SIAM J. APPL. MATH.
Vol. 61, No. 2, pp. 633–658

LOCAL STRONG HOMOGENEITY OF A REGULARIZED ESTIMATOR *

MILA NIKOLOVA†

Abstract. This paper deals with regularized pointwise estimation of discrete signals which contain large *strongly homogeneous zones*, where typically they are constant, or linear, or more generally satisfy a linear equation. The estimate is defined as the minimizer of an objective function combining a quadratic data-fidelity term and a regularization prior term. The latter term is the sum of the values obtained by applying a potential function (PF) to each component, called a *difference*, of a linear transform of the signal. Minimizers of functions of this form arise in various settings in statistics and optimization.

The features exhibited by such an estimate are closely related to the shape of the PF. Our goal is to determine estimators providing solutions which involve large *strongly homogeneous zones*—where more precisely the differences are null—in spite of the noise corrupting the data. To this end, we require that the strongly homogeneous zones, recovered by the estimator, be insensitive to any variation of the data inside a small open ball. More generally, this requirement is addressed to any *local* or *global* minimizer of the objective function whose local behavior with respect to the data gives rise to a locally continuous minimizer function. On the one hand, we show that if the PF is *smooth at zero*, then all the data, yielding minimizers with large, strongly homogeneous zones, are contained in a closed, negligible set. The chance that noisy data generate such minimizers is null. In contrast, if the PF is *nonsmooth at zero*, then for almost all data, the strongly homogeneous zones recovered by a minimizer function are preserved constant under any small perturbation of the data. The data domain is thus organized into volumes whose elements yield minimizers which share the same strongly homogeneous zones. This explains why the solutions, obtained using nonsmooth-at-zero PFs, exhibit strongly homogeneous zones.

These theoretical results are illustrated using a numerical example. Our analysis can be extended to general functions combining smooth and nonsmooth terms.

Key words. inverse problems, MAP estimation, nonsmooth analysis, perturbation analysis, proximal point, reconstruction, regularization, soft thresholding, total variation, variational methods

AMS subject classifications. 49J52, 49Q20, 26B05, 58F10, 62H12, 94A12

PII. S0036139997327794

1. Introduction.

Let an unknown signal[1] $\boldsymbol{x} \in \mathbb{R}^M$ be observed through a system $\boldsymbol{y} = \tilde{A}\boldsymbol{x} + \boldsymbol{n} \in \mathbb{R}^N$, where $\tilde{A} \in \mathbb{R}^{N\times M}$ is a linear operator and $\boldsymbol{n}$ represents the observation noise. Bayesian MAP estimation defines the inverse solution $\hat{\boldsymbol{x}}$ as the minimizer of the posterior energy $\mathcal{E}_{\boldsymbol{y}}(\boldsymbol{x}) \propto -\ln P(\boldsymbol{x}|\boldsymbol{y})$, which combines log-likelihood $-\ln P(\boldsymbol{y}|\boldsymbol{x})$ and prior energy $\tilde{\Phi}(\boldsymbol{x}) \propto -\ln P(\boldsymbol{x})$. In a regularization framework, $\mathcal{E}_{\boldsymbol{y}}$ is an *objective function* and its global minimizer $\hat{\boldsymbol{x}}$ is a *regularized estimate*:

$$\hat{\boldsymbol{x}} := \arg\min_{\boldsymbol{x}} \mathcal{E}_{\boldsymbol{y}}(\boldsymbol{x}). \tag{1.1}$$

*Received by the editors September 25, 1997; accepted for publication (in revised form) February 10, 1999; published electronically August 9, 2000. This paper contains the demonstration and development of some ideas that were summarized in a short note, published in *C.R. Acad. Sci.,* 6 (1997).

http://www.siam.org/journals/siap/61-2/32779.html

†UFR Mathématiques et Informatique, Université René Descartes, 45, rue des Saints-Pères, 75270 Paris Cédex 06 (nikolova@math-info.univ-paris5.fr).

[1] Vectors are denoted with bold lowercase letters: let $\boldsymbol{v}$ be a vector; its components are denoted by v_k, or equivalently by $\boldsymbol{v}[k]$. Functions are denoted with calligraphic letters; these are in bold for vector-valued functions.

We confine our attention to objective functions of the form[2]

$$\mathcal{E}_{\boldsymbol{y}}(\boldsymbol{x}) := \|\tilde{A}\boldsymbol{x} - \boldsymbol{y}\|^2 + \beta\tilde{\Phi}(\boldsymbol{x}), \tag{1.2}$$

$$\tilde{\Phi}(\boldsymbol{x}) := \sum_{k\in\mathcal{S}^\circ} \varphi(\boldsymbol{g}_k^T\boldsymbol{x}), \tag{1.3}$$

