Thresholding Implied by Truncated Quadratic Regularization

Mila Nikolova

***Abstract*—We address the problem of the estimation of an unknown signal that is known to involve sharp edges, from noisy data given at the output of a linear system. The sought solution is defined to be the global minimizer of an objective function combining a quadratic data-fidelity term and a regularization term. The latter term is a sum whose entries are obtained by applying a truncated quadratic potential function to every difference between adjacent samples. Such objective functions are naturally formulated either in a statistical framework, or in a variational framework, and they are customarily used in signal and image reconstruction. However, these objective functions are nonsmooth and highly nonconvex, and many questions related to their minimization, as well as to the features of the resulting solutions, remain open.**

In this paper, we present some new facts characterizing the features exhibited by the minimizers of such objective functions. Our main result states that the magnitude of the differences between adjacent samples of a global minimizer are either smaller than a first threshold or larger than a second, strictly larger threshold. Conversely, no difference corresponding to a global minimizer of the objective function can be placed among these thresholds for any data. This explains how edges are recovered in a signal and estimated using truncated quadratic regularization. These thresholds are independent of the data but are related to the observation system and to the regularization parameters. They can be used to derive necessary conditions for the choice of the regularization parameters. We also show that the chance to get data for which the objective function has two or more global minimizers is null. Numerical experiments corroborate the obtained theoretical results.

***Index Terms*—Denoising, edge-detection, inverse problems, MAP estimation, Mumford–Shah functional, reconstruction, regularized estimation, segmentation, stability, weak string.**

I. INTRODUCTION

WE ADDRESS the problem of the estimation of a signal $\boldsymbol{x} \in \mathbb{R}^n$ that is known to contain locally homogeneous zones separated by sharp edges from observed data $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}+\boldsymbol{n}$, where $\boldsymbol{y} \in \mathbb{R}^m$, $\boldsymbol{A}$: $\mathbb{R}^n \mapsto \mathbb{R}^m$ is a linear operator, and $\boldsymbol{n} \in \mathbb{R}^m$ is observation noise. Since [1], [2] the following estimator for $\boldsymbol{x}$

$$\hat{\boldsymbol{x}} := \arg\min_{\boldsymbol{x}} \mathcal{E}(\boldsymbol{x}, \boldsymbol{y}) \tag{1}$$

$$\mathcal{E}(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|^2 + \Phi(\boldsymbol{x}) \tag{2}$$

$$\Phi(\boldsymbol{x}) = \sum_{k=1}^{n-1} \varphi(x_k - x_{k+1}) \tag{3}$$

where φ: $\mathbb{R} \mapsto \mathbb{R}$ is a truncated quadratic potential function (PF) parameterized by $(\alpha > 0, \lambda > 0)$

$$\varphi(t) = \begin{cases} \lambda^2 t^2, & \text{if } |t| < \sqrt{\alpha}/\lambda \\ \alpha, & \text{if } |t| \geq \sqrt{\alpha}/\lambda \end{cases} \tag{4}$$

provokes a considerable theoretical and practical interest. Its rationale can be summarized as follows. The PF φ is applied to the differences between adjacent samples $|x_k - x_{k+1}|$ for $k = 1, \cdots, n-1$. Being quadratic for $|t| < \sqrt{\alpha}/\lambda$ and constant elsewhere, φ expresses a wish to smooth small differences ($|x_k - x_{k+1}| < \sqrt{\alpha}/\lambda$) while suspending smoothing over large differences ($|x_k - x_{k+1}| \geq \sqrt{\alpha}/\lambda$). A binary line variable that is equal to zero in the first case and to one in the second is naturally associated with each difference; see [2]. The estimation method (1)–(4) thus involves a segmentation level. In the sequel, we systematically assume that α and λ are strictly positive parameters. We will also identify $\boldsymbol{A}$ with a matrix of $\mathbb{R}^{m\times n}$.

In a statistical framework, the quadratic data-fidelity term in (2) can be seen as a log-likelihood of the data under the hypothesis that $\boldsymbol{n}$ is Gaussian white random noise. The regularization term Φ can be seen as the energy of a piecewise Gaussian Markov random chain [1], [3]–[6]. After [1], numerous works were dedicated to both problems, computing the estimate $\hat{\boldsymbol{x}}$ by using stochastic algorithms and selecting the parameters (α, λ) [3]–[5], [7], [8]. Various techniques for deterministic minimization of $\mathcal{E}(., \boldsymbol{y})$ have also been investigated [2], [9]–[17], [6]. Objective functions of the form (2)–(4) are equivalently derived in a variational framework [2], [9], [19]. In the case when $\boldsymbol{A}$ is the identity matrix, the shape of $\hat{\boldsymbol{x}}$ was considered for several synthetic noise-free data sets $\boldsymbol{y}$: a step, a characteristic function, and a ramp [2]. The objective function in (2)–(4) was transposed in a continuous setting in [20], which gave rise to important questions and to numerous works [21], [22]. More generally, solving estimation problems by minimizing objective functions of the form (1)–(3) by introducing various PFs φ in (3) is currently a very active research field [3], [21], [23]–[25]–[27]. Beyond all the experience that has been acquired, controlling the features of the estimate $\hat{\boldsymbol{x}}$ by means of the shape of φ is an intricate open question. This work can be seen as an attempt to address the latter question in the context of truncated quadratic PFs.

Our approach is to analyze some properties of the minimizer $\hat{\boldsymbol{x}}$ in (1) that are entailed by the specific form of the objective function (2)–(4). This paper presents some new facts characterizing such an estimator in the context of general observation systems $\boldsymbol{A}$, arbitrary original signals $\boldsymbol{x}$, and random noise perturbations. Recall that $\mathcal{E}(., \boldsymbol{y})$ is nonsmooth and highly nonconvex, and that up to now, very little is known about the behavior of its

Manuscript received June 14, 1999; revised August 25, 2000. This work was supported by the Société de Secours des Amis des Sciences. The associate editor coordinating the review of this paper and approving it for publication was Dr. Paulo J. S. G. Ferreira.

The author is with the Département Traitement du Signal et des Images, ENST, Paris, France. She is also with UFR Mathématiques et Informatique, Université René Descartes, Paris, France (e-mail: nikolova@tsi.enst.fr).

Publisher Item Identifier S 1053-587X(00)10163-1.

minimizers in such a general context. Our study points out several attractive properties exhibited by such minimizers. We start by showing that $\mathcal{E}(., \boldsymbol{y})$ is $\mathcal{C}^2$-continuous on a neighborhood surrounding each one of its local minimizers, in spite of the nondifferentiability of φ at $\pm\sqrt{\alpha}/\lambda$. Then, we show that any global minimum of $\mathcal{E}(., \boldsymbol{y})$ is strict for every $\boldsymbol{A}$ and for every $\boldsymbol{y}$. Moreover, the local minimizers of $\mathcal{E}(., \boldsymbol{y})$ are strict and are under some pretty general conditions. These facts are essential for the conception of minimization algorithms. Our main result states that given $\boldsymbol{A}$ and (α, λ), with every $k \in \{1, \cdots, n-1\}$, there are associated two different thresholds such that the magnitude of the difference $|\hat{x}_k - \hat{x}_{k+1}|$ relevant to a global minimizer of $\mathcal{E}(., \boldsymbol{y})$ can never be placed between these thresholds for any $\boldsymbol{y} \in \mathbb{R}^m$. Notice that these thresholds are independent of $\boldsymbol{y}$. In other words, if for $\boldsymbol{y} \in \mathbb{R}^m$ $\hat{x}$ is a global minimizer of $\mathcal{E}(., \boldsymbol{y})$, then $|\hat{x}_k - \hat{x}_{k+1}|$ is either smaller than the smallest threshold, or it is larger than the largest threshold. This fact explains *how* an estimator involving a truncated quadratic regularization performs edge detection and gives rise to estimates exhibiting sharp edges and how this behavior is related to $\boldsymbol{A}$ and (α, λ). On the other hand, the smoothness of φ at zero entails that the zones in $\hat{\boldsymbol{x}}$ beyond the edges exhibit weak variations without being constant [28]. The edge-detection thresholds evoked above admit a simple explicit form, which is given in Section IV. They can be used to derive necessary conditions for the choice of the parameters (α, λ). Next, we show that the chance to get data $\boldsymbol{y}$ for which $\mathcal{E}(., \boldsymbol{y})$ has two or more global minimizers is null. Finally, we show that a global minimizer of the objective function is almost surely continuous under small perturbations of $\boldsymbol{y}$.

The results presented in this paper can be extended to regularizers of the form $\Phi(\boldsymbol{x}) = \sum_k \varphi(\boldsymbol{g}_k^T \boldsymbol{x})$, where $\{\boldsymbol{g}_k, k \geq 1\}$ is any collection of vectors. However, deriving explicit expressions for the edge-detection phenomenon in such a situation seems more difficult. We leave this question for future works.

A. Organization of the Paper

Section II focuses on the local behavior of $\mathcal{E}$ in the vicinity of its minimizers. The core of the thresholding effect—the existence of regions where the differences of a global minimizer cannot be placed—is developed in Section III. The behavior of a global minimizer, as a function of the data, is discussed in Section IV. Numerical experiments are presented in Section V. Concluding remarks are summarized in Section VI.

The proofs of the assertions in this paper are outlined in the Appendixes B–D. Most of them use an equivalent representation of $\mathcal{E}(., \boldsymbol{y})$, which is given in Appendix A.

II. Local Continuity of the Minimizers of $\mathcal{E}(., \boldsymbol{y})$

The objective function $\mathcal{E}(., \boldsymbol{y})$ in (2)–(4) is nonsmooth on the union of hyperplanes

$$\bigcup_{k=1}^{n-1} \left\{ \left[x_k - x_{k+1} = \sqrt{\alpha}/\lambda\right] \cup \left[x_k - x_{k+1} = -\sqrt{\alpha}/\lambda\right] \right\}. \tag{5}$$

Our first question is to see whether a global or a local minimizer of $\mathcal{E}(., \boldsymbol{y})$ can belong to this union. Its issue gives a first characterization of $\hat{\boldsymbol{x}}$ and determines the tools that can be used to pursue this study.

Proposition 1: Given $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, let $\mathcal{E}$: $\mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}$ be as introduced in (2)–(4).

Then, for every $\boldsymbol{y} \in \mathbb{R}^m$, each local or global minimizer of $\mathcal{E}(., \boldsymbol{y})$: $\mathbb{R}^n \mapsto \mathbb{R}$, say $\hat{\boldsymbol{x}}$, satisfies

$$|\hat{x}_k - \hat{x}_{k+1}| \neq \frac{\sqrt{\alpha}}{\lambda} \qquad \text{for any } k = 1, \cdots, n-1 \tag{6}$$

or equivalently, $\hat{\boldsymbol{x}}$ does not belong to the union of hyperplanes given in (5).

Recall that $\sqrt{\alpha}/\lambda$ is the point of truncation of φ; see (4). Let us emphasize that (6) holds for any $\boldsymbol{y}$ and for any $\boldsymbol{A}$. Intuitively, it may seem unlikely that the minimizers of an objective function corresponding to noisy data belong to the special set of hyperplanes given in (5). However, recall that whenever the left and the right derivatives of φ at zero satisfy $\varphi'(0^-) < \varphi'(0^+)$ [e.g. $\varphi(t) = |t|$], then the minimizers of the relevant objective function *do* belong to the intersection of a large number of hyperplanes since $\hat{x}_k = \hat{x}_{k+1}$ for many indexes k—see [28]. Therefore, Proposition 1 provides an useful precision. Moreover, it asserts that $|\hat{x}_k - \hat{x}_{k+1}| = \theta$ is not only unlikely, but that it is impossible, even for specially chosen data.

Hereafter, $B(\boldsymbol{x}; \rho) := \{\boldsymbol{x}' \colon \|\boldsymbol{x}' - \boldsymbol{x}\| < \rho\}$ denotes an open ball with radius $\rho > 0$, which is defined with respect to the Euclidian norm $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^T\boldsymbol{x}}$, whereas T stands for transposition. Proposition 1 means that any local minimizer $\hat{\boldsymbol{x}}$ is contained in a ball $B(\hat{\boldsymbol{x}}; \rho)$, where $\mathcal{E}(., \boldsymbol{y})$ is $\mathcal{C}^2$-continuous, where

$$\rho := \frac{1}{2} \min \left\{ \left| |\hat{x}_k - \hat{x}_{k+1}| - \frac{\sqrt{\alpha}}{\lambda} \right| \text{ for } k = 1, \cdots, n-1 \right\}. \tag{7}$$

Then, having $\boldsymbol{x} \in B(\hat{\boldsymbol{x}}; \rho)$ implies that

$$\begin{aligned} \varphi(x_k - x_{k+1}) &= \lambda^2 (x_k - x_{k+1})^2 \\ &\qquad \text{if } \varphi(\hat{x}_k - \hat{x}_{k+1}) = \lambda^2(\hat{x}_k - \hat{x}_{k+1})^2 \\ \varphi(x_k - x_{k+1}) &= \alpha \qquad \text{if } \varphi(\hat{x}_k - \hat{x}_{k+1}) = \alpha. \end{aligned} \tag{8}$$

Let $\mathcal{J}$ be the mapping that, for each signal $\boldsymbol{x} \in \mathbb{R}^n$, gives $\mathcal{J}(\boldsymbol{x})$ the set of its *jumps* (or edges), which are the indexes of all differences whose magnitude is larger than θ:

$$\mathcal{J}(\boldsymbol{x}) := \{k \in \{1, \cdots, n-1\} \colon |x_k - x_{k+1}| \geq \theta\}. \tag{9}$$

Consequently, $B(\hat{\boldsymbol{x}}; \rho)$ is composed of signals whose edges are located in the same way, that is, $\mathcal{J}(\boldsymbol{x}) = \hat{J}$ for all $\boldsymbol{x} \in B(\hat{\boldsymbol{x}}; \rho)$, where we put $\hat{J} := \mathcal{J}(\hat{\boldsymbol{x}})$.

Let $\mathbb{I}$ stand for the characteristic function $\mathbb{I}(T) = 1$ if T is true and $\mathbb{I}(T) = 0$ otherwise. The letter $\boldsymbol{I}$ will denote the identity matrix of whatever size appropriate to the context. Furthermore, $\mathbf{0}$ and $\mathbf{1}$ will denote vectors or matrices composed of zeros and of ones, respectively. When necessary, their size will be indicated in superscript. Given a subset $J \subset \{1, \cdots, n-1\}$, we define the following diagonal $n-1 \times n-1$ matrix:

$$\boldsymbol{\mathcal{U}}_J := \boldsymbol{I} - \text{Diag}\{\mathbb{I}(1 \in J), \mathbb{I}(2 \in J), \cdots, \mathbb{I}(n-1 \in J)\}. \tag{10}$$

Thus, we determine a mapping $J \mapsto \boldsymbol{\mathcal{U}}_J$. The next Proposition is a straightforward consequence of (8).

Proposition 2: For $\boldsymbol{y} \in \mathbb{R}^m$ and $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, consider $\mathcal{E}(., \boldsymbol{y})$: $\mathbb{R}^n \mapsto \mathbb{R}$ as given in (2)–(4).

Then, a point $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ is a local or a global minimizer of $\mathcal{E}(., \boldsymbol{y})$ if and only if

$$2\boldsymbol{A}^T(\boldsymbol{A}\hat{\boldsymbol{x}} - \boldsymbol{y}) + \nabla\Phi(\hat{\boldsymbol{x}}) = \boldsymbol{0}$$

where

$$\nabla\Phi(\hat{\boldsymbol{x}}) = 2\lambda^2 \boldsymbol{D}^T \boldsymbol{\mathcal{U}}_{\hat{J}} \boldsymbol{D}\hat{\boldsymbol{x}}. \tag{11}$$

In the last expression, $\hat{J} := \mathcal{J}(\hat{\boldsymbol{x}})$, where $\mathcal{J}$ is the mapping defined in (9), $\boldsymbol{\mathcal{U}}$ is as defined in (10), and $\boldsymbol{D}$ is the $(n-1)\times n$ difference matrix whose kth row $\boldsymbol{d}_k^T$ reads $\boldsymbol{d}_k[k] = 1$, $\boldsymbol{d}_k[k+1] = -1$ and $\boldsymbol{d}_k[i] = 0$ otherwise.

The local behavior of an estimator is tightly dependent on the shape of the objective function in the vicinity of its global minimizers. In particular, a crucial point is to know whether a minimizer is strict or not. The cases when $\boldsymbol{A}$ is injective (i.e., rank$\boldsymbol{A} = n \le m$) or noninjective (then rank$\boldsymbol{A} < n$) need to be considered separately.

Proposition 3: Suppose that $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is injective, and consider $\mathcal{E}$: $\mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}$, as given in (2)–(4).

Then for every $\boldsymbol{y} \in \mathbb{R}^m$, all local and global minimizers of $\mathcal{E}(., \boldsymbol{y})$: $\mathbb{R}^n \mapsto \mathbb{R}$ are *strict*.

