MULTISCALE MODEL. SIMUL.
Vol. 4, No. 3, pp. 960–991

ANALYSIS OF THE RECOVERY OF EDGES IN IMAGES AND SIGNALS BY MINIMIZING NONCONVEX REGULARIZED LEAST-SQUARES*

MILA NIKOLOVA†

Abstract. We consider the restoration of discrete signals and images using least-squares with nonconvex regularization. Our goal is to find important features of the (local) minimizers of the cost function in connection with the shape of the regularization term. This question is of paramount importance for a relevant choice of regularization term. The main point of interest is the restoration of edges. We show that the differences between neighboring pixels in homogeneous regions are smaller than a small threshold, while they are larger than a large threshold at edges: we can say that the former are shrunk, while the latter are enhanced. This naturally entails a neat classification of differences as belonging to smooth regions or to edges. Furthermore, if the original signal or image is a scaled characteristic function of a subset, we show that the global minimizer is smooth everywhere if the contrast is low, whereas edges are correctly recovered at higher (finite) contrast. Explicit expressions are derived for the truncated quadratic and the "0-1" regularization function. It is seen that restoration using nonconvex regularization is fundamentally different from edge-preserving convex regularization. Our theoretical results are illustrated using a numerical experiment.

Key words. image restoration, signal restoration, regularization, variational methods, edge restoration, inverse problems, nonconvex analysis, nonsmooth analysis, optimization

AMS subject classifications. 26B99, 35A15, 49K30, 49K99, 49N45, 49N60, 49Q12, 62H12, 90C26, 94A08, 94A12

DOI. 10.1137/040619582

1. Introduction.

We consider the classical inverse problem of the finding of an estimate $\hat{x} \in \mathbb{R}^p$ of an unknown image or signal $x \in \mathbb{R}^p$, based on data $y = Ax + n \in \mathbb{R}^q$, where $A \in \mathbb{R}^{q \times p}$ accounts for the data-acquisition system and n for the noise. For instance, A can be a point-spread function modelling optical blurring, a distortion wavelet in seismic imaging and nondestructive evaluation, a Radon transform in X-ray tomography, a Fourier transform in diffraction tomography, or the identity in denoising and segmentation problems. To solve such a problem, we focus on regularized least-squares methods where $\hat{x} \in \mathbb{R}^p$ minimizes a cost-function $\mathcal{F}_y : \mathbb{R}^p \to \mathbb{R}$ of the form

$$\mathcal{F}_y(x) = \|Ax - y\|^2 + \beta\Phi(x), \tag{1.1}$$

where Φ is a regularization term and $\beta > 0$ is a parameter which controls the trade-off between fidelity to data and regularization. Such cost-functions are classical in variational methods and in Bayesian estimation; an overview of these approaches can be found in [4, 10, 3]. In a statistical setting, the quadratic data-fidelity term above supposes that the noise n is white and Gaussian. The role of Φ is to push $\hat{x}$ to exhibit some a priori expected features, such as the presence of edges and smooth regions. A useful class of regularization functions is [4, 14, 9, 3]

$$\Phi(x) = \sum_{i \in J} \varphi(g_i^T x), \quad J = \{1, \ldots, r\}, \tag{1.2}$$

*Received by the editors November 24, 2004; accepted for publication (in revised form) March 16, 2005; published electronically September 8, 2005.
http://www.siam.org/journals/mms/4-3/61958.html

†Centre de Mathématiques et de Leurs Applications (CNRS-UMR 8536)–ENS de Cachan, 61 avenue du Président Wilson, 94235 Cachan Cedex, France (nikolova@cmla.ens-cachan.fr).

TABLE 1.1
Commonly used PFs φ where $\alpha > 0$ is a parameter.

	Convex PFs				
	Smooth at zero PFs			Nonsmooth at zero PFs	
(f1)	$\varphi(t) = \lvert t\rvert^\alpha,\ 1 < \alpha \le 2$	[6]	(f3)	$\varphi(t) = \lvert t\rvert$	[4, 36]
(f2)	$\varphi(t) = \sqrt{\alpha + t^2}$	[41]			
	Nonconvex PFs				
	Smooth at zero PFs			Nonsmooth at zero PFs	
(f4)	$\varphi(t) = \min\{\alpha t^2, 1\}$	[26, 5]	(f8)	$\varphi(t) = \lvert t\rvert^\alpha,\ 0 < \alpha < 1$	[37]
(f5)	$\varphi(t) = \dfrac{\alpha t^2}{1 + \alpha t^2}$	[18]	(f9)	$\varphi(t) = \dfrac{\alpha\lvert t\rvert}{1 + \alpha\lvert t\rvert}$	[15]
(f6)	$\varphi(t) = \log(\alpha t^2 + 1)$	[20]	(f10)	$\varphi(t) = \log(\alpha\lvert t\rvert + 1)$	
(f7)	$\varphi(t) = 1 - \exp(-\alpha t^2)$	[22, 34]	(f11)	$\varphi(0) = 0,\ \varphi(t) = 1$ if $t \neq 0$	[22]

where $g_i \in \mathbb{R}^p$, for $i \in J$, are difference operators and $\varphi : \mathbb{R} \to \mathbb{R}$ is called a potential function (PF). In the following, the letter G will denote the $r \times p$ matrix whose rows are g_i^T for $i \in J$. A basic requirement to have regularization is

$$\ker(A) \cap \ker(G) = \{0\}. \tag{1.3}$$

Many different PFs have been used in the literature; some relevant examples are given in Table 1.1. Although PFs differ in convexity, boundedness, differentiability, etc., they share some common features. A general assumption is the following.

H1. *$\varphi(t) = \varphi(-t)$, φ is $\mathcal{C}^2$ on $(0, +\infty)$, and $\varphi'(t) \ge 0$, for all $t > 0$, and $\varphi(0) = 0$ is a strict minimum.*

Edges in images and breaking points in signals concentrate critically important information. Hence we have the requirement that φ leads to minimizers $\hat{x}$ involving large differences $|g_i^T \hat{x}|$ at the location of edges in the original signal or image and smooth differences elsewhere. The very first regularized cost-function was introduced in [40] and corresponds to $\varphi(t) = t^2$; it is well known that this PF entails oversmoothing of edges. Since the pioneering work of Geman and Geman [17], different nonconvex functions φ have been considered in either a statistical or variational framework [26, 18, 4, 34, 15, 16, 23, 3]. The relevant minimizers provide solutions with neat edges and well-smoothed homogeneous regions. However, they are awkward to compute, to control, and to analyze. In order to alleviate these intricacies, a considerable effort has been made to derive *convex edge-preserving* functions φ; see, for instance, [39, 19, 21, 36, 6, 9]. These PFs have an almost linear growth beyond an interval surrounding the origin—see (f1), (f2), and (f3) in Table 1.1—and they realize a considerable improvement with respect to $\varphi(t) = t^2$. Nevertheless, possibilities are limited with respect to nonconvex PFs. Research on nonconvex PFs is mainly dedicated to the Mumford–Shah model for signals and images defined on $\mathbb{R}$ and $\mathbb{R}^2$; see, e.g., [27, 25, 24]; its discrete equivalent is (f4). For general nonconvex PFs, various necessary conditions and heuristics have been formulated; let us cite [15, 23, 9]. In this paper we derive formal results characterizing the (local) minimizers $\hat{x}$ of $\mathcal{F}_y$ when φ is nonconvex according to the assumptions listed below.

H2. *There is $\theta > 0$ such that $\varphi''(\theta) < 0$ and $\lim_{t\to\infty} \varphi''(t) = 0$.*

A critical distinction between PFs is their smoothness at zero since $\varphi'(0^+) > 0$ gives rise to (local) minimizers $\hat{x}$ such that $g_i^T \hat{x} = 0$ for some indexes $i \in J$ [30, 32]. For smooth regularization we assume the following.

Table 1.2
Second derivatives φ'' for the nonconvex PFs in Table 1.1.

φ'' on $\mathbb{R}^*$ for nonconvex PFs			
Smooth at zero PFs		Nonsmooth at zero PFs	
(f4)	$\varphi''(t) = \begin{Bmatrix} 2\alpha & \text{if} & \lvert t\rvert < 1/\sqrt{\alpha} \\ 0 & \text{if} & \lvert t\rvert > 1/\sqrt{\alpha} \end{Bmatrix}$	(f8)	$\varphi''(t) = \alpha(\alpha-1)\lvert t\rvert^{\alpha-2}$, $0 < \alpha < 1$
(f5)	$\varphi''(t) = \dfrac{2\alpha(1-3\alpha t^2)}{(1+\alpha t^2)^3}$	(f9)	$\varphi''(t) = \dfrac{-2\alpha^2}{(1+\alpha\lvert t\rvert)^3}$
(f6)	$\varphi''(t) = \dfrac{1-2\alpha^2 t^2}{(1+\alpha t^2)^2}$	(f10)	$\varphi''(t) = \dfrac{-\alpha^2}{(1+\alpha\lvert t\rvert)^2}$
(f7)	$\varphi''(t) = 2\alpha(1-2\alpha t^2)\exp(-\alpha t^2)$	(f11)	$\varphi''(t) = 0$

H3. *φ'' is $\mathcal{C}^2$, and there are $\tau > 0$ and $\mathcal{T} \in (\tau, \infty)$ such that $\varphi''(t) \geq 0$ if $t \in [0, \tau]$ and $\varphi''(t) \leq 0$ if $t \geq \tau$, where φ'' is decreasing on $(\tau, \mathcal{T})$ and increasing on $(\mathcal{T}, \infty)$.*

For nonsmooth regularization, the equivalent counterpart of H3 is the following.

H4. *$\varphi'(0^+) > 0$ and φ'' is increasing on $(0, \infty)$ with $\varphi''(t) \leq 0$ for all $t > 0$.*

As seen from Tables 1.1 and 1.2, these assumptions are satisfied for almost all nonconvex PFs used in practice. They can be extended to other classes of functions, too. However, all of them fail to hold for the truncated quadratic PF (f4) and the "0-1" PF (f11); these PFs are considered in separate statements.

The objective of this paper is to exhibit important properties of the minimizers $\hat{x}$ of cost-functions $\mathcal{F}_y$ of the form (1.1)–(1.2) when φ is *nonconvex* as specified above. Let us notice that although solutions to various applied problems are usually defined as the minimizers of cost-functions, the features of the minimizers have seldom been the focus of systematic analysis. And yet, this question is of critical importance for a pertinent choice of cost-function. Generic stability of the (local) minimizers of $\mathcal{F}_y$, when φ is nonconvex as specified above, has been studied in [12, 13]. The question of the properties of minimizers for some *convex* cost-functions has been addressed by [1, 11, 2, 7, 35, 38, 8]. For more general cost-functions, it has been considered by the author in [28, 30, 31, 33].

Outline of the paper. The simple case when $\mathcal{F}_y$ is a scalar function, studied in section 2, gives an instructive insight into the features of the minimizers relevant to nonconvex PFs as specified above. In section 3 we analyze how differences $g_i^T\hat{x}$ at a (local) minimizer $\hat{x}$ of $\mathcal{F}_y$ either are *enhanced* and form edges or are *shrunk* and form homogeneous regions. More precisely, we show that there are two thresholds, $\theta_0 \geq 0$ and $\theta_1 > \theta_0$, such that shrunk differences satisfy $|g_i^T\hat{x}| \leq \theta_0$, while enhanced edges satisfy $|g_i^T\hat{x}| \geq \theta_1$. Equivalently, this result says that $|g_i^T\hat{x}| \notin (\theta_0, \theta_1)$ for all $i \in J$. Given a (local) minimizer $\hat{x}$ of $\mathcal{F}_y$, the subsets $\widehat{J}_0$ and $\widehat{J}_1$,

$$\widehat{J}_0 = \left\{ i \in J : \left|g_i^T\hat{x}\right| \leq \theta_0 \right\} \quad \text{and} \quad \widehat{J}_1 = \left\{ i \in J : \left|g_i^T\hat{x}\right| \geq \theta_1 \right\}, \tag{1.4}$$

satisfy $J = \widehat{J}_0 \cup \widehat{J}_1$, and they address *the homogeneous regions* and *the edges* in $\hat{x}$, respectively. It turns out that if φ is smooth at zero, we have $\theta_0 > 0$, so homogeneous regions are smoothly varying. If φ is nonsmooth at zero, we find $\theta_0 = 0$, which means that differences satisfy either $g_i^T\hat{x} = 0$ or $|g_i^T\hat{x}| \geq \theta_1$, where $\theta_1 > 0$. In such a case,

$$\widehat{J}_0 = \left\{ i \in J : g_i^T\hat{x} = 0 \right\} \quad \text{and} \quad \widehat{J}_1 = \left\{ i \in J : \left|g_i^T\hat{x}\right| \geq \theta_1 \right\} = J \setminus \widehat{J}_0. \tag{1.5}$$

If $\{g_i : i \in J\}$ are first-order differences between neighboring samples, homogeneous regions are constant: regularization using nonsmooth nonconvex PFs entails an enhanced stair-casing effect!

Let us denote

$$\Omega = \{1, \ldots, p\},$$

which is the domain of the signal or the image x. In section 4 we study how an original image or signal of the form $h\mathbb{1}_\Sigma$, where $h > 0$, the sets $\Sigma \subset \Omega$ and $\Sigma^c = \Omega \setminus \Sigma$ are nonempty, and $\mathbb{1}_\Sigma \in \mathbb{R}^p$ reads

$$\mathbb{1}_\Sigma[i] = \begin{cases} 1 & \text{if} \quad i \in \Sigma, \\ 0 & \text{if} \quad i \in \Omega \setminus \Sigma, \end{cases} \tag{1.6}$$

is recovered at the *global* minimizer $\hat{x}$ of $\mathcal{F}_y$ when $y = A\ h\mathbb{1}_\Sigma$ and $\{g_i : i \in J\}$ correspond to first-order difference operators. We show that there is $h_0 > 0$ such that if $h \in (0, h_0)$, we have $|g_i^T \hat{x}| \leq \theta_0$ for all $i \in \Omega$, so the *global* minimizer $\hat{x}$ of $\mathcal{F}_y$ does not involve edges and is constant if φ is nonsmooth at zero. Furthermore, there is $h_1 \geq h_0$ such that if $h > h_1$, the global minimizer $\hat{x}$ is a good approximation of the original $h\mathbb{1}_\Sigma$ since $|g_i^T \hat{x}| \geq \theta_1$ for all i such that $|g_i^T h\mathbb{1}_\Sigma| = h$, whereas $|g_i^T \hat{x}| \leq \theta_0$ for all i such that $|g_i^T h\mathbb{1}_\Sigma| = 0$.

Our theoretical results are illustrated using a numerical experiment in section 5. By way of conclusion, in section 6 we provide a further interpretation of the obtained results. We also compare the minimizers relevant to nonconvex PFs φ with those corresponding to convex edge-preserving PFs. The proofs of all propositions and lemmas are outlined in the appendix.

Notation. The components of a vector $x \in \mathbb{R}^p$ read $x[i]$, for $i \in \Omega$, and its support is $\text{supp}(x) = \{i \in \Omega : x[i] \neq 0\}$. We denote by $\|.\|$ the ℓ_2-norm; so $\|x\| = (\sum_{i=1}^p x[i]^2)^{\frac{1}{2}}$ for $x \in \mathbb{R}^p$. If A is a real-valued matrix, A^T is its transpose, and we recall that the largest eigenvalue of $A^T A$ is $\|A^T A\|$; the smallest eigenvalue of $A^T A$ will be denoted by $\alpha_{\min}$. The letter I will stand for identity matrix. If K is a vector (sub)space, we write $K^\perp$ for its orthogonal complement and define $B(\tilde{x}, \rho) = \{x \in K : \|x - \tilde{x}\| < \rho\}$. The cardinality of a discrete set L is denoted $\sharp L$. We write e_i for the ith vector of the canonical basis of $\mathbb{R}^p$, that is, $e_i[j] = 1$ if $j = i$ and $e_i[j] = 0$ if $j \neq i$. To simplify the notation, we set $\mathbb{1} = \mathbb{1}_\Omega$, i.e., $\mathbb{1}[i] = 1$, for all $i \in \Omega$.