where $\beta > 0$ is a parameter and $\{\boldsymbol{g}_k,\ k \in \mathcal{S}^\circ\}$ *is any collection of linearly independent operators*, $\boldsymbol{g}_k : \mathbb{R}^M \mapsto \mathbb{R}$, and ^{T} denotes transpose. Let $\mathcal{S}$ be the ordered set of the sites of $\boldsymbol{x}$; then $\mathcal{S}^\circ \subseteq \mathcal{S}$. Typically, $\mathcal{S} = \{1, \ldots, M\}$ and $\boldsymbol{g}_k^T\boldsymbol{x}$ are finite differences; then $\mathcal{S}^\circ$ is composed of those k for which $\boldsymbol{g}_k^T\boldsymbol{x}$ involves only elements x_i with $i \in \mathcal{S}$ (for instance, $\boldsymbol{g}_k^T\boldsymbol{x} = x_k - x_{k+1}$ and $\mathcal{S}^\circ = \{1, \ldots, M-1\}$). By a slight abuse of language, hereafter $\boldsymbol{g}_k^T\boldsymbol{x}$ are called *differences*. The regularizer $\tilde{\Phi}$ results from applying a *potential function* (PF) φ to each difference $\boldsymbol{g}_k^T\boldsymbol{x}$ for $k \in \mathcal{S}^\circ$. The PF φ in (1.3) is symmetric, increasing on $[\,0, \infty\,[$, twice differentiable except at several points where it can be nonsmooth and even discontinuous; it can be nonconvex. Minimizers of objective functions of the form (1.2)–(1.3) arise in different settings: Bayesian estimation [5, 6], regularization [29, 14], variational methods [27, 3, 8, 9], proximal point optimization [25, 20, 12].

This work addresses any strict *local* or *global* minimizer $\hat{\boldsymbol{x}}$ whose local behavior with respect to $\boldsymbol{y}$ gives rise to a locally continuous *minimizer function* $\boldsymbol{\mathcal{X}}$, yielding in particular $\hat{\boldsymbol{x}} = \boldsymbol{\mathcal{X}}(\boldsymbol{y})$. Although $\boldsymbol{\mathcal{X}}$ is a random function with respect to the original unknown signal, the features of any particular solution $\hat{\boldsymbol{x}} = \boldsymbol{\mathcal{X}}(\boldsymbol{y})$ are inherently related to the shape of $\tilde{\Phi}$. Our approach is to consider the behavior of $\boldsymbol{\mathcal{X}}$ in connection with the shape of $\tilde{\Phi}$. This work is focused on the possibility of obtaining estimates $\hat{\boldsymbol{x}} = \boldsymbol{\mathcal{X}}(\boldsymbol{y})$ which exhibit large *strongly homogeneous zones*—namely, zones where the differences are null—in spite of the noise corrupting $\boldsymbol{y}$. According to the form of $\boldsymbol{g}_k$, a strongly homogeneous zone may be constant, linear, quadratic, etc. Recovering strongly homogeneous zones "in spite of the noise" means that these zones are invariably recovered from data corrupted by arbitrary noise samples of weak amplitude. *We establish that strongly homogeneous zones in $\hat{\boldsymbol{x}}$ are both recovered from noisy data and preserved intact from small variations of the data, if and only if the PF φ in* (1.3) *is nonsmooth at zero.* Such a behavior is *local* in two different senses: it is independent of the shape of φ beyond 0 and it is exhibited by almost any strict local minimizer of $\mathcal{E}_{\boldsymbol{y}}$. Among the most popular nonsmooth-at-zero PFs, we cite the following [23, 17, 19, 27, 2]:

$$\text{modulus:}\quad \varphi(t) = |t|, \tag{1.4}$$

$$\text{concave:}\quad \varphi(t) = \alpha|t|/(1+\alpha|t|), \tag{1.5}$$

$$\text{“0-1”:}\quad \varphi(t) = 1 \ \text{ if } \ t \neq 0, \quad \varphi(0) = 0. \tag{1.6}$$

In order to simplify the presentation, we suppose φ is twice differentiable everywhere except at 0. However, our results can be extended to more general objective functions combining smooth and nonsmooth terms.

To our knowledge, the *generic* problem of obtaining solutions involving large strongly homogeneous zones using regularization has never been formalized previously. Nonetheless, the ability of the modulus PF (1.4) to recover noncorrelated ($\boldsymbol{g}_k^T\boldsymbol{x} = x_k$) "nearly black images" is interpreted in [17] using minimax decision theory. In total variation methods, pioneered in [27], the regularization is an ℓ_1-norm of the derivatives

[2] Recall that $-\ln P(\boldsymbol{y}|\boldsymbol{x}) \propto \|\tilde{A}\boldsymbol{x} - \boldsymbol{y}\|^2$ when $\boldsymbol{n}$ is white Gaussian noise.

of the unknown signal. Such regularizations have been observed to produce "blocky estimates" [15, 11]. The concave PF (1.5) is shown in [19] to give rise to a step-shaped estimate from ramp-shaped data, and this PF is called "strictly noninterpolating."

Our study can be seen as an attempt to understand what regularization using nonsmooth-at-zero PFs accomplishes on the estimate, in comparison with smooth-at-zero PFs. It provides some new results in the analysis of nonsmooth functions. Nonsmooth analysis has been developed widely for the purpose of optimization. Different properties of the minimum of an objective function, $\boldsymbol{y} \mapsto \mathcal{E}_{\boldsymbol{y}}(\boldsymbol{\mathcal{X}}(\boldsymbol{y}))$, in the case when φ is convex and nonsmooth, are examined in the framework of Moreau–Yosida regularization [20, 24]. Many studies are concerned with the optimization of nonsmooth objective functions [28, 20, 12].