However, if $\boldsymbol{A}$ is noninjective, $\mathcal{E}(., \boldsymbol{y})$ can exhibit nonstrict local minima, as can be seen in the next example.

Example 1: Consider

$$\mathcal{E}(\boldsymbol{x}, y) = (x_1+x_2-y)^2 + \varphi(x_1-x_2) \text{ with either } y < -7\theta \text{ or } y > -\theta, \quad \text{for } \theta := \frac{\sqrt{\alpha}}{\lambda}. \tag{12}$$

Now, $\boldsymbol{A} = [1, 1]$, and it is noninvective. Let us check that the point $(\hat{x}_1, \hat{x}_2)$, where $\hat{x}_1 = 2\theta + y$ and $\hat{x}_2 = -2\theta$ is a nonstrict local minimizer of $\mathcal{E}(\boldsymbol{x}, y)$. First, we remark that $0 = (\hat{x}_1 + \hat{x}_2 - y)^2 = (\hat{x}_1 + v_1 + \hat{x}_2 + v_2 - y)^2$ for any $\boldsymbol{v}$ of the form $\boldsymbol{v} = [v_1, -v_1]^T$ with $v_1 \in \mathbb{R}$. Next, for any $\|\boldsymbol{v}\| < \theta$, we have

$$\begin{aligned} &|\hat{x}_1 - \hat{x}_2 + (v_1 - v_2)| \\ &\quad = |4\theta + y + (v_1 - v_2)| \ge |4\theta + y| - |v_1 - v_2| \\ &\quad > 3\theta - 2\|\boldsymbol{v}\| \ge \theta \end{aligned}$$

in which case, $\varphi[(\hat{x}_1 + v_1) - (\hat{x}_2 + v_2)] = \varphi(\hat{x}_1 - \hat{x}_2) = \alpha$. Thus, we deduce that $\mathcal{E}(\hat{\boldsymbol{x}} + \boldsymbol{v}, y) = \mathcal{E}(\hat{\boldsymbol{x}}, y) = \alpha$ for any $\boldsymbol{v} = [v_1, -v_1]^T$ with $|v_1| < \theta/\sqrt{2}$. △

A naturally arising concern is to see the shape of the connected set of points yielding an isolated nonstrict minimum of $\mathcal{E}(., \boldsymbol{y})$. These kinds of sets can be "quite irregular" for general nonconvex objective functions.

Proposition 4: For $\boldsymbol{y} \in \mathbb{R}^m$ and $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, suppose that $\mathcal{E}(., \boldsymbol{y})$, as defined in (2)–(4), reaches a minimum at $\hat{\boldsymbol{x}}$.

Then, there is a *convex* set $\Xi \subset \mathbb{R}^n$ containing $\hat{\boldsymbol{x}}$ such that $\mathcal{E}(\boldsymbol{z}, \boldsymbol{y}) = \mathcal{E}(\hat{\boldsymbol{x}}, \boldsymbol{y})$ for all $\boldsymbol{z} \in \Xi$.

In fact, (8) shows that $\mathcal{E}(., \boldsymbol{y})$ is quadratic in the vicinity of any local minimizer, hence, the proposition. If $\hat{\boldsymbol{x}}$ is a strict minimizer, then trivially, $\Xi = \{\hat{\boldsymbol{x}}\}$. More generally, the expressions (58), arising in the proof of Theorem 1, show that Ξ contains elements $\boldsymbol{z}$ of the form

$$\boldsymbol{z} = \hat{\boldsymbol{x}} + \ell\boldsymbol{v} \quad \text{where } |\ell| < \rho \quad \text{and } \begin{cases} \boldsymbol{v} \in \text{Ker}\,(\boldsymbol{A}^T\boldsymbol{A}) \\ v_k = v_{k+1}, & \text{if } k \notin \hat{J} \end{cases} \tag{13}$$

with ρ, as defined in (7). It follows that Ξ is strictly larger than $\{\hat{\boldsymbol{x}}\}$ only if Ker$\,(\boldsymbol{A}^T\boldsymbol{A})$ contains nonzero vectors $\boldsymbol{v}$ composed of constant segments. However, the latter is a quite an atypical situation. Otherwise, if Ker$\,(\boldsymbol{A}^T\boldsymbol{A})$ does not contain such vectors, all minimizers of $\mathcal{E}(., \boldsymbol{y})$ are strict for every $\boldsymbol{y} \in \mathbb{R}^m$. Whenever Ξ is strictly larger than $\{\hat{\boldsymbol{x}}\}$, then it is composed of signals $\boldsymbol{z}$ that share the same homogeneous regions ($z_k - z_{k+1} = \hat{x}_k - \hat{x}_{k+1}$ if $|\hat{x}_k - \hat{x}_{k+1}| < \sqrt{\alpha}/\lambda$) but whose jumps may have different magnitudes.

In any case, there is a stronger result stating that a global minimizer is always strict. In Example 1, the global minimizer of $\mathcal{E}(., y)$ reads $\hat{x}_1 = y/2 = \hat{x}_2$ and it is easy to see that it is strict.

Theorem 1: Assume that $\boldsymbol{A1} \neq \boldsymbol{0}$, and consider $\mathcal{E}$ as defined in (2)–(4).

Then, for any $\boldsymbol{y} \in \mathbb{R}^m$, any *global* minimum of $\mathcal{E}(., \boldsymbol{y})$: $\mathbb{R}^n \mapsto \mathbb{R}$ is *strict*.

The assumption $\boldsymbol{A1} \neq \boldsymbol{0}$ means that $\boldsymbol{A}$ preserves the mean of the original signal. Such are most of the observation operators encountered in practice. Conversely, if $\boldsymbol{A1} = \boldsymbol{0}$, then $\mathcal{E}(\hat{\boldsymbol{x}} + \ell\boldsymbol{1}, \boldsymbol{y}) = \mathcal{E}(\hat{\boldsymbol{x}}, \boldsymbol{y})$ for any real ℓ, and hence, the mean of the signal is undetermined. In such a situation, the problem should be reformulated as suggested in Remark 2 of Appendix A. Henceforth, we systematically assume that $\boldsymbol{A1} \neq \boldsymbol{0}$. Let us emphasize that Theorem 1 also addresses the situations where $\boldsymbol{A}$ is noninjective and when Ker$\,(\boldsymbol{A}^T\boldsymbol{A})$ can contain locally constant vectors.

The next Theorem formalizes the main conclusion of this section.

Theorem 2: For $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, let $\mathcal{E}$ be as in (2)–(4). Given $\boldsymbol{y} \in \mathbb{R}^m$, assume that $\hat{\boldsymbol{x}}$ is a *strict* local or global minimizer of $\mathcal{E}(., \boldsymbol{y})$.

Then, there exist $\mu > 0$ and a differentiable mapping $\boldsymbol{\mathcal{X}}$: $B(\boldsymbol{y}; \mu) \mapsto \mathbb{R}^n$ such that $\boldsymbol{\mathcal{X}}(\boldsymbol{y}) = \hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{x}}' := \boldsymbol{\mathcal{X}}(\boldsymbol{y}')$ is a strict local or global minimizer of $\mathcal{E}(., \boldsymbol{y}')$ if $\boldsymbol{y}' \in B(\boldsymbol{y}; \mu)$.

In other words, $\boldsymbol{\mathcal{X}}$ is a local minimizer function relevant to $\mathcal{E}$. Theorem 2 only states that $\boldsymbol{\mathcal{X}}$ is continuous on $B(\boldsymbol{y}; \mu)$, but if $\boldsymbol{\mathcal{X}}(\boldsymbol{y})$ is a global minimizer of $\mathcal{E}(., \boldsymbol{y})$, it remains unknown whether $\boldsymbol{\mathcal{X}}(\boldsymbol{y}')$, relevant to $\boldsymbol{y}' \in B(\boldsymbol{y}; \mu)$, is still a global minimizer of $\mathcal{E}(., \boldsymbol{y}')$ or not. This question is considered in Section IV.

III. Zones Where a Global Minimizer Cannot Lie

The result found next is guaranteed to hold at any global minimizer of $\mathcal{E}(., \boldsymbol{y})$ for any $\boldsymbol{y} \in \mathbb{R}^m$. It exhibits the presence of intervals of positive length that can *never* contain the differences of any global minimizer of $\mathcal{E}(., \boldsymbol{y})$ for any $\boldsymbol{y} \in \mathbb{R}^m$. It can be expected that the same kind of behavior is exhibited by some or many among the local minimizers of $\mathcal{E}(., \boldsymbol{y})$.

Let $\boldsymbol{M} \in \mathbb{R}^{m\times m}$ and $\boldsymbol{u}_k \in \mathbb{R}^n$ be the following projection matrix and unit step-sequence:

$$\boldsymbol{M} := \boldsymbol{I} - \frac{\boldsymbol{A}\boldsymbol{1}\boldsymbol{1}^T\boldsymbol{A}^T}{\|\boldsymbol{A}\boldsymbol{1}\|^2} \tag{14}$$

$$\boldsymbol{u}_k[i] := \begin{cases} 0, & \text{if } i=1,\cdots,k, \\ 1, & \text{if } i=k+1,\cdots,n, \end{cases} \qquad \text{for } k=1,\cdots,n-1. \tag{15}$$

The subsequent analysis reveal an essential distinction in the behavior of the differences $\hat{x}_k - \hat{x}_{k+1}$ whose indexes k are such that $\boldsymbol{M}\boldsymbol{A}\boldsymbol{u}_k \neq \boldsymbol{0}$ and the differences for which $\boldsymbol{M}\boldsymbol{A}\boldsymbol{u}_k = \boldsymbol{0}$. The differences of the first kind will be called *observable*, whereas those of the second kind are *unobservable*.

Theorem 3: For $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, consider $\mathcal{E}$ as given in (2)–(4). Let $k \in \{1, \cdots, n-1\}$ and assume that $\boldsymbol{M}\boldsymbol{A}\boldsymbol{u}_k \neq \boldsymbol{0}$, with $\boldsymbol{M}$ and $\boldsymbol{u}_k$ as given in (14)–(15).

Then there exists a constant $\Gamma_k \in]0, 1[$ such that for any $\boldsymbol{y} \in \mathbb{R}^m$, any *global* minimizer $\hat{\boldsymbol{x}}$ of the relevant $\mathcal{E}(., \boldsymbol{y})$, satisfies the alternative:

$$\text{either} \quad |\hat{x}_k - \hat{x}_{k+1}| \leq \theta\Gamma_k \quad \text{or} \quad |\hat{x}_k - \hat{x}_{k+1}| \geq \frac{\theta}{\Gamma_k}$$
$$\text{where } \theta := \frac{\sqrt{\alpha}}{\lambda}. \tag{16}$$

More precisely, Γ_k reads

$$\Gamma_k = \sqrt{\frac{\xi_k}{\lambda^2 + \xi_k}} \quad \text{with} \quad \xi_k = \boldsymbol{u}_k^T\boldsymbol{A}^T\boldsymbol{M}\boldsymbol{A}\boldsymbol{u}_k > 0. \tag{17}$$

Moreover, the inequalities in (16) are *strict* whenever $\mathcal{E}(., \boldsymbol{y})$ has only one global minimizer.

If k is such that $\boldsymbol{M}\boldsymbol{A}\boldsymbol{u}_k \neq \boldsymbol{0}$, then for any $\boldsymbol{y} \in \mathbb{R}^m$, the difference $\hat{x}_k - \hat{x}_{k+1}$, relevant to a global minimizer of $\mathcal{E}(., \boldsymbol{y})$ lies *beyond* two intervals, namely, $\hat{x}_k - \hat{x}_{k+1} \notin \{[-\theta/\Gamma_k, -\theta\Gamma_k] \bigcup [\theta\Gamma_k, \theta/\Gamma_k]\}$. By (17), these intervals are independent of $\boldsymbol{y} \in \mathbb{R}^m$, and they have a strictly positive length since $\Gamma_k \in]0, 1[$.

Example 2: Let $\mathcal{E}$ read

$$\mathcal{E}(\boldsymbol{x}, \boldsymbol{y}) = (x_1 - y_1)^2 + (x_2 - y_2)^2 + \varphi(x_1 - x_2). \tag{18}$$

In this case, $k = 1$, $\boldsymbol{A} = \boldsymbol{I}$, and hence

$$\boldsymbol{u}_1 = [0, 1]^T, \quad \boldsymbol{M} = \tfrac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$
$$\xi_1 = \frac{1}{2}, \quad \Gamma_1 = \frac{1}{\sqrt{1+2\lambda^2}}.$$

Let θ be as defined in (16). We have the following situations:

a) $\boldsymbol{y}$ satisfies $|y_1 - y_2| < \theta/\Gamma^2$; then, $\mathcal{E}(., \boldsymbol{y})$ has a local minimizer $\hat{\boldsymbol{x}}_0$ without edge, whose entries read

$$\hat{\boldsymbol{x}}_0[1] = \Gamma_1^2[y_1(1+\lambda^2) + \lambda^2 y_2]$$

and

$$\hat{\boldsymbol{x}}_0[2] = \Gamma_1^2[\lambda^2 y_1 + y_2(1+\lambda^2)]. \tag{19}$$

Let us verify that $\hat{\boldsymbol{x}}_0$ is a local minimizer of $\mathcal{E}(., \boldsymbol{y})$. From $|\hat{\boldsymbol{x}}_0[1] - \hat{\boldsymbol{x}}_0[2]| = \Gamma_1^2|y_1 - y_2| < \theta$, we get $\hat{J}_0 := \mathcal{J}(\hat{\boldsymbol{x}}_0) = \{\emptyset\}$. It remains to be checked that $\hat{\boldsymbol{x}}_0$ satisfies (11) with respect to $\boldsymbol{\mathcal{U}}_{\hat{J}_0}$. Notice that

$$\mathcal{E}(\hat{\boldsymbol{x}}_0, \boldsymbol{y}) = \Gamma_1^2\lambda^2(y_1 - y_2)^2.$$

However, if $|y_1 - y_2| \geq \theta/\Gamma^2$, then $\mathcal{J}(\hat{\boldsymbol{x}}_0) = \{1\}$ and $\hat{\boldsymbol{x}}_0$ in (19) is not a minimizer of $\mathcal{E}(., \boldsymbol{y})$.

b) $\boldsymbol{y}$ satisfies $|y_1 - y_2| \geq \theta$, and then, $\mathcal{E}(., \boldsymbol{y})$ has a local minimizer $\hat{\boldsymbol{x}}_1$ involving an edge

$$\hat{\boldsymbol{x}}_1[1] = y_1 \quad \text{and} \quad \hat{\boldsymbol{x}}_1[2] = y_2. \tag{20}$$

Since $|\hat{\boldsymbol{x}}_1[1] - \hat{\boldsymbol{x}}_1[2]| \geq \theta$, $\hat{J}_1 := \mathcal{J}(\hat{\boldsymbol{x}}_1) = \{1\}$, and (11) shows that $\hat{\boldsymbol{x}}_1$ is a local minimizer of $\mathcal{E}(\boldsymbol{x}, \boldsymbol{y})$. Now, $\mathcal{E}(\hat{\boldsymbol{x}}_1, \boldsymbol{y}) = \alpha$.

However, if $|y_1 - y_2| < \theta$, then $|\hat{\boldsymbol{x}}_1[1] - \hat{\boldsymbol{x}}_1[2]| < \theta$ and $\mathcal{J}(\hat{\boldsymbol{x}}_1) = \{\emptyset\}$, and hence, $\hat{\boldsymbol{x}}_1$ is not a minimizer of $\mathcal{E}(., \boldsymbol{y})$.

We draw the following conclusions.

i) If $|y_1 - y_2| < \theta$, then $\mathcal{E}(., \boldsymbol{y})$ has a unique minimizer that is $\hat{\boldsymbol{x}}_0$. We check that $|\hat{\boldsymbol{x}}_0[1] - \hat{\boldsymbol{x}}_0[2]| = \Gamma_1^2|y_1 - y_2| < \Gamma_1^2\theta < \Gamma_1\theta$, which corroborates (16).

ii) If $|y_1 - y_2| \geq \theta/\Gamma^2$, then $\mathcal{E}(., \boldsymbol{y})$ has a unique minimizer, which is $\hat{\boldsymbol{x}}_1$. Now, clearly, $|\hat{\boldsymbol{x}}_1[1] - \hat{\boldsymbol{x}}_1[2]| \geq \theta/\Gamma^2 > \theta/\Gamma$, as stated in (16).

iii) If $\theta/\Gamma^2 < |y_1 - y_2| \leq \theta$, then $\mathcal{E}(., \boldsymbol{y})$ has two local minimizers, which are $\hat{\boldsymbol{x}}_0$ and $\hat{\boldsymbol{x}}_1$. The global minimizer of $\mathcal{E}(., \boldsymbol{y})$, which is denoted by $\hat{\boldsymbol{x}}$, is determined by comparing $\mathcal{E}(\hat{\boldsymbol{x}}_0, \boldsymbol{y})$ and $\mathcal{E}(\hat{\boldsymbol{x}}_1, \boldsymbol{y})$. Consequently

$$|y_1 - y_2| \leq \theta\Gamma^{-1} \quad \text{leads to} \quad \hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}_0 \tag{21}$$
$$|y_1 - y_2| \geq \theta\Gamma^{-1} \quad \text{leads to} \quad \hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}_1. \tag{22}$$

In the first case, $|\hat{x}_1 - \hat{x}_2| \leq \theta\Gamma$, whereas in the second, $|\hat{x}_1 - \hat{x}_2| \geq \theta\Gamma^{-1}$, which is the alternative asserted in (16). △

Therefore, the magnitude of an observable difference arising at a global minimizer of the objective function is either smaller than a *small* first threshold or larger than a second threshold that is larger then the first one. These thresholds are independent of the data. The differences that are smaller than the first threshold are *smooth*, and they form the homogeneous zones in $\hat{\boldsymbol{x}}$, whereas the differences larger than the second threshold correspond to *edges*. This neat classification of the differences at a global minimizer is nicely observed in the experiments presented in Fig. 1. In a global plan, the bounds in (16) and (17) are specific for each position k of the difference along the signal. The latter is seen in the simulations in Section V and especially in Fig. 5.