2. Illustration using a cost-function on $\mathbb{R}$.

Let us consider the simple case when $y \in \mathbb{R}_+$ and $\mathcal{F}_y : \mathbb{R} \to \mathbb{R}$ reads

$$\mathcal{F}_y(x) = (x - y)^2 + \beta\varphi(x), \tag{2.1}$$

where φ satisfies H1 and H2, along with one of the assumptions H3 (if φ is smooth) or H4 (if φ is nonsmooth at zero). Consider that

$$\beta > -\frac{2}{\varphi''(\mathcal{T})} \text{ under H3} \quad \text{or} \quad \beta > -\frac{2}{\varphi''(0^+)} \text{ under H4}, \tag{2.2}$$

where $\varphi''(0^+) = \lim_{t \searrow 0} \varphi''(t)$; if $\varphi''(0^+) = -\infty$, we find $\beta > 0$. Define

$$\theta_0 = \inf C_\beta \quad \text{and} \quad \theta_1 = \sup C_\beta,$$

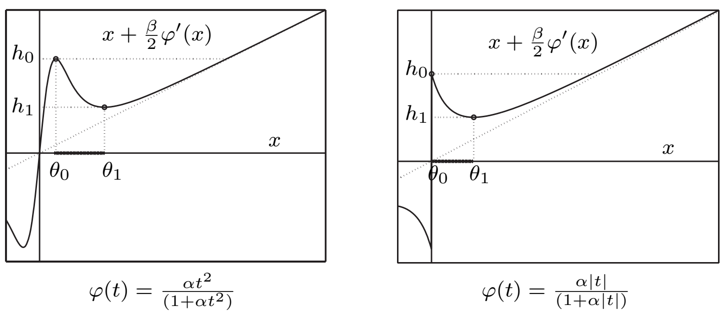


FIG. 2.1. *Plots of* $\frac{1}{2}\mathcal{F}'_y(x) - y = x + \frac{\beta}{2}\varphi'(x)$ *on* $\mathbb{R} \setminus \{0\}$ *for a PF satisfying* H1, H2, *and* H3 *on the left and a PF satisfying* H1, H2, *and* H4 *on the right. These plots suggest how to solve* (2.6) *graphically.*

where

$$C_\beta = \left\{ t \in (0, \infty) : \varphi''(t) < -\frac{2}{\beta} \right\}. \tag{2.3}$$

Notice that $\theta_0 = 0$ if H4 holds and that $\mathcal{T} \in (\theta_0, \theta_1)$ under H3. In both cases,

$$\mathcal{F}''_y(x) = 2 + \beta\varphi''(x) < 0 \quad \text{if} \quad \theta_0 < |x| < \theta_1. \tag{2.4}$$

It follows that *for any* $y \in \mathbb{R}_+$, *no local minimizer* $\hat{x}$ *of* $\mathcal{F}_y$ *lies in* $(-\theta_1, -\theta_0) \cup (\theta_0, \theta_1)$. Conversely, minimizers $\hat{x}$ satisfy either $|\hat{x}| \in [0, \theta_0]$ or $|\hat{x}| \in [\theta_1, \infty)$. This observation underlies the property of recovering either shrunk or enhanced differences at the (local) minimizers of $\mathcal{F}_y$, developed in section 3. It is worth noticing that θ_0 decreases with β, while θ_1 increases with β.

Let us now focus on the global minimization of $\mathcal{F}_y$. Without loss of generality, suppose that

$$\left\{ t > 0 : \varphi''(t) = -\frac{2}{\beta} \right\} = \begin{cases} \{\theta_0\} \cup \{\theta_1\} & \text{under H3,} \\ \{\theta_1\} & \text{under H4.} \end{cases} \tag{2.5}$$

This assumption is satisfied by all PFs used in practice; see, e.g., Table 1.2. By the first-order necessary condition for a minimum,

$$\hat{x} + \frac{\beta}{2}\varphi'(\hat{x}) = y \text{ if } \hat{x} \neq 0 \text{ or } \varphi \text{ is } \mathcal{C}^2 \text{ on } \mathbb{R}, \tag{2.6}$$

$$\frac{\beta}{2}\varphi'(0^+) \geq |y| \text{ if } \hat{x} = 0 \text{ and } \varphi \text{ satisfies H4.} \tag{2.7}$$

Since $y \geq 0$, if we had $\hat{x} < 0$, then $\varphi'(\hat{x}) \leq 0$, and (2.6) cannot hold. By (2.6) yet again, $y - \hat{x} = \frac{\beta}{2}\varphi'(\hat{x}) \geq 0$. It follows that for any $y \geq 0$, any (local) minimizer $\hat{x}$ of $\mathcal{F}_y$ satisfies

$$0 \leq \hat{x} \leq y. \tag{2.8}$$

The analysis presented below is illustrated in Figure 2.1. Define h_1 and h_0 by

$$h_1 = \theta_1 + \frac{\beta}{2}\varphi'(\theta_1) \quad \text{and} \quad h_0 = \begin{cases} \theta_0 + \dfrac{\beta}{2}\varphi'(\theta_0) & \text{under H3,} \\ \dfrac{\beta}{2}\varphi'(0^+) & \text{under H4.} \end{cases} \tag{2.9}$$

Using (2.4), $x \to x + \frac{\beta}{2}\varphi'(x)$ is $\mathcal{C}^1$ and strictly decreasing on (θ_0, θ_1); thus $0 < h_1 < h_0 \leq \infty$. If $y \in [0, h_0)$, the function $\mathcal{F}_y$ admits a strict (local) minimizer $\hat{x}_0$ that satisfies either $0 \leq \hat{x}_0 < \theta_0$ under H3 or $\hat{x}_0 = \theta_0 = 0$ under H4. If $y > h_1$, (2.6) defines a strict (local) minimizer $\hat{x}_1$ satisfying $\hat{x}_1 > \theta_1$. These two statements are developed in the appendix. They show that if $y \in (h_1, h_0)$, there are two strict local minimizers, $\hat{x}_0 \in [0, \theta_0]$ and $\hat{x}_1 > \theta_1$. Let $\chi_0 : [0, h_0) \to [0, \theta_0]$ and $\chi_1 : (h_1, \infty) \to (\theta_1, \infty)$ denote the minimizer functions corresponding to $\hat{x}_0$ and $\hat{x}_1$, respectively.[1] Notice that these functions are $\mathcal{C}^1$, that $\chi_0 = 0$ if H4 holds, and that

$$\chi_1(y) - \chi_0(y) > \theta_1 - \theta_0 > 0 \quad \forall y \in (h_1, h_0).$$

Combining this with (2.8) allows us to write that

$$|y - \chi_0(y)| > |y - \chi_1(y)| + (\theta_1 - \theta_0) \quad \forall y \in (h_1, h_0).$$

We can say that $\hat{x}_0 = \chi_0(y)$ incurs strong smoothing, while smoothing for $\hat{x}_1 = \chi_1(y)$ is weak.

Assume that in addition $\lim_{t\to\infty} \varphi'(t) = 0$, since this holds for all nonconvex PFs used in practice (e.g., see Table 1.1). Using that $x \to x + \frac{\beta}{2}\varphi'(x)$ defines a one-to-one mapping of (θ_1, ∞) onto (h_1, ∞), (2.6) shows that $\lim_{y\to\infty} \chi_1(y) = \infty$. This entails that $\lim_{y\to\infty} \varphi'(\chi_1(y)) = 0$, and then

$$\lim_{y\to\infty} |y - \chi_1(y)| = 0.$$

We can say that smoothing for $\hat{x}_1 = \chi_1(y)$ is vanishing.

Under H3, constant h_0 in (2.9) is finite. For definiteness, assume that $h_0 < \infty$ under H4 as well. Put $\chi_0(h_0) = \lim_{y \nearrow h_0} \chi_0(y)$ and $\chi_1(h_1) = \lim_{y \searrow h_1} \chi_1(y)$; then $\chi_0(h_0) = \theta_0$ and $\chi_1(h_1) = \theta_1$. Define Δ by

$$\Delta(y) = \mathcal{F}_y(\chi_0(y)) - \mathcal{F}_y(\chi_1(y)) \quad \text{for} \quad h_1 \leq y \leq h_0.$$

It is shown in the appendix that $\mathcal{F}_{h_1}$ (respectively, $\mathcal{F}_{h_0}$) does not have any (local) minimum at $\chi_1(h_1) = \theta_1$ (respectively, at $\chi_0(h_0) = \theta_0$). Combining this with the fact that $\lim_{|x|\to\infty} \mathcal{F}_y(x) = +\infty$ allows us to write that

$$\Delta(h_1) < 0 \quad \text{and} \quad \Delta(h_0) > 0. \tag{2.10}$$

Furthermore, for any $y \in (h_1, h_0)$ we can write that (cf. the appendix)

$$\frac{d\mathcal{F}_y(\chi(y))}{dy} = 2(y - \chi(y)), \quad \text{where} \quad \chi = \chi_0 \text{ or } \chi = \chi_1. \tag{2.11}$$

Using this expression, it is seen that if $y \in (h_1, h_0)$, then

$$\Delta'(y) = \frac{d\mathcal{F}_y(\chi_0(y))}{dy} - \frac{d\mathcal{F}_y(\chi_1(y))}{dy} = 2(\chi_1(y) - \chi_0(y)) > 2(\theta_1 - \theta_0). \tag{2.12}$$

Hence Δ is strictly increasing on (h_1, h_0). If $h_0 = +\infty$ under H4, (2.12) shows that $\Delta(y) \to +\infty$ as $y \to h_0$. This, combined with (2.10), shows that there is a unique $\overline{h} \in (h_1, h_0)$ such that $\Delta(y) < 0$ if $y \in (h_1, \overline{h})$ and $\Delta(y) > 0$ if $y \in (\overline{h}, h_1)$, with $\Delta(\overline{h}) = 0$. Consequently,

[1] Minimizer functions χ_0 and χ_0 are defined next. For every $y \in [0, h_0)$, $\mathcal{F}_y$ has a strict local minimum at $\chi_0(y)$ such that $\chi_0(y) \in [0, \theta_0)$ under H3 and $\chi_0(y) = 0$ under H4. Furthermore, for every $y > h_1$, $\mathcal{F}_y$ has a strict local minimum at $\chi_1(y) > \theta_1$.

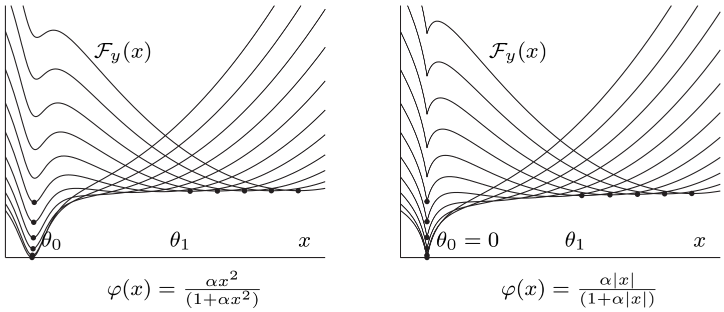


FIG. 2.2. *Each curve represents $\mathcal{F}_y(x) = (x-y)^2 + \beta\varphi(x)$ for a different y in (h_1, h_0). The global minimizer of each $\mathcal{F}_y$ is marked with a "•." Observe also that no local minimizer belongs to (θ_0, θ_1).*

- if $y \in [0, \overline{h})$, the *global* minimizer is $\hat{x} = \chi_0(y) \in [0, \theta_0]$ because $\mathcal{F}_y(\chi_0(y)) < \mathcal{F}_y(\chi_1(y))$,
- if $y > \overline{h}$, the *global* minimizer is $\hat{x} = \chi_1(y) > \theta_1$ because $\mathcal{F}_y(\chi_0(y)) > \mathcal{F}_y(\chi_1(y))$,

whereas $\mathcal{F}_{\overline{h}}$ has two global minimizers, $\chi_0(\overline{h})$ and $\chi_1(\overline{h})$. This behavior is illustrated in Figure 2.2. Clearly, the *global minimizer function* is discontinuous at $y = \overline{h}$. The critical value $y = \overline{h}$ can be seen as a threshold to deciding whether or not the global minimizer $\hat{x}$ of $\mathcal{F}_y$ incurs strong smoothing. In the context of signals and images, this amounts to deciding whether a difference belongs to a homogeneous region or to an edge. These ideas are pursued in section 4.

3. Either shrinkage or enhancement of the differences. In this section we show that nonconvex PFs give rise to (local) minimizers $\hat{x}$ whose differences $g_i^T\hat{x}$ have magnitudes which are either smaller than a (small) threshold $\theta_0 \geq 0$ or larger than a larger threshold $\theta_1 > \theta_0$. The cases when φ is smooth or nonsmooth at zero are analyzed separately.

3.1. Smooth at zero potential functions. The theorem below involves two statements. First, if β is not too small, there are θ_0 and θ_1 as mentioned above. Reciprocally, if we fix either θ_0 or θ_1, we can find a suitable β such that the property holds for an appropriate θ_1 or θ_0, respectively.

THEOREM 3.1. *Let $\mathcal{F}_y : \mathbb{R}^p \to \mathbb{R}$ be of the form* (1.1)–(1.2), *where φ satisfies* H1, H2, *and* H3. *Assume that the set $\{g_i : i \in J\}$ is linearly independent, and put* $\mu = \max_{i\in J} \|G^T(GG^T)^{-1}e_i\|$.

(i) *If $\beta > \beta_0$ for*

$$\beta_0 = \frac{2\mu^2\,\|A^TA\|}{|\varphi''(\mathcal{T})|}, \tag{3.1}$$

there exist $\theta_0 \in (\tau, \mathcal{T})$ and $\theta_1 \in (\mathcal{T}, \infty)$ such that for every $y \in \mathbb{R}^q$, every local minimizer $\hat{x}$ of $\mathcal{F}_y$ satisfies

$$\text{either} \quad |g_i^T\hat{x}| \leq \theta_0 \quad \text{or} \quad |g_i^T\hat{x}| \geq \theta_1 \quad \forall i \in J. \tag{3.2}$$

(ii) *Let $\theta_1 > \mathcal{T}$ (respectively, $\theta_0 \in (\tau, \mathcal{T})$) be such that $\varphi''(\theta_1) < 0$ (respectively, $\varphi''(\theta_0) < 0$) and φ'' is strictly monotonous near θ_1 (respectively, θ_0).*

Consider that $\beta \geq \beta_1$ *for*

$$\beta_1 = \frac{2\mu^2 \, \|A^T A\|}{|\varphi''(\theta_1)|} \quad \left(\textit{respectively,} \quad \beta_1 = \frac{2\mu^2 \, \|A^T A\|}{|\varphi''(\theta_0)|} \right). \tag{3.3}$$

Then there is $\theta_0 \in (\tau, \mathcal{T})$ *(respectively,* $\theta_1 > \mathcal{T}$*) such that for every* $y \in \mathbb{R}^q$*, every local minimizer* $\hat{x}$ *of* $\mathcal{F}_y$ *satisfies* (3.2).

Remark 1. Clearly, θ_0 and θ_1 depend on the shape of φ and are controlled by β. Since $\theta_1 > \mathcal{T}$ and $|\varphi''|$ is decreasing on $(\mathcal{T}, \infty)$, (3.3) shows that β_1 is increasing with θ_1. Since $\theta_0 \in (\tau, \mathcal{T})$ and $|\varphi''|$ increases on $(\tau, \mathcal{T})$, then θ_0 decreases with β_1. It is worth emphasizing that β_0, β_1, θ_0, and θ_1 are independent of y.

Proof. Since $\mathcal{F}_y$ has a minimum at $\hat{x}$, then $D\mathcal{F}_y(\hat{x}) = 0$ and

$$D^2 \mathcal{F}_y(\hat{x})(v, v) \geq 0 \quad \forall v \in \mathbb{R}^p, \tag{3.4}$$

where the second derivative of $\mathcal{F}_y$ at $\hat{x}$ in the direction of v reads

$$D^2 \mathcal{F}_y(x)(v, v) = 2\|Av\|^2 + \beta \sum_{i \in J} \varphi''(g_i^T x)(g_i^T v)^2.$$

Statement (i). For $\beta > \beta_0$, H3 shows that $\varphi''(\mathcal{T})\frac{\beta_0}{\beta} \in \{\varphi''(t) : t \geq \tau\}$. Then the constants θ_0 and θ_1,

$$\begin{aligned} \theta_0 &= \sup \left\{ t \in (\tau, \mathcal{T}) : \varphi''(t) = \varphi''(\mathcal{T}) \frac{\beta_0}{\beta} \right\}, \\ \theta_1 &= \inf \left\{ t \in (\mathcal{T}, \infty) : \varphi''(t) = \varphi''(\mathcal{T}) \frac{\beta_0}{\beta} \right\}, \end{aligned} \tag{3.5}$$

are well defined and satisfy $\tau < \theta_0 < \mathcal{T} < \theta_1 < \infty$. Consequently,

$$t \in (\theta_0, \theta_1) \quad \Rightarrow \quad \varphi''(t) < \varphi''(\mathcal{T}) \frac{\beta_0}{\beta}. \tag{3.6}$$

The proof of the statement consists in showing that no difference of $\hat{x}$ has its magnitude in (θ_0, θ_1). So, suppose that there is $j \in J$ such that $|g_j^T \hat{x}| \in (\theta_0, \theta_1)$. Let us choose $v = G^T (G G^T)^{-1} e_j$; then

$$\begin{aligned} g_j^T v &= 1, \\ g_i^T v &= 0 \quad \forall i \in J \setminus \{j\}, \end{aligned}$$

and $\|v\|^2 \leq \mu^2$. Using successively (3.6) and (3.1) we find that

$$\begin{aligned} D^2 \mathcal{F}_y(\hat{x})(v, v) &= 2\|Av\|^2 + \beta \varphi''(g_j^T \hat{x})(g_j^T v)^2 \\ &< 2\|A^T A\|\mu^2 + \beta \, \varphi''(\mathcal{T}) \frac{\beta_0}{\beta} = 0, \end{aligned}$$

which contradicts (3.4). Consequently, $|g_j^T \hat{x}| \notin (\theta_0, \theta_1)$. The same result holds for every $j \in J$.

