$. (b) Total-variation restoration: $\mathcal{F}_v$ as in (1) with $\varphi(t) = t$.

Figure 15: Classical denoising methods.

(a) Wavelets thresholding with $T = 50$ (b) Restoration of (a) by the proposed method ($\varphi(t) = t$)

Figure 16: The proposed method.

(a) Restoration of Fig. 16 (a) by the fast method

(b) Outliers detected by the fast method

Figure 17: Fast approximation of the proposed method.

(a) Translation invariant wavelet thresholding ($T = 50$)

(b) Fast method adapted to translation invariant wavelets

Figure 18: Translation invariant wavelets.

# References


[1] R. Acar and C. Vogel, *Analysis of bounded variation penalty methods for ill-posed problems*, IEEE Transactions on Image Processing, 10 (1994), pp. 1217–1229.

[2] F. Andreu, V. Caselles, and J. M. Mazón, *Parabolic Quasilinear Equations Minimizing Linear Growth Functionals*, Progress in Mathematics 223, Birkhauser Verlag, 2004.

[3] A. Antoniadis and J. Fan, *Regularization of wavelet approximations*, Journal of Acoustical Society America, 96 (2001), pp. 939–967.

[4] A. Antoniadis, D. Leporini, and J.-C. Pesquet, *Wavelet thresholding for some classes of non-gaussian noise*, Statistica Neerlandica, 56 (2002), pp. 434–453.

[5] G. Aubert and P. Kornprobst, *Mathematical problems in images processing*, Springer-Verlag, Berlin, 2002.

[6] M. Belge, M. Kilmer, and E. Miller, *Wavelet domain image restoration with adaptive edge-preserving regularization*, IEEE Transactions on Image Processing, 9 (2000), pp. 597–608.

[7] J. E. Besag, *On the statistical analysis of dirty pictures (with discussion)*, Journal of the Royal Statistical Society B, 48 (1986), pp. 259–302.

[8] ———, *Digital image processing : Towards Bayesian image analysis*, Journal of Applied Statistics, 16 (1989), pp. 395–407.

[9] 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, 19 (1996), pp. 57–91.

[10] Y. Bobichon and A. Bijaoui, *Regularized multiresolution methods for astronomical image enhancement*, Exper. Astron., (1997), pp. 239–255.

[11] C. Bouman and K. Sauer, *A generalized Gaussian image model for edge-preserving* map *estimation*, IEEE Transactions on Image Processing, 2 (1993), pp. 296–310.

[12] E. J. Candès and F. Guo, *New multiscale transforms, minimum total variation synthesis. Applications to edge-preserving image reconstruction*, Signal Processing, 82 (2002).

[13] E. J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, IEEE Transactions on Information Theory, 52 (2006), pp. 489 – 509.

[14] A. Chambolle, R. DeVore, N.-y. Lee, and B. Lucier, *Nonlinear wavelet image processing: variational problems, compression, and noise removval through wavelet shrinkage*, IEEE Transactions on Image Processing, 7 (1998), pp. 319–335.

[15] T. Chan and H. Zhou, *Total variation improved wavelet thresholding in image compression*, in Proceedings of the IEEE International Conference on Image Processing, vol. 2, IEEE, 2000, pp. 391–394.

[16] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, *Deterministic edge-preserving regularization in computed imaging*, IEEE Transactions on Image Processing, 6 (1997), pp. 298–311.

[17] A. Cohen, R. DeVore, P. Petrushev, and H. Xu, *Nonlinear approximation and the space $BV(\mathbb{R}^2)$*, American Journal of Mathematics, 121 (1999).

[18] R. R. Coifman and D. Donoho, *Translation-invariant de-noising*, Tech. Report Report 475, Stanford University, Dept. of Statistics, 1995.

[19] R. R. Coifman and A. Sowa, *Combining the calculus of variations and wavelets for image enhancement*, Applied and Computational Harmonic Analysis, 9 (2000).

[20] P. Combettes and J. Luo, *An adaptive level set method for nondifferentiable constrained image recovery*, IEEE Transactions on Image Processing, 11 (2002), pp. 1295–1304.

[21] G. Demoment, *Image reconstruction and restoration : Overview of common estimation structure and problems*, IEEE Transactions on Acoustics Speech and Signal Processing, ASSP-37 (1989), pp. 2024–2036.