Theorem 4: Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ satisfy $\boldsymbol{M}\boldsymbol{A}\boldsymbol{u}_k = \boldsymbol{0}$. Consider $\mathcal{E}$ as given in (2)–(4).

Then, for any $\boldsymbol{y} \in \mathbb{R}^m$, any global minimizer $\hat{\boldsymbol{x}}$ of $\mathcal{E}(., \boldsymbol{y})$ satisfies

$$\hat{x}_k = \hat{x}_{k+1}. \tag{23}$$

Let us now come back to Example 1. It is easy to see that for any $y \in \mathbb{R}$, the point $\hat{\boldsymbol{x}} = (y/2, y/2)$ is a global minimizer of the relevant $\mathcal{E}(., y)$ since it yields $\mathcal{E}(\hat{\boldsymbol{x}}, y) = 0$. The latter result is an application of (23).

Remark 1: Speaking more loosely, $\boldsymbol{M}\boldsymbol{A}\boldsymbol{u}_k = \boldsymbol{0}$ means that the operator $\boldsymbol{A}$ is *blind* to catch any information relevant to the

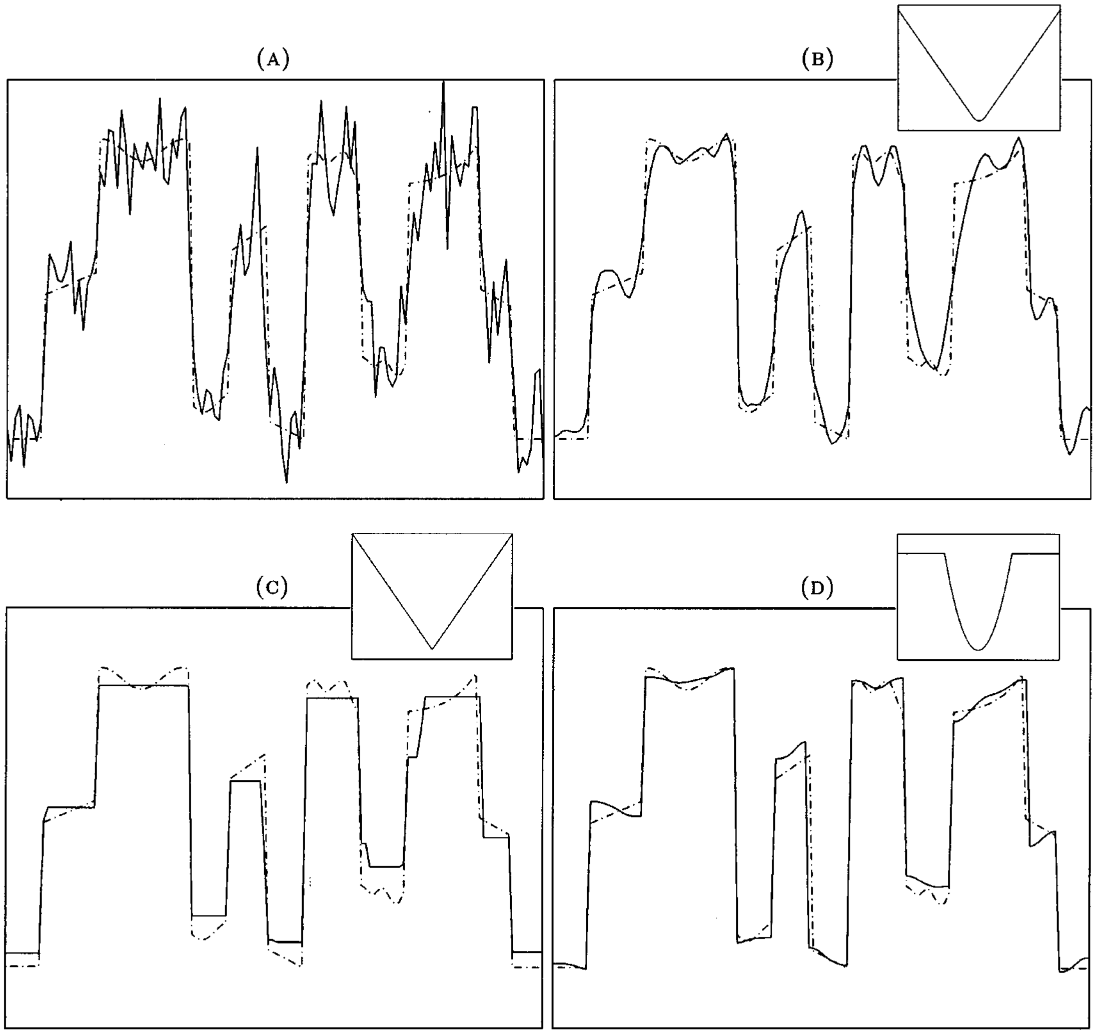

Fig. 1. Deconvolution. (a) Noisy data $\boldsymbol{y} = \boldsymbol{x} * \boldsymbol{h} + \boldsymbol{n}$ with $h_k = \exp(-0.4k^2)/1.41$ for $|k| \leq 5$ and $\boldsymbol{n}$ white Gaussian noise, 10 dB SNR. Reconstructions using differents PFs ε in (1)–(3), where the latter are plotted above the relevant solutions. (b) Huber PF with $\alpha = 0.1$ and $\lambda = 1$. (c) Modulus PF $\lambda = 2$. (d) Truncated quadratic PF $(\alpha, \lambda) = (0.7, 7.7)$.

magnitude of the original $|x_k - x_{k+1}|$. Next, we explain this assertion. Since Ker$\boldsymbol{M}$ is spanned by $\boldsymbol{A1}$, the equality $\boldsymbol{MAu}_k = \boldsymbol{0}$ means that

- either $\boldsymbol{Au}_k \neq \boldsymbol{0}$ and $\boldsymbol{Au}_k \in \mathrm{Ker}\boldsymbol{M}$, hence, there exists $c \neq 0$ such that $\boldsymbol{Au}_k = c\boldsymbol{A1}$;
- or $\boldsymbol{Au}_k = \boldsymbol{0}$, which amounts to the previous case if we take $c = 0$.

Given an original signal $\boldsymbol{x}$, let us consider the family all signals $\boldsymbol{x}_\beta$, $\beta \in \mathbb{R}$, which is obtained from $\boldsymbol{x}$ by modifying only its kth difference

$$\boldsymbol{x}_\beta = \boldsymbol{x} + \beta \boldsymbol{u}_k.$$

(Clearly, $\boldsymbol{x}_\beta[i] - \boldsymbol{x}_\beta[i+1] = \boldsymbol{x}[i] - \boldsymbol{x}[i+1]$ for $i \neq k$, whereas $\boldsymbol{x}_\beta[k] - \boldsymbol{x}_\beta[k+1] = \boldsymbol{x}[k] - \boldsymbol{x}[k+1] - \beta$.) When transformed by the observation operator $\boldsymbol{A}$, this family yields

$$\boldsymbol{Ax}_\beta = \boldsymbol{Ax} + \beta \boldsymbol{Au}_k = \boldsymbol{Ax} + \beta c \boldsymbol{A1}.$$

In the data $\boldsymbol{y}$, the contribution of the kth difference is thus confounded with the contribution of $\boldsymbol{1}$. In particular, $\boldsymbol{Ax}_\beta = \boldsymbol{Ax}$ if $\boldsymbol{Au}_k = \boldsymbol{0}$. This is the reason why a difference $x_k - x_{k+1}$, whose index k is such that $\boldsymbol{MAu}_k \neq \boldsymbol{0}$, is said to be *unobservable*. △

The operators $\boldsymbol{A}$, such that $\boldsymbol{MAu}_k = \boldsymbol{0}$ for some k, are rare since their columns $\boldsymbol{a}_i$, $i = 1, \cdots, n$ have to satisfy the equation $(1 - c)\sum_{i=k+1}^{n} \boldsymbol{a}_i = c\sum_{i=1}^{k} \boldsymbol{a}_i$, where c is the constant given above. This is a very special requirement that is not satisfied customarily.

Both Theorems 3 and 4 show that there are regions in $\mathbb{R}^n$ that cannot be reached by the differences of any global minimizer of the objective function for any data $\boldsymbol{y} \in \mathbb{R}^m$. Such "prohibited" regions exist for any $\boldsymbol{A}$ and for any (α, λ); their extent is fixed by $\boldsymbol{A}$ and (α, λ). Notice that the bounds found in (16) and (17) are *necessary*, whereas (23) is both necessary and sufficient.

IV. Thresholding and Local Stability at a Global Minimizer

In this section, we focus on the behavior of the global minimizers of $\mathcal{E}(., \boldsymbol{y})$ when $\boldsymbol{y}$ ranges over $\mathbb{R}^m$. We begin these considerations by emphasizing that the global minimizer in Example 2 involves a typical *detection* stage since it satisfies $|\hat{x}_1 - \hat{x}_2| \leq \theta\Gamma$ if $|y_1 - y_2| \leq \theta\Gamma^{-1}$ and $|\hat{x}_1 - \hat{x}_2| \geq \theta\Gamma^{-1}$ if

$|y_1 - y_2| \geq \theta\Gamma^{-1}$. In the first case, the data reflect a weak contribution of the original difference $|x_1 - x_2|$, and its estimate $|\hat{x}_1 - \hat{x}_2|$ is subject to smoothing, whereas in the second case, the contribution of $|x_1 - x_2|$ is strong, and its estimate $|\hat{x}_1 - \hat{x}_2|$ is large and corresponds to a jump (an edge). Qualitatively speaking, the estimation of signals by the method in (1)–(4) is based on the same effect. This question is considered below in several steps.

The necessary and sufficient condition for minimum, which is given in Proposition 2, can be reformulated as $(\boldsymbol{A}^T\boldsymbol{A} + \lambda^2\boldsymbol{D}^T\boldsymbol{\mathcal{U}}_{\hat{J}}\boldsymbol{D})\hat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{y}$, where $\hat{J} = \mathcal{J}(\hat{\boldsymbol{x}})$. Moreover, a (local or global) minimizer $\hat{\boldsymbol{x}}$ is strict if and only if $\boldsymbol{A}^T\boldsymbol{A} + \lambda^2\boldsymbol{D}^T\boldsymbol{\mathcal{U}}_{\hat{J}}\boldsymbol{D}$ is invertible. Therefore, we have the following:

$$\mathcal{M} := \{J \subset \{1, \cdots, n-1\}\colon \operatorname{rank}(\boldsymbol{A}^T\boldsymbol{A} + \lambda^2 \boldsymbol{D}^T\boldsymbol{\mathcal{U}}_J\boldsymbol{D}) = n\}. \qquad (24)$$

If $J' \notin \mathcal{M}$, no strict minimizer of $\mathcal{E}(., \boldsymbol{y})$, for any $\boldsymbol{y} \in \mathbb{R}^m$, has J' as the set of its jumps. Reciprocally, all edge configurations that may arise at a strict local or global minimizer of $\mathcal{E}(., \boldsymbol{y})$ for any $\boldsymbol{y} \in \mathbb{R}^m$ are elements of $\mathcal{M}$. If $\hat{\boldsymbol{x}}$ is a strict minimizer of $\mathcal{E}(., \boldsymbol{y})$, then $\mathcal{J}(\hat{\boldsymbol{x}}) \in \mathcal{M}$.

Proposition 5: With every $J \in \mathcal{M}$, let there be associated with it the following matrix $\boldsymbol{\mathcal{X}}_J \in \mathbb{R}^{n\times m}$ and set $Y_J \subset \mathbb{R}^m$:

$$\boldsymbol{\mathcal{X}}_J := (\boldsymbol{A}^T\boldsymbol{A} + \lambda^2\boldsymbol{D}^T\boldsymbol{\mathcal{U}}_J\boldsymbol{D})^{-1}\boldsymbol{A}^T \qquad (25)$$

$$Y_J = \left\{\boldsymbol{y} \in \mathbb{R}^m\colon |\boldsymbol{d}_k^T\boldsymbol{\mathcal{X}}_J\boldsymbol{y}| > \frac{\sqrt{\alpha}}{\lambda}, \text{ if } k \in J \text{ and } |\boldsymbol{d}_k^T\boldsymbol{\mathcal{X}}_J\boldsymbol{y}| < \frac{\sqrt{\alpha}}{\lambda}, \text{ otherwise}\right\} \qquad (26)$$

where $\boldsymbol{d}_k$ and $\boldsymbol{D}$ are as in Proposition 2. Suppose Y_J is non empty.

Then, for any $\boldsymbol{y} \in Y_J$, the point

$$\hat{\boldsymbol{x}} := \boldsymbol{\mathcal{X}}_J\boldsymbol{y} \qquad (27)$$

is a strict local minimizer of $\mathcal{E}(., \boldsymbol{y})$ that satisfies $\mathcal{J}(\hat{\boldsymbol{x}}) = J$.

From (26), Y_J, which is the domain of $\boldsymbol{\mathcal{X}}_J$, is the union of several polyhedrons of $\mathbb{R}^m$. An identification with Theorem 2 allows us to write that $\boldsymbol{\mathcal{X}}(\boldsymbol{y}) = \boldsymbol{\mathcal{X}}_J\boldsymbol{y}$ for every $\boldsymbol{y} \in Y_J$.

Let the entries of $\mathcal{M}$ be J_p for $p = 0, \cdots, \#\mathcal{M} - 1$, where $\#$ means cardinality. Given $\boldsymbol{y} \in \mathbb{R}^m$, the set of *all* strict (local or global) minimizers of $\mathcal{E}(., \boldsymbol{y})$ reads $\{\boldsymbol{\mathcal{X}}_{J_p}\boldsymbol{y}\colon p \in S_{\boldsymbol{y}}\}$, where

$$S_{\boldsymbol{y}} := \{p \in \{0, \cdots, \#\mathcal{M} - 1\}\colon \boldsymbol{y} \in Y_{J_p}\}. \qquad (28)$$

Given $\boldsymbol{y}$, the set $S_{\boldsymbol{y}}$ contains the indexes p of all jump configurations J_p such that $\mathcal{E}(., \boldsymbol{y})$ has a strict minimizer $\hat{\boldsymbol{x}}_p$ with $\mathcal{J}(\hat{\boldsymbol{x}}_p) = J_p$. Since for every $\boldsymbol{y}$, $S_{\boldsymbol{y}}$ is discrete and finite, $\mathcal{E}(., \boldsymbol{y})$ has only a finite number of strict minimizers.

Given a set of edges $J \in \mathcal{M}$, let $\boldsymbol{\mathcal{G}}_J\colon Y_J \mapsto \mathbb{R}$ be a local minimum-value function

$$\boldsymbol{\mathcal{G}}_J(\boldsymbol{y}) := \mathcal{E}(\boldsymbol{\mathcal{X}}_J\boldsymbol{y}, \boldsymbol{y}) = \|\boldsymbol{y}\|^2 - \boldsymbol{y}^T\boldsymbol{A}\boldsymbol{\mathcal{X}}_J\boldsymbol{y} + \alpha\#J. \qquad (29)$$

We wish to check whether for some $\boldsymbol{y} \in \mathbb{R}^m$, $\mathcal{E}(., \boldsymbol{y})$ can exhibit strict minimizers at which $\mathcal{E}(., \boldsymbol{y})$ takes the same value. To this end, we seek data $\boldsymbol{y}$ leading to $\boldsymbol{\mathcal{G}}_{J_p}(\boldsymbol{y}) = \boldsymbol{\mathcal{G}}_{J_q}(\boldsymbol{y})$ for some $J_p \neq J_q$. A first remark is that if $Y_{J_p} \cap Y_{J_q} = \emptyset$, $\mathcal{E}(., \boldsymbol{y})$ cannot have simultaneously a strict minimizer $\hat{\boldsymbol{x}}_p$ with edges J_p and another minimizer $\hat{\boldsymbol{x}}_q$ with edges J_q. However, $Y_{J_p} \cap Y_{J_q}$ can be nonempty for numerous indexes p and q.