Organization of the paper. The notion of strong homogeneity, in connection with the minimizers of $\mathcal{E}_{\boldsymbol{y}}$, is introduced in section 2. The inability of a minimizer function $\boldsymbol{\mathcal{X}}$, corresponding to a *smooth-at-zero* PF, to retrieve strongly homogeneous zones from noisy data, is shown in section 3. Henceforth, nonsmooth-at-zero PFs are considered in (1.2)–(1.3). The objective function $\mathcal{E}_{\boldsymbol{y}}$ is then nonsmooth, and necessary conditions for minimum are derived in section 4. Sufficient conditions for a strict local minimum are given in section 5. In section 6, we show how the variations of the data inside an open ball yield a family of minimizers that are strongly homogeneous over the same zones. This behavior induces a specific organization of the data space $\mathbb{R}^N$ which is discussed in section 7. Numerical illustrations are given in section 8, with concluding remarks in section 9.

2. Strong homogeneity of local minimizers.

2.1. Notion of strong homogeneity.

Different kinds of homogeneity, i.e., of smoothness, are encountered in real-world signals. These are frequently modeled using linear operators $\boldsymbol{g}_k : \mathbb{R}^M \mapsto \mathbb{R}$, $k \in \mathcal{S}^\circ$, in such a way that the homogeneous zones in $\boldsymbol{x}$ are the locations k of weak differences, $\boldsymbol{g}_k^T \boldsymbol{x} \approx 0$. Usually $\{\boldsymbol{g}_k,\ k \in \mathcal{S}^\circ\}$ provide a discrete approximation of differential operators. Such a notion of homogeneity is loose and it concerns a wide range of features. In this work, we introduce the precise notion of *strong homogeneity* which can be applied to *any* collection $\{\boldsymbol{g}_k,\ k \in \mathcal{S}^\circ\}$.

DEFINITION 2.1. *The set of strong homogeneity $\mathcal{J}(\boldsymbol{x})$ of a signal $\boldsymbol{x}$, with respect to a family of difference operators $\{\boldsymbol{g}_k : k \in \mathcal{S}^\circ\}$, is composed of the indices k of all zero-valued differences:*

$$\mathcal{J}(\boldsymbol{x}) := \left\{ k \in \mathcal{S}^\circ : \boldsymbol{g}_k^T \boldsymbol{x} = 0 \right\}.$$

Thus $\mathcal{J} : \mathbb{R}^M \mapsto \mathcal{P}(\mathcal{S}^\circ)$, where $\mathcal{P}(\mathcal{S}^\circ)$ is the set of all possible subsets of $\mathcal{S}^\circ$.

The values of $\boldsymbol{x}$, corresponding to a strongly homogeneous zone, satisfy a homogeneous system of linear equations. For example, if $\boldsymbol{g}_k^T \boldsymbol{x} := x_k - x_{k+1}$ for $k \in \{1, \ldots, M-1\}$, then $\mathcal{J}$ addresses the constant zones in $\boldsymbol{x}$; if $\boldsymbol{g}_k^T \boldsymbol{x} := x_{k-1} - 2x_k + x_{k+1}$ for $k \in \{2, \ldots, M-1\}$, then $\mathcal{J}$ addresses the linear zones in $\boldsymbol{x}$, third-order differences correspond to parabolic zones, and so on.

2.2. Minimizers of an objective function.

In order to narrow the context of this study, we recall several facts about the minimizers of an objective function $\mathcal{E}_{\boldsymbol{y}}$ of the form (1.2)–(1.3). If $\mathcal{E}_{\boldsymbol{y}}$ is strictly convex, it has a unique minimizer $\hat{\boldsymbol{x}}$. A nonstrictly convex function has a unique minimum that is either strict (i.e., it is reached at a unique minimizer point) or nonstrict (when it is reached at an infinite set of points). If φ is nonconvex, $\mathcal{E}_{\boldsymbol{y}}$ usually exhibits numerous local minima, some of which can be nonstrict. Our analysis concerns all strict *local* and *global* minimizers of $\mathcal{E}_{\boldsymbol{y}}$, so *we suppose that* $\mathcal{E}_{\boldsymbol{y}}$ *exhibits strict minimizers.* Then we focus on the behavior of any such local minimizer of $\mathcal{E}_{\boldsymbol{y}}$ entailed by small, arbitrary variations of $\boldsymbol{y}$.

DEFINITION 2.2. *Suppose* $\mathcal{E}_{\boldsymbol{y}}$ *admits a* strict local minimum *for any* $\boldsymbol{y} \in U$, *where* $U \subseteq \mathbb{R}^N$ *is a domain. An application* $\boldsymbol{\mathcal{X}} : U \mapsto \mathbb{R}^M$ *is said to be a* (local) minimizer function *relevant to* $\mathcal{E}_{\boldsymbol{y}}$ *if for any* $\boldsymbol{y} \in U$, $\mathcal{E}_{\boldsymbol{y}}$ *reaches a* strict local minimum *at* $\boldsymbol{\mathcal{X}}(\boldsymbol{y})$.

Minimizer functions are generally *implicit*. If $\mathcal{E}_{\boldsymbol{y}}$ is strictly convex, it admits a unique minimizer function $\boldsymbol{\mathcal{X}}$ and then $U = \mathbb{R}^N$. Otherwise, $\mathcal{E}_{\boldsymbol{y}}$ may give rise to numerous local and global minimizer functions. Notice that $\boldsymbol{\mathcal{X}}$ in Definition 2.2 is *any local or global* minimizer function. The properties established in this paper address minimizer functions that are continuous at some points or on some domains.