Statement (ii). Let $\theta_1 > \mathcal{T}$ be as specified in (ii). We can define θ_0 by

$$\theta_0 = \sup \left\{ t \in (\tau, \mathcal{T}) : \varphi''(t) = \varphi''(\theta_1) \right\}.$$

Combining H2, H3, and the assumption that φ'' is strictly monotonous near θ_1 entails that

$$\varphi''(t) < \varphi''(\theta_1) \quad \forall t \in (\theta_0, \theta_1).$$

If there was $j \in J$ such that $|g_j^T \hat{x}| \in (\theta_0, \theta_1)$, then for $v = G^T(GG^T)^{-1}e_j$ we would find

$$D^2\mathcal{F}_y(\hat{x})(v,v) < 2\|A^TA\|\mu^2 + \beta\,\varphi''(\theta_1) \leq 2\|A^TA\|\mu^2 + \beta_1\,\varphi''(\theta_1) = 0.$$

It follows that $|g_j^T \hat{x}| \notin (\theta_0, \theta_1)$. This conclusion clearly holds for every $j \in J$.

The proof of the statement when θ_0 is fixed follows the same lines and is omitted. ∎

From Table 1.2 it is seen that φ'' is strictly monotonous at every θ such that $\varphi''(\theta) < 0$; hence the assumption involved in statement (ii) is not restrictive. This assumption can be omitted if we systematically consider $\inf\{t \in (\tau, \mathcal{T}) : \varphi''(t) = \varphi''(\theta_1)\}$ in place of θ_1 and $\sup\{t \in (\tau, \mathcal{T}) : \varphi''(t) = \varphi''(\theta_0)\}$ in place of θ_0.

The thresholds θ_0 and θ_1 exhibited in the theorem delimit only the regions in $\mathbb{R}^p$ where $D^2\mathcal{F}_y(x)$ is not nonnegative definite. They do not account for the fact that minimizers $\hat{x}$ satisfy $D\mathcal{F}_y(\hat{x}) = 0$ as well. We can expect that the bounds exhibited here are pessimistic.

The assumption that $\{g_i : i \in J\}$ is linearly independent fails in usual image restoration problems where for each pixel we consider the difference with several neighbors; hence $\sharp J > p$. Nevertheless, the analysis above is easy to extend to all situations where a (local) minimizer $\hat{x}$ is homogeneous on some connected regions. Let us examine this question in more detail. Trivial assumptions on $\{g_i : i \in J\}$ are that $\mathbb{1} \in \ker G$ and that

$$\Sigma \subset \Omega \ \text{ such that } \ \mathrm{supp}(g_i) \cap \Sigma \neq \emptyset \text{ and } \mathrm{supp}(g_i) \cap \Sigma^{\mathrm{c}} \neq \emptyset \quad \Rightarrow \quad g_i^T \mathbb{1}_\Sigma \neq 0.$$

Given $\Sigma \subset \Omega$ such that $\Sigma \neq \emptyset$ and $\Sigma^{\mathrm{c}} \neq \emptyset$, the constant

$$\gamma_\Sigma = \min\left\{|g_i^T \mathbb{1}_\Sigma| : i \in J \text{ such that } \mathrm{supp}(g_i) \cap \Sigma \neq \emptyset \text{ and } \mathrm{supp}(g_i) \cap \Sigma^{\mathrm{c}} \neq \emptyset\right\}$$

is strictly positive. Put $\mu = \sqrt{p}(\min\{\gamma_\Sigma : \Sigma \subset \Omega,\ \Sigma \neq \emptyset, \text{ and } \Sigma^{\mathrm{c}} \neq \emptyset\})^{-1}$; then $\mu \in (0, \infty)$. For example, in the most usual case when $\{g_i : i \in J\}$ yield the first-order differences between each pixel and its nearest neighbors we find $\gamma_\Sigma = 1$ and $\mu = \sqrt{p}$. Consider that $\beta > \beta_0$, where β_0 is of the form (3.1) for μ defined above. Define θ_0 and $\theta_1 > \theta_0$ according to (3.5). Suppose now that $\mathcal{F}_y$ has a (local) minimizer $\hat{x}$ that is homogeneous with respect to $\{g_i : i \in J\}$ on a nonempty subset $\Sigma \subset \Omega$, $\Sigma \neq \Omega$, i.e., that

$$|g_i^T \hat{x}| \leq \theta_0 \quad \forall i \in \{j \in J : \mathrm{supp}(g_j) \subset \Sigma\} \subset \widehat{J}_0, \tag{3.7}$$

$$|g_i^T \hat{x}| > \theta_0 \quad \forall i \in I_\Sigma, \tag{3.8}$$

where $I_\Sigma \neq \emptyset$ corresponds to the boundary of Σ, namely

$$I_\Sigma = \{j \in J : \mathrm{supp}(g_j) \cap \Sigma \neq \emptyset \text{ and } \mathrm{supp}(g_j) \cap \Sigma^{\mathrm{c}} \neq \emptyset\}.$$

Using the reasoning of Theorem 3.1, one can see that in fact

$$|g_i^T \hat{x}| \geq \theta_1 \quad \text{for every} \quad i \in I_\Sigma. \tag{3.9}$$

On the contrary, suppose that there is $j \in I_\Sigma$ such that $|g_j^T \hat{x}| \in (\theta_0, \theta_1)$. Let us choose $v = \frac{\mathbb{1}_\Sigma}{g_j^T \mathbb{1}_\Sigma}$; noticing that $\|\mathbb{1}_\Sigma\| < \sqrt{p}$ and that $|g_j^T \mathbb{1}_\Sigma| \geq \gamma_\Sigma$, we find $\|v\| < \mu$. On the other hand, $g_i^T v = 0$ if $\mathrm{supp}(g_i) \subset \Sigma$ or if $\mathrm{supp}(g_i) \subset \Sigma^c$. Combining these observations with the fact that

$$\varphi''(g_i^T \hat{x}) \leq \varphi''(\theta_0) < 0 \quad \text{for every} \quad i \in I_\Sigma$$

leads to the following:

$$\begin{aligned} D^2 \mathcal{F}_y(\hat{x})(v,v) &= 2\|Av\|^2 + \beta \sum_{i \in I_\Sigma} \varphi''(g_i^T \hat{x})(g_i^T v)^2 \\ &\leq 2\|A^T A\| \, \|v\|^2 + \beta \varphi''(g_j^T \hat{x}) \\ &< 2\|A^T A\| \, \mu^2 + \beta_0 \varphi''(\mathcal{T}) < 0. \end{aligned}$$

Such an $\hat{x}$ cannot be a local minimizer. Hence (3.9) is true, and we can then write that $I_\Sigma \subset \widehat{J}_1$.

This analysis is hard to extend to an arbitrary $\hat{x} \in \mathbb{R}^p$, as far as there is no guarantee[2] to have a subset $\Sigma \subset \Omega$ such that (3.7)–(3.8) hold. Without such a Σ, there is no general way to find a direction $v \in \mathbb{R}^p$ such that $D^2\Phi(\hat{x})(v,v) < 0$ in case there is $j \in J$ such that $|g_j^T \hat{x}| \in (\theta_0, \theta_1)$. Let us emphasize that $D^2\Phi(\hat{x})(v,v) < 0$ is a strong sufficient condition for "nonminimum." It is reasonable to expect that at a minimizer $\hat{x}$, differences $|g_i^T \hat{x}|$ "avoid" the vicinity of $\mathcal{T}$ since φ is very concave there.

Truncated quadratic PF. This important PF, given in (f4) in Table 1.1, fails to satisfy H1, H2, and H3. Because of its nonsmoothness at $\pm 1/\sqrt{\alpha}$, there is no guarantee that its local minimizers satisfy a property of the form (3.2); however, its *global* minimizers do so. Before examining this question in detail, we need some additional notation. We consider that $\{g_i : i \in J\}$ is linearly independent, in which case $r = \sharp J \leq p$. If $r < p$, by the assumption in (1.3), we can take a $p - r \times p$ matrix G_b for which there are $H \in \mathbb{R}^{p \times r}$ and $H_b \in \mathbb{R}^{p \times (p-r)}$ such that[3]

$$\begin{matrix} z = Gx, \\ z_b = G_b x \end{matrix} \quad \Leftrightarrow \quad x = Hz + H_b z_b \tag{3.10}$$

and that $\mathrm{rank}(AH_b) = p - r$. Then we introduce the matrices

$$B = AH, \quad B_b = AH_b, \tag{3.11}$$

$$P = I - B_b \left(B_b^T B_b\right)^{-1} B_b^T. \tag{3.12}$$

If $r = p$, we have $H = G^{-1}$, and hence $P = I$.

[2]For instance, consider $\hat{x}$ as given below for $\{g_i\}$ the differences of each pixel with its adjacent neighbors and $\theta_0 = 1$:

$$\hat{x} = \begin{bmatrix} 0 & 1 & 1 & 2 \\ 0 & 0 & 2 & 2 \\ 0 & 1 & 1 & 2 \end{bmatrix}.$$

We have $|\hat{x}[2,3] - \hat{x}[2,2]| = 2 > \theta_0$, while all other differences are $\leq \theta_0$.

[3]Consider that the rows of G read $g_i[i] = -1$, $g_i[i+1] = 1$, and $g_i[j] = 0$ otherwise for $i = 1, \dots, p-1 = r$. Then $G_b \in \mathbb{R}^{1 \times p}$, and we can choose $G_b[i] = 0$ if $i < p$ and $G_b[p] = -1$. Then $[H, H_b]$—the matrix whose first $p-1$ columns are those of H and whose pth is H_b—is upper triangular composed of -1; then $H_b = -\mathbb{1}$. Using (1.3), we have $AH_b = -A\mathbb{1} \neq 0$. Furthermore, $P = I - \frac{1}{p}\mathbb{1}\mathbb{1}^T$.

PROPOSITION 3.2. *Given $y \in \mathbb{R}^q$, let $\mathcal{F}_y$ read as in* (1.1)–(1.2)*, where* $\{g_i : i \in J\}$ *is linearly independent and*

$$\varphi(t) = \begin{cases} \alpha t^2 & \text{if} \quad |t| \le 1/\sqrt{\alpha}, \\ 1 & \text{if} \quad |t| > 1/\sqrt{\alpha}. \end{cases} \tag{3.13}$$

If $\mathcal{F}_y$ reaches its global minimum at $\hat{x}$, then for every $i \in J$ the following holds:

(i) *if $PAHe_i = 0$, then $g_i^T \hat{x} = 0$;*

(ii) *if $PAHe_i \neq 0$, then*

$$\text{either} \quad |g_i^T \hat{x}| \le \frac{1}{\sqrt{\alpha}} \Gamma_i \quad \text{or} \quad |g_i^T \hat{x}| \ge \frac{1}{\sqrt{\alpha}\,\Gamma_i}, \tag{3.14}$$

where

$$\Gamma_i = \sqrt{\frac{\|PAHe_i\|^2}{\|PAHe_i\|^2 + \alpha\beta}} < 1.$$

Moreover, the inequalities in (3.14) *are strict if $\mathcal{F}_y$ has a unique global minimizer.*

Proposition 3.2 furnishes a useful *necessary condition for a global minimum* of $\mathcal{F}_y$. It provides quite a fine result since thresholds are adapted to each difference individually. In particular, (3.2) holds for

$$\theta_0 = \frac{\gamma}{\sqrt{\alpha}} \quad \text{and} \quad \theta_1 = \frac{1}{\sqrt{\alpha}\gamma}, \quad \text{where} \quad \gamma = \max_{i \in J} \Gamma_i < 1.$$

Clearly, $\theta_0 < \theta_1$ as stated in Theorem 3.1.

3.2. Nonsmooth at zero potential functions. Let us introduce the set $\mathcal{J}_1$ as

$$\mathcal{J}_1 = \left\{ J_1 \subseteq J : \exists v \in \mathbb{R}^p \text{ such that } \begin{bmatrix} g_i^T v = 0 & \text{if} \quad i \in J_0 \stackrel{def}{=} J \setminus J_1, \\ g_i^T v \neq 0 & \text{if} \quad i \in J_1. \end{bmatrix} \right\}. \tag{3.15}$$

Notice that $\{\emptyset\} \in \mathcal{J}_1$ and $J \in \mathcal{J}_1$. Let K denote the application which for every $J_1 \subset \mathcal{J}_1$ yields the subspace $K(J_1)$ defined by

$$K(J_1) = \{u \in \mathbb{R}^p : g_i^T u = 0 \ \forall i \in J_0\} \quad \text{for} \quad J_0 = J \setminus J_1. \tag{3.16}$$

Given $J_1 \in \mathcal{J}_1$, for every $j \in J_1$, let $v_j(J_1) \in \mathbb{R}^p$ be the solution to the problem

$$\text{minimize} \quad \|v\|^2 \quad \text{subject to} \quad v \in K(J_1) \ \text{ and } \ g_j^T v = 1. \tag{3.17}$$

Because of the last constraint, $v_j(J_1) \neq 0$. Then define $\mu > 0$ by

$$\mu = \max \left\{ \max_{j \in J_1} \|v_j(J_1)\| : J_1 \in \mathcal{J}_1 \right\}. \tag{3.18}$$

In general, it is difficult to get an explicit solution for μ. Notice that no assumptions on $\{g_i : i \in J\}$ are made in Theorem 3.3. Its first statement says that if β is not too small, there is $\theta_1 > 0$ such that nonzero differences have a magnitude larger than θ_1. Reciprocally, if we fix $\theta_1 > 0$, we can find β such that this property holds for our θ_1.

THEOREM 3.3. *Let $\mathcal{F}_y$ be of the form* (1.1)–(1.2)*, where φ satisfies* H1, H2, *and* H4. *Let μ read as in* (3.18).

(i) *If* $\beta > \beta_0$ *for*

$$\beta_0 = \frac{2\mu^2 \, \|A^T A\|}{|\varphi''(0^+)|}, \tag{3.19}$$

then there exists $\theta_1 > 0$ *such that for every* $y \in \mathbb{R}^q$*, every (local) minimizer* $\hat{x}$ *of* $\mathcal{F}_y$ *satisfies*

$$\text{either} \quad |g_i^T \hat{x}| = 0 \quad \text{or} \quad |g_i^T \hat{x}| \geq \theta_1 \quad \forall i \in J. \tag{3.20}$$

(ii) *Given* $\theta_1 > 0$ *such that* $\varphi''(\theta_1) < 0$ *and* φ'' *is strictly monotonous near* θ_1*, if* $\beta \geq \beta_1$ *for*

$$\beta_1 = \frac{2\mu^2 \, \|A^T A\|}{|\varphi''(\theta_1)|}, \tag{3.21}$$

then for every $y \in \mathbb{R}^q$*, every (local) minimizer* $\hat{x}$ *of* $\mathcal{F}_y$ *satisfies* (3.20).

In particular, if $|\varphi''(0^+)| = \infty$, we find $\beta_0 = 0$ in (3.19).

Remark 2. The magnitude of θ_1 depends on φ and is controlled by β. Indeed, since $|\varphi''|$ is decreasing on $(0, +\infty)$, (3.21) shows that β_1 is increasing with θ_1.

Given a (local) minimizer $\hat{x}$ of $\mathcal{F}_y$, let us define $\widehat{J}_0$ and $\widehat{J}_1$ by

$$\widehat{J}_0 = \left\{ i \in J : g_i^T \hat{x} = 0 \right\} \quad \text{and} \quad \widehat{J}_1 = J \setminus \widehat{J}_0. \tag{3.22}$$

Clearly, $\widehat{J}_1 \in \mathcal{J}_1$. Since φ is nonsmooth at zero, $\widehat{J}_0$ is usually nonempty [30, 33]. Theorem 3.3 says that the sets $\widehat{J}_0$ and $\widehat{J}_1$ above are equivalent to (1.5).