Theorem 5: Consider J_p and J_q elements of $\mathcal{M}$, as given in (24), and the domains Y_{J_p} and Y_{J_q},which are defined in (26).

A) If rank$\boldsymbol{A} < n$ and if $Y_{J_p} \bigcap Y_{J_p} \neq \emptyset$, assume that rank$\boldsymbol{A}(\boldsymbol{\mathcal{X}}_{J_q} - \boldsymbol{\mathcal{X}}_{J_p}) \geq 1$.

Then, all the $\boldsymbol{y} \in \mathbb{R}^m$ leading to $\mathcal{E}(\boldsymbol{\mathcal{X}}_{J_p}\boldsymbol{y}, \boldsymbol{y}) = \mathcal{E}(\boldsymbol{\mathcal{X}}_{J_q}\boldsymbol{y}, \boldsymbol{y})$ are contained in a set $\mathcal{N}_{J_p, J_q}$, which is closed and negligible with respect to the Lebesgue measure on $\mathbb{R}^m$.

By using the notation introduced in (29), we can specify that

$$\begin{aligned}\mathcal{N}_{J_p, J_q} &= \{\boldsymbol{y} \in \mathbb{R}^m\colon \boldsymbol{\mathcal{G}}_{J_p}(\boldsymbol{y}) = \boldsymbol{\mathcal{G}}_{J_q}(\boldsymbol{y})\} \\ &= \{\boldsymbol{y} \in Y_{J_p} \cap Y_{J_q}\colon \boldsymbol{y}^T\boldsymbol{A}(\boldsymbol{\mathcal{X}}_{J_q} - \boldsymbol{\mathcal{X}}_{J_p})\boldsymbol{y} - \alpha(\#J_q - \#J_p) = 0\}.\end{aligned} \qquad (30)$$

Provided that A) is true for any $J \in \mathcal{M}$, all the data $\boldsymbol{y} \in \mathbb{R}^m$, for which $\mathcal{E}(., \boldsymbol{y})$ can take the same value at two ore more *different* strict local or global minimizers, are hence contained in the union of a few nonempty closed sets of the form (30). Such a union is a closed negligible subset of $\mathbb{R}^m$. The chance that noisy data come on such an union, leading to multiple minimizers at which $\mathcal{E}(., \boldsymbol{y})$ takes the same value, is null. In particular, the probability that $\mathcal{E}(., \boldsymbol{y})$ has more than one global minimizers is null.

There may exist situations where A) fails to hold, as seen in the next example.

Example 3: Consider the following objective function:

$$\mathcal{E}(\boldsymbol{x}, \boldsymbol{y}) = (x_1 - y_1)^2 + (x_3 - y_2)^2 + \varphi(x_1 - x_2) + \varphi(x_2 - x_3). \qquad (31)$$

We have

$$\boldsymbol{A} = \begin{bmatrix}1 & 0 & 0\\ 0 & 0 & 1\end{bmatrix}, \quad \begin{aligned}\boldsymbol{\mathcal{U}}_{J_0} &= \text{Diag}\{1, 1\}\\ \boldsymbol{\mathcal{U}}_{J_1} &= \text{Diag}\{1, 0\}\\ \boldsymbol{\mathcal{U}}_{J_2} &= \text{Diag}\{0, 1\}\end{aligned}$$

$$\boldsymbol{\mathcal{X}}_{J_0} = \frac{1}{2(1+\lambda^2)}\begin{bmatrix}2+\lambda^2 & \lambda^2\\ 1+\lambda^2 & 1+\lambda^2\\ \lambda^2 & 2+\lambda^2\end{bmatrix}$$

$$\boldsymbol{\mathcal{X}}_{J_1} = \begin{bmatrix}1 & 0\\ 1 & 0\\ 0 & 1\end{bmatrix}, \quad \boldsymbol{\mathcal{X}}_{J_2} = \begin{bmatrix}1 & 0\\ 0 & 1\\ 0 & 1\end{bmatrix}.$$

Recall that $J_3 = \{1, 1\} \notin \mathcal{M}$, and hence, no strict minimizer of $\mathcal{E}(., \boldsymbol{y})$ for any $\boldsymbol{y}$ can involve two edges. Notice that $\boldsymbol{A}(\boldsymbol{\mathcal{X}}_{J_1} - \boldsymbol{\mathcal{X}}_{J_2}) = \boldsymbol{0}$ and that $\#J_1 = \#J_2$. After some calculations similar to those outlined in Example 2, we see that if $|y_1 - y_2| > \sqrt{\alpha}/\lambda\sqrt{1+\lambda^2}$, then both $\hat{\boldsymbol{x}}_1 = (y_1, y_1, y_2)$ and $\hat{\boldsymbol{x}}_2 = (y_1, y_2, y_2)$ are strict global minimizers, yielding $\mathcal{E}(\hat{\boldsymbol{x}}_1, \boldsymbol{y}) = \mathcal{E}(\hat{\boldsymbol{x}}_2, \boldsymbol{y}) = \alpha$. In other words, the jump can equivalently be placed either between x_1 and x_2 or between x_2 and x_3. Neither data nor prior gives any reason to choose the one among these minimizers. △

More generally, A) is false *only* if all the columns of $\boldsymbol{\mathcal{X}}_{J_p} - \boldsymbol{\mathcal{X}}_{J_q}$ are in the null space of $\boldsymbol{A}$. Then, if in addition $\#J_p = \#J_q$, the set $\mathcal{N}_{J_p, J_q}$ in (30) reads $\mathcal{N}_{J_p, J_q} = Y_{J_p} \bigcap Y_{J_q}$, whereas $\mathcal{N}_{J_p, J_q} = \emptyset$ if $\#J_p \neq \#J_q$. In any case, having $\boldsymbol{A}(\boldsymbol{\mathcal{X}}_{J_p} - \boldsymbol{\mathcal{X}}_{J_q}) = \boldsymbol{0}$ is a truly pathological situation. Assumption A) is generally satisfied by the observation operators used in practice. Note that the latter is easy to test off line.

Now, we would like to have some more facts about the global minimizers of $\mathcal{E}(., \boldsymbol{y})$. Given $J \in \mathcal{M}$, Theorem 3 shows that every $\boldsymbol{y}$ for which $\mathcal{E}(., \boldsymbol{y})$ has a *global* minimizer $\hat{\boldsymbol{x}}$ with $\mathcal{J}(\hat{\boldsymbol{x}}) = J$ belongs to the following set:

$$\tilde{Y}_J = \left\{ \boldsymbol{y} \in \mathbb{R}^m \colon |\boldsymbol{d}_k^T \boldsymbol{\mathcal{X}}_J \boldsymbol{y}| \geq \frac{\theta}{\Gamma_k}, \text{ if } k \in J \right.$$
$$\left. \text{and } |\boldsymbol{d}_k^T \boldsymbol{\mathcal{X}}_J \boldsymbol{y}| \leq \theta\Gamma_k, \text{ otherwise} \right\}$$
$$\text{where } \theta = \frac{\sqrt{\alpha}}{\lambda}. \tag{32}$$

Clearly, $\tilde{Y}_J \subset Y_J$, where the inclusion is strict. Since each $\tilde{Y}_J$ contains an open subset of $\mathbb{R}^m$, the probability to get noisy data belonging to such a set $\tilde{Y}_J$ is strictly positive, and hence, we can really get minimizers whose set of edges is J. In a global plan, when $\boldsymbol{y}$ ranges over $\mathbb{R}^m$, we get strict minimizers involving all edge configurations in $\mathcal{M}$.

Next , we focus on the way in which $\hat{J} = \mathcal{J}(\hat{\boldsymbol{x}})$ behaves under small variations of the data $\boldsymbol{y}$.

Theorem 6: Let A) of Theorem 5 hold for every J_p, J_q in $\mathcal{M}$, as defined in (24).

Then, for almost any $\boldsymbol{y} \in \mathbb{R}^m$, if $\hat{\boldsymbol{x}}$ is a global minimizer of $\mathcal{E}(., \boldsymbol{y})$, then there exists $\xi > 0$ such that $\hat{\boldsymbol{x}}' = \boldsymbol{\mathcal{X}}_{\hat{J}} \boldsymbol{y}'$, where $\hat{J} = \mathcal{J}(\hat{\boldsymbol{x}})$, and $\boldsymbol{\mathcal{X}}_{\hat{J}}$ is defined according to (25) and is a global minimizer of $\mathcal{E}(., \boldsymbol{y}')$ if $\boldsymbol{y}' \in B(\boldsymbol{y}, \xi)$.

Given $\boldsymbol{y} \in \mathbb{R}^m$, consider $\hat{\boldsymbol{x}}$ to be a global minimizer of $\mathcal{E}(., \boldsymbol{y})$ and the relevant edge estimate $\hat{J} = \mathcal{J}(\hat{\boldsymbol{x}})$. By Theorem 6, it is almost sure that this edge estimate will be kept constant $\mathcal{J}(\boldsymbol{\mathcal{X}}_{\hat{J}} \boldsymbol{y}') = \hat{J}$ under small data variations $\boldsymbol{y}' \in B(\boldsymbol{y}, \xi)$, which can be due to noise contamination. This behavior can be seen as a *local stability* property exhibited by an estimator of the form of (1)–(4). However, stronger data variations—which reverse the sign of $\boldsymbol{\mathcal{G}}_{J_q} - \boldsymbol{\mathcal{G}}_{\hat{J}}$ for some $q \in S_{\boldsymbol{y}}$—reverse the configuration of the jumps at the global minimizer. By Theorem 3, the magnitudes of all differences whose indexes belong to $\{k \in \hat{J} \colon k \notin J_q\}$ are modified *discontinuously* with a jump from a value larger than θ/Γ_k to a value smaller then $\theta\Gamma_k$. Inversely, the magnitudes of differences relevant to $\{k \notin \hat{J} \colon k \in J_q\}$ jump from a value smaller then $\theta\Gamma_k$ to a value larger than θ/Γ_k. This effect is the core of the edge detection performed by an estimator of the form (1)–(4).

V. Numerical Experiments

The first experiment, which is presented in Fig. 1, concerns the deconvolution of noisy data. Its goal is to illustrate the ability of different PFs φ, involved in (3), to recover both smoothly varying zones and sharp edges. The data, which are processed, are presented with a solid line in Fig. 1(a); they are obtained from the original signal $\boldsymbol{x}$ and plotted with a dashed line in all figures from (a) to (d) using $\boldsymbol{y} = \boldsymbol{x} * \boldsymbol{h} + \boldsymbol{n}$, where $h_k = \exp(-0.4k^2)/1.41$ for $|k| \leq 5$ and $h_k = 0$ otherwise, and $\boldsymbol{n}$ is white Gaussian noise leading to 10 dB SNR. The shape of the PF that is used in each reconstruction is plotted above the obtained solution.

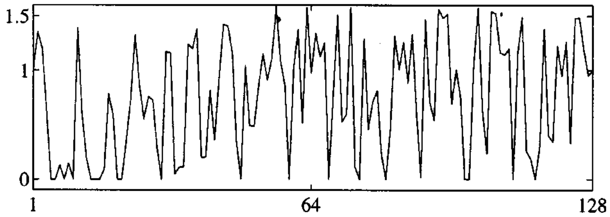

Fig. 2. Pointwise linear operator applied to the original signals whose differnces are depicted in Fig. 3(a).

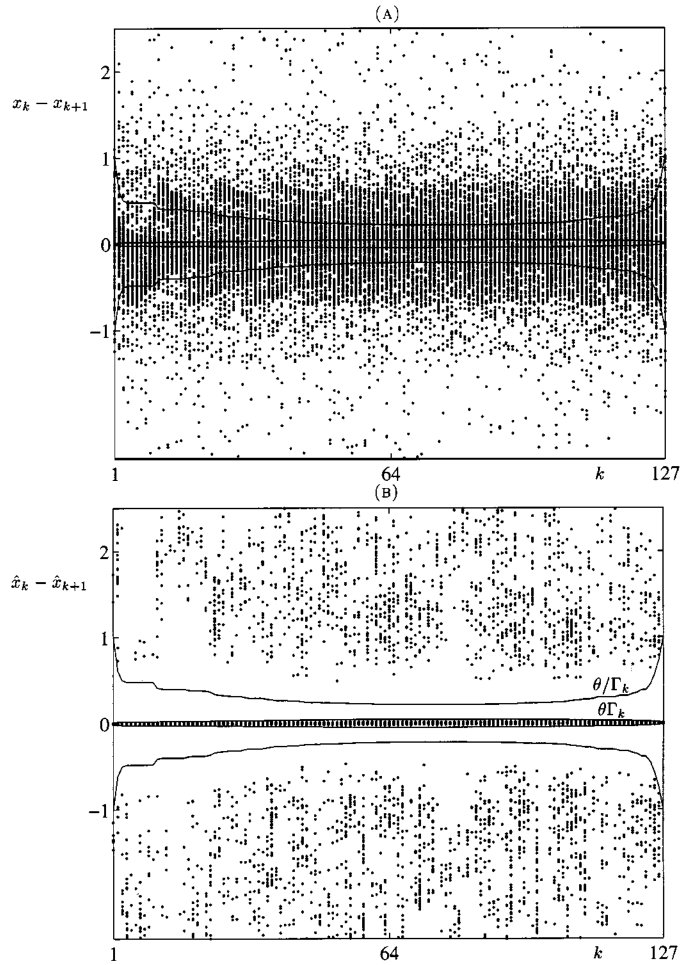

Fig. 3. Distribution of the differences. In (a) and (b), the thresholds $\pm\theta\Gamma_k, \pm\theta/\Gamma_k$ for $k = 1, \ldots, 127$ are plotted with a solid line (—). X—axis: positions of differneces $k = 1, \ldots, 127$. Y-axis: a dot at position$k$ is the value of the kth difference of a signal. Thus, the differences of 100 signals (each having 128 samples) are represented. (a) Differences of the original signals. It is worth noting that numerous differences are placed in the intervals $] - \theta/\Gamma_k, -\theta/\Gamma_k[and]\theta/\Gamma_k, \theta/\Gamma_k[$. (b) Differences of the global minimizers of the objective functions. As predicted, *no* difference is placed in $] - \theta/\Gamma_k, -\theta\Gamma_k[\cup]\theta\Gamma_k, \theta/\Gamma_k$

Fig. 1(b) shows an estimation using a *Huber* PF that is quadratic near the origin [$\varphi(t) = \alpha\lambda t^2/2$ if $|t| < 1/\alpha$] and affine beyond it [$\varphi(t) = \lambda(|t| - 1/2\alpha)$] if $|t| \geq 1/\alpha$. The relevant objective function is convex. This PF smoothes small differences while it adds a constant bias to large differences. The solution in Fig. 1(b) corresponds to $\alpha = 0.1$ and $\lambda = 1$. Fig. 1(c) illustrates an estimation using a *modulus* PF $\varphi(t) = \lambda|t|$ with $\lambda = 2$. This PF is nonsmooth at zero, which leads to estimates that are constant over large zones [28]. The solution thus involves several large differences, but the fact that the local variations beyond them are reduced to constant segments is often undesirable. In contrast, the solution reached using a *truncated quadratic* PF, which is given in Fig. 1(d), exhibits a nice reconstruction of both edges and smooth regions. It corresponds to $(\alpha, \lambda) = (0.7, 7.7)$ and was calculated using the generalized GNC algorithm proposed in [6].

The next experiment, which is given in Figs. 2 and 3, shows the distribution of a large set of estimated differences with respect to the thresholds found in Section III. To this end, 100 original 128-length signals were generated. The data corresponding to an original signal $\boldsymbol{x}$ are obtained by multiplying, pointwise, each one of its samples with the relevant element of the vector $\boldsymbol{h}$ presented in Fig. 2 and by adding white Gaussian noise: $y_k = x_k h_k + n_k$. In Fig. 3(a), the positions $k = 1, \cdots, 127$ of the differences $x_k - x_{k+1}$ of these original signals $\boldsymbol{x}$ are placed along the x axis, whereas their values are represented along the y axis. The subsequent reconstructions of these original signals from the noisy data are calculated for $(\alpha, \lambda) = (1, 10)$. The relevant thresholds $\pm\theta\Gamma_k$, $\pm\theta/\Gamma_k$, with θ as given in (16), are calculated using (17) and are plotted with a solid line in both Fig. 3(a) and (b).

For any k, the original signals have *numerous* differences $x_k - x_{k+1}$ belonging to $[\theta\Gamma_k, \theta/\Gamma_k]$ or to $[-\theta/\Gamma_k, -\theta\Gamma_k]$; see Fig. 3(a). The estimates, which are used to plot Fig. 3(b), are calculated by means of a Viterbi algorithm [9], which guarantees that they yield a global minimum of the objective function. It is striking to observe how the differences of the obtained estimates $\hat{x}_k - \hat{x}_{k+1}$ *avoid* the intervals $[-\theta/\Gamma_k, -\theta\Gamma_k]$ and $[\theta\Gamma_k, \theta/\Gamma_k]$ for all k and for all $\boldsymbol{y}$.