Global or only local continuity of estimators of the form (1.1) is a theme which has no direct impact on this work; we assume the weaker among these requirements, so our statements hold in both situations. It is worth recalling that global continuity of the minimizer of a strictly convex objective function is a classical result [22]. Global continuity of the unique minimizer of a strictly unimodal objective function is considered in [7]. Stability of total variation methods is analyzed in [1, 10]. Local continuity for some special unbounded nonsmooth regularizers is considered in [30]. Bounded PFs give rise to locally smooth estimates involving sharp edges [19]: although the relevant estimators are not globally continuous [7], it can be observed that in general $\mathcal{E}_{\boldsymbol{y}}$ has strict minimizers which are locally continuous with respect to $\boldsymbol{y}$.

2.3. Strong homogeneity of minimizer functions. The set of strong homogeneity of a minimizer $\hat{\boldsymbol{x}}$ is denoted by $\hat{J} := \mathcal{J}(\hat{\boldsymbol{x}})$. Given a minimizer function $\boldsymbol{\mathcal{X}}$, we will focus on the reciprocal function $\mathcal{J} \circ \boldsymbol{\mathcal{X}}$ with argument $\boldsymbol{y}$ which yields the set of strong homogeneity of $\boldsymbol{\mathcal{X}}(\boldsymbol{y})$:

$$\left(\mathcal{J} \circ \boldsymbol{\mathcal{X}}\right)(\boldsymbol{y}) := \mathcal{J}(\boldsymbol{\mathcal{X}}(\boldsymbol{y}));$$

hence $(\mathcal{J} \circ \boldsymbol{\mathcal{X}}) : U \mapsto \mathcal{P}(\mathcal{S}^\circ)$.

Since the data are noisy, estimating a set of strong homogeneity $\hat{J}$ is not meaningful unless this set can be recovered from slightly perturbed data $\boldsymbol{y} + \boldsymbol{n}$, where $\boldsymbol{n}$ has a "small" amplitude. That is, the set of strong homogeneity, yielded by a minimizer function, should be insensitive to *any* small variation of the data.

DEFINITION 2.3. *Let* $\boldsymbol{\mathcal{X}}$ *be a minimizer function relevant to* $\mathcal{E}_{\boldsymbol{y}}$. *Suppose* $\hat{\boldsymbol{x}} = \boldsymbol{\mathcal{X}}(\boldsymbol{y})$ *involves a nonempty set of strong homogeneity* $\hat{J} = \mathcal{J}(\hat{\boldsymbol{x}})$. *The minimizer function* $\boldsymbol{\mathcal{X}}$ *is said to be* locally strongly homogeneous *if there exists* $\xi > 0$ *such that*

$$\left(\mathcal{J} \circ \boldsymbol{\mathcal{X}}\right)(\boldsymbol{y}') = \hat{J} \quad \textit{for all} \quad \boldsymbol{y}' \in B(\boldsymbol{y}; \xi),$$

i.e., such that $\hat{\boldsymbol{x}}$ *and* $\hat{\boldsymbol{x}}' := \boldsymbol{\mathcal{X}}(\boldsymbol{y}')$ *share the same set of strong homogeneity* $\hat{J}$ *whenever* $\boldsymbol{y}'$ *ranges over a neighborhood of* $\boldsymbol{y}$.

Ball $B(\boldsymbol{y}; \xi) := \{\, \boldsymbol{y}' : \|\boldsymbol{y}' - \boldsymbol{y}\| < \xi \,\}$ is defined with respect to the ℓ_2-norm $\|\boldsymbol{v}\| := (\sum_k v_k^2)^{\frac{1}{2}}$. The relevant closed ball is denoted by $\overline{B(\boldsymbol{y}; \xi)}$. The centered, unit sphere in $\mathbb{R}^N$ is $\mathbb{I}^N := \{\boldsymbol{v} \in \mathbb{R}^N : \|\boldsymbol{v}\| = 1\}$. The cardinality of a set $\hat{J}$ is denoted by $\#\{\hat{J}\}$.

2.4. Equivalent formulation of $\mathcal{E}_{\boldsymbol{y}}$. It is convenient to transform $\mathcal{E}_{\boldsymbol{y}}$ into a function of the differences $\boldsymbol{g}_k^T \boldsymbol{x}$. Since $\{\boldsymbol{g}_k, k \in \mathcal{S}^\circ\}$ are linearly independent, we can find a family of operators $\tilde{\boldsymbol{g}}_k : \mathbb{R}^M \mapsto \mathbb{R}$ for $k \in \mathcal{S} \setminus \mathcal{S}^\circ$, so that the matrix

$G := [\boldsymbol{g}_1, \ldots, \boldsymbol{g}_{\#\{\mathcal{S}^\circ\}}, \tilde{\boldsymbol{g}}_1, \ldots, \tilde{\boldsymbol{g}}_{\#\{\mathcal{S}\setminus\mathcal{S}^\circ\}}]$ is invertible. Define

$$\begin{cases} t_k := \boldsymbol{g}_k^T \boldsymbol{x} & \text{for} \quad k \in \mathcal{S}^\circ, \\ t_k := \tilde{\boldsymbol{g}}_k^T \boldsymbol{x} & \text{for} \quad k \in \mathcal{S} \setminus \mathcal{S}^\circ, \end{cases} \quad \text{i.e.,} \quad \boldsymbol{t} = G^T \boldsymbol{x}, \quad \text{then} \quad \boldsymbol{x} = \left(G^T\right)^{-1} \boldsymbol{t}.$$