Proof. Consider that $\mathcal{F}_y$ has a (local) minimum at $\hat{x}$. Let $\widehat{J}_0$ and $\widehat{J}_1$ be defined by (3.22). If $\widehat{J}_1 = \emptyset$, then $g_i^T \hat{x} = 0$ for all $i \in J = \widehat{J}_0$, so statement (3.20) holds. Next, consider that $\widehat{J}_1$ is nonempty. Put

$$\rho = \min_{i \in \widehat{J}_1} |g_i^T \hat{x}| \, \frac{1}{\max\limits_{j \in J} \|g_i\|};$$

then $\rho > 0$ according to (3.22). For every $v \in B(0, \rho)$ we have

$$|g_i^T (\hat{x} + v)| \geq |g_i^T \hat{x}| - |g_i^T v| \geq (\rho - \|v\|) \max_{j \in J} \|g_i\| > 0 \quad \forall i \in \widehat{J}_1.$$

Then the function $\widehat{\mathcal{F}}_y$,

$$\widehat{\mathcal{F}}_y(x) = \|Ax - y\|^2 + \beta \sum_{i \in \widehat{J}_1} \varphi(g_i^T x),$$

is $\mathcal{C}^2$ on $B(\hat{x}, \rho)$. Moreover, using that $\varphi(0) = 0$ by H1, we have

$$v \in K(\widehat{J}_1) \cap B(\hat{x}, \rho) \quad \Rightarrow \quad \widehat{\mathcal{F}}_y(\hat{x} + v) = \mathcal{F}_y(\hat{x} + v), \tag{3.23}$$

where K is as introduced in (3.16). Since $\mathcal{F}_y$ has a (local) minimum at $\hat{x}$, (3.23) shows that $\hat{x}$ is a (local) minimizer of $\widehat{\mathcal{F}}_y$ over $K(\widehat{J}_1) \cap B(\hat{x}, \rho)$. Consequently,

$$D^2 \widehat{\mathcal{F}}_y(\hat{x})(v, v) = 2\|Av\|^2 + \beta \sum_{i \in \widehat{J}_1} \varphi''(g_i^T \hat{x})(g_i^T v)^2 \geq 0 \quad \forall v \in K(\widehat{J}_1). \tag{3.24}$$

Statement (i). Since $\beta > \beta_0$, we have $\frac{2\mu^2\ \|A^TA\|}{\beta} < |\varphi''(0^+)|$. Then the constant θ_1 given by

$$\theta_1 = \inf\left\{t > 0 : \varphi''(t) = -\frac{2\mu^2\ \|A^TA\|}{\beta}\right\} \tag{3.25}$$

is well defined and finite. Using H4,

$$0 < t < \theta_1 \quad \Rightarrow \quad \varphi''(t) < \varphi''(\theta_1). \tag{3.26}$$

Our reasoning is conducted by contradiction. So suppose that there is $j \in \widehat{J}_1$ such that

$$0 < |g_j^T \hat{x}| < \theta_1. \tag{3.27}$$

Combining H2, H4, and (3.26) shows that

$$\varphi''(g_j^T \hat{x}) < \varphi''(\theta_1). \tag{3.28}$$

Let us consider $D^2\mathcal{F}_y(\hat{x})$ in the direction of the vector $v = v_j(\widehat{J}_1)$ defined by (3.17). Using successively (3.24), H4, (3.17), (3.28), and (3.25), we find the following:

$$\begin{aligned} D^2\widehat{\mathcal{F}}_y(\hat{x})(v,v) &\le 2\|Av\|^2 + \beta(g_j^T v)^2 \varphi''(g_j^T \hat{x}) \\ &\le 2\|A^TA\|\|v\|^2 + \beta\varphi''(g_j^T \hat{x}) \\ &< 2\|A^TA\|\ \mu^2 + \beta\ \varphi''(\theta_1) = 0. \end{aligned}$$

This result contradicts (3.24). It follows that (3.27) cannot be true. This conclusion holds for every $j \in \widehat{J}_1$; hence $|g_i^T \hat{x}| \ge \theta_1$ for all $i \in \widehat{J}_1$. Since θ_1 in (3.25) is independent of $\widehat{J}_1$, the same holds for every $\widehat{J}_1 \in \mathcal{J}_1$.

Statement (ii). Assume that there is $j \in \widehat{J}_1$ such that (3.27) holds. Since φ'' is strictly monotonous near θ_1, (3.26) is satisfied, and hence (3.28) holds too. For $v = v_j(\widehat{J}_1)$ defined by (3.17), we get

$$\begin{aligned} D^2\widehat{\mathcal{F}}_y(\hat{x})(v,v) &\le 2\|A^TA\|\mu^2 + \beta\varphi''(g_j^T \hat{x}) \\ &< 2\|A^TA\|\ \mu^2 + \beta\ \varphi''(\theta_1) \\ &\le 2\|A^TA\|\ \mu^2 + \beta_1\ \varphi''(\theta_1) = 0. \end{aligned}$$

The obtained inequality shows that (3.27) cannot be true for any $j \in \widehat{J}_1$. The same result holds for every minimizer $\hat{x}$ of $\mathcal{F}_y$, for any $y \in \mathbb{R}^q$, since the relevant $\widehat{J}_1$ belongs to $\mathcal{J}_1$. $\square$

Table 1.2 shows that in practice, nonconvex, nonsmooth at zero PFs have φ'' *strictly* increasing on $\{t > 0 : \varphi''(t) \neq 0\}$. So, the assumption in (ii) on the strict increase of φ'' near θ_1 is reasonable. It can be avoided if in (3.21) we replace θ_1 by $\inf\{t > 0 : \varphi''(t) = -\frac{2\mu^2\ \|A^TA\|}{\beta} + \varepsilon\}$ for $\varepsilon \gtrapprox 0$.

Here again, θ_1 delimits only the regions in $\mathbb{R}^p$ where $\mathcal{F}_y$ does not satisfy the second-order necessary condition for a local minimum. The first-order necessary condition for a (local) minimum is not taken into consideration. This suggests that our bounds may be pessimistic.

"0-1" PF. This function, given in (f11) in Table 1.1, is discontinuous at 0 and does not satisfy H1, H2, and H4. It can also be seen that the relevant local minimizers of $\mathcal{F}_y$ can fail to satisfy a property of the form (3.20). However, such a property takes place for the global minimizers of $\mathcal{F}_y$.

PROPOSITION 3.4. *Given* $y \in \mathbb{R}^q$, *let* $\mathcal{F}_y$ *be defined by* (1.1)–(1.2), *where* $\{g_i : i \in J\}$ *is linearly independent and*

$$\varphi(t) = \begin{cases} 0 & \text{if} \quad t = 0, \\ 1 & \text{if} \quad t \neq 0. \end{cases} \tag{3.29}$$

If $\mathcal{F}_y$ *has a global minimum at* $\hat{x}$, *then for every* $i \in J$,

(i) *if* $PAHe_i = 0$, *then* $g_i^T \hat{x} = 0$;

(ii) *if* $PAHe_i \neq 0$, *then*

$$\text{either} \quad g_i^T \hat{x} = 0 \quad \text{or} \quad |g_i^T \hat{x}| \geq \frac{\sqrt{\beta}}{\|PAHe_i\|}, \tag{3.30}$$

where H *and* P *are given in* (3.10) *and* (3.12). *The last inequality is strict if* $\mathcal{F}_y$ *has a unique global minimizer.*

This proposition provides a simple *necessary condition for a global minimum* of $\mathcal{F}_y$. We can notice that (3.30) is finely adapted to each difference $g_i^T \hat{x}$ for $i \in J$. It is readily seen that (3.20) is true if we put

$$\theta_1 = \min_{i \in J} \frac{\sqrt{\beta}}{\|PAHe_i\|}.$$

4. Selection for the global minimizer.

We can observe in Table 1.1 that most of the nonconvex PFs used in practice are bounded on $\mathbb{R}$ by a constant. In this section, we will consider the following.

H5. $\varphi(t) \leq 1$ *for all* $t \in \mathbb{R}$.

Notice that[4] then $\lim_{t\to\infty} \varphi'(t) = 0$. We will often use the fact that by H1, $\varphi(t) \geq 0$ on $\mathbb{R}$, and that φ is increasing on $\mathbb{R}_+$. Furthermore, we will consider that $\{g_i : i \in J\}$ yields first-order differences.

H6. *With every* $i \in J$ *there are associated* $N_i = (i_1, i_2) \subset \Omega$ *and* $\gamma_i > 0$ *so that* $g_i^T x = \gamma_i (x_{i_1} - x_{i_2})$, *for all* $x \in \mathbb{R}^p$, *and the null space of* G *is spanned by* $\mathbb{1}$. *We will denote* $\gamma_{\min} = \min_{i \in J} \gamma_i$.

Usually $\gamma_i = 1$ for all $i \in J$; in some models, $\gamma_i = 1/\sqrt{2}$ if $g_i^T x$ corresponds to differences between diagonal pixels in an image. An additional assumption taken in this section is that $A^T A$ is invertible. Then $\alpha_{\min}$—the smallest eigenvalue of $A^T A$—satisfies $\alpha_{\min} > 0$.

Remark 3 (existence of a global minimizer). When $A^T A$ is invertible, for every $y \in \mathbb{R}^q$, the function $\mathcal{F}_y$ defined by (1.1)–(1.2) and H1 is bounded below by 0 and coercive; hence it admits a global minimizer, and the latter is bounded.

Our goal now is to study how an original image or signal of the form $h\mathbb{1}_\Sigma$, where

$$\Sigma \subset \Omega, \quad \text{with} \quad \Sigma \neq \emptyset \ \text{ and } \ \Sigma^c = \Omega \setminus \Sigma \neq \emptyset,$$

---

[4] Since $\lim_{t\to\infty} \varphi''(t) = 0$ and $\varphi''(t) \geq 0$ for all $t > 0$, there is $c \geq 0$ such that $\lim_{t\to\infty} \varphi'(t) = c$. Consider that $c > 0$. By the mean-value theorem, $\varphi(t + \frac{2}{c}) - \varphi(t) = \varphi'(\mu(t))\frac{2}{c}$, where $\mu(t) \in (t, t + \frac{2}{c})$. Noticing that $\lim_{t\to\infty} \mu(t) = +\infty$, we find that $\lim_{t\to\infty}(\varphi(t + \frac{2}{c}) - \varphi(t)) = \lim_{t\to\infty} \varphi'(\mu(t))\frac{2}{c} = 2$, which is impossible because by H5, $\varphi(t + \frac{2}{c}) - \varphi(t) \leq 1$. It follows that $c = 0$.

is recovered at the *global* minimizer $\hat{x}$ of $\mathcal{F}_y$ when $y \in \mathbb{R}^q$ is of the form

$$y = hA\mathbb{1}_\Sigma \quad \text{for} \quad h \in \mathbb{R}_+. \tag{4.1}$$

From now on, we systematically denote

$$J_0 = \{i \in J : g_i^T \mathbb{1}_\Sigma = 0\} \quad \text{and} \quad J_1 = J \setminus J_0. \tag{4.2}$$

It will be convenient to put $F_h = \mathcal{F}_{h\mathbb{1}_\Sigma}$ for every $h \in \mathbb{R}_+$, i.e.,

$$F_h(x) = \|A(x - h\mathbb{1}_\Sigma)\|^2 + \beta \sum_{i \in J} \varphi(g_i^T x), \tag{4.3}$$

and to denote by $\hat{x}_h$ a *global minimizer* of the latter function,

$$F_h(\hat{x}_h) \le F_h(x) \quad \forall x \in \mathbb{R}^p.$$

Remark 4 (upper bound on the global minimum). If y is given by (4.1), we have

$$F_h(h\mathbb{1}_\Sigma) = \beta \sum_{i \in J_1} \varphi(\gamma_i h) \le \beta \sharp J_1, \tag{4.4}$$

where the inequality comes from the assumption that $\varphi(t) \le 1$ on $\mathbb{R}$. Since $\hat{x}_h$ is a global minimizer,

$$F_h(\hat{x}_h) \le \beta \sharp J_1 \quad \forall h \in \mathbb{R}_+.$$

This constitutes a simple necessary condition for a global minimum.

The cases when φ is smooth at zero, and when it is nonsmooth at zero, are analyzed separately.

4.1. Smooth at zero potential functions. The next theorem addresses functions F_h of the form (4.3) which corroborate the conclusions of Theorem 3.1.

THEOREM 4.1. *Let* $F_h : \mathbb{R}^p \to \mathbb{R}$ *be of the form* (4.3), *where* $\{g_i : i \in J\}$ *satisfies* H6 *and* $A^T A$ *is invertible. Let* φ *satisfy* H1, H2, H3, *and* H5. *For every* $h \ge 0$, *suppose that every (local) minimizer* $\hat{x}$ *of* F_h *satisfies* (3.2), *where* $0 < \theta_0 < \theta_1$, *and denote by* $\hat{x}_h$ *a global minimizer of* F_h. *Then we have the following:*

(i) *There is a constant* $h_0 > 0$ *such that*

$$h \in [0, h_0) \quad \Rightarrow \quad |g_i^T \hat{x}_h| \le \theta_0 \quad \forall i \in J. \tag{4.5}$$

(ii) *Assume in addition that* θ_1 *is such that* $\frac{\sharp J_1}{\sharp J_1 + 1} \le \varphi(\theta_1) < 1$. *Then there is* $h_1 > 0$ *such that*

$$h \ge h_1 \quad \Rightarrow \quad \begin{array}{ll} |g_i^T \hat{x}_h| \le \theta_0 & \forall i \in J_0, \\ |g_i^T \hat{x}_h| \ge \theta_1 & \forall i \in J_1. \end{array}$$

This theorem corroborates the interpretation of θ_0 and θ_1 as thresholds for the detection of smooth differences and edges, respectively. Equivalently, the sets $\widehat{J}_0$ and $\widehat{J}_1$ in (1.4) address the homogeneous regions and the edges in $\hat{x}$, respectively.

Proof. The first and second differentials of F_h at any $x \in \mathbb{R}^p$ are well defined and read

$$DF_h(x) = 2A^T A(x - h\mathbb{1}_\Sigma) + \beta G^T \left[\varphi'(g_i^T x)\right]_{i \in J}, \tag{4.6}$$

$$D^2 F_h(x) = 2A^T A + \beta G^T \operatorname{diag}\left(\left[\varphi''(g_i^T x)\right]_{i \in J}\right) G. \tag{4.7}$$

Statement (i). For $h = 0$, the function F_0 reaches its global minimum at $\hat{x}_0 = 0$. Hence $DF_0(\hat{x}_0) = 0$ and $D^2F_0(\hat{x}_0)$ is positive definite because $\varphi''(0) \geq 0$ by H1 and A^TA is invertible. By the implicit function theorem, there are $\rho_0 > 0$ and a unique $\mathcal{C}^1$-function $\chi : [0, \rho_0) \to \mathbb{R}^p$ such that

$$DF_h(\chi(h)) = 0 \quad \forall h \in [0, \rho_0)$$

and that $\chi(0) = 0$. By the continuity of D^2F_h and χ, there is $\rho \in (0, \rho_0]$ such that

$$h \in [0, \rho) \quad \Rightarrow \quad D^2F_h(\chi(h)) \quad \text{is positive definite.}$$

Consequently, if $h \in [0, \rho)$, the function F_h has a strict (local) minimum at $\chi(h)$. Using that $\chi(0) = 0$, that $h \to F_h(\chi(h))$ is continuous, and that $F_0(\chi(0)) = 0$, there is $h_0 \in (0, \rho]$ such that

$$h \in [0, h_0) \quad \Rightarrow \quad F_h(\chi(h)) < \beta\varphi(\theta_1) \quad \text{and} \quad |g_i^T\chi(h)| \leq \theta_0 \quad \forall i \in J.$$

Suppose that for $h \in (0, h_0)$ there is another (local) minimizer $\tilde{x} \neq \hat{x}$ such that $|g_i^T\tilde{x}| \geq \theta_1$ for some $i \in J$. Then $\varphi(g_i^T\tilde{x}) \geq \varphi(\theta_1)$, and we can write that

$$F_h(\tilde{x}) \geq \beta\varphi(\theta_1) > F_h(\chi(h)).$$

This shows that for any $h \in [0, h_0)$, the function F_h reaches its global minimum at an $\hat{x}_h$ satisfying (4.5).

Statement (ii). We will consider that $h \geq h_1$ for

$$h_1 = \frac{\theta_0}{\gamma_{\min}} + \sqrt{\frac{2\beta\sharp J_1}{\alpha_{\min}}}. \tag{4.8}$$

To simplify, we will write $\hat{x}$ for $\hat{x}_h$ to denote a global minimizer of F_h. For $\hat{x}$, let the sets $\widehat{J}_0$ and $\widehat{J}_1$ be defined by (1.4); since (3.2) holds, $\widehat{J}_0 \cup \widehat{J}_1 = J$. Let us examine the possibility that $\widehat{J}_1 \neq J_1$. Two cases arise according to the relationship between $\widehat{J}_1$ and J_1.

(C1) $\widehat{J}_0 \cap J_1$ is nonempty. Let $i \in \widehat{J}_0 \cap J_1$ and $N_i = \{i_1, i_2\}$, according to H6. For definiteness, assume that $\mathbb{1}_\Sigma[i_1] = 1$ and $\mathbb{1}_\Sigma[i_2] = 0$. It is easy to see that[5] $\hat{x} \neq c\mathbb{1}$ for any $c \in \mathbb{R}$. Then $\Phi(\hat{x}) > 0$, and we have

$$\begin{aligned} F_h(\hat{x}) &> \alpha_{\min} \ \|\hat{x} - h\mathbb{1}_\Sigma\|^2 \\ &\geq \alpha_{\min} \ \left((\hat{x}[i_1] - h)^2 + (\hat{x}[i_2])^2\right). \end{aligned}$$

Noticing that

$$\gamma_i \mid \hat{x}[i_1] - \hat{x}[i_2] \mid = \left|g_i^T\hat{x}\right| \leq \theta_0,$$

we find that[6]

$$(\hat{x}[i_1] - h)^2 + (\hat{x}[i_2])^2 \geq \frac{1}{2}\left(h - \frac{\theta_0}{\gamma_i}\right)^2.$$

[5] Suppose that for $h > 0$, F_h has a (local) minimizer of the form $\hat{x} = c\mathbb{1}$ for $c \in \mathbb{R}$. Using that $DF_h(\hat{x}) = 0$, (4.6) leads to $A^TA(c\mathbb{1} - h\mathbb{1}_\Sigma) = 0$. Since A^TA is invertible and Σ and Σ^c are nonempty, there is no $c \in \mathbb{R}$ satisfying this equation.