The last experiment in Figs. 4 and 5 concerns the inversion of a discrete Laplace transform. The data $\boldsymbol{y}$ plotted in Fig. 4(a) are obtained according to $y_k = \sum_{l=1}^{128} \exp[-2\pi i(k-1)(l-1)/128] h_l x_l + n_k$ for $k = 1, \cdots, 128$ and $h_l = \exp[-0.2(l-1)]$, whereas $\boldsymbol{n}$ is white Gaussian noise yielding 15 dB SNR. The attenuation factor involved in this transform h_l, $l = 1, \cdots, 128$ is given in Fig. 4(b). The data $\boldsymbol{y}$ are plotted with a dashed line in Fig. 5(a).

Because of the attenuation involved in $\boldsymbol{A}$, the threshold θ/Γ_k is rapidly increasing with k, whereas $\theta\Gamma_k$ is decreasing at the same time. Consequently, it will be increasingly difficult to recover edges when k increases. This effect is obvious in the reconstruction presented in Fig. 5(a). The obtained estimate corresponds to $(\alpha, \lambda) = (1, 10.5)$. The edges, which are located in the first part of the signal, are well recovered. However, those located in its second part are not detected at all and are approximated by an almost constant piece. Since the above Laplace transform can be represented as a discrete Fourier transform applied to the attenuated signal $h_l x_l$, $l = 1, \cdots, 128$, it was possible to calculate the estimate by using a Viterby algorithm. In Fig. 5(b), the differences of the original signal (plotted with "o") and the differences of the reconstructed signal (plotted with "$*$") are compared with the magnitude of the thresholds $\pm\theta\Gamma_k$, $\pm\theta/\Gamma_k$. The sought signal containing constant zones θ is about 0.1, and $\pm\theta\Gamma_k$ are close to zero. The upper threshold θ/Γ_k rapidly increases with k, which underlies the impossibility for the reconstructed signal to contain edges for k large, whereas the original one does contain such large differences.

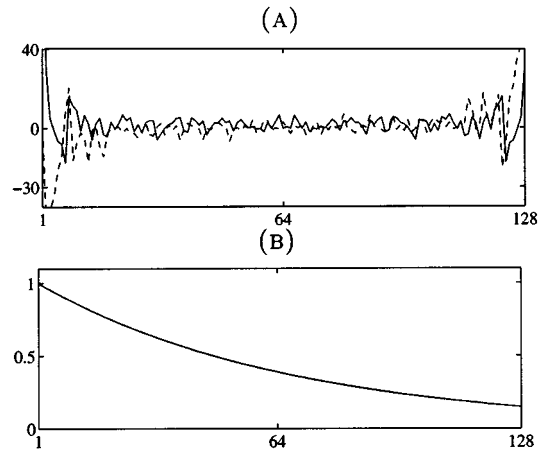

Fig. 4. Discrete Laplace transform. (a) Real part (—) and imaginary part (- -) of the data. (b) Attenuation factor $h_l = \exp[-0.2(l-1)]$ for $l = 1, \ldots, 128$.

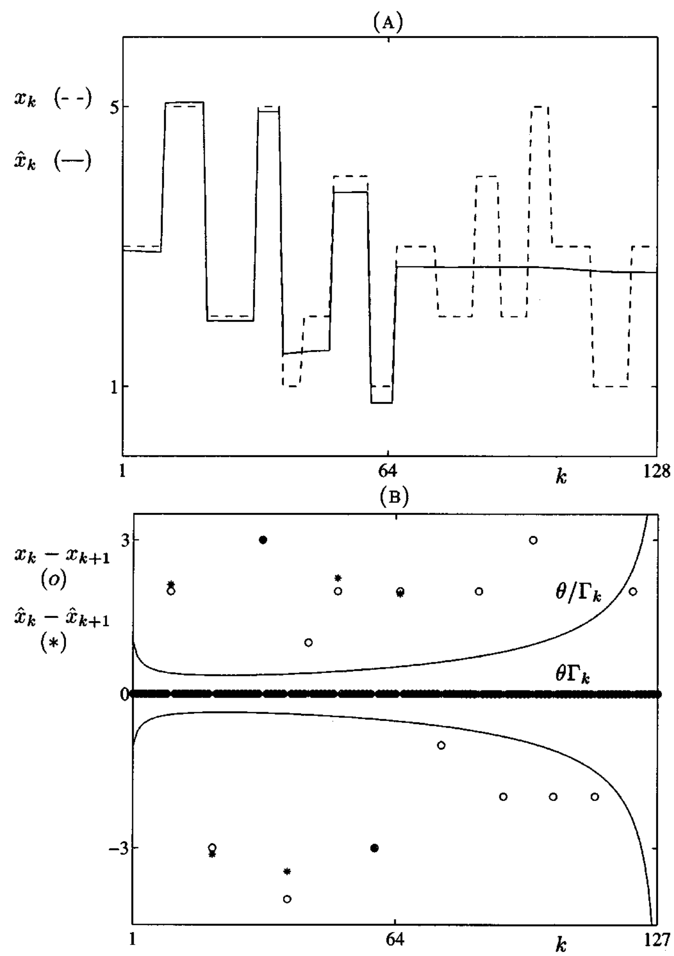

Fig. 5. Reconstruction. (a) Original signal (- -), reconstruction (—), corresponding to $(\alpha, \lambda) = (1, 10.5)$. (b) Differences of the original signal (o) differnces of the solution ($*$). Thresholds $\pm\theta\Gamma_k, \pm\theta/\Gamma_k$ plotted with (—). The increase of θ/Γ_k with k is worth noting.

VI. Conclusion

In this paper, we analyzed an estimator involving truncated quadratic regularization. The relevant objective function is nonsmooth and highly nonconvex, and we examined the regularity of its minimizers. It was shown that this function is differentiable in the vicinity of each one of its local minimizers and that its global minimizers are strict; therefore, they determine locally continuous minimizer functions.

Then, we considered the way in which edges are recovered in an estimated signal. It was exhibited that with each difference between adjacent samples, there is associated with it an interval of positive length so that the magnitude of this difference at any global minimizer, for any data $\boldsymbol{y}$, cannot lie in this interval. When the data vary, the differences of the global minimizers jump over these intervals, which corresponds to an edge detection. This is the mathematical cause that underlies the edge-enhancement effect induced by regularization with a truncated quadratic PF. We demonstrated that at the same time, almost all global minimizers are continuous on some neighborhood, which can be interpreted as a local stability property of the estimator.

APPENDIX A
FAMILY OF EQUIVALENT OBJECTIVE FUNCTIONS

The regularization term Φ in (2) involves $n-1$ differences of the form $x_k - x_{k+1}$. Then, $\mathcal{E}(., \boldsymbol{y})$ can equivalently be expressed as a function of all these differences plus an auxiliary component

$$t_k = x_k - x_{k+1} \text{ for } k = 1, \cdots, n-1 \tag{33}$$
$$t_n = \boldsymbol{h}^T \boldsymbol{x} \tag{34}$$

where $\boldsymbol{h} \in \mathbb{R}^n$ is such that the mapping $\boldsymbol{x} \mapsto \boldsymbol{t}$ defined in (33) and (34) is invertible. Putting $\boldsymbol{t}^{n-1} := [t_1, \cdots, t_{n-1}]^T$, $\boldsymbol{t}^{n-1} = \boldsymbol{D}\boldsymbol{x}$, where $\boldsymbol{D}$ is the matrix given in Proposition 2. For $\boldsymbol{h}$ as specified above, we have

$$\mathcal{F}(\boldsymbol{t}, \boldsymbol{y}) := \mathcal{E}[\boldsymbol{\mathcal{P}}\boldsymbol{t}, \boldsymbol{y}] = \|\boldsymbol{B}\boldsymbol{t} - \boldsymbol{y}\|^2 + \Psi(\boldsymbol{t}) \tag{35}$$
$$\text{where } \boldsymbol{B} = \boldsymbol{A}\boldsymbol{\mathcal{P}}, \quad \text{with } \boldsymbol{\mathcal{P}} = \begin{bmatrix} \boldsymbol{D} \\ \boldsymbol{h}^T \end{bmatrix}^{-1} \tag{36}$$
$$\Psi(\boldsymbol{t}) = \sum_{k=1}^{n-1} \varphi(t_k). \tag{37}$$

$\mathcal{F}(., \boldsymbol{y})$ is equivalent to $\mathcal{E}(., \boldsymbol{y})$ in the sense that any local minimizer $\hat{\boldsymbol{t}}$ of $\mathcal{F}(., \boldsymbol{y})$ is related to the corresponding local minimizer $\hat{\boldsymbol{x}}$ of $\mathcal{E}(., \boldsymbol{y})$ through

$$\begin{array}{rcl} \hat{\boldsymbol{t}}^{n-1} & = & \boldsymbol{D}\hat{\boldsymbol{x}} \\ \hat{t}_n & = & \boldsymbol{h}^T\hat{\boldsymbol{x}} \end{array} \Leftrightarrow \hat{\boldsymbol{x}} = \boldsymbol{\mathcal{P}}\hat{\boldsymbol{t}}. \tag{38}$$

In the following, we will need to know the form of the columns of $\boldsymbol{B}$.

Lemma 1: For any $\boldsymbol{h} \in \mathbb{R}^n$ for which (33) and (34) is invertible, there exists $\boldsymbol{c} \in \mathbb{R}^n$ with $c_n \neq 0$ so that the columns of $\boldsymbol{B}$, as defined in (36), read

$$\boldsymbol{b}_k = c_k \boldsymbol{A}\boldsymbol{1} - \boldsymbol{A}\boldsymbol{u}_k \text{ for } k = 1, \cdots, n-1 \tag{39}$$
$$\boldsymbol{b}_n = c_n \boldsymbol{A}\boldsymbol{1} \tag{40}$$

with $\boldsymbol{u}_k \in \mathbb{R}^n$ as given in (15).

Proof: We start by calculating the matrix $\boldsymbol{\mathcal{P}}$, which is given in the right side of (36). This matrix is partitioned as $\boldsymbol{\mathcal{P}} = [\boldsymbol{\mathcal{P}}^{n-1} \ \vdots \ \boldsymbol{p}_n]$, where $\boldsymbol{\mathcal{P}}^{n-1}$ contains the first $n-1$ columns of $\boldsymbol{\mathcal{P}}$. Then, the identity in the right of (36) can be reformulated as the following system:

$$\begin{aligned} \boldsymbol{D}\boldsymbol{\mathcal{P}}^{n-1} &= \boldsymbol{I}^{n-1} \\ \boldsymbol{D}\boldsymbol{p}_n &= \boldsymbol{0}^{n-1} \\ \boldsymbol{h}^T\boldsymbol{p}_n &= 1 \\ \boldsymbol{h}^T\boldsymbol{\mathcal{P}}^{n-1} &= (\boldsymbol{0}^{n-1})^T. \end{aligned}$$

These conditions are examined next.

- $\boldsymbol{D}\boldsymbol{\mathcal{P}}^{n-1} = \boldsymbol{I}^{n-1}$. Then, each column $\boldsymbol{p}_k$ of $\boldsymbol{\mathcal{P}}^{n-1}$ must satisfy $\boldsymbol{D}\boldsymbol{p}_k = \boldsymbol{e}_k$, where $\boldsymbol{e}_k$ is the kth vector of the standard basis of $\mathbb{R}^{n-1}$ (i.e., $\boldsymbol{e}_k[k] = 1$ and $\boldsymbol{e}_k[i] = 0$ for $i \neq k$). The matrix $\boldsymbol{D}$ is partitioned as[1]

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{D}' & \vdots & \boldsymbol{0}^{k-1 \times n-k} \\ 1 & \vdots & -1 \\ \boldsymbol{0}^{n-k-1 \times k} & \vdots & \boldsymbol{D}'' \end{bmatrix} (\text{row } k)$$

where $\boldsymbol{D}'$ is of size $k-1 \times k$, and $\boldsymbol{D}''$ is $n-k-1 \times n-k$. Vector $\boldsymbol{e}_k$ is partitioned correspondingly as $\boldsymbol{e}_k = [(\boldsymbol{0}^{k-1})^T, \ 1, \ (\boldsymbol{0}^{n-k-1})^T]^T$ and $\boldsymbol{p}_k$ as $\boldsymbol{p}_k = [(\boldsymbol{p}')^T, p_k, (\boldsymbol{p}'')^T]^T$.

Using that the null space of both $\boldsymbol{D}'$ and $\boldsymbol{D}''$ is composed of the constant vectors, the requirement $\boldsymbol{D}\boldsymbol{p}_k = \boldsymbol{e}_k$ is equivalent to the system

$$\boldsymbol{D}'\boldsymbol{p}' = \boldsymbol{0}^{k-1} \quad \text{then } \boldsymbol{p}' = c'\boldsymbol{1} \text{ with } c' \in \mathbb{R} \tag{41}$$
$$\boldsymbol{D}''\boldsymbol{p}'' = \boldsymbol{0}^{n-k-1} \quad \text{then } \boldsymbol{p}'' = c''\boldsymbol{1} \text{ with } c'' \in \mathbb{R} \tag{42}$$
$$\boldsymbol{d}_k^T\boldsymbol{p}_k = 1, \quad \text{then } \begin{cases} \boldsymbol{p}_k[k-1] - \boldsymbol{p}_k[k] = 0 \\ \boldsymbol{p}_k[k] - \boldsymbol{p}_k[k+1] = 1. \end{cases} \tag{43}$$

From the first expression in (43), $\boldsymbol{p}[k] = c'$, whereas from the second, $\boldsymbol{p}[k] = 1 - c''$, and hence, $c'' = c' - 1$. Combining this with (41) and (42) shows that the columns of $\boldsymbol{\mathcal{P}}^{n-1}$ have the form

$$\boldsymbol{p}_k = c_k\boldsymbol{1} - \boldsymbol{u}_k \text{ for } k = 1, \cdots, n-1 \tag{44}$$

where c_k are real constants.
- $\boldsymbol{D}\boldsymbol{p}_n = \boldsymbol{0}^{n-1}$. Then, $\boldsymbol{p}_n = c_n\boldsymbol{1}$, where c_n is a real number.
- $\boldsymbol{h}^T\boldsymbol{p}_n = 1$. By using the last result, we have

$$c_n\boldsymbol{h}^T\boldsymbol{1} = 1. \tag{45}$$

Note that $\boldsymbol{h}^T\boldsymbol{1} \neq 0$, since otherwise, $[\boldsymbol{D}^T \ \vdots \ \boldsymbol{h}]^T$ is singular. Hence

$$c_n = \frac{1}{\boldsymbol{h}^T\boldsymbol{1}} \neq 0. \tag{46}$$

[1]More precisely, $\boldsymbol{D}'$ and $\boldsymbol{D}''$ are the following submatrices of $\boldsymbol{D}$:

$$\boldsymbol{D}'(i, j) = \boldsymbol{D}(i, j) \quad \text{for } i = 1, \cdots, k-1 \text{ and } j = 1, \cdots, k$$
$$\boldsymbol{D}''(i, j) = \boldsymbol{D}(k+i, k+j) \quad \text{for } i = 1, \cdots, n-k-1 \text{ and } j = 1, \cdots, n-k.$$

- $\boldsymbol{h}^T\boldsymbol{\mathcal{P}}^{n-1} = \mathbf{0}^T$. Since rank$\boldsymbol{\mathcal{P}}^{n-1} = n-1$, we get $\boldsymbol{h} \in \operatorname{Ker}\boldsymbol{\mathcal{P}}^{n-1^T}$, or equivalently, $\boldsymbol{h}^T\boldsymbol{p}_k = 0$ for every $k = 1, \cdots, n-1$. Introducing (44) into the last system yields
$$c_k\boldsymbol{h}^T\mathbf{1} = \boldsymbol{h}^T\boldsymbol{u}_k, \quad \text{hence } c_k = \frac{\boldsymbol{h}^T\boldsymbol{u}_k}{\boldsymbol{h}^T\mathbf{1}} \quad \text{for } k = 1, \cdots, n-1. \tag{47}$$
It is easy to check that the obtained matrix $\boldsymbol{\mathcal{P}}$ is the sought inverse. The columns of $\boldsymbol{B}$ come from $\boldsymbol{B} = \boldsymbol{A}\boldsymbol{\mathcal{P}}$.