Put $A := \tilde{A}\left(G^T\right)^{-1}$. We get an equivalent objective function $\mathcal{F}_{\boldsymbol{y}}(\boldsymbol{t}) := \mathcal{E}_{\boldsymbol{y}}[\left(G^T\right)^{-1}\boldsymbol{t}]$:

$$\mathcal{F}_{\boldsymbol{y}}(\boldsymbol{t}) = \|A\boldsymbol{t} - \boldsymbol{y}\|^2 + \beta\Phi(\boldsymbol{t}), \tag{2.1}$$

$$\text{where} \qquad \Phi(\boldsymbol{t}) = \sum_{k \in \mathcal{S}^\circ} \varphi(t_k). \tag{2.2}$$

In this formulation, $\boldsymbol{\mathcal{T}} := G^T \boldsymbol{\mathcal{X}}$ is a minimizer function relevant to $\mathcal{F}_{\boldsymbol{y}}$ which meets Definition 2.2 and is defined on the same domain U. Let $\boldsymbol{y} \in U$; then $\hat{\boldsymbol{t}} := \boldsymbol{\mathcal{T}}(\boldsymbol{y})$ is a strict minimizer of $\mathcal{F}_{\boldsymbol{y}}$ and its set of strong homogeneity reads

$$\hat{J} = \mathcal{J}(\hat{\boldsymbol{t}}) = (\mathcal{J} \circ \boldsymbol{\mathcal{T}})(\boldsymbol{y}) = \left\{k \in \mathcal{S}^\circ : \hat{t}_k = \mathcal{T}_k(\boldsymbol{y}) = 0\right\}. \tag{2.3}$$

In the last expression, $\mathcal{T}_k : U \mapsto \mathbb{R}$ is the kth entry of the vector-valued function $\boldsymbol{\mathcal{T}}$.

3. Estimation using a smooth-at-zero PF.

3.1. Minimizers with no strong homogeneity. We wish to know whether a minimizer $\hat{\boldsymbol{t}}$, recovered from noisy data $\boldsymbol{y}$, can involve a large set of strong homogeneity $\hat{J} = \mathcal{J}(\hat{\boldsymbol{t}})$ when φ is smooth at zero. To simplify the presentation, the next theorem is stated for PFs which are twice differentiable everywhere.

THEOREM 3.1. *Suppose* φ *in* (2.2) *is a twice differentiable function on* $\mathbb{R}$. *Let* $\mathcal{F}_{\boldsymbol{y}}$, *as given in* (2.1), *reach a strict local minimum at* $\hat{\boldsymbol{t}}$. *Assume the following:*

(a) $\hat{\boldsymbol{t}}$ *involves a large set of strong homogeneity* $\hat{J} = \mathcal{J}(\hat{\boldsymbol{t}})$ *with* $\#\{\hat{J}\} > M - N$;

(b) $\mathcal{F}_{\boldsymbol{y}}$ *admits a (local) minimizer function* $\boldsymbol{\mathcal{T}}$, *such that* $\hat{\boldsymbol{t}} = \boldsymbol{\mathcal{T}}(\boldsymbol{y})$, *and which is defined and differentiable on a neighborhood of* $\boldsymbol{y}$;

(c) $A \in \mathbb{R}^{N \times M}$ *in* (2.1) *satisfies* Rank$(A) = N \leq M$.

Then there exists $\eta > 0$ *such that the set*

$$\mathcal{N}_{\hat{J}} = \left\{\boldsymbol{y}' \in \overline{B(\boldsymbol{y}; \eta)} : (\mathcal{J} \circ \boldsymbol{\mathcal{T}})(\boldsymbol{y}') = \hat{J}\right\} \tag{3.1}$$

is closed and negligible with respect to the Lebesgue measure on $\mathbb{R}^N$. *In other words, any* $\boldsymbol{y}' \in B(\boldsymbol{y}; \eta)$, *yielding a minimizer* $\hat{\boldsymbol{t}}' = \boldsymbol{\mathcal{T}}(\boldsymbol{y}')$ *which has the same set of strong homogeneity as* $\hat{\boldsymbol{t}}$, *i.e.,* $(\mathcal{J} \circ \boldsymbol{\mathcal{T}})(\boldsymbol{y}') = \hat{J}$, *belongs to a closed set of measure zero in* $\mathbb{R}^N$.

If an estimate involves strongly homogeneous zones, these are usually quite large, $\#\{\hat{J}\} \gg M - N$ (see the experiments in section 8). So the assumption (a) is not restrictive, while it reduces to $\#\{\hat{J}\} > 0$ when A is invertible. By the implicit functions theorem we see that (b) is satisfied when $\mathcal{F}_{\boldsymbol{y}}$ is $\mathcal{C}^2$-continuous on a neighborhood of $\hat{\boldsymbol{t}}$ and its Hessian at $\hat{\boldsymbol{t}}$ is positive definite. As to (c), note that the remaining cases, namely Rank$(A) < N \leq M$ and Rank$(A) \leq M \leq N$, can be reduced to (c).