[6] Here we consider the following problem:

$$\text{minimize} \quad f(t, s) \quad \text{subject to} \quad \gamma_i|t - s| \leq \theta_0,$$

where $f : \mathbb{R}^2 \to \mathbb{R}$ reads $f(t, s) = (t - h)^2 + s^2$ for $h > \theta_0/\gamma_i$. Using Kuhn–Tucker conditions, the minimum is reached for $\hat{t} = \frac{1}{2}(h + \frac{\theta_0}{\gamma_i})$ and $\hat{s} = \frac{1}{2}(h - \frac{\theta_0}{\gamma_i})$, and its value is $f(\hat{t}, \hat{s}) = \frac{1}{2}(h - \frac{\theta_0}{\gamma_i})^2$.

Using that $h \geq h_1$, (4.4) shows that

$$F_h(\hat{x}) > \frac{\alpha_{\min}}{2}\left(h - \frac{\theta_0}{\gamma_{\min}}\right)^2 \geq \beta \sharp J_1 \geq F_h(h\mathbb{1}_\Sigma).$$

Since $\hat{x}$ is a global minimizer, it follows that $\widehat{J}_0 \cap J_1$ is empty.

(C2) $\widehat{J}_1 \supset J_1$ and $\widehat{J}_1 \neq J_1$; hence $\sharp\widehat{J}_1 \geq \sharp J_1 + 1$. Since $\hat{x} \neq h\mathbb{1}_\Sigma$ and A^TA is invertible, we have $\|A(\hat{x} - h\mathbb{1}_\Sigma)\| > 0$. On the other hand, $\varphi(g_i^T\hat{x}) \geq \varphi(\theta_1)$ for every $i \in \widehat{J}_1$. Then

$$F_h(\hat{x}) > \beta\,(\sharp J_1 + 1)\,\varphi(\theta_1).$$

Combining this with the assumption on $\varphi(\theta_1)$ in (ii) and with (4.4) shows that

$$F_h(\hat{x}) > \beta \sharp J_1 \geq F_h(h\mathbb{1}_\Sigma). \tag{4.9}$$

Since $\hat{x}$ is global minimizer, J_1 cannot be strictly included in $\widehat{J}_1$.

The conclusions of (C1) and (C2) show that the global minimizer $\hat{x}$ is such that $\widehat{J}_1 = J_1$ and $\widehat{J}_0 = J_0$. Hence we have proved the statement. □

This theorem focuses on functions φ satisfying $\lim_{t\to\infty}\varphi(t) = 1$. This, combined with H1, H2, and H3, shows that if $\varphi(\theta_1) < 1$, then $\varphi''(\theta_1) < 0$. Under the conditions of Theorem 3.1, its statement (ii) says that there are β and θ_0 such that all minimizers of F_h satisfy (3.2).

The values of h_0, θ_1, and h_1 used in the proof of the theorem correspond to strong sufficient conditions for a global minimum. We can suppose that in practice statements (i) and (ii) hold for a larger h_0 and for smaller θ_1 and h_1, respectively.

Truncated quadratic PF. As in section 3, this function—see (f4) in Table 1.1—needs a separate analysis. In this case, the global minimizer $\hat{x}_h$ of F_h can be derived explicitly.

PROPOSITION 4.2 (truncated quadratic PF). *Let F_h be of the form* (4.3)*, where A^TA is invertible,* H6 *holds, and φ is given by* (3.13)*. Define $\chi_\Sigma \in \mathbb{R}^p$ by*

$$\chi_\Sigma = \left(A^TA + \beta\alpha G^TG\right)^{-1} A^TA\mathbb{1}_\Sigma. \tag{4.10}$$

For every $h \in \mathbb{R}_+$ let $\hat{x}_h$ denote a global minimizer of F_h. Then there are $h_0 > 0$ and $h_1 > h_0$ such that

$$h \in [0, h_0) \Rightarrow \hat{x}_h = h\,\chi_\Sigma, \tag{4.11}$$

$$h \geq h_1 \Rightarrow \hat{x}_h = h\,\mathbb{1}_\Sigma. \tag{4.12}$$

Moreover, $\hat{x}_h$ in (4.11) *and* (4.12) *is the unique global minimizer of the relevant F_h.*

Observe that $h\chi_\Sigma$ is the regularized least-squares solution, i.e., the minimizer of F_h corresponding to $\varphi(t) = t^2$. So, the global minimizer in (4.11) does not involve edges.

4.2. Nonsmooth at zero potential function. Since now $\theta_0 = 0$ and $\{g_i\}$ satisfies H6, we have to deal with images and signals which are constant on some regions. To this end, we introduce the following definition.

DEFINITION 4.3. *A subset $\Sigma \subset \Omega$ is* connected *with respect to $\{g_i : i \in J\}$ either if Σ is a singleton or if for every $i, j \in \Sigma$ there is a sequence $k_1, \ldots, k_n$ with elements* *of J such that $i \in N_{k_1}$ and $j \in N_{k_n}$, $N_{k_\ell} \subset \Sigma$, for all $\ell = 1, \ldots, n$ and $N_{k_\ell} \cap N_{k_{\ell+1}} \neq \emptyset$, for all $\ell = 1, \ldots, n-1$.*

THEOREM 4.4. *Let $F_h : \mathbb{R}^p \to \mathbb{R}$ be of the form (4.3), where $\{g_i : i \in J\}$ satisfies* H6 *and $A^T A$ is invertible. Let φ satisfy* H1, H2, H4, *and* H5. *For every $h \geq 0$, suppose that every minimizer $\hat{x}$ of F_h satisfies* (3.20), *where $\theta_1 > 0$, and denote by $\hat{x}_h$ a global minimizer of F_h. Then we have the following:*

(i) *There is a constant $h_0 > 0$ such that*

$$h \in [0, h_0) \quad \Rightarrow \quad \hat{x}_h = h\zeta\, \mathbb{1},$$

where

$$\zeta = \frac{(A\mathbb{1})^T A \mathbb{1}_\Sigma}{\|A\mathbb{1}\|^2}. \tag{4.13}$$

(ii) *Assume in addition that θ_1 is such that $\frac{\sharp J_1}{\sharp J_1 + 1} \leq \varphi(\theta_1) < 1$. Then there is $h_1 > 0$ such that*

$$h > h_1 \quad \Rightarrow \quad \begin{array}{ll} g_i^T \hat{x}_h = 0 & \forall i \in J_0, \\ |g_i^T \hat{x}_h| \geq \theta_1 & \forall i \in J_1. \end{array}$$

(iii) *For θ_1 as in* (ii), *let $h > h_1$. If Σ and Σ^c are connected with respect to $\{g_i : i \in J\}$, there are $\hat{s}_h \in (0, h]$ and $\hat{c}_h \in \mathbb{R}$ such that*

$$\hat{x}_h = \hat{s}_h \mathbb{1}_\Sigma + \hat{c}_h \mathbb{1}. \tag{4.14}$$

Moreover, $\hat{s}_h \to h$ and $\hat{c}_h \to 0$ as $h \to \infty$.

If $A = I$, we have $\zeta = \frac{\sharp\Sigma}{p}$ in (i). Statement (ii) shows that for h large enough, the global minimizer $\hat{x}_h$ of F_h has the same edges and the same constant regions as the original $h\mathbb{1}_\Sigma$. Furthermore, (iii) indicates that $\hat{x}_h$ provides a faithful restoration of the original $h\mathbb{1}_\Sigma$.

Remark 5. Statement (iii) can be extended to arbitrary subsets $\Sigma \subset \Omega$ in the following way. Let us represent Σ and Σ^c as unions of subsets which are connected with respect to $\{g_i : i \in J\}$, say Σ_i, $1 \leq i \leq m$, and Σ_i^c, $1 \leq i \leq n$, respectively. In such a case, we will find that there are reals $\hat{s}_i$, $1 \leq i \leq m$, and $\hat{c}_i$, $1 \leq i \leq n$, such that for $h > 0$ large enough, $\hat{x}_h = \sum_{i=1}^m \hat{s}_i \mathbb{1}_{\Sigma_i} + \sum_{i=1}^n \hat{c}_i \mathbb{1}_{\Sigma_i^c}$.

Proof. The constant below will be used several times in what follows:

$$\xi = (A\mathbb{1}_\Sigma)^T \left(I - \frac{(A\mathbb{1})(A\mathbb{1})^T}{\|A\mathbb{1}\|^2} \right) A\mathbb{1}_\Sigma = \|A\mathbb{1}_\Sigma\|^2 - \zeta^2 \|A\mathbb{1}\|^2. \tag{4.15}$$

Clearly, $\xi \geq 0$. Furthermore, the null space of $I - \frac{1}{\|A\mathbb{1}\|^2}(A\mathbb{1})(A\mathbb{1})^T$ being spanned by $A\mathbb{1}$, it does not contain $A\mathbb{1}_\Sigma$ since $A^T A$ is invertible, and $\mathbb{1}_\Sigma \neq \mathbb{1}$ and $\mathbb{1}_\Sigma \neq 0$. Hence $\xi > 0$.

Statement (i). We will consider that $h \in (0, h_0)$, where

$$h_0 = \sqrt{\frac{\beta\varphi(\theta_1)}{\xi}}.$$

If there is $\hat{c} \in \mathbb{R}$ such that F_h has a (local) minimum at $\hat{x} = \hat{c}\mathbb{1}$, then $\hat{c}$ minimizes the function $c \to F_h(c\mathbb{1})$,

$$F_h(c\mathbb{1}) = \|A(c\mathbb{1} - h\mathbb{1}_\Sigma)\|^2 = c^2\|A\mathbb{1}\|^2 - 2ch(A\mathbb{1}_\Sigma)^T A\mathbb{1} + h^2\|A\mathbb{1}_\Sigma\|^2.$$

This function has a unique minimizer. It is easy to calculate that $\hat{c} = h\zeta$ for ζ as in (4.13) and that

$$F_h(\hat{c}\mathbb{1}) = h^2\,\xi.$$

Let $\tilde{x} \neq \hat{x}$ be a (local) minimizer of F_h. Then H6 shows that there is $j \in J$ such that $g_j^T\tilde{x} \neq 0$, in which case (3.20) entails that $|g_j^T\tilde{x}| \geq \theta_1$. Using that $h \in (0, h_0)$, we obtain

$$F_h(\tilde{x}) \geq \beta\varphi(g_j^T\tilde{x}) \geq \beta\varphi(\theta_1) = h_0^2\,\xi > F_h(\hat{c}\mathbb{1}).$$

It follows that for every $h \in (0, h_0)$, the function F_h reaches its global minimum at $\hat{x}_h$ as given in (i).

Statement (ii). Next we consider that $h > h_1$, where

$$h_1 = \sqrt{\frac{2\beta\sharp J_1}{\alpha_{\min}}}. \tag{4.16}$$

Let $\hat{x}$ denote a global minimizer of F_h. With $\hat{x}$, we associate the subsets $\widehat{J}_0$ and $\widehat{J}_1$ as given in (1.5). Since (3.20) holds, $\widehat{J}_0 \cup \widehat{J}_1 = J$. Let us analyze the possibility that $\widehat{J}_1 \neq J_1$. Two cases can then arise.

(C1) $J_1 \cap \widehat{J}_0$ is nonempty. For $i \in J_1 \cap \widehat{J}_0$, let $N_i = (i_1, i_2)$, according to H6. Since $g_i^T\hat{x} = 0$, there is $c \in \mathbb{R}$ such that $\hat{x}[i_1] = \hat{x}[i_2] = c$. Using that $|g_i^T\mathbb{1}_\Sigma| = \gamma_i|\mathbb{1}_\Sigma[i_1] - \mathbb{1}_\Sigma[i_2]| = \gamma_i$, we find

$$\begin{aligned} F_h(\hat{x}) &\geq \alpha_{\min}\|\hat{x} - h\mathbb{1}_\Sigma\|^2 + \beta\sum_{j\in J}\varphi(g_j^T\hat{x}) \\ &\geq \alpha_{\min}\left((\hat{x}[i_1] - h\mathbb{1}_\Sigma[i_1])^2 + (\hat{x}[i_2] - h\mathbb{1}_\Sigma[i_2])^2\right) \\ &= \alpha_{\min}\left((c-h)^2 + c^2\right) \\ &\geq \alpha_{\min}\frac{h^2}{2}, \end{aligned}$$

because the function $c \to (c-h)^2 + c^2$ reaches its minimum for $c = h/2$. Since $h > h_1$, Remark 4 shows that

$$F_h(\hat{x}) > \alpha_{\min}\frac{h_1^2}{2} = \beta\sharp J_1 \geq F_h(h\mathbb{1}_\Sigma). \tag{4.17}$$

It follows that $J_1 \cap \widehat{J}_0$ is empty.

(C2) $J_1 \subset \widehat{J}_1$ with $J_1 \neq \widehat{J}_1$. Applying the reasoning behind item (C2) in the proof of Theorem 4.1(ii) shows that J_1 cannot be strictly included in $\widehat{J}_1$.

It follows that any global minimizer $\hat{x}_h$ of F_h is such that $\widehat{J}_0 = J_0$ and, equivalently, $\widehat{J}_1 = J_1$.

Statement (iii). Since Σ and Σ^{c} are connected, (ii) and H6 show that[7] there are $\hat{s}_h \in \mathbb{R}$ and $\hat{c}_h \in \mathbb{R}$ such that $\hat{x}_h$ is of the form (4.14). Moreover, $\hat{s}_h \neq 0$ because

[7] For $i, j \in \Sigma$, let $\{k_\ell : \ell = 1, \ldots, n\}$ be as in the definition for connectedness; then $k_\ell \in J_0$ for all $\ell = 1, \ldots, n$. Hence, $\hat{x}[i] = \hat{x}[j]$ for all $j \in N_{k_\ell}$, for all $\ell = 1, \ldots, n$. It follows that $\hat{x}[i] = \hat{x}[j]$ for all $i, j \in \Sigma$.

In a similar way it is found that $\hat{x}[i] = \hat{x}[j]$ for all $i, j \in \Sigma^{\mathrm{c}}$.

$\hat{s}_h = 0$ would entail $\widehat{J}_0 = J \neq J_0$. Furthermore, $(\hat{s}_h, \hat{c}_h)$ minimizes on $\mathbb{R} \setminus \{0\} \times \mathbb{R}$ the function $(s, c) \to F_h(s\mathbb{1}_\Sigma + c\mathbb{1})$, namely

$$F_h(s\mathbb{1}_\Sigma + c\mathbb{1}) = \|A(s\mathbb{1}_\Sigma + c\mathbb{1}) - hA\mathbb{1}_\Sigma\|^2 + \beta \sum_{i \in J} \varphi(g_i^T(s\mathbb{1}_\Sigma + c\mathbb{1})).$$

Using that $\varphi(g_i^T(s\mathbb{1}_\Sigma + c\mathbb{1})) = \varphi(s|g_i^T\mathbb{1}_\Sigma|) = \varphi(\gamma_i s)$ for all $i \in J_1$,

$$F_h(s\mathbb{1}_\Sigma + c\mathbb{1}) = \|(s-h)A\mathbb{1}_\Sigma + cA\mathbb{1}\|^2 + \beta \sum_{i \in J_1} \varphi(s\gamma_i).$$

Noticing that $(s, c) \to F_h(s\mathbb{1}_\Sigma + c\mathbb{1})$ is $\mathcal{C}^2$ on $\mathbb{R} \setminus \{0\} \times \mathbb{R}$, if this function has a (local) minimum at $(\hat{s}_h, \hat{c}_h)$, then

$$\frac{\partial F_h}{\partial c}(\hat{s}_h\mathbb{1}_\Sigma + c\mathbb{1})\bigg|_{c=\hat{c}_h} = 0.$$

Hence $\hat{c}_h = \sigma(\hat{s}_h)$, where

$$\sigma(s) = -(s-h)\zeta \tag{4.18}$$

for ζ as in (4.13). Then $\hat{s}_h$ minimizes the function $f(s) = F_h(s\mathbb{1}_\Sigma + \sigma(s)\mathbb{1})$ which reads

$$f(s) = (s-h)^2\xi + \beta \sum_{i \in J_1} \varphi(s\gamma_i),$$

where $\xi > 0$ is given in (4.15). Noticing that $\hat{s}_h\gamma_i \neq 0$ for all $i \in J$, we have $f'(\hat{s}_h) = 0$, that is,

$$\hat{s}_h + \frac{\beta}{2\xi} \sum_{i \in J_1} \gamma_i \varphi'(\hat{s}_h\gamma_i) = h.$$

Using that φ is symmetric, this equation equivalently reads

$$\operatorname{sign}(\hat{s}_h) \left(|\hat{s}_h| + \frac{\beta}{2\xi} \sum_{i \in J_1} \gamma_i \varphi'(|\hat{s}_h|\gamma_i) \right) = h > 0.$$

It follows that $0 < \hat{s}_h \leq h$ for every $h \geq h_1$ and that $\hat{s}_h \to \infty$ as $h \to \infty$. Noticing that $\lim_{t\to\infty} \varphi'(t) = 0$, it is seen that $\hat{s}_h \to h$ as $h \to \infty$. Inserting this into (4.18) shows that $\hat{c}_h \to 0$ as $h \to \infty$. $\square$

Recall that (3.20) holds for $\theta_1 > 0$ if $\beta > \beta_0$, where β_0 is given in (3.19). The assumption in (ii) was used in Theorem 4.1 and discussed after the end of the proof. Using those same arguments, we arrive at $\varphi''(\theta_1) < 0$. Then Theorem 3.3(ii) indicates how to choose β. The magnitudes of h_0, h_1, and θ_1 used in this proof guarantee strong sufficient conditions for global minimum. It is reasonable to suppose that the statements remain true for a larger h_0 and for smaller h_1 and θ_1.