Remark 2: It may occur that $\boldsymbol{A}\mathbf{1} = \mathbf{0}$. In such a case, the columns of $\boldsymbol{B}$ read
$$\boldsymbol{b}_k = -\boldsymbol{A}\boldsymbol{u}_k \text{ for } k = 1, \cdots, n-1, \quad \text{while } \boldsymbol{b}_n = \mathbf{0}.$$
Since $\boldsymbol{b}_n t_n = 0$, the quadratic term in (35) can be expressed as $\|\boldsymbol{B}\boldsymbol{t} - \boldsymbol{y}\|^2 = \|\boldsymbol{B}^{n-1}\boldsymbol{t}^{n-1} - \boldsymbol{y}\|^2$, where $\boldsymbol{B}^{n-1}$ is formed by the first $n-1$ columns of $\boldsymbol{B}$. Therefore, t_n is not involved in $\mathcal{F}(\boldsymbol{t}, \boldsymbol{y})$. The minimization problem should be reformulated in terms of $\boldsymbol{t}^{n-1}$ only:
$$\tilde{\mathcal{F}}(\boldsymbol{t}^{n-1}, \boldsymbol{y}) = \|\boldsymbol{B}_1^{n-1}\boldsymbol{t}^{n-1} - \boldsymbol{y}\|^2 + \sum_{k=1}^{n-1}\varphi(t_k).$$
All results established in this paper hold for $\tilde{\mathcal{F}}(., \boldsymbol{y})$ as well, but they take a simpler form. △

APPENDIX B
PROOFS RELEVANT TO SECTION II

A. Proof of Proposition 1

Since (33) and (34) is an isomorphism, this proposition is proven by checking whether there exists $\boldsymbol{y} \in \mathbb{R}^m$ such that $\mathcal{F}(., \boldsymbol{y})$ has local minimizers belonging to the union of hyperplanes
$$\bigcup_{k=1}^{n-1}\{[t_k = \theta] \cup [t_k = -\theta]\} \text{ where } \theta := \frac{\sqrt{\alpha}}{\lambda}.$$
Suppose the contrary, namely, that there exists $\boldsymbol{y} \in \mathbb{R}^m$ for which $\mathcal{F}(., \boldsymbol{y})$ has a local or a global minimizer $\hat{\boldsymbol{t}}$ such that $|\hat{t}_k| = \theta$ for $k \in \mathcal{K}$, where $\mathcal{K} \subseteq \{1, \cdots, n-1\}$ is a subset of indices. For some $k \in \mathcal{K}$, we will analyze $\mathcal{F}(., \boldsymbol{y})$ at $\hat{\boldsymbol{t}}$ along the direction $\boldsymbol{e}_k$. Without loss of generality, suppose that $\hat{t}_k = \theta$ (the reasoning is the same for $\hat{t}_k = -\theta$). Define the function $f: \mathbb{R} \mapsto \mathbb{R}$ to read
$$\ell \mapsto f(\ell) = \mathcal{F}(\hat{\boldsymbol{t}} + \ell\boldsymbol{e}_k, \boldsymbol{y}). \tag{48}$$
The fact that $\hat{\boldsymbol{t}}$ is a local minimizer of $\mathcal{F}(., \boldsymbol{y})$ implies that f has a local minimum at $\hat{\ell} = 0$. Consequently, its side derivatives at zero satisfy both inequalities:[2]
$$f'_-(0) \le 0 \le f'_+(0). \tag{49}$$
We will take $s \ge 0$. The left and the right derivatives of f at zero read
$$f'_-(0) := \lim_{s\downarrow 0}\frac{\mathcal{F}(\hat{\boldsymbol{t}} - s\boldsymbol{e}_k, \boldsymbol{y}) - \mathcal{F}(\hat{\boldsymbol{t}}, \boldsymbol{y})}{-s}$$
and
$$f'_+(0) := \lim_{s\downarrow 0}\frac{\mathcal{F}(\hat{\boldsymbol{t}} + s\boldsymbol{e}_k, \boldsymbol{y}) - \mathcal{F}(\hat{\boldsymbol{t}}, \boldsymbol{y})}{s}$$

[2]Numerous references can be cited about (49), e.g., [29]. Let us recall its proof. Since f has a minimum at zero, for any sufficiently small $s > 0$, we have $f(-s) - f(0) \ge 0$ and $f(s) - f(0) \ge 0$. Then, $[f(-s) - f(0)]/(-s) \le 0$, and $[f(s) - f(0)]/s \ge 0$. The necessary condition (49) is obtained at the limit when $s \to 0$.

respectively. Then
$$f'_-(0) = \lim_{s\downarrow 0} \frac{\|\boldsymbol{B}(\hat{\boldsymbol{t}} - s\boldsymbol{e}_k) - \boldsymbol{y}\|^2 + \Psi(\hat{\boldsymbol{t}} - s\boldsymbol{e}_k) - \|\boldsymbol{B}\hat{\boldsymbol{t}} - \boldsymbol{y}\|^2 - \Psi(\hat{\boldsymbol{t}})}{-s}.$$
Noticing that $\hat{t}_k = \theta$ and that $\varphi(\theta - s) = \lambda^2(\theta - s)^2$ for $s \in]0, 2\theta[$, we get
$$\begin{aligned}\Psi(\hat{\boldsymbol{t}} - s\boldsymbol{e}_k) &= \sum_{i=1}^{n-1}\varphi(\hat{t}_i - s\boldsymbol{e}_k[i]) \\ &= \sum_{i\ne k}\varphi(\hat{t}_i) + \varphi(\hat{t}_k - s) \\ &= \Psi(\hat{\boldsymbol{t}}) - \varphi(\hat{t}_k) + \varphi(\hat{t}_k - s) = \Psi(\hat{\boldsymbol{t}}).\end{aligned}$$
Consequently
$$\begin{aligned}f'_-(0) &= \lim_{s\downarrow 0} \cdot \frac{s^2\|\boldsymbol{B}\boldsymbol{e}_k\|^2 - 2s\boldsymbol{e}_k^T\boldsymbol{B}^T(\boldsymbol{B}\hat{\boldsymbol{t}} - \boldsymbol{y}) + \lambda^2(\theta - s)^2 - \lambda^2\theta^2}{-s} \\ &= 2\boldsymbol{b}_k^T(\boldsymbol{B}\hat{\boldsymbol{t}} - \boldsymbol{y}) + 2\lambda^2\theta.\end{aligned}$$
On the other hand, for any $s > 0$, we have $\varphi(\theta + s) = \alpha = \varphi(\theta) = \lambda^2\theta^2$, and hence, $\Psi(\hat{\boldsymbol{t}} + s\boldsymbol{e}_k) = \Psi(\hat{\boldsymbol{t}})$. In a similar way, we find
$$\begin{aligned}f'_+(0) &= \lim_{s\downarrow 0}\frac{s^2\|\boldsymbol{B}\boldsymbol{e}_k\|^2 - 2s\boldsymbol{e}_k^T\boldsymbol{B}^T(\boldsymbol{B}\hat{\boldsymbol{t}} - \boldsymbol{y})}{s} \\ &= 2\boldsymbol{b}_k^T(\boldsymbol{B}\hat{\boldsymbol{t}} - \boldsymbol{y}).\end{aligned}$$
Clearly, $f'_-(0) > f'_+(0)$, which contradicts the necessary condition (49). Hence, we have the result.

B. Proof of Proposition 2

Necessary Condition: By (38), we consider equivalently $\hat{\boldsymbol{t}}$ to be a local minimizer of $\mathcal{F}(., \boldsymbol{y})$. Clearly, $\mathcal{F}(., \boldsymbol{y})$ is $\mathcal{C}^2$ on $B(\hat{\boldsymbol{t}}, \rho)$, where ρ is as given in (7). Then, the gradient of $\mathcal{F}(., \boldsymbol{y})$ at $\hat{\boldsymbol{t}}$ is null, that is, $\nabla\mathcal{F}(\hat{\boldsymbol{t}}, \boldsymbol{y}) = \mathbf{0}$:
$$2\boldsymbol{B}^T(\boldsymbol{B}\hat{\boldsymbol{t}} - \boldsymbol{y}) + \nabla\Psi(\hat{\boldsymbol{t}}) = \mathbf{0}. \tag{50}$$
Using that $|\hat{t}_k| \ne \theta$ for all k and that $\varphi'(t) = 2\lambda^2 t\mathbb{1}(|t| < \theta)$ if $|t| \ne \theta$, we get
$$\nabla\Psi(\hat{\boldsymbol{t}}) = 2\lambda^2\tilde{\boldsymbol{\mathcal{U}}}_{\hat{J}}\hat{\boldsymbol{t}} \text{ with } \hat{J} = \mathcal{J}(\hat{\boldsymbol{t}}) \tag{51}$$
where we use the equivalent forms of (9) and (10):
$$\mathcal{J}(\boldsymbol{t}) = \{k \in \{1, \cdots, n-1\}: |t_k| > \theta\} \tag{52}$$
$$\tilde{\boldsymbol{\mathcal{U}}}_J = \operatorname{Diag}\{\mathbb{1}(|t_1| < \theta), \cdots, \mathbb{1}(|t_{n-1}| < \theta), 0\}. \tag{53}$$
Notice that $\tilde{\boldsymbol{\mathcal{U}}}_J$ is $n \times n$.

Sufficient Condition: Let $\hat{\boldsymbol{t}}$ satisfy $\nabla\mathcal{F}(\hat{\boldsymbol{t}}, \boldsymbol{y}) = \mathbf{0}$. The equivalent form of (8) shows that $\|\boldsymbol{v}\| < \rho$ ensures that
$$\begin{aligned}&\varphi(\hat{t}_k + v_k) = \lambda^2(\hat{t}_k + v_k)^2, \qquad \text{if } |\hat{\boldsymbol{t}}_k| < \theta \\ &\varphi(\hat{t}_k + v_k) = \alpha, \qquad \text{if } |\hat{\boldsymbol{t}}_k| \ge \theta.\end{aligned}$$

Next, we consider $\mathcal{F}(\hat{\boldsymbol{t}}+\boldsymbol{v}, \boldsymbol{y}) - \mathcal{F}(\hat{\boldsymbol{t}}, \boldsymbol{y})$ for $0 < \|\boldsymbol{v}\| < \rho$:

$$\begin{aligned}\mathcal{F}(\hat{\boldsymbol{t}}+\boldsymbol{v}, \boldsymbol{y}) &- \mathcal{F}(\hat{\boldsymbol{t}}, \boldsymbol{y})\\ &= \|\boldsymbol{B}(\hat{\boldsymbol{t}}+\boldsymbol{v})-\boldsymbol{y}\|^2 - \|\boldsymbol{B}\hat{\boldsymbol{t}}-\boldsymbol{y}\|^2\\ &\quad+\sum_{k=1}^{n-1}[\lambda^2(\hat{t}_k+v_k)^2-\lambda^2 h t_k^2]\mathbb{1}(|\hat{t}_k|<\theta)\\ &= \boldsymbol{v}^T\boldsymbol{B}^T\boldsymbol{B}\boldsymbol{v}+\boldsymbol{v}^T\boldsymbol{B}^T(\boldsymbol{B}\boldsymbol{t}-\boldsymbol{y})\\ &\quad+\sum_{k=1}^{n-1}\lambda^2 v_k^2\mathbb{1}(|\hat{t}_k|<\theta)+2\sum_{k=1}^{n-1}\lambda^2 v_k\hat{t}_k\mathbb{1}(|\hat{t}_k|<\theta)\\ &= \boldsymbol{v}^T\boldsymbol{B}^T\boldsymbol{B}\boldsymbol{v}+\lambda^2\boldsymbol{v}^T\tilde{\boldsymbol{\mathcal{U}}}_{\hat{\mathcal{J}}}\boldsymbol{v}+\boldsymbol{v}^T\nabla\mathcal{F}(\hat{\boldsymbol{t}}, \boldsymbol{y})\geq 0. \qquad (54)\end{aligned}$$

Hence, $\hat{\boldsymbol{t}}$ is a local minimizer of $\mathcal{F}(., \boldsymbol{y})$.

C. *Proof of Proposition 3*

Equivalently, we will see that all minimizers of $\mathcal{F}(., \boldsymbol{y})$ are strict whenever $\boldsymbol{B}^T\boldsymbol{B}$ is invertible. Using the same arguments as in the proof of Proposition 2, we see that now, $\boldsymbol{v}^T\boldsymbol{B}^T\boldsymbol{B}\boldsymbol{v}>0$ ensures that the expression in (54) is strictly positive. Hence, we have the result.

D. *Proof of Proposition 4*

Note that $\hat{\boldsymbol{t}}$ satisfies (50), and we get $\hat{J}=\mathcal{J}(\hat{\boldsymbol{t}})$. By assumption, there is $\hat{\boldsymbol{t}}'\in B(\hat{\boldsymbol{t}}, \rho)$, with ρ as in (7), so that

$$\mathcal{F}(\hat{\boldsymbol{t}}', \boldsymbol{y})=\mathcal{F}(\hat{\boldsymbol{t}}, \boldsymbol{y}) \text{ and } \nabla\mathcal{F}(\hat{\boldsymbol{t}}', \boldsymbol{y})=\boldsymbol{0}.$$

Using that for all $\boldsymbol{t}\in B(\hat{\boldsymbol{t}}; \rho)$ we have $\mathcal{J}(\boldsymbol{t})=\hat{J}$, and hence $\tilde{\boldsymbol{\mathcal{U}}}_{\mathcal{J}(\boldsymbol{t})}=\tilde{\boldsymbol{\mathcal{U}}}_{\hat{\mathcal{J}}}$, we can see that

$$\nabla\mathcal{F}(\hat{\boldsymbol{t}}', \boldsymbol{y})=2\boldsymbol{B}^T(\boldsymbol{B}\hat{\boldsymbol{t}}'-\boldsymbol{y})+2\lambda^2\tilde{\boldsymbol{\mathcal{U}}}_{\hat{\mathcal{J}}}\hat{\boldsymbol{t}}'.$$

Consider $\boldsymbol{z}=\beta\hat{\boldsymbol{t}}+(1-\beta)\hat{\boldsymbol{t}}'$ with $\beta\in[0, 1]$; then, $\boldsymbol{z}\in B(\hat{\boldsymbol{t}}, \rho)$. We get

$$\begin{aligned}\nabla\mathcal{F}(\boldsymbol{z}, \boldsymbol{y}) &= 2\boldsymbol{B}^T(\boldsymbol{B}\boldsymbol{z}-\boldsymbol{y})+2\lambda^2\tilde{\boldsymbol{\mathcal{U}}}_{\hat{\mathcal{J}}}\boldsymbol{z}\\ &= 2\boldsymbol{B}^T[\boldsymbol{B}(\beta\hat{\boldsymbol{t}}+(1-\beta)\hat{\boldsymbol{t}}')-\boldsymbol{y}]\\ &\quad+2\lambda^2\tilde{\boldsymbol{\mathcal{U}}}_{\hat{\mathcal{J}}}[\beta\hat{\boldsymbol{t}}+(1-\beta)\hat{\boldsymbol{t}}']\\ &= \beta 2\boldsymbol{B}^T(\boldsymbol{B}\hat{\boldsymbol{t}}-\boldsymbol{y})+(1-\beta)2\boldsymbol{B}^T(\boldsymbol{B}\hat{\boldsymbol{t}}'-\boldsymbol{y})\\ &\quad+2\lambda^2\beta\tilde{\boldsymbol{\mathcal{U}}}_{\hat{\mathcal{J}}}\hat{\boldsymbol{t}}+2\lambda^2(1-\beta)\tilde{\boldsymbol{\mathcal{U}}}_{\hat{\mathcal{J}}}\hat{\boldsymbol{t}}'\\ &= \beta\nabla\mathcal{F}(\hat{\boldsymbol{t}}, \boldsymbol{y})+(1-\beta)\nabla\mathcal{F}(\hat{\boldsymbol{t}}', \boldsymbol{y})=\boldsymbol{0}.\end{aligned}$$

By Proposition 2, $\mathcal{F}(., \boldsymbol{y})$ has a local minimum at any $\boldsymbol{z}$ placed on the segment that links $\hat{\boldsymbol{t}}$ and $\hat{\boldsymbol{t}}'$.

E. *Proof of Theorem 1*

From (40), assuming that $\boldsymbol{A}\mathbf{1}\neq\mathbf{0}$ means that $\boldsymbol{b}_n\neq\mathbf{0}$. To prove this theorem, we will check whether the global minimizers of $\mathcal{F}(., \boldsymbol{y})$ are strict when $\boldsymbol{B}^T\boldsymbol{B}$ is singular with $\boldsymbol{b}_n\neq\mathbf{0}$. We prove this by contradiction.

Suppose we are given $\hat{\boldsymbol{t}}$, which is a *nonstrict global* minimizer of $\mathcal{F}(., \boldsymbol{y})$. From Proposition 4, there exists a normalized direction $\boldsymbol{v}\in\mathbb{R}^n$ with $\|\boldsymbol{v}\|=1$, and there exist ρ_1, ρ_2, where $\rho_1\leq 0<\rho_2$, such that

$$\mathcal{F}(\hat{\boldsymbol{t}}+\ell\boldsymbol{v}, \boldsymbol{y})=\mathcal{F}(\hat{\boldsymbol{t}}, \boldsymbol{y}) \text{ for all } \ell\in[\rho_1, \rho_2] \qquad (55)$$

whereas $\{\hat{\boldsymbol{t}}+\ell\boldsymbol{v}: \rho_1\leq\ell\leq\rho_2\}\subset B(\hat{\boldsymbol{t}}; \rho)$ with ρ as the radius specified in (7).