Proof. Let $\varepsilon > 0$ be such that (b) is valid on $B(\boldsymbol{y}; \varepsilon)$. Let D refer to differential operator. Being a minimizer of $\mathcal{F}_{\boldsymbol{y}'}$, any $\hat{\boldsymbol{t}}' = \boldsymbol{\mathcal{T}}(\boldsymbol{y}')$, corresponding to $\boldsymbol{y}' \in B(\boldsymbol{y}; \varepsilon)$, satisfies the equation

$$\begin{aligned} &\mathrm{D}\mathcal{F}_{\boldsymbol{y}'}(\hat{\boldsymbol{t}}') = \mathbf{0}^T, \\ \text{where} \quad &\mathrm{D}\mathcal{F}_{\boldsymbol{y}'}(\boldsymbol{t}') = 2(A\boldsymbol{t}' - \boldsymbol{y}')^T A + \beta \mathrm{D}\Phi(\boldsymbol{t}'). \end{aligned} \tag{3.2}$$

From now on, $\mathbf{0}$ and $\mathbf{1}$ will denote vectors or matrices with zero-valued and one-valued entries, respectively, of whatever size appropriate to the context. Differentiating both sides of (3.2) with respect to $\boldsymbol{y}'$ yields

$$\mathrm{D}^2\mathcal{F}_{\boldsymbol{y}'}(\boldsymbol{\mathcal{T}}(\boldsymbol{y}'))\,\mathrm{D}\boldsymbol{\mathcal{T}}(\boldsymbol{y}') = 2A^T \quad \text{for any} \quad \boldsymbol{y}' \in B(\boldsymbol{y};\varepsilon),$$

where we recall that $\mathrm{D}^2F\boldsymbol{y}'(\boldsymbol{t}) = 2A^TA + \beta\mathrm{D}^2\Phi(\boldsymbol{t})$ for any $\boldsymbol{y}' \in \mathbb{R}^N$. We determine now the rank of $\mathrm{D}\boldsymbol{\mathcal{T}}(\boldsymbol{y}')$. On the one hand,

$$\min\left\{\text{Rank}\ \left[\mathrm{D}^2\mathcal{F}_{\boldsymbol{y}'}(\boldsymbol{\mathcal{T}}(\boldsymbol{y}'))\right],\ \text{Rank}\ [\mathrm{D}\boldsymbol{\mathcal{T}}(\boldsymbol{y}')]\right\} \geq \text{Rank}\ \left(2A^T\right) = N,$$

then $\text{Rank}\,[\mathrm{D}\boldsymbol{\mathcal{T}}(\boldsymbol{y}')] \geq N$ on $B(\boldsymbol{y};\varepsilon)$. On the other hand, $\boldsymbol{\mathcal{T}} : \mathbb{R}^N \mapsto \mathbb{R}^M$, then $\text{Rank}\,[\mathrm{D}\boldsymbol{\mathcal{T}}(\boldsymbol{y}')] \leq \min\{M,N\} = N$. Hence

$$\text{Rank}\,[\mathrm{D}\boldsymbol{\mathcal{T}}(\boldsymbol{y}')] = N \quad \text{whenever} \quad \boldsymbol{y}' \in B(\boldsymbol{y};\varepsilon). \tag{3.3}$$

More generally, the equality above is true at any $\boldsymbol{y}'$ where $\mathrm{D}\boldsymbol{\mathcal{T}}$ is defined.

Let us suggest now a value for the radius η in (3.1). For $\mathcal{S}^\circ \setminus \hat{J}$ nonempty, put

$$\mu := \min_{k\in\mathcal{S}^\circ\setminus\hat{J}} |\hat{t}_k|, \quad \text{then } \mu > 0. \tag{3.4}$$

By (b), there exists $\eta \in]\,0,\varepsilon\,[$ such that

$$\boldsymbol{y}' \in \overline{B(\boldsymbol{y};\eta)} \ \text{ leads to } \ \left|\mathcal{T}_k(\boldsymbol{y}') - \hat{t}_k\right| \leq \frac{\mu}{2} \ \text{ for all } \ k \in \mathcal{S}^\circ \setminus \hat{J}. \tag{3.5}$$

Then any $\boldsymbol{\mathcal{T}}(\boldsymbol{y}')$ yielded by $\boldsymbol{y}' \in \overline{B(\boldsymbol{y};\eta)}$ satisfies

$$|\mathcal{T}_k(\boldsymbol{y}')| \geq \left|\hat{t}_k\right| - \frac{\mu}{2} \geq \frac{\mu}{2} > 0 \quad \text{for any} \quad k \in \mathcal{S}^\circ \setminus \hat{J}.$$

The latter inequality means that

$$\mathcal{J}(\boldsymbol{\mathcal{T}}(\boldsymbol{y}')) \subseteq \hat{J} \ \text{ for any } \ \boldsymbol{y}' \in \overline{B(\boldsymbol{y};\eta)}. \tag{3.6}$$

If $\mathcal{S}^\circ = \hat{J}$, (3.6) is trivially satisfied for $\eta = \varepsilon/2$.