"0-1" PF. This PF, given in (f11) in Table 1.1, is discontinuous at zero. First, we derive a necessary and sufficient condition for a local minimum of $\mathcal{F}_y$ as defined by (1.1)–(1.2) with φ the "0-1" PF. Lemma 4.5 does not involve any assumptions on $y \in \mathbb{R}^q$, A, or $\{g_i : i \in J\}$.

LEMMA 4.5. *For $y \in \mathbb{R}^q$, let $\mathcal{F}_y$ be of the form* (1.1)–(1.2)*, where φ is defined in* (3.29).

(i) *We have the following equivalence:*

$$\hat{x} \in \mathbb{R}^p \text{ is a local minimizer of } \mathcal{F}_y \quad \Leftrightarrow \quad \left[\begin{array}{l} \hat{x} \text{ is a solution to the} \\ \text{following problem:} \\ \text{minimize} \quad \|Ax - y\|^2 \\ \qquad\qquad \text{subject to} \;\; x \in K(\widehat{J}_1), \end{array} \right.$$

where $K(\widehat{J}_1)$ *is defined according to* (3.16) *for* $\widehat{J}_1 = \{i \in J : g_i^T \hat{x} \neq 0\}$.

(ii) *Given* $y \in \mathbb{R}^q$, *if* $\hat{x}$ *is a (local) minimizer of* $\mathcal{F}_y$, *then for every* $h \in \mathbb{R}$, *the function* $\mathcal{F}_{hy}$ *has a (local) minimum at* $h\hat{x}$.

Notice that by (i), if $A^T A$ is invertible, every local minimizer of $\mathcal{F}_y$ is strict.

When φ in (4.3) is the "0-1" PF, the global minimizer $\hat{x}_h$ of F_h can be determined explicitly.

PROPOSITION 4.6 ("0-1" PF). *Let* F_h *be of the form* (4.3), *where* $A^T A$ *is invertible,* $\{g_i : i \in J\}$ *satisfies* H6, *and* φ *reads as in* (3.29). *For every* $h \in \mathbb{R}_+$ *let* $\hat{x}_h$ *denote a global minimizer of* F_h. *Then there are* $h_0 > 0$ *and* $h_1 > h_0$ *such that the global minimizer* $\hat{x}_h$ *of* F_h *reads*

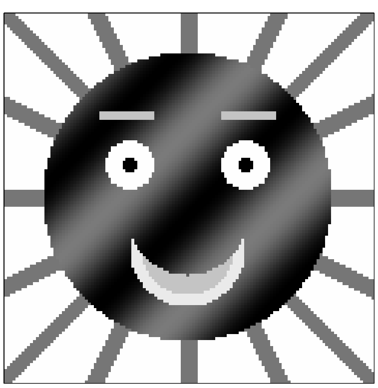
$$h \in [0, h_0) \Rightarrow \hat{x}_h = h\zeta \, \mathbb{1}, \tag{4.19}$$

$$h > h_1 \Rightarrow \hat{x}_h = h \, \mathbb{1}_\Sigma, \tag{4.20}$$

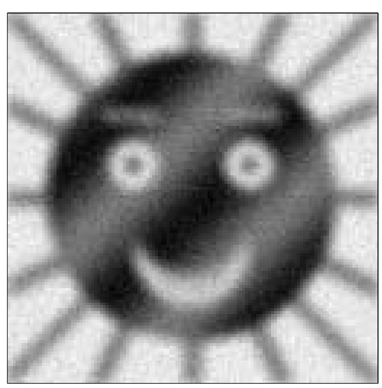
where ζ *is defined in* (4.13). *Moreover,* $\hat{x}_h$ *in* (4.19) *and* (4.20) *is the unique global minimizer of* F_h.

Observe that $\hat{x}_h$ in (4.19) is the same as in Theorem 4.4(i).

5. Experiments. First, we present the restoration of a blurred, noisy 128×128 synthetic image using both convex and nonconvex PFs φ. The original image in Figure 5.1(a) presents smoothly varying regions, constant regions, and sharp edges. Data in Figure 5.1(b) correspond to $y = a * x + n$, where a is a blur with entries $a_{i,j} = \exp(-(i^2 + j^2)/12.5)$ for $-4 \leq i, j \leq 4$, and n is white Gaussian noise yielding 20 dB of SNR. All restored images are obtained by minimizing a cost-function $\mathcal{F}_y$ of the form (1.1)–(1.2), where $\{g_i : i \in J\}$ correspond to the first-order differences of each pixel with its eight nearest neighbors (then $\gamma_i = 1$ in H6) for different functions

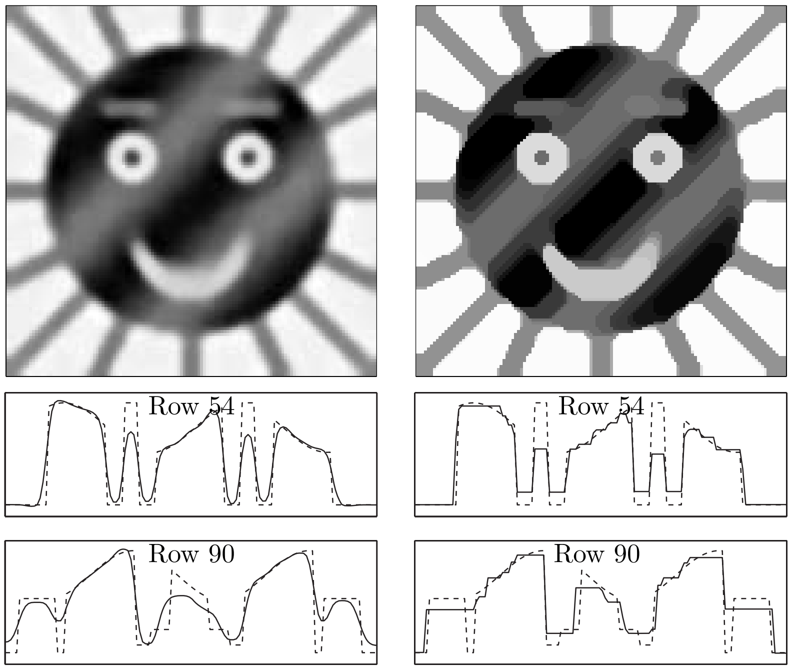


(a) Original image



(b) Data $y = \text{blur} + \text{noise}$

FIG. 5.1. *Data* $y = a \star x + n$, *where* a *is a blur and* n *is white Gaussian noise, with* 20 *dB of SNR.*

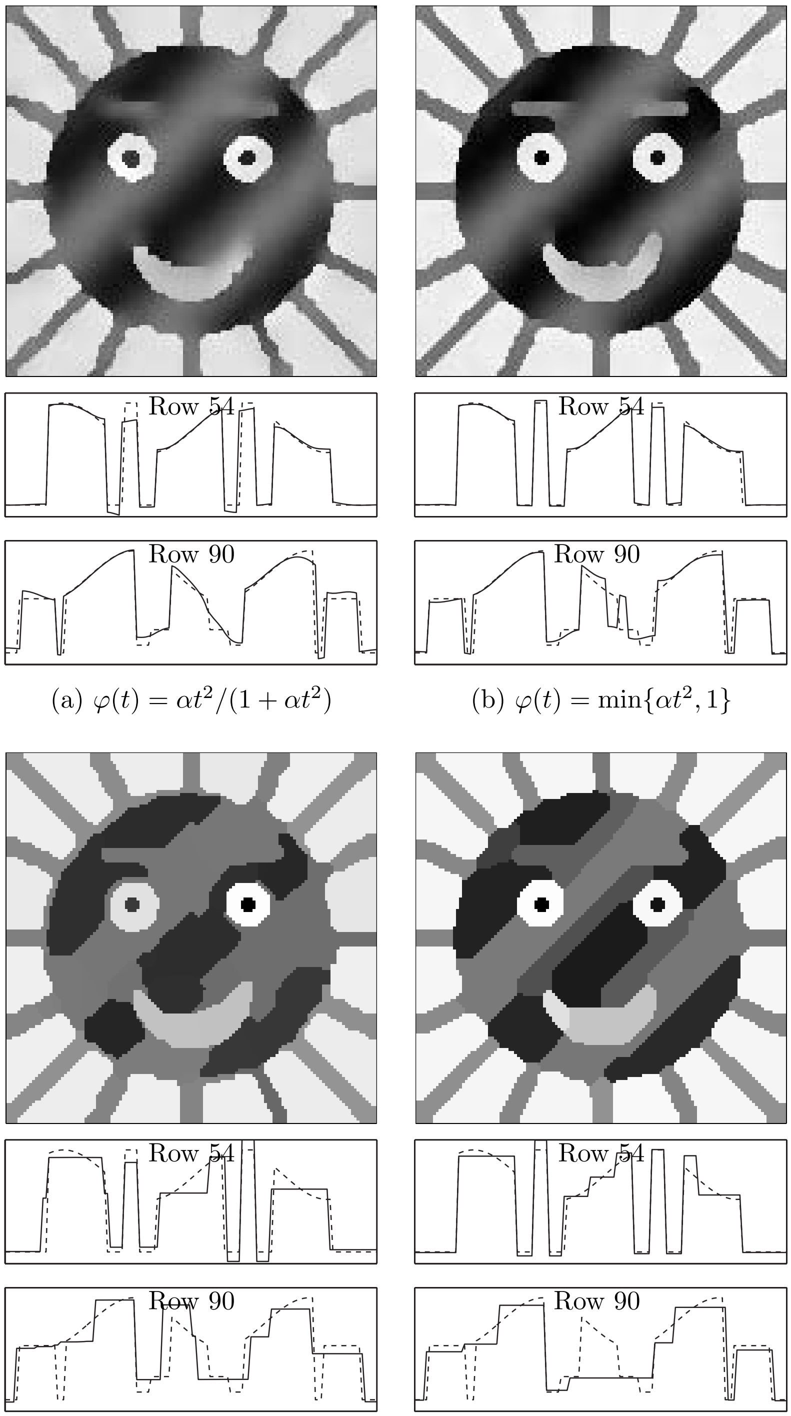


(a) $\varphi(t) = t^\alpha$ (b) $\varphi(t) = |t|$

FIG. 5.2. *Restoration using convex PFs. Left: smooth at zero PF. Right: nonsmooth at zero PF.*

φ. In all figures, the obtained minimizers are displayed on the top. Below we give two sections of the restored images, corresponding to rows 54 and 90, where the relevant sections of the original image are plotted with a dotted line. The minimizers corresponding to nonconvex PFs are calculated using a generalized graduated nonconvexity method [29].

The restorations in Figure 5.2(a) and (b) correspond to convex PFs, namely $\varphi(t) = |t|^\alpha$ for $\alpha = 1.4$, $\beta = 40$ and $\alpha = 1$, $\beta = 100$, respectively. In (a), edges are slightly blurred and underestimated. In (b), the stair-caising effect is very visible: there are numerous spurious edges, whereas important edges are underestimated. The restorations in Figure 5.3 are calculated using nonconvex PFs. The restorations in the first row correspond to smooth at zero PFs, while those in the second row correspond to nonsmooth at zero PFs. On the average, the important edges are very neat, and their amplitude is correct. The image in (a) corresponds to PF (f5) (see Table 1.1) for $\alpha = 25$, $\beta = 35$. The image in (b) is obtained using the PF (f4) for $\alpha = 60$, $\beta = 10$. Both images have neat edges and smoothly varying homogeneous regions. Some fine features in (a) are underestimated, and others are skipped. The image in (b) provides a faithful restoration. The image in (c) corresponds to the PF (f9) for $\alpha = 20$, $\beta = 100$, while the one in (d) corresponds to the PF (f11) for $\beta = 25$. These PFs are nonsmooth at zero, and the restored images are piecewise constant: planar-shaped features are fitted using several constant patches, and some fine features are skipped. The results in Figure 5.2(b) and Figure 5.3(c) and (d) clearly show that nonsmooth at zero PFs are not adapted to the restoration of smoothly varying regions.



(a) $\varphi(t) = \alpha t^2/(1 + \alpha t^2)$ (b) $\varphi(t) = \min\{\alpha t^2, 1\}$

(c) $\varphi(t) = \alpha|t|/(1 + \alpha|t|)$ (d) $\varphi(t) = 1 - \mathbb{1}_{(t=0)}$

FIG. 5.3. *Restoration using nonconvex PFs. First row (*(a) *and* (b)*): smooth at zero PFs. Second row (*(c) *and* (d)*): nonsmooth at zero PFs.*

6. Conclusion: Further interpretation of the results. Each local minimizer $\hat{x}$ of $\mathcal{F}_y$ can be seen as resulting from a local minimizer function $y \to \chi(y)$ defined on a subset of $U \subset \mathbb{R}^q$, i.e., $\hat{x} = \chi(y)$. It has been established in [12] that when $A^T A$ is invertible, local minimizer functions are $\mathcal{C}^1$-continuous on their domains. Consequently, a local minimizer function $\chi : U \to \mathbb{R}^p$ produces minimizers $\hat{x} = \chi(y)$, for $y \in U$, that have the same set of edges $\widehat{J}_1$, that is, $\{i \in J : g_i^T \chi(y) \geq \theta_1\} = \widehat{J}_1$, for all $y \in U$. (To prove it, notice that the converse would contradict the continuity of χ on U.)

Given $y \in \mathbb{R}^q$, let $\mathcal{F}_y$ reach its global minimum at $\hat{x} = \chi(y)$ with edges indexed by $\widehat{J}_1$ and homogeneous regions indexed by $\widehat{J}_0$. When data vary in a neighborhood of y in such a way that noticeable edges either appear or disappear in the original signal or image, the *global* minimum will jump from the (local) minimizer function χ with edges $\widehat{J}_1$ to another (local) minimizer function χ' whose edges are $\widehat{J}_1' \neq \widehat{J}_1$. This discontinuity of the *global minimizer function* is the property that allows edges to be detected or removed at the global minimum of $\mathcal{F}_y$. Using the results of [13], such discontinuities occur only at data points included in a negligible subset of $\mathbb{R}^q$.

In contrast, if $\mathcal{F}_y$ is strictly convex, there is a unique minimizer function $\chi : \mathbb{R}^q \to \mathbb{R}^p$, and the latter is continuous. In particular, differences $g_i^T \hat{x}$ can take any value on $\mathbb{R}$. The edge-preservation properties of $\varphi(t) = |t|$—the famous total-variation regularization—have been extensively discussed in the literature. We should emphasize that they are based on a totally different property. As explained in [30, 33], the relevant minimizers $\hat{x}$ exhibit *stair-casing*: for many differences, $g_i^T \hat{x} = 0$, so $\hat{x}$ contains constant regions. The nonzero differences that separate the constant regions in $\hat{x}$ then naturally appear as edges. This effect is observed in Figure 5.2(b), where numerous spurious edges appear on planar-shaped regions.

Thus, image and signal restoration using nonconvex regularization is fundamentally different from restoration using convex regularization. The main difference is related to the (dis)continuity of the global minimizers with respect to the data.

7. Appendix.

A cost-function on $\mathbb{R}$ (section 2)—details. The considerations below are nicely illustrated in Figure 2.1, on the left for H3 and on the right for H4.

Local minimizer $\hat{x}_0 \in [0, \theta_0]$ for $y \in [0, h_0)$. Consider that H3 holds. Using (2.9), the equation in (2.6) has a solution $\hat{x}_0 \in [0, \theta_0]$. However, for no $y \in [0, h_0)$ it can be satisfied by $\hat{x}_0 = \theta_0$. Hence $\hat{x}_0 < \theta_0$. Combining this with (2.5) and H3 shows that $\varphi''(\hat{x}_0) > -\frac{2}{\beta}$, and hence $\mathcal{F}_y''(\hat{x}_0) > 0$. Hence $\mathcal{F}_y$ has a strict minimum at $\hat{x}_0$.