Let $\mathcal{J}^C$ denote the complement function that extracts the indexes of the differences in the homogeneous zones:

$$\mathcal{J}^C(\boldsymbol{t}) := \{k\in\{1, \cdots, n-1\}: |t_k|<\theta\}. \qquad (56)$$

Make $\hat{J}^C := \mathcal{J}^C(\hat{\boldsymbol{t}})$. The identity (55) can be reformulated as

$$\begin{aligned}\|\boldsymbol{B}(\hat{\boldsymbol{t}}+\ell\boldsymbol{v})-\boldsymbol{y}\|^2 &+ \sum_{k\in\hat{J}^C}\lambda^2(\hat{t}_k+\ell v_k)^2+\alpha\#\hat{J}\\ &= \|\boldsymbol{B}\hat{\boldsymbol{t}}-\boldsymbol{y}\|^2+\sum_{k\in\hat{J}^C}\lambda^2\hat{t}_k^2+\alpha\#\hat{J} \text{ for all } \ell\in[\rho_1, \rho_2]\end{aligned} \qquad (57)$$

where $\#$ denotes cardinality. It follows that

$$\begin{aligned}&\ell^2\left(\|\boldsymbol{B}\boldsymbol{v}\|^2+\lambda^2\sum_{\hat{J}^C}v_k^2\right)\\ &\quad+2\ell\left[\boldsymbol{v}^T\boldsymbol{B}^T(\boldsymbol{B}\hat{\boldsymbol{t}}-\boldsymbol{y})+\lambda^2\sum_{\hat{J}^C}v_k t_k\right]=0\\ &\quad\text{for all } \ell\in[\rho_1, \rho_2].\end{aligned}$$

Hence, the coefficients of the above second-order polynomial with respect to ℓ must be null. In particular, $\|\boldsymbol{B}\boldsymbol{v}\|^2+\lambda^2\sum_{k\in\hat{J}^C}v_k^2=0$ implies that there exists $\boldsymbol{v}$ such that

$$\begin{aligned}&\boldsymbol{v}\in\operatorname{Ker}\boldsymbol{B}^T\boldsymbol{B},\\ &v_k=0 \text{ for all } k\in\hat{J}^C.\end{aligned} \qquad (58)$$

The result is trivial if no $\boldsymbol{v}$ satisfies (58). Now, suppose that there exists $\boldsymbol{v}$ satisfying (58). Note that by $\|\boldsymbol{v}\|=1$ and by (58), there exists $j\in\hat{J}$ such that $v_j\neq 0$. For such a j, make $\ell^o=-\hat{t}_j/v_j$. Then, $\varphi(\hat{t}_j+\ell^o v_j)=0<\varphi(\hat{t}_j)=\alpha$. For any other $k\in\hat{J}\setminus\{j\}$, two situations can occur:

— If $|\hat{t}_k+\ell^o v_k|\geq\theta$, then $\varphi(\hat{t}_k+\ell^o v_k)=\alpha=\varphi(\hat{t}_k)$.

— If $|\hat{t}_k+\ell^o v_k|<\theta$, then $\varphi(\hat{t}_k+\ell^o v_k)=\lambda^2(\hat{t}_k+\ell^o v_k)^2<\alpha=\varphi(\hat{t}_k)$.

In all cases

$$\sum_{k\in\hat{J}}\varphi(\hat{t}_k+\ell^o v_k)\leq\sum_{k\in\hat{J}}\varphi(\hat{t}_k)-\alpha.$$

Then, we have

$$\begin{aligned}\mathcal{F}(\hat{\boldsymbol{t}}+\ell^o\boldsymbol{v}, \boldsymbol{y}) &= \|\boldsymbol{B}\hat{\boldsymbol{t}}-\boldsymbol{y}\|^2+\sum_{k\in\hat{J}^C}(\lambda\hat{t}_k)^2\\ &\quad+\sum_{k\in\hat{J}}\varphi(\hat{t}_k+\ell^o v_k)<\mathcal{F}(\hat{\boldsymbol{t}}, \boldsymbol{y}).\end{aligned}$$

However, this contradicts the hypothesis that $\mathcal{F}(., \boldsymbol{y})$ has a global minimum at $\hat{\boldsymbol{t}}$. Hence, we have the result.

F. Proof of Theorem 2

Recall first that $\nabla^2\mathcal{E}(\boldsymbol{x}, \boldsymbol{y})$ is positive definite on $B(\hat{\boldsymbol{x}}, \rho)$, i.e., it determines an isomorphism on this ball. The implicit function theorem can then be applied. Then, there exists $\mu > 0$ and a unique function $\boldsymbol{\mathcal{X}}$: $B(\boldsymbol{y}, \mu) \mapsto \mathbb{R}^n$ so that $\nabla\mathcal{E}[\boldsymbol{\mathcal{X}}(\boldsymbol{y}'), \boldsymbol{y}'] = \boldsymbol{0}$ on $B(\boldsymbol{y}, \mu)$. From the preceding, $\hat{\boldsymbol{x}}' = \boldsymbol{\mathcal{X}}(\boldsymbol{y}')$ is a strict minimizer.

APPENDIX C
PROOFS RELEVANT TO SECTION III

A. Proof of Theorem 3

As previously, we will consider $\mathcal{F}(., \boldsymbol{y})$. Let $\hat{\boldsymbol{t}}$ be a global minimizer of $\mathcal{F}(., \boldsymbol{y})$.

The last equation in the system (50) reads

$$\boldsymbol{b}_n^T\left(\sum_{i=1}^{n-1}\boldsymbol{b}_i\hat{t}_i + \boldsymbol{b}_n\hat{t}_n - \boldsymbol{y}\right) = 0.$$

It allows us to express $\hat{t}_n$ as a function of $\hat{\boldsymbol{t}}^{n-1}$, that is

$$\hat{t}_n = \frac{\boldsymbol{b}_n^T}{\|\boldsymbol{b}_n\|^2}\left(\boldsymbol{y} - \sum_{i=1}^{n-1}\boldsymbol{b}_i\hat{t}_i\right). \tag{59}$$

Focus on the kth component of $\hat{\boldsymbol{t}}^{n-1}$ and decompose (59) into terms involving $\hat{t}_k$ and terms not involving $\hat{t}_k$:

$$\hat{t}_n = -\frac{\boldsymbol{b}_n^T\mathcal{Z}_k}{\|\boldsymbol{b}_n\|^2} - \frac{\boldsymbol{b}_n^T\boldsymbol{b}_k\hat{t}_k}{\|\boldsymbol{b}_n\|^2}$$

where

$$\mathcal{Z}_k = \sum_{i=1}^{k-1}\boldsymbol{b}_i\hat{t}_i + \sum_{i=k+1}^{n-1}\boldsymbol{b}_i\hat{t}_i - \boldsymbol{y}. \tag{60}$$

Notice that $\mathcal{Z}_k$ is a function of $t_1, \cdots, t_{k-1}, t_{k+1}, \cdots, t_{n-1}$ and of $\boldsymbol{y}$. The following expression will simplify further calculations

$$\begin{aligned}\boldsymbol{B}\hat{\boldsymbol{t}} - \boldsymbol{y} &= \sum_{i=1}^{k-1}\boldsymbol{b}_i\hat{t}_i + \boldsymbol{b}_k\hat{t}_k + \sum_{i=k+1}^{n-1}\boldsymbol{b}_i\hat{t}_i + \boldsymbol{b}_n\hat{t}_n - \boldsymbol{y}\\ &= \mathcal{Z}_k + \boldsymbol{b}_k\hat{t}_k - \frac{\boldsymbol{b}_n\boldsymbol{b}_n^T}{\|\boldsymbol{b}_n\|^2}\mathcal{Z}_k - \frac{\boldsymbol{b}_n\boldsymbol{b}_n^T}{\|\boldsymbol{b}_n\|^2}\boldsymbol{b}_k\hat{t}_k\\ &= \boldsymbol{M}\mathcal{Z}_k + \boldsymbol{M}\boldsymbol{b}_k\hat{t}_k \end{aligned} \tag{61}$$

with $\boldsymbol{M}$ as given in (14). The last equality is obtained by using the fact that $\boldsymbol{b}_n = c_n\boldsymbol{A}\boldsymbol{1}$, as found in Lemma 1. Combining (39) and (14) shows that

$$\boldsymbol{M}\boldsymbol{b}_k = c_k\boldsymbol{M}\boldsymbol{A}\boldsymbol{1} - \boldsymbol{M}\boldsymbol{A}\boldsymbol{u}_k = -\boldsymbol{M}\boldsymbol{A}\boldsymbol{u}_k.$$

Using the above expression, (61) yields

$$\boldsymbol{B}\hat{\boldsymbol{t}} - \boldsymbol{y} = \boldsymbol{M}\mathcal{Z}_k - \boldsymbol{M}\boldsymbol{A}\boldsymbol{u}_k\hat{t}_k \tag{62}$$

$$\boldsymbol{b}_k^T(\boldsymbol{B}\hat{\boldsymbol{t}} - \boldsymbol{y}) = -\boldsymbol{u}_k^T\boldsymbol{A}^T\boldsymbol{M}\mathcal{Z}_k + \boldsymbol{u}_k^T\boldsymbol{A}^T\boldsymbol{M}\boldsymbol{A}\boldsymbol{u}_k\hat{t}_k \tag{63}$$

where we recall that $\boldsymbol{M}^T = \boldsymbol{M}$ and $\boldsymbol{M}^T\boldsymbol{M} = \boldsymbol{M}$. The kth equation in the system (50), where $k \le n-1$, can be reformulated as

$$-\boldsymbol{u}_k^T\boldsymbol{A}^T\boldsymbol{M}\mathcal{Z}_k + \xi_k\hat{t}_k + \lambda^2\hat{t}_k\mathbb{1}(|\hat{t}_k| < \theta) = 0 \tag{64}$$

with $\xi_k > 0$, as defined in (17). For the moment, we only know that $\hat{t}_k \ne \theta$, but we do not know whether $|\hat{t}_k| > \theta$ or $|\hat{t}_k| < \theta$. Let us examine both situations.

- $\hat{t}_k = \tau_0$ satisfies (64) for some $|\tau_0| < \theta$. Then (64) yields

$$\tau_0 = \frac{\boldsymbol{u}_k^T\boldsymbol{A}^T\boldsymbol{M}\mathcal{Z}_k}{\lambda^2 + \xi_k}. \tag{65}$$

- $\hat{t}_k = \tau_1$ satisfies (64) with $|\tau_1| > \theta$. By (64)

$$\tau_1 = \frac{\boldsymbol{u}_k^T\boldsymbol{A}^T\boldsymbol{M}\mathcal{Z}_k}{\xi_k}. \tag{66}$$

Define the vectors $\boldsymbol{t}_0$ and $\boldsymbol{t}_1$ as follows:

$$\begin{aligned}\boldsymbol{t}_0 &:= [\hat{t}_1, \cdots, \hat{t}_{k-1}, \tau_0, \hat{t}_{k+1}, \cdots, \hat{t}_{n-1}]^T\\ \boldsymbol{t}_1 &:= [\hat{t}_1, \cdots, \hat{t}_{k-1}, \tau_1, \hat{t}_{k+1}, \cdots, \hat{t}_{n-1}]^T\end{aligned}$$

with τ_0 and τ_1, as given in (65) and (66), respectively. Since the kth entry of $\hat{\boldsymbol{t}}$ is either $|\hat{t}_k| < \theta$ or $|\hat{t}_k| > \theta$, we have either $\hat{\boldsymbol{t}} = \boldsymbol{t}_0$ or $\hat{\boldsymbol{t}} = \boldsymbol{t}_1$. The fact that $\hat{\boldsymbol{t}}$ is a global minimizer yields that

$$\mathcal{F}(\hat{\boldsymbol{t}}, \boldsymbol{y}) \le \mathcal{F}(\boldsymbol{t}_0, \boldsymbol{y}) \text{ and } \mathcal{F}(\hat{\boldsymbol{t}}, \boldsymbol{y}) \le \mathcal{F}(\boldsymbol{t}_1, \boldsymbol{y}) \tag{67}$$

and that equality is reached for one of the vectors $\boldsymbol{t}_0$ or $\boldsymbol{t}_1$. More precisely, whenever $\mathcal{F}(., \boldsymbol{y})$ has a unique global minimizer, equality is reached for exactly one among the vectors $\boldsymbol{t}_0$ and $\boldsymbol{t}_1$, and this is the vector equal to $\hat{\boldsymbol{t}}$, whereas the other vector may or may not be a stationary point of $\mathcal{F}(., \boldsymbol{y})$. However, it may happen that both equalities are reaches in (67), which means that both $\boldsymbol{t}_0$ and $\boldsymbol{t}_1$ are global minimizers.

In order to test whether $\mathcal{F}(\boldsymbol{t}_0, \boldsymbol{y}) \le \mathcal{F}(\boldsymbol{t}_1, \boldsymbol{y})$ or not, we will compare, with zero

$$\Delta = \mathcal{F}(\boldsymbol{t}_0, \boldsymbol{y}) - \mathcal{F}(\boldsymbol{t}_1, \boldsymbol{y}).$$

Let us split $\mathcal{F}(., \boldsymbol{y})$ into terms involving $\hat{t}_k$ and terms not involving $\hat{t}_k$:

$$\begin{aligned}\mathcal{F}(\tilde{\boldsymbol{t}}, \boldsymbol{y}) &= \|\boldsymbol{M}\mathcal{Z}_k - \boldsymbol{M}\boldsymbol{A}\boldsymbol{u}_k\hat{t}_k\|^2 + \sum_{i\ne k}\varphi(\hat{t}_i) + \varphi(\hat{t}_k)\\ &= \hat{t}_k^2\xi_k - 2\hat{t}_k\boldsymbol{u}_k^T\boldsymbol{A}^T\boldsymbol{M}\mathcal{Z}_k + \varphi(\hat{t}_k) + \mathcal{Z}_k^T\boldsymbol{M}\mathcal{Z}_k\\ &\quad + \sum_{i=1}^{k-1}\varphi(\hat{t}_i) + \sum_{i=k+1}^{n-1}\varphi(\hat{t}_i)\end{aligned}$$

where we used (62)–(64) combined with (17). Then, Δ reads

$$\begin{aligned}\Delta &= \tau_0^2\xi_k - 2\tau_0\boldsymbol{u}_k^T\boldsymbol{A}^T\boldsymbol{M}\mathcal{Z}_k + \varphi(\tau_0)\\ &\quad - \tau_1^2\xi_k + 2\tau_1\boldsymbol{u}_k^T\boldsymbol{A}^T\boldsymbol{M}\mathcal{Z}_k - \varphi(\tau_1).\end{aligned}$$

Substituting the expressions for τ_0 and τ_1 into Δ while recalling that $\varphi(\tau_1) = \alpha$, it is found that

$$\begin{aligned}\Delta &= -\frac{(\boldsymbol{u}_k^T\boldsymbol{A}^T\boldsymbol{M}\mathcal{Z}_k)^2}{\lambda^2 + \xi_k} + \frac{(\boldsymbol{u}_k^T\boldsymbol{A}^T\boldsymbol{M}\mathcal{Z}_k)^2}{\xi_k} - \alpha\\ &= \frac{\lambda^2(\boldsymbol{u}_k^T\boldsymbol{A}^T\boldsymbol{M}\mathcal{Z}_k)^2}{(\lambda^2 + \xi_k)\xi_k} - \alpha.\end{aligned}$$

According to the sign of Δ, two situations arise.