Let $\hat{J}$ read $\hat{J} = \{j_1,\dots,j_{\#\{\hat{J}\}}\}$; for definiteness, assume that $j_1 < j_2 < \cdots < j_{\#\{\hat{J}\}}$. Define $C_{\hat{J}}$ to be the following $\#\{\hat{J}\} \times M$ matrix:

$$\begin{cases} C_{\hat{J}}\,[i,j_i] = 1 & \text{for} \ \ i = 1,\dots,\#\{\hat{J}\}, \\ C_{\hat{J}}\,[i,j] = 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

Clearly, $C_{\hat{J}}\boldsymbol{t} = \mathbf{0}$ is a compact way to say that $t_k = 0$ for any $k \in \hat{J}$, i.e., that $\mathcal{J}(\boldsymbol{t}) \supseteq \hat{J}$. This remark, combined with (3.6), suggests expressing the set $\mathcal{N}_{\hat{J}}$ in (3.1) as

$$\mathcal{N}_{\hat{J}} = \left\{\boldsymbol{y}' \in \overline{B(\boldsymbol{y};\eta)} : C_{\hat{J}}\boldsymbol{\mathcal{T}}(\boldsymbol{y}') = \mathbf{0}\right\}. \tag{3.8}$$

Next, we characterize the rank of $\mathrm{D}\left(C_{\hat{J}}\boldsymbol{\mathcal{T}}(\boldsymbol{y}')\right) = C_{\hat{J}}\mathrm{D}\boldsymbol{\mathcal{T}}(\boldsymbol{y}')$. Using both Sylvester's theorem [21] and (3.3), and by noticing that $\text{Rank}\,(C_{\hat{J}}) = \#\{\hat{J}\}$, we get

$$\text{Rank}\ \left[\mathrm{D}\left(C_{\hat{J}}\boldsymbol{\mathcal{T}}(\boldsymbol{y}')\right)\right] \geq \#\{\hat{J}\} + N - M \geq 1 \ \text{ for all } \ \boldsymbol{y}' \in B(\boldsymbol{y};\varepsilon). \tag{3.9}$$

The set $\mathcal{N}_{\hat{J}}$ in (3.8) is clearly closed in $\mathbb{R}^N$, since $C_{\hat{J}}\boldsymbol{\mathcal{T}}$ is continuous on $B(\boldsymbol{y},\varepsilon)$. Let m_N denote the Lebesgue measure on $\mathbb{R}^N$. Next we check whether $m_N(\mathcal{N}_{\hat{J}}) = 0$ or not. For suppose $m_N(\mathcal{N}_{\hat{J}}) > 0$. The latter, combined with the closeness of $\mathcal{N}_{\hat{J}}$, indicates that $\mathcal{N}_{\hat{J}}$ contains a nonempty open N-cell, say, $\widetilde{\mathcal{N}_{\hat{J}}} \subseteq \mathcal{N}_{\hat{J}}$. By (3.8) we have $C_{\hat{J}}\boldsymbol{\mathcal{T}}(\boldsymbol{y}') = \mathbf{0}$ for any $\boldsymbol{y}' \in \widetilde{\mathcal{N}_{\hat{J}}}$. Differentiating both sides of the latter identity yields

$$C_{\hat{J}}\mathrm{D}\boldsymbol{\mathcal{T}}(\boldsymbol{y}') = \mathbf{0} \quad \text{for all} \quad \boldsymbol{y}' \in \widetilde{\mathcal{N}_{\hat{J}}}.$$

Hence Rank $\left[C_{\hat{J}}\mathrm{D}\boldsymbol{\mathcal{T}}(\boldsymbol{y}')\right] = 0$ for all $\boldsymbol{y}' \in \widetilde{\mathcal{N}_{\hat{J}}}$; but this contradicts (3.9) since we have $\widetilde{\mathcal{N}_{\hat{J}}} \subset \overline{B(\boldsymbol{y};\eta)} \subset B(\boldsymbol{y};\varepsilon)$. It follows that $m_N(\mathcal{N}_{\hat{J}}) = 0$. This completes the proof. ◻

The set $\mathcal{N}_{\hat{J}}$ in (3.8) can equivalently be expressed as

$$\mathcal{N}_{\hat{J}} = \overline{B(\boldsymbol{y};\eta)} \bigcap \left(C_{\hat{J}}\boldsymbol{\mathcal{T}}\right)^{-1}(\mathbf{0}). \tag{3.10}$$

Let us focus on the particular case when Rank $(A) = M = N$. Now, (3.9) yields

$$\text{Rank } \left[\mathrm{D}\left(C_{\hat{J}}\boldsymbol{\mathcal{T}}(\boldsymbol{y}')\right)\right] = \#\{\hat{J}\} \quad \text{for all} \quad \boldsymbol{y}' \in B(\boldsymbol{y};\varepsilon).$$

The rank of $\mathrm{D}C_{\hat{J}}\boldsymbol{\mathcal{T}}$ being constant on $\overline{B(\boldsymbol{y};\eta)}$, then $\mathcal{N}_{\hat{J}}$ in (3.10) determines a *continuous manifold* [4] of dimension

$$\dim(\mathcal{N}_{\hat{J}}) = N - \text{Rank } \left[\mathrm{D}\left(C_{\hat{J}}\boldsymbol{\mathcal{T}}(\boldsymbol{y}')\right)\right] = N - \#\{\hat{J}\}.$$

But the assumption (c) of the same Theorem 3.1 is more general and does not guarantee that Rank $[\mathrm{D}\left(C_{\hat{J}}\boldsymbol{\mathcal{T}}(\boldsymbol{y}')\right)]$ in (3.9) remains constant near to $\boldsymbol{y}$.