Now let H4 hold. By (2.7) and (2.9), $\mathcal{F}_y$ may have a (local) minimum at $\hat{x}_0 = 0$. In order to check this possibility, let us consider

$$\mathcal{K}(u) = \mathcal{F}_y(u) - \mathcal{F}_y(0) = u^2 - 2uy + \beta\varphi(u).$$

Suppose $\varphi'(0^+) < +\infty$. Since $\frac{2y}{\beta} \in [0, \varphi'(0^+))$, by the definition of $\varphi'(0^+)$, there is $\varepsilon > 0$ such that

$$u \in (0, \varepsilon) \quad \Rightarrow \quad \varphi(u) \geq \frac{2y}{\beta} u.$$

Then for every $u \in (0, \varepsilon)$ we find that $\mathcal{K}(u) \geq u^2 > 0$.

If $\varphi'(0^+) = +\infty$, there is $\varepsilon > 0$ such that $\frac{\varphi(u)}{u} \geq \frac{2}{\beta} y$ for all $u \in (0, \varepsilon)$. Then $\mathcal{K}(u) = u^2 + \beta(\frac{\varphi(u)}{u} - \frac{2}{\beta} y)u > 0$ for all $u \in (0, \varepsilon)$. Combining these results with the observation that $\mathcal{K}(u) > 0$ for all $u < 0$ shows that $\mathcal{F}_y$ has a strict (local) minimum at $\hat{x}_0 = 0$.

Local minimizer $\hat{x}_1 > \theta_1$ *for* $y > h_1$. Now the equation in (2.6) admits a solution $\hat{x}_1 \in [\theta_1, \infty)$. Since (2.6) cannot be satisfied for θ_1, we have $\hat{x}_1 > \theta_1$. This, combined with (2.5) and with either H3 or H4, shows that $\varphi''(\hat{x}_1) > -\frac{2}{\beta}$, and hence $\mathcal{F}_y''(\hat{x}_1) > 0$. Thus $\mathcal{F}_y$ admits a strict local minimum at $\hat{x}_1$.

Proof of (2.10). We first consider the possibility that $\mathcal{F}_{h_1}$ has a local minimum at $\chi_1(h_1) = \theta_1$. Noticing that for any $u > 0$, φ'' is continuous and satisfies $\varphi''(\theta_1 + u) > \varphi''(\theta_1)$, we can write that

$$\begin{aligned}\varphi(\theta_1 + u) - \varphi(\theta_1) - u\varphi'(\theta_1) &= u^2 \int_0^1 (1-t)\varphi''(\theta_1 + tu)dt \\ &< \frac{u^2}{2}\varphi''(\theta_1) = -\frac{u^2}{\beta},\end{aligned}$$

where the last equality comes from (2.5). It follows that for any $u > 0$ we have

$$\mathcal{F}_{h_1}(\theta_1 + u) - \mathcal{F}_{h_1}(\theta_1) = u^2 + \beta\left(\varphi(\theta_1 + u) - \varphi(\theta_1) - u\varphi'(\theta_1)\right) < 0.$$

Hence $\mathcal{F}_{h_1}$ does not have any local minimum at θ_1.

Now we focus on the possibility that $\mathcal{F}_{h_0}$ has a local minimum at $\chi_0(h_0) = \theta_0$. When H3 holds, similar reasoning shows that $\mathcal{F}_{h_0}$ does not have any local minimum at $\chi_0(h_0) = \theta_0$.

Consider next that H4 holds. Let us consider the function $\mathcal{K} : [0, \varepsilon) \to \mathbb{R}$,

$$\mathcal{K}(u) = \mathcal{F}_{h_0}(u) - \mathcal{F}_{h_0}(0) = u^2 + \beta(\varphi(u) - u\varphi'(0^+)),$$

where the second equality comes from (2.9). Using (2.2), $-\frac{2}{\beta\varphi''(0^+)} \in (0, 1)$. Let us choose $\eta \in (-\frac{2}{\beta\varphi''(0^+)}, 1)$. Then there is $\varepsilon > 0$ such that

$$0 < u < \varepsilon \quad \Rightarrow \quad \varphi''(u) \le \eta\varphi''(0^+).$$

Using that $\beta\eta\varphi''(0^+) < -2$, for any $u \in (0, \varepsilon)$ we have

$$\mathcal{K}''(u) = 2 + \beta\varphi''(u) \le 2 + \beta\eta\varphi''(0^+) < 0.$$

Hence $\mathcal{K}'$ is strictly decreasing on $(0, \varepsilon)$. Combining the latter with $\mathcal{K}'(0^+) = 0$ shows that $\mathcal{K}'(u) < 0$ if $0 < u < \varepsilon$. Hence $\mathcal{K}$ is strictly decreasing on $(0, \varepsilon)$ as well. Combining this with $\mathcal{K}(0) = 0$ shows that $\mathcal{K}(u) < 0$ if $0 < u < \varepsilon$. The latter shows that $\mathcal{F}_{h_0}$ does not have any local minimum at $\chi_0(h_0) = 0$.

Proof of (2.11). For either $\chi = \chi_1$ or $\chi = \chi_0$ if H3 holds, we can write that

$$\mathcal{F}_y(\chi(y)) = (\chi(y) - y)^2 + \beta\varphi(\chi(y)).$$

Using that $\chi(y)$ satisfies (2.6), we get

$$\frac{d\mathcal{F}_y}{dy}(\chi(y)) = 2(\chi(y) - y)(\chi'(y) - 1) + \beta\varphi'(\chi(y))\chi'(y) = 2(y - \chi(y)).$$

For $\chi = \chi_0 = 0$ under H4 we find $\mathcal{F}_y(\chi(y)) = y^2$ for all $y \in [0, h_0)$; hence (2.11) holds again.

Proof of Proposition 3.2. Consider first that $r < p$. Using the change of variables $x \to (z, z_b)$ given in (3.10), we consider $\widehat{F}_y(z, z_b) = \mathcal{F}_y(Hz + H_b z_b)$, namely

$$\widehat{F}_y(z, z_b) = \|Bz + B_b z_b - y\|^2 + \beta \sum_{i \in J} \varphi(z[i]), \tag{7.1}$$

where B and B_b are given in (3.11). Let $\hat{x}$ be a minimizer of $\mathcal{F}_y$. Equivalently, $(\hat{z}, \hat{z}_b)$, where $\hat{z} = G\hat{x}$ and $\hat{z}_b = G_b\hat{x}$, is a minimizer of F_y. Using that in particular the derivative of $\widehat{F}_y$ with respect to z_b is zero, we have $\hat{z}_b = \sigma_y(\hat{z})$, where $\sigma_y : \mathbb{R}^r \to \mathbb{R}^{p-r}$ reads

$$\sigma_y(z) = -\left(B_b^T B_b\right)^{-1} B_b^T (Bz - y).$$

Put $F_y(z) = \widehat{F}_y(z, \sigma_y(z))$, i.e.,

$$F_y(z) = \|P(Bz - y)\|^2 + \beta \sum_{i=1}^{r} \varphi(z[i]), \tag{7.2}$$

where P is as given in (3.12). If $p = r$, then z_b is empty, and (7.2) holds with $P = I$. Clearly, $\hat{x}$ is a global minimizer of $\mathcal{F}_y$ if and only if $\hat{z} = G\hat{x}$ is a global minimizer of F_y. In the following we focus on F_y. Our reasoning relies on the observation that if $\hat{z}$ is a global minimizer of F_y, then for any $i \in J$, the function $f_i : \mathbb{R} \to \mathbb{R}$,

$$f_i(t) = F_y\left(\hat{z} + (t - \hat{z}[i])e_i\right),$$

has a global minimum at $\hat{t} = \hat{z}[i]$. After some elementary calculations, we can write that

$$f_i(t) = t^2 \, \|Pb_i\|^2 + 2t \, w_i^T Pb_i + \beta\varphi(t) + \kappa_i, \tag{7.3}$$

$$\text{where} \qquad \begin{aligned} b_j &= Be_j, \quad j = 1, \dots, r, \\ w_i &= \sum_{j \in J \setminus \{i\}} b_j \hat{z}[j] - y, \\ \kappa_i &= \|Pw_i\|^2 + \beta \sum_{j \in J \setminus \{i\}} \varphi(\hat{z}[j]). \end{aligned}$$

If $Pb_i = 0$, the function f_i has a unique minimum at $\hat{t} = 0$, which entails that $g_i^T \hat{x} = 0$. Hence we have proved (i).

Next, we consider that $Pb_i \neq 0$. Put

$$\chi_0 = -\frac{w_i^T Pb_i}{\|Pb_i\|^2 + \alpha\beta} \quad \text{and} \quad \chi_1 = -\frac{w_i^T Pb_i}{\|Pb_i\|^2}. \tag{7.4}$$

If $|w_i^T Pb_i| < \frac{\|Pb_i\|^2}{\sqrt{\alpha}}$, then f_i has a unique minimizer which reads $\hat{t} = \chi_0$ and satisfies

$$|\hat{t}| < \frac{\|Pb_i\|^2}{\sqrt{\alpha}\,\left(\|Pb_i\|^2 + \alpha\beta\right)} < \frac{1}{\sqrt{\alpha}} \Gamma_i.$$

If $|w_i^T Pb_i| > \frac{\|Pb_i\|^2 + \alpha\beta}{\sqrt{\alpha}}$, then f_i has a unique minimizer which reads $\hat{t} = \chi_1$ and satisfies

$$|\hat{t}| > \frac{\|Pb_i\|^2 + \alpha\beta}{\sqrt{\alpha}\,\left(\|Pb_i\|^2\right)} > \frac{1}{\sqrt{\alpha}\,\Gamma_i}.$$

If $\frac{\|Pb_i\|^2}{\sqrt{\alpha}} \le |w_i^T Pb_i| \le \frac{\|Pb_i\|^2+\alpha\beta}{\sqrt{\alpha}}$, then f_i has two local minimizers reached for χ_0 and χ_1. So, the global minimizer $\hat{z}$ of F_y is either $\hat{z} = \hat{z}_0$ or $\hat{z} = \hat{z}_1$, where

$$\hat{z}_0 = \hat{z} + (\chi_0 - \hat{z}[i])e_i \quad \text{and} \quad \hat{z}_1 = \hat{z} + (\chi_1 - \hat{z}[i])e_i. \tag{7.5}$$

Let us denote $\Delta = F_y(\hat{z}_0) - F_y(\hat{z}_1)$. We have $\hat{z} = \hat{z}_0$ if $\Delta \le 0$ and $\hat{z} = \hat{z}_1$ if $\Delta \ge 0$. Using that

$$f_i(\chi_0) = -\frac{(w_i^T Pb_i)^2}{\|Pb_i\|^2 + \alpha\beta} + \kappa_i \quad \text{and} \quad f_i(\chi_1) = -\frac{(w_i^T Pb_i)^2}{\|Pb_i\|^2} + \beta + \kappa_i,$$

it is easily found that

$$\Delta = f_i(\chi_0) - f_i(\chi_1) = \frac{\alpha\beta(w_i^T Pb_i)^2}{\|Pb_i\|^2(\|Pb_i\|^2 + \alpha\beta)} - \beta.$$

Consequently,

$$\Delta \le 0 \quad \Rightarrow \quad |\hat{z}[i]| = \chi_0 \le \sqrt{\frac{\|Pb_i\|^2}{\alpha(\|Pb_i\|^2 + \alpha\beta)}} = \frac{1}{\sqrt{\alpha}}\,\Gamma_i,$$

$$\Delta \ge 0 \quad \Rightarrow \quad |\hat{z}[i]| = \chi_1 \ge \sqrt{\frac{\|Pb_i\|^2 + \alpha\beta}{\alpha\|Pb_i\|^2}} = \frac{1}{\sqrt{\alpha}\,\Gamma_i}.$$

Hence we have proved (3.14). If F_y has a unique global minimizer, we have either $\Delta < 0$ or $\Delta > 0$, which implies that the inequalities in (3.14) are strict.

Proof of Proposition 3.4. The reasoning here is similar to the proof of Proposition 3.2. We first obtain the equivalent cost-function given in (7.2) and then check the minimizers of $f_i : \mathbb{R} \to \mathbb{R}$ as introduced in (7.3). Define now χ_0 and χ_1 by

$$\chi_0 = 0 \quad \text{and} \quad \chi_1 = -\frac{w_i^T Pb_i}{\|Pb_i\|^2} \ne 0.$$

If $PBe_i = 0$, (7.3) shows that $f_i(t) = \beta\varphi(t) + \kappa_i$: the global minimum is reached for $\hat{t} = 0$, which entails that $g_i^T \hat{x} = 0$, as stated in (i). Consider next that $PBe_i \ne 0$. In all cases, f_i has a local minimum at $\hat{t} = 0$; notice that if $w_i^T Pb_i = 0$, this is the unique minimizer of f_i. If $w_i^T Pb_i \ne 0$, there is another local minimum at $\hat{t} = \chi_1$. So, the global minimizer of F_y is either $\hat{z} = \hat{z}_0$ or $\hat{z} = \hat{z}_1$, where $\hat{z}_0$ and $\hat{z}_1$ read as in (7.5). Using that

$$f_i(0) = \kappa_i \quad \text{and} \quad f_i(\chi_1) = -\frac{(w_i^T Pb_i)^2}{\|Pb_i\|^2} + \beta + \kappa_i,$$

it is found that

$$\Delta = F_y(z_1) - F_y(z_2) = f_i(0) - f_i(\chi_1) = \frac{(w_i^T Pb_i)^2}{\|Pb_i\|^2} - \beta.$$

It follows that

$$\Delta \le 0 \quad \Rightarrow \quad |\hat{z}[i]| = \chi_0 = 0,$$

$$\Delta \ge 0 \quad \Rightarrow \quad |\hat{z}[i]| = \chi_1 \ge \frac{\sqrt{\beta}}{\|Pb_i\|}.$$

Hence we have proved (3.30). If F_y has a unique global minimizer, we have either $\Delta < 0$ or $\Delta > 0$; hence the inequalities in (3.30) are strict.

Proof of Proposition 4.2. For any $x \in \mathbb{R}^p$ satisfying $|g_i^T x| \neq \frac{1}{\sqrt{\alpha}}$ for all $i \in J$, the function F_h is $\mathcal{C}^2$. Its second differential $D^2 F_h(x)$ is of the form (4.7), where

$$\varphi''(g_i^T x) = \begin{cases} 2\alpha & \text{if} \quad |g_i^T x| < \frac{1}{\sqrt{\alpha}}, \\ 0 & \text{if} \quad |g_i^T x| > \frac{1}{\sqrt{\alpha}}. \end{cases}$$

This, combined with the invertibility of $A^T A$, entails that $D^2 F_h(x)$ is positive definite. It has been shown in [33] that if $\hat{x}$ is a (local) minimizer of F_h, then

$$|g_i^T \hat{x}| \neq \frac{1}{\sqrt{\alpha}} \quad \forall i \in J. \tag{7.6}$$

Hence $\hat{x}$ satisfies $DF_h(\hat{x}) = 0$, and the minimum of F_h at $\hat{x}$ is strict.

Implication (4.11). We start with analyzing the sign of the constant κ,

$$\kappa = (A\mathbb{1}_\Sigma)^T A\,(\mathbb{1}_\Sigma - \chi_{\Sigma}) = (A\mathbb{1}_\Sigma)^T \left(I - A(A^T A + \beta\alpha G^T G)^{-1} A^T\right) A\mathbb{1}_\Sigma, \tag{7.7}$$

where χ_Σ is given in (4.10). Let us notice first[8] that all the eigenvalues of $A(A^T A + \beta\alpha G^T G)^{-1} A^T$ are in $[0, 1]$. Then it follows that $\kappa \geq 0$. If $\kappa = 0$, then

$$A(A^T A + \beta\alpha G^T G)^{-1} A^T A\mathbb{1}_\Sigma = A\mathbb{1}_\Sigma.$$

Since $A^T A$ is invertible, this is equivalent to $(A^T A + \beta\alpha G^T G)^{-1} A^T A\mathbb{1}_\Sigma = \mathbb{1}_\Sigma$. Consequently, $G^T G\mathbb{1}_\Sigma = 0$. By H6, this is impossible unless Σ is empty. It follows that $\kappa > 0$.

We will consider that $h \in (0, h_0)$, where

$$h_0 = \min\left\{ \left(\sqrt{\alpha}\ \max_{i \in J} \left|g_i^T \chi_\Sigma\right|\right)^{-1},\ \sqrt{\frac{\beta}{\kappa}}\ \right\}. \tag{7.8}$$

Let us examine the possibility that F_h has a (local) minimizer $\hat{x}$ satisfying

$$|g_i^T \hat{x}| < \frac{1}{\sqrt{\alpha}} \quad \forall i \in J. \tag{7.9}$$

[8] Denote $M = \beta\alpha G^T G$. Let λ and v be such that $A(A^T A + M)^{-1} A^T v = \lambda v$; then clearly $\lambda \geq 0$. If $A^T v = 0$, then $\lambda = 0$. In the following, consider that $A^T v \neq 0$. Using that $A^T A$ is invertible, we deduce that

$$\left(A^T A + M\right)^{-1} A^T v = \lambda (A^T A)^{-1} A^T v.$$

Multiplying both sides of this equation by $v^T A(A^T A)^{-1}(A^T A + M)$ yields

$$\begin{aligned} v^T A(A^T A)^{-1} A^T v &= \lambda v^T A(A^T A)^{-1} \left(A^T A + M\right) (A^T A)^{-1} A^T v \\ &= \lambda v^T A(A^T A)^{-1} A^T v + \lambda v^T A(A^T A)^{-1} M (A^T A)^{-1} A^T v. \end{aligned}$$

If we denote $c_1 = v^T A(A^T A)^{-1} A^T v$ and $c_2 = v^T A(A^T A)^{-1} M(A^T A)^{-1} A^T v$, the latter equation becomes

$$(1 - \lambda)c_1 = \lambda c_2.$$

Since $A^T v \neq 0$, we have $c_1 > 0$. Combining this with the facts that $c_2 \geq 0$ and $\lambda \geq 0$ shows that $1 - \lambda \geq 0$.