[22] D. L. Donoho and I. M. Johnstone, *Ideal spatial adaptation by wavelet shrinkage*, Biometrika, 81 (1994), pp. 425–455.

[23] ———, *Adapting to unknown smoothness via wavelet shrinkage*, Journal of Acoustical Society America, 90 (1995).

[24] S. Durand and J. Froment, *Reconstruction of wavelet coefficients using total variation minimization*, SIAM Journal on Scientific Computing, 24 (2003), pp. 1754–1767.

[25] J. Froment and S. Durand, *Artifact free signal denoising with wavelets*, in Proceedings of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing, vol. 6, 2001.

[26] H. Fu, M. Ng, Michael K.and Nikolova, and J. L. Barlow, *Efficient minimization methods of mixed $\ell_1 - \ell_l$ and $\ell_2 - \ell_1$ norms for image restoration*, SIAM Journal on Scientific Computing, 27 (2005), pp. 1881–1902.

[27] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6 (1984), pp. 721–741.

[28] P. J. Green, *Bayesian reconstructions from emission tomography data using a modified* em *algorithm*, IEEE Transactions on Medical Imaging, MI-9 (1990), pp. 84–93.

[29] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and Minimization Algorithms, vol. I*, Springer-Verlag, Berlin, 1996.

[30] D. Keren and M. Werman, *Probabilistic analysis of regularization*, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-15 (1993), pp. 982–995.

[31] S. Li, *Markov Random Field Modeling in Computer Vision*, Springer-Verlag, New York, 1 ed., 1995.

[32] F. Malgouyres, *Mathematical analysis of a model which combines total variation and wavelet for image restoration*, Journal of information processes, 2 (2002), pp. 1–10.

[33] F. Malgouyres, *Minimizing the total variation under a general convex constraint for image restoration*, IEEE Transactions on Image Processing, 11 (2002), pp. 1450–1456.

[34] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, London, 1999.

[35] P. Moulin and J. Liu, *Analysis of multiresolution image denoising schemes using generalized gaussian and complexity priors*, IEEE Transactions on Image Processing, 45 (1999), pp. 909–919.

[36] P. Mrázek, J. Weickert, and G. Steidl, *Diffusion-inspired shrinkage functions and stability results for wavelet denoising*, International Journal of Computer Vision, 64 (2005).

[37] M. Nikolova, *Estimées localement fortement homogènes*, Comptes-Rendus de l'Académie des Sciences, t. 325, série 1 (1997), pp. 665–670.

[38] ———, *Local strong homogeneity of a regularized estimator*, SIAM Journal on Applied Mathematics, 61 (2000), pp. 633–658.

[39] ———, *Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers*, SIAM Journal on Numerical Analysis, 40 (2002), pp. 965–994.

[40] ———, *A variational approach to remove outliers and impulse noise*, Journal of Mathematical Imaging and Vision, 20 (2004).

[41] W. Ring, *Structural properties of solutions of total variation regularization problems*, ESSAIM: Mathematical Modelling and Numerical Analysis, 34 (2000), pp. 799–810.

[42] L. Rudin, S. Osher, and C. Fatemi, *Nonlinear total variation based noise removal algorithm*, Physica, 60 D (1992), pp. 259–268.

[43] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*, vol. 3, Springer-Verlag, 1985.

[44] E. P. Simoncelli and E. H. Adelson, *Noise removal via Bayesian wavelet coding*, in Proceedings of the IEEE International Conference on Image Processing, Lausanne, Switzerland, Sep. 1996, pp. 379–382.

[45] G. Steidl, J. Weickert, P. Mrázek, and M. Welk, *On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and sides*, SIAM Journal on Numerical Analysis, 42 (2004), pp. 686–713.

[46] A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems*, Winston, Washington DC, 1977.

[47] L. Vese, *A study in the BV space of a denoising-deblurring variational problem*, Journal of Applied Mathematics and Optimization, (2001).

[48] C. R. Vogel and M. E. Oman, *Iterative method for total variation denoising*, SIAM Journal on Scientific Computing, 17 (1996), pp. 227–238.

[49] C. R. Vogel and M. E. Oman, *Fast, robust total variation-based reconstruction of noisy, blurred images*, IEEE Transactions on Image Processing, 7 (1998), pp. 813–824.

[50] G. Wang, J. Zhang, and G.-W. Pan, *Solution of inverse problems in image processing by wavelet expansion*, IEEE Transactions on Image Processing, 4 (1995), pp. 579–593.