The fact that $\mathcal{N}_{\hat{J}}$ is a closed, negligible subset of $\mathbb{R}^N$ is of crucial importance. The possible configurations of $\hat{J}$ being of finite number, *any* data $\boldsymbol{y}'$ in the vicinity of $\boldsymbol{y}$, leading to minimizers which are strongly homogeneous over large zones, belong to a finite union of closed negligible sets. Such a union is a set of measure zero in $\mathbb{R}^N$. So, even if we trap a minimizer that is strongly homogeneous over large zones, the latter's zones are destroyed under almost any perturbation of the data, due to the noise. In a global plan, the chance that noisy data belong to the special set of the data in $\mathbb{R}^N$ which lead to minimizers with large strongly homogeneous zones is *almost null.* In other words, *if φ is smooth at zero, we should not expect to find minimizers that involve large strongly homogeneous zones.* For illustration, see Figure 8.2 in section 8.

3.2. Example. Let $\varphi(t) = t^2$ and Rank $(A) = N = M$. Then $\boldsymbol{\mathcal{T}}$ is explicit:

$$\hat{\boldsymbol{t}} = \boldsymbol{\mathcal{T}}(\boldsymbol{y}) = \left(A^T A + \beta I^\circ\right)^{-1} A^T \boldsymbol{y}, \quad \text{where} \quad I^\circ_{k,l} = \begin{cases} 1 & \text{if } \ k = l \in \mathcal{S}^\circ, \\ 0 & \text{otherwise.} \end{cases}$$

Reciprocally, $\boldsymbol{y}$ can be expressed as a function of $\hat{\boldsymbol{t}}$:

$$\boldsymbol{y} = H\hat{\boldsymbol{t}} \quad \text{with} \quad H = A + \beta\left(A^T\right)^{-1} I^\circ. \tag{3.11}$$

Let $\hat{t}_k = 0$ for $k \in \hat{J}$ with $\hat{J} \neq \emptyset$. Any $\hat{\boldsymbol{t}}'$, satisfying $\mathcal{J}(\hat{\boldsymbol{t}}') = \hat{J}$, has the form

$$\hat{\boldsymbol{t}}' = \hat{\boldsymbol{t}} + \boldsymbol{w}, \quad \text{where} \quad w_k = 0 \quad \text{if} \quad k \in \hat{J} \quad \text{and} \quad w_k \neq 0 \quad \text{otherwise.}$$

Hence, $\hat{\boldsymbol{t}}'$ belongs to an affine subspace of dimension $M - \#\{\hat{J}\}$. Since H is invertible, any $\boldsymbol{y}'$, yielding $\hat{\boldsymbol{t}}' = \boldsymbol{\mathcal{T}}(\boldsymbol{y}')$ with $(\mathcal{J} \circ \boldsymbol{\mathcal{T}})(\boldsymbol{y}') = \hat{J}$, necessarily belongs to the affine subspace spanned by those columns $\boldsymbol{h}_k$ of H which are indexed by $\mathcal{S} \setminus \hat{J}$:

$$\boldsymbol{y}' = H(\hat{\boldsymbol{t}} + \boldsymbol{w}) = \boldsymbol{y} + \sum_{k \in \mathcal{S} \setminus \hat{J}} \boldsymbol{h}_k w_k.$$

This affine subspace is of dimension $M - \#\{\hat{J}\} \leq M - 1$, so it is a closed, negligible subset of $\mathbb{R}^N$.

Take now a nonempty $\hat{J}$. We will determine the set of all $\boldsymbol{y}$ for which $(\mathcal{J} \circ \boldsymbol{\mathcal{T}})(\boldsymbol{y}) = \hat{J}$. By (3.11), any such $\boldsymbol{y}$ must have the form

$$\boldsymbol{y} = \sum_{k \in \mathcal{S} \setminus \hat{J}} \boldsymbol{h}_k v_k \quad \text{with} \quad v_k \in \mathbb{R} \setminus \{0\};$$

i.e., it belongs to the subspace spanned by $\{\boldsymbol{h}_k,\ k \in \mathcal{S} \setminus \hat{J}\}$. Being of dimension $N - \#\{\hat{J}\}$, this subspace is a closed, negligible subset of $\mathbb{R}^N$. The same holds for *any* nonempty $\hat{J} \in \mathcal{P}(\mathcal{S}^\circ)$. Any $\boldsymbol{y}$, leading to $(\mathcal{J} \circ \boldsymbol{\mathcal{T}})(\boldsymbol{y}) \neq \emptyset$, therefore belongs to the set

$$\bigcup_{\hat{J} \in \mathcal{P}(\mathcal{S}^\circ) \setminus \{\emptyset\}} \operatorname{Span}\left\{\boldsymbol{h}_k,\ k \in \mathcal{S} \setminus \hat{J}\right\}.$$

The above union forms a closed set of measure zero in $\mathbb{R}^N$. The presence of noise in the data makes it extremely unlikely to come across data $\boldsymbol{y}$ placed in this set.

4. Regularization using a PF that is nonsmooth at zero.

4.1. General relations. *Henceforth, φ is nonsmooth at zero.* Suppose also that

(H1) *φ is twice differentiable on $\mathbb{R} \setminus \{0\}$; φ is symmetric and $\varphi(0) = 0$.*

The first assumption can be relaxed, as mentioned in section 1. The second is a technical assumption which simplifies the presentation.

Furthermore, we shall focus on two types of PFs:

(H2) *φ is continuous on $\mathbb{R}$ and admits a positive right*[3] *derivative at zero, say, $\varphi'_+(0) = \lim_{h \downarrow 0} \varphi(h)/h > 0$, which can be finite or infinite;*

(H3) *φ is discontinuous at zero and there are $\gamma > 0$ and $\eta > 0$ such that $