In such a case, $\varphi'(g_i^T\hat{x}) = 2\alpha\, g_i^T\hat{x}$ for all $i \in J$. Introducing this into (4.6) shows that the equation $DF_h(\hat{x}) = 0$ has a unique solution which reads $\hat{x} = h\chi_{\Sigma}$. Since $h \in (0, h_0)$, it is easy to check that (7.9) is satisfied and that

$$F_h(\hat{x}) = h^2(A\mathbb{1}_\Sigma)^T A\,(\mathbb{1}_\Sigma - \chi_\Sigma) = h^2\kappa < \beta.$$

If $\tilde{x} \neq \hat{x}$ is another (local) minimizer, there is $i \in J$ such that $|g_i^T\tilde{x}| > \frac{1}{\sqrt{\alpha}}$; hence

$$F_h(\tilde{x}) \geq \beta > F_h(\hat{x}).$$

It follows that if $h \in [0, h_0)$, then F_h reaches its global minimum at $\hat{x}_h = h\chi_\Sigma$.

Implication (4.12). In the following, suppose that $h \geq h_1$, where

$$h_1 = \frac{1}{\sqrt{\alpha}\ \ \gamma_{\min}} + \sqrt{\frac{2\beta\sharp J_1}{\alpha_{\min}}}. \tag{7.10}$$

Let $\hat{x}$ be a global minimizer of F_h. With $\hat{x}$ we associate the subsets $\widehat{J}_0$ and $\widehat{J}_1$ defined by

$$\widehat{J}_0 = \left\{ i \in J : \left|g_i^T\hat{x}\right| < \frac{1}{\sqrt{\alpha}} \right\} \quad \text{and} \quad \widehat{J}_1 = \left\{ i \in J : \left|g_i^T\hat{x}\right| > \frac{1}{\sqrt{\alpha}} \right\}.$$

Using (7.6), $J = \widehat{J}_0 \cup \widehat{J}_1$. Then we apply the reasoning behind the proof of Theorem 4.1(ii) with the modifications explained next. In item (C1) we take $\theta_0 = \frac{1}{\sqrt{\alpha}}$. In item (C2) we consider $\varphi(g_i^T\hat{x}) = 1$ for all $i \in \widehat{J}_1$, which leads directly to (4.9). In this way we find that $\widehat{J}_0 = J_0$ and $\widehat{J}_1 = J_1$.

If $\hat{x} \neq h\mathbb{1}_\Sigma$, using that A^TA is invertible we find $\|A(\hat{x} - h\mathbb{1}_\Sigma)\| > 0$, and hence $F_h(\hat{x}) > F_h(h\mathbb{1}_\Sigma)$. It follows that the global minimum is reached for $\hat{x} = h\mathbb{1}_\Sigma$.

Proof of Lemma 4.5.

Statement (i). Let us define ρ by

$$\rho = \min\left\{ \min_{i\in\widehat{J}_1} \left|g_i^T\hat{x}\right| \frac{1}{\max\limits_{i\in J}\|g_i\|},\ \frac{\beta}{2\|A^T(A\hat{x} - y)\| + 1} \right\}.$$

For any $u \in \mathbb{R}^p$ such that $0 < \|u\| < \rho$, define $\Delta(u)$ by

$$\begin{aligned}\Delta(u) &= \mathcal{F}_y(\hat{x} + u) - \mathcal{F}_y(\hat{x}) \\ &= \|A(\hat{x} + u) - y\|^2 - \|A\hat{x} - y\|^2 \\ &\quad + \beta\sum_{i\in\widehat{J}_1}\varphi(g_i^T(\hat{x} + u)) - \beta\sum_{i\in\widehat{J}_1}\varphi(g_i^T\hat{x}) + \beta\sum_{i\in\widehat{J}_0}\varphi(g_i^T(\hat{x} + u)) - \beta\sum_{i\in\widehat{J}_0}\varphi(g_i^T\hat{x}).\end{aligned}$$

Since $g_i^T\hat{x} = 0$ for all $i \in \widehat{J}_0$, the last term in the expression above vanishes. Furthermore,

$$\begin{aligned} 0 \leq \|u\| < \rho \ &\Rightarrow\ \left|g_i^T(\hat{x} + u)\right| > 0 && \forall i \in \widehat{J}_1, \\ &\Rightarrow\ \varphi\left(g_i^T(\hat{x} + u)\right) = 1 = \varphi\left(g_i^T\hat{x}\right) && \forall i \in \widehat{J}_1. \end{aligned}$$

It follows that

$$\Delta(u) = \|A(\hat{x} + u) - y\|^2 - \|A\hat{x} - y\|^2 + \beta\sum_{i\in\widehat{J}_0}\varphi(g_i^Tu). \tag{7.11}$$

($\Rightarrow$) Since $\hat{x}$ is a (local) minimizer of $\mathcal{F}_y$, there is $\rho_0 \in (0, \rho)$ such that $\Delta(u) \geq 0$ for all $u \in B(0, \rho_0)$. Noticing that $\hat{x} \in K(\widehat{J}_1)$ and that the last term in (7.11) vanishes if $u \in K(\widehat{J}_1)$, it is seen that

$$\|A(\hat{x}+u) - y\|^2 - \|A\hat{x} - y\|^2 \geq 0 \quad \forall u \in K(\widehat{J}_1) \cap B(0, \rho_0). \tag{7.12}$$

Clearly, the inequality above holds for every $u \in K(\widehat{J}_1)$.

($\Leftarrow$) Let $u \in \mathbb{R}^p$ be such that $0 < \|u\| < \rho$. We can decompose u into

$$u = u_0 + u_1, \quad \text{where} \quad u_0 \in K(\widehat{J}_1) \ \text{ and } \ u_1 \in \left(K(\widehat{J}_1)\right)^{\perp}.$$

Consider first that $u_1 \neq 0$. Then there is $j \in \widehat{J}_0$ such that $g_i^T u_1 \neq 0$, and hence $\varphi(g_i^T u) = 1$. It follows that

$$\Delta(u) \geq \|Au\|^2 + 2(A\hat{x} - y)^T Au + \beta \geq -2\|A^T(A\hat{x} - y)\| \, \|u\| + \beta > 0. \tag{7.13}$$

If $u_1 = 0$, then $u \in K(\widehat{J}_1) \cap B(0, \rho_0)$, in which case $\Delta(u) = \|A(\hat{x}+u) - y\|^2 - \|A\hat{x} - y\|^2 \geq 0$ since (7.12) holds by assumption. It follows that $\Delta(u) \geq 0$ for all $u \in B(0, \rho_0)$.

Statement (ii). The case when $h = 0$ is trivial. Consider next that $h \neq 0$. Since $\hat{x} \in K(\widehat{J}_1)$, then $h\hat{x} \in K(\widehat{J}_1)$ as well. Using (i), we have to show that $h\hat{x}$ minimizes $x \to \|Ax - hy\|^2$ on $K(\widehat{J}_1)$. For any $u \in K(\widehat{J}_1)$, we have

$$\|A(h\hat{x}+u) - hy\|^2 - \|A\hat{x} - hy\|^2 = h^2 \left(\left\| A\left(\hat{x} + \frac{u}{h}\right) - y \right\|^2 - \|A\hat{x} - y\|^2 \right) \geq 0,$$

where the inequality comes from the observation that $\hat{x}$ minimizes $x \to \|Ax - y\|^2$ on $K(\widehat{J}_1)$.

Proof of Proposition 4.6.

Implication (4.19). We will show the statement for h_0 given by

$$h_0 = \sqrt{\frac{\beta}{\xi}}, \tag{7.14}$$

where $\xi > 0$ is defined in (4.15). Using Lemma 4.5, F_h has a minimizer $\hat{x} \in K(\emptyset)$; using H6, it reads $\hat{x} = \hat{c}\,\mathbb{1}$, where $\hat{c}$ minimizes the function

$$c \to \|A(c\mathbb{1} - h\mathbb{1}_\Sigma)\|^2;$$

hence $\hat{c} = h\zeta$ and $F_h(\hat{c}\mathbb{1}) = h^2\xi$. This is the unique minimizer of F_h belonging to $K(\emptyset)$.

Let $\tilde{x} \neq \hat{x}$ be another (local) minimizer of F_h. Then $\tilde{x} \notin K(\emptyset)$. Using H6, there is $j \in J$ such that $g_j^T \tilde{x} \neq 0$, and hence $\varphi(g_j^T \tilde{x}) = 1$. It follows that for $h \in (0, h_0)$,

$$F_h(\tilde{x}) \geq \beta = h_0^2 \xi > F_h(\hat{c}\mathbb{1}).$$

Implication (4.20). Given $\mathrm{J}_1 \subset \mathcal{J}_1$, let $u(\mathrm{J}_1) \in \mathbb{R}^p$ be the unique solution to

$$\text{minimize} \quad \|A(x - \mathbb{1}_\Sigma)\|^2 \quad \text{subject to} \quad x \in K(\mathrm{J}_1). \tag{7.15}$$

Then define

$$\kappa = \min\left\{\|A\,(u(\mathrm{J}_1) - 1\!\!1_\Sigma)\|^2 \ : \ \mathrm{J}_1 \subset \mathcal{J}_1 \setminus \{J_1\}\right\}.$$

Since $A^T A$ is invertible, $\|A(u(\mathrm{J}_1) - 1\!\!1_\Sigma)\|^2 > 0$ if $\mathrm{J}_1 \neq J_1$; hence $\kappa > 0$. We will consider that $h > h_1$ for

$$h_1 = \sqrt{\frac{\beta \sharp J_1}{\kappa}}.$$

Let $\hat{x}$ be a (local) minimizer of $\mathcal{F}_h$, and let $\widehat{J}_0$ and $\widehat{J}_1$ be defined according to (3.22). Consider first the possibility that $\widehat{J}_1 \neq J_1$. Using Lemma 4.5(ii),

$$\hat{x} = h\; u(\widehat{J}_1),$$

where $u(\widehat{J}_1)$ is the solution to (7.15). Since $h > h_1$,

$$\mathcal{F}_h(\hat{x}) = h^2 \left\|A\left(u(\widehat{J}_1) - 1\!\!1_\Sigma\right)\right\|^2 + \beta\sharp\widehat{J}_1 \geq h^2\kappa > \beta\sharp J_1 = \mathcal{F}_h(h1\!\!1_\Sigma).$$

It follows that no $\hat{x}$ such that $\widehat{J}_1 \neq J_1$ is a global minimizer of $\mathcal{F}_h$. Reciprocally, if $\mathcal{F}_h$ reaches its global minimum at $\hat{x}$, then $\widehat{J}_1 = J_1$ (and, equivalently, $\widehat{J}_0 = J_0$). Using Lemma 4.5, we find that $\hat{x} = h1\!\!1_\Sigma$.

REFERENCES

[1] S. Alliney, *Digital filters as absolute norm regularizers*, IEEE Trans. Signal Process., 40 (1992), pp. 1548–1562.

[2] S. Alliney, *A property of the minimum vectors of a regularizing functional defined by means of absolute norm*, IEEE Trans. Signal Process., 45 (1997), pp. 913–917.

[3] G. Aubert and P. Kornprobst, *Mathematical Problems in Images Processing*, Springer-Verlag, Berlin, 2002.

[4] J. E. Besag, *Digital image processing: Towards Bayesian image analysis*, J. Appl. Statist., 16 (1989), pp. 395–407.

[5] A. Blake and A. Zisserman, *Visual Reconstruction*, MIT Press, Cambridge, MA, 1987.

[6] C. Bouman and K. Sauer, *A generalized Gaussian image model for edge-preserving* MAP *estimation*, IEEE Trans. Image Process., 2 (1993), pp. 296–310.

[7] A. Chambolle, *Finite-differences discretizations of the Mumford-Shah functional*, M2AN Math. Model. Numer. Anal., 33 (1999), pp. 261–288.

[8] T. F. Chan and S. Esedoḡlu, *Aspects of total variation regularized L_1 function approximation*, SIAM J. Appl. Math., 65 (2005), pp. 1817–1837.

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

[10] G. Demoment, *Image reconstruction and restoration: Overview of common estimation structure and problems*, IEEE Trans. Acoust. Speech Signal Process., 37 (1989), pp. 2024–2036.

[11] D. Donoho, I. Johnstone, J. Hoch, and A. Stern, *Maximum entropy and the nearly black object*, J. Roy. Statist. Soc. Ser. B, 54 (1992), pp. 41–81.

[12] S. Durand and M. Nikolova, *Stability of minimizers of regularized least squares objective functions* I: *Study of the local behaviour*, Appl. Math. Optim., to appear.

[13] S. Durand and M. Nikolova, *Stability of minimizers of regularized least squares objective functions* II: *Study of the global behaviour*, Appl. Math. Optim., to appear.

[14] D. Geman, *Random fields and inverse problems in imaging*, in École d'Été de Probabilités de Saint-Flour XVIII—1988, Lecture Notes in Math. 1427, Springer-Verlag, Berlin, 1990, pp. 113–193.

[15] D. Geman and G. Reynolds, *Constrained restoration and recovery of discontinuities*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 14 (1992), pp. 367–383.

[16] D. Geman and C. Yang, *Nonlinear image recovery with half-quadratic regularization*, IEEE Trans. Image Process., 4 (1995), pp. 932–946.

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

[18] S. Geman and D. E. McClure, *Statistical methods for tomographic image reconstruction*, Bull. Inst. Internat. Statist., 52 (1987), pp. 5–21.

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

[20] T. Hebert and R. Leahy, *A generalized EM algorithm for* 3*-D Bayesian reconstruction from Poisson data using Gibbs priors*, IEEE Transactions on Medical Imaging, 8 (1989), pp. 194–202.

[21] K. Lange, *Convergence of EM image reconstruction algorithms with Gibbs priors*, IEEE Transactions on Medical Imaging, 9 (1990), pp. 439–446.

[22] Y. Leclerc, *Constructing simple stable description for image partitioning*, International Journal of Computer Vision, 3 (1989), pp. 73–102.

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

[24] J.-M. Morel and S. Solimini, *Segmentation of images by variational methods: A constructive approach*, Rev. Mat. Univ. Complut. Madrid, 1 (1988), pp. 169–182.

[25] J.-M. Morel and S. Solimini, *Variational Methods in Image Segmentation*, Birkhäuser, Basel, 1995.

[26] D. Mumford and J. Shah, *Boundary detection by minimizing functionals*, in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 1985, pp. 22–26.

[27] D. Mumford and J. Shah, *Optimal approximations by piecewise smooth functions and associated variational problems*, Comm. Pure Appl. Math., 42 (1989), pp. 577–685.

[28] M. Nikolova, *Estimées localement fortement homogènes*, C. R. Acad. Sci. Paris Sér. I Math., 325 (1997), pp. 665–670.

[29] M. Nikolova, *Markovian reconstruction using a GNC approach*, IEEE Trans. Image Process., 8 (1999), pp. 1204–1220.

[30] M. Nikolova, *Local strong homogeneity of a regularized estimator*, SIAM J. Appl. Math., 61 (2000), pp. 633–658.

[31] M. Nikolova, *Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers*, SIAM J. Numer. Anal., 40 (2002), pp. 965–994.

[32] M. Nikolova, *A variational approach to remove outliers and impulse noise*, J. Math. Imaging Vision, 20 (2004), pp. 99–120.

[33] M. Nikolova, *Weakly constrained minimization. Application to the estimation of images and signals involving constant regions*, J. Math. Imaging Vision, 21 (2004), pp. 155–175.

[34] P. Perona and J. Malik, *Scale-space and edge detection using anisotropic diffusion*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12 (1990), pp. 629–639.

[35] W. Ring, *Structural properties of solutions of total variation regularization problems*, M2AN Math. Model. Numer. Anal., 34 (2000), pp. 799–810.

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

[37] S. Saquib, C. Bouman, and K. Sauer, *ML parameter estimation for Markov random fields, with applications to Bayesian tomography*, IEEE Trans. Image Process., 7 (1998), pp. 1029–1044.

[38] G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Welk, *On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs*, SIAM J. Numer. Anal., 42 (2004), pp. 686–713.

[39] R. Stevenson and E. Delp, *Fitting curves with discontinuities*, in Proceedings of the 1st International Workshop on Robust Computer Vision, Seattle, WA, 1990, pp. 127–136.

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

[41] C. R. Vogel and M. E. Oman, *Iterative method for total variation denoising*, SIAM J. Sci. Comput., 17 (1996), pp. 227–238.