- If $\Delta \leq 0$, then $\boldsymbol{t}_0$ is global minimizer. The inequality $\Delta \leq 0$ yields
$$\frac{(\boldsymbol{u}_k^T \boldsymbol{A}^T \boldsymbol{M} \mathcal{Z}_k)^2}{(\lambda^2+\xi_k)\xi_k} = \frac{(\boldsymbol{u}_k^T \boldsymbol{A}^T \boldsymbol{M} \mathcal{Z}_k)^2(\lambda^2+\xi_k)}{(\lambda^2+\xi_k)^2\xi_k} \leq \frac{\alpha}{\lambda^2}.$$
After identification with (65), it is equivalent to
$$\tau_0^2 \frac{(\lambda^2+\xi_k)}{\xi_k} \leq \frac{\alpha}{\lambda^2}.$$
Hence
$$|\tau_0| \leq \theta \sqrt{\frac{\xi_k}{\lambda^2+\xi_k}} = \theta\Gamma_k$$
with $\Gamma_k \in]0, 1[$, as introduced in (17).
- If $\Delta \geq 0$, then $\boldsymbol{t}_1$ is a global minimizer. From $\Delta \geq 0$
$$\frac{(\boldsymbol{u}_k^T \boldsymbol{A}^T \boldsymbol{M} \mathcal{Z}_k)^2}{(\lambda^2+\xi_k)\xi_k} = \frac{(\boldsymbol{u}_k^T \boldsymbol{A}^T \boldsymbol{M} \mathcal{Z}_k)^2 \xi_k}{(\lambda^2+\xi_k)\xi_k^2} \geq \frac{\alpha}{\lambda^2}$$
which, by identification with (66), is the same as
$$\tau_1^2 \frac{\xi_k}{\lambda^2+\xi_k} \geq \frac{\alpha}{\lambda^2}.$$
Then
$$|\tau_1| \geq \theta \sqrt{\frac{\lambda^2+\xi_k}{\xi_k}} = \theta\Gamma_k^{-1}.$$

Since $\hat{\boldsymbol{t}}$ is either $\hat{\boldsymbol{t}} = \boldsymbol{t}_0$ or $\hat{\boldsymbol{t}} = \boldsymbol{t}_1$, it follows that its kth entry $\hat{t}_k$ satisfies the following:
$$\text{either } |\hat{t}_k| \leq \theta\Gamma_k \text{ or } |\hat{t}_k| \geq \theta\Gamma_k^{-1}.$$
Notice that $\Delta = 0$ is impossible if $\mathcal{F}(., \boldsymbol{y})$ has a unique global minimizer. In the latter case, the kth entry of $\hat{\boldsymbol{t}}$ satisfies
$$\text{either } |\hat{t}_k| < \theta\Gamma_k \text{ or } |\hat{t}_k| > \theta\Gamma_k^{-1}.$$

B. Proof of Theorem 4

For c as in Remark 1, define the following vector and function:
$$\begin{aligned}\boldsymbol{v} &:= c\mathbf{1} - \boldsymbol{u}_k\\ f(\ell) &:= \mathcal{E}(\hat{\boldsymbol{x}} + \ell\boldsymbol{v}, \boldsymbol{y}) - \mathcal{E}(\hat{\boldsymbol{x}}, \boldsymbol{y})\\ &= \|\boldsymbol{A}(\hat{\boldsymbol{x}} + \ell\boldsymbol{v}) - \boldsymbol{y}\|^2 - \|\boldsymbol{A}\hat{\boldsymbol{x}} - \boldsymbol{y}\|^2\\ &\quad \cdot \sum_{i=1}^{n-1} \varphi[\hat{x}_i - \hat{x}_{i+1} + \ell(v_i - v_{i+1})]\\ &\quad - \sum_{i=1}^{n-1} \varphi(\hat{x}_i - \hat{x}_{i+1}).\end{aligned}$$
From the comments given in the beginning of Remark 1, $\boldsymbol{A}\boldsymbol{v} = \mathbf{0}$. On the other hand, by construction, $v_i = v_{i+1}$ for every $i \neq k$, whereas $v_k - v_{k+1} = 1$. These facts show that
$$f(\ell) = \varphi(\hat{x}_k - \hat{x}_{k+1} + \ell) - \varphi(\hat{x}_k - \hat{x}_{k+1}).$$
The fact that $\mathcal{E}(., \boldsymbol{y})$ has a strict global minimum at $\hat{\boldsymbol{x}}$ implies that f reaches a strict minimum at $\hat{\ell} = 0$. Two situations arise according to the magnitude of $\hat{x}_k - \hat{x}_{k+1}$.

- If $|\hat{x}_k - \hat{x}_{k+1}| > \theta$, then $f(\ell) = 0$ whenever $|\ell| < |\hat{x}_k - \hat{x}_{k+1}| - \theta$; hence, $\hat{\boldsymbol{x}}$ is not a strict minimizer. By Theorem 1, $\hat{\boldsymbol{x}}$ cannot be a global minimizer of $\mathcal{E}(., \boldsymbol{y})$.
- It remains that $|\hat{x}_k - \hat{x}_{k+1}| < \theta$. If $\hat{x}_k \neq \hat{x}_{k+1}$, then f reaches its minimum for $\tilde{\ell} = -\hat{x}_k + \hat{x}_{k+1} \neq 0$, but then, $\varphi(\hat{x}_k - \hat{x}_{k+1}) > 0$, and $f(\tilde{\ell}) = -\varphi(\hat{x}_k - \hat{x}_{k+1}) < 0$ since $\varphi(t) > 0$ strictly, whenever $t \neq 0$. The latter conclusion contradicts the fact that $\hat{\boldsymbol{x}}$ is a global minimizer. It remains that $\hat{x}_k = \hat{x}_{k+1}$.

Appendix D
Proofs Relevant to Section IV

A. Proof of Proposition 5

By (25), $\hat{\boldsymbol{x}}$ in (27) satisfies $(\boldsymbol{A}^T\boldsymbol{A} + \lambda^2 \boldsymbol{D}^T \boldsymbol{\mathcal{U}}_J \boldsymbol{D})\hat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{y}$, whereas (26) shows that $\mathcal{J}(\hat{\boldsymbol{x}}) = J$. Then, Proposition 2 allows us to make the conclusion.

B. Proof of Theorem 5

Take p and q in $\{0, \cdots, \#\mathcal{M} - 1\}$ with $p \neq q$. The result is trivial if $Y_{J_p} \cap Y_{J_q} = \emptyset$ since then, $\mathcal{N}_{J_p, J_q} = \emptyset$.

Consider the case when $Y_{J_p} \cap Y_{J_q}$ is nonempty. In the following, we consider $\mathcal{F}(., \boldsymbol{y})$, as introduced in (35). Note that (24)–(26) and (29) are equivalently reformulated as
$$\begin{aligned}\mathcal{M} &= \{J \subset \{1, \cdots, n-1\}\text{: rank}(\boldsymbol{B}^T\boldsymbol{B} + \lambda^2\tilde{\boldsymbol{\mathcal{U}}}_J) = n\}\\ \boldsymbol{T}_J &= (\boldsymbol{B}^T\boldsymbol{B} + \lambda^2\tilde{\boldsymbol{\mathcal{U}}}_J)^{-1}\boldsymbol{B}^T \text{ for } J \in \mathcal{M}\\ Y_J &= \{\boldsymbol{y} \in \mathbb{R}^m\text{: } |\boldsymbol{T}_J\boldsymbol{y}| > \theta \text{ if } k \in J\\ &\qquad \text{and } |\boldsymbol{T}_J\boldsymbol{y}| < \theta \text{ otherwise}\}\\ \boldsymbol{\mathcal{G}}_J(\boldsymbol{y}) &= \mathcal{F}(\boldsymbol{T}_J\boldsymbol{y}, \boldsymbol{y}) = \|\boldsymbol{y}\|^2 - \boldsymbol{y}^T\boldsymbol{B}\boldsymbol{T}_J\boldsymbol{y} + \alpha\#J\end{aligned}$$
where $\tilde{\boldsymbol{\mathcal{U}}}_J$ is given in (53). Similarly to (30), $\mathcal{N}_{J_p, J_q}$ reads
$$\mathcal{N}_{J_p, J_q} = \Big\{\boldsymbol{y} \in Y_{J_p} \bigcap Y_{J_q}\text{: } \boldsymbol{y}^T\boldsymbol{B}(\boldsymbol{T}_{J_q} - \boldsymbol{T}_{J_p})\boldsymbol{y} - \alpha(\#J_q - \#J_p) = 0\Big\}. \quad (68)$$
Since $J_q \neq J_p$, $\text{rank}(\boldsymbol{T}_{J_q} - \boldsymbol{T}_{J_p}) \geq 1$. If $\text{rank}\boldsymbol{B} = n$, then $\boldsymbol{B}(\boldsymbol{T}_{J_q} - \boldsymbol{T}_{J_p})$ is not identically null. Otherwise, the latter matrix is nonzero by assumption (A). In both cases, the second-order polynomial with respect to $\boldsymbol{y}$ in (68) involves nonzero coefficients; therefore, it cannot remain zero on any open subset of $\mathbb{R}^m$. Hence, we have the result.

C. Proof of Theorem 6

Since every Y_{J_q}, as introduced in (26), is open, there exists $\eta > 0$ such that
$$B(\boldsymbol{y}, \eta) \subset \bigcap_{q \in S_{\boldsymbol{y}}} Y_{J_q}$$
where $S_{\boldsymbol{y}}$ is the set given in (28). Each minimum-value function $\boldsymbol{\mathcal{G}}_{J_q}$, as given in (29), corresponding to $q \in S_{\boldsymbol{y}}$, is well defined and continuous on $B(\boldsymbol{y}, \eta)$. In the following, we compare the value of $\boldsymbol{\mathcal{G}}_{\mathcal{J}}$ with the value of every $\boldsymbol{\mathcal{G}}_{J_q}$, $q \in S_{\boldsymbol{y}}$ when $\boldsymbol{y}'$ ranges over $B(\boldsymbol{y}, \eta)$.

By the fact that $\hat{\boldsymbol{x}}$ is a strict minimizer of $\mathcal{E}(., \boldsymbol{y})$, there exists $p \in S_{\boldsymbol{y}}$ such that $\mathcal{J}(\hat{\boldsymbol{x}}) = J_p$. Put $\tilde{S}_{\boldsymbol{y}} := S_{\boldsymbol{y}} \setminus \{p\}$. The statement of this theorem being valid for almost any $\boldsymbol{y}$, we consider
$$\boldsymbol{y} \notin \tilde{\mathcal{N}}_{\mathcal{J}} \text{ where } \tilde{\mathcal{N}}_{\mathcal{J}} := \bigcup_{q \in \tilde{S}_{\boldsymbol{y}}} \mathcal{N}_{\mathcal{J}, J_q}$$
since by Theorem 5, the set $\tilde{\mathcal{N}}_{\mathcal{J}}$ is closed and negligible in $\mathbb{R}^m$.

In such a case, saying that $\hat{\boldsymbol{x}} = \boldsymbol{\mathcal{X}}_{\hat{J}}\boldsymbol{y}$ is a global minimizer is equivalent to saying that

$$\boldsymbol{\mathcal{G}}_{\hat{J}}(\boldsymbol{y}) < \boldsymbol{\mathcal{G}}_{J_q}(\boldsymbol{y}) \text{ for every } q \in \tilde{S}_{\boldsymbol{y}}.$$

Put

$$\gamma := \min\{\boldsymbol{\mathcal{G}}_{J_q}(\boldsymbol{y}) - \boldsymbol{\mathcal{G}}_{\hat{J}}(\boldsymbol{y}) \colon q \in \tilde{S}_{\boldsymbol{y}}\}.$$

Then, $\boldsymbol{y} \notin \tilde{\mathcal{N}}_{\hat{J}}$ ensures that $\gamma > 0$. Furthermore, since $\boldsymbol{y} \notin \tilde{\mathcal{N}}_{\hat{J}}$, there exists $\eta' \in]0, \eta]$ such that $B(\boldsymbol{y}; \eta') \cap \tilde{\mathcal{N}}_{\hat{J}} = \emptyset$. By the continuity of the minimum-value functions $\boldsymbol{\mathcal{G}}_{J_q}$ and $\boldsymbol{\mathcal{G}}_{\hat{J}}$, and there are radii $\xi_q \in]0, \eta']$ for $q \in \tilde{S}_{\boldsymbol{y}}$ such that

$$\boldsymbol{y}' \in B(\boldsymbol{y}; \xi_q) \text{ leads to } \boldsymbol{\mathcal{G}}_{J_q}(\boldsymbol{y}') - \boldsymbol{\mathcal{G}}_{\hat{J}}(\boldsymbol{y}') > \frac{\gamma}{2}.$$

Put $\xi := \min\{\xi_q \colon q \in \tilde{S}_{\boldsymbol{y}}\}$. Then, we see that for every $q \in \tilde{S}_{\boldsymbol{y}}$, we have

$$\boldsymbol{\mathcal{G}}_{\hat{J}}(\boldsymbol{y}') < \boldsymbol{\mathcal{G}}_{J_q}(\boldsymbol{y}') \text{ for all } \boldsymbol{y}' \in B(\boldsymbol{y}; \xi).$$

Hence, we have the result.

Acknowledgment

The author wishes to express her gratitude to E. Hansen for his criticism and suggestions about Remark 1.

References

[1] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, Nov. 1984.

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

[3] J. E. Besag, "Digital image processing: Toward Bayesian image analysis," *J. Appl. Statist.*, vol. 16, no. 3, pp. 395–407, 1989.

[4] F. Jeng and J. Woods, "Simulated annealing in compound Gaussian random fields," *IEEE Trans. Inform. Theory*, vol. 36, pp. 94–107, Jan. 1990.

[5] ——, "Compound Gauss–Markov random fields for image estimation," *IEEE Trans. Signal Processing*, vol. 39, pp. 683–697, Mar. 1991.

[6] M. Nikolova, J. Idier, and A. Mohammad-Djafari, "Inversion of ill-posed operators using a piecewise Gaussian MRF," *IEEE Trans. Image Processing*, vol. 7, pp. 571–585, Apr. 1998.

[7] C. Yang, "Efficient stochastic algorithms on locally bounded image space," *CVGIP: Graph. Models Image Process.*, vol. 55, no. 6, pp. 494–506, Nov. 1993.

[8] H. R. Künsch, "Robust priors for smoothing and image restoration," *Ann. Inst. Statist. Math.*, vol. 46, no. 1, pp. 1–19, 1994.

[9] A. Blake, "Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 2–12, Jan. 1989.

[10] H. Hiriyannaiah, G. Bilbro, M. Snyder, and R. Mann, "Restoration of piecewise constant images by mean field annealing," *J. Opt. Soc. Amer.*, vol. 6, no. 12, pp. 1901–1989, 1989.

[11] J. Zerubia and R. Chellappa, "Mean field approximation using compound Gauss Markov random field for edge detection and image restoration," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1990, pp. 2193–2196.

[12] D. Geiger and F. Girosi, "Parallel and deterministic algorithms from MRF's: Surface reconstruction," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 401–412, May 1991.

[13] G. Bilbro, W. Snyder, S. Garnier, and J. Gault, "Mean-field annealing: A formalism for constructing gnc-like algorithms," *IEEE Trans. Neural Networks*, vol. 3, pp. 131–138, Jan. 1992.

[14] D. Geiger and A. Yuille, "A common framework for image segmentation," *Int. J. Comput. Vis.*, vol. 6, no. 3, pp. 227–243, 1991.

[15] J. Marroquin, "Deterministic interactive particle models for image processing and computer graphics," *Comput. Vis. Graph. Image Process.*, vol. 55, no. 5, pp. 408–417, 1993.

[16] M. Figueiredo and J. Leitao, "Simulated tearing: An algorithm for discontinuity-preserving visual surface recconstruction," in *Proc. IEEE CVPR*, 1993, pp. 28–33.

[17] W. Snyder, Y.-S. Han, G. Bilbro, R. Whitaker, and S. Pizer, "Image relaxation: restoration and feature extraction," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, pp. 620–624, June 1995.

[18] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Processing*, vol. 5, pp. 1–44, Dec. 1996.

[19] D. Terzopoulos, "Regularization of inverse visual problems involving discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 413–424, July 1986.

[20] D. Mumford and J. Shah, "Boundary detection by minimizing functionals," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, pp. 22–26, 1985.

[21] A. Chambolle, "Image segmentation by variational methods: Mumford and Shah functional and the discrete approximation," *SIAM J. Appl. Math.*, vol. 55, no. 3, pp. 827–863, 1997.

[22] A. Chambolle, "Finite-differences discretizations of the Mumford–Shah functional," *RAIRO-Model. Math. Anal. Numer.*, 1997.

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

[24] L. Alvarez, P. L. Lions, and J. M. Morel, "Image selective smoothing and edge detection by nonlinear diffusion (II)," *SIAM J. Numer. Anal.*, vol. 29, no. 2, pp. 845–866, 1992.

[25] F. Catte, T. Coll, P. L. Lions, and J. M. Morel, "Image selective smoothing and edge detection by nonlinear diffusion (I)," *SIAM J. Numer. Anal.*, vol. 29, no. 1, pp. 182–193, 1992.

[26] M. Black, G. Sapiro, D. Marimont, and D. Heeger, "Robust anisotropic diffusion," *IEEE Trans. Image Processing*, vol. 7, pp. 421–432, Mar. 1998.

[27] S. Teboul, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Variational approach for edge-preserving regularization using coupled pde's," *IEEE Trans. Image Processing*, vol. 7, pp. 387–397, Mar. 1998.

[28] M. Nikolova, "Local strong homogeneity of a regularized estimator," SIAM J. Appl. Math., 1999, to be published.

[29] F. Clarke, *Optimization and Nonsmooth Analysis*. Philadelphia, PA: SIAM, 1990.

Mila Nikolova received the Ph.D. degree in signal processing from the Université Paris-Sud, Paris, France, in 1995.

Currently, she is a Researcher with the National Center for Scientific Research (CNRS), Paris. Her research interests are in inverse problems and reconstruction of images and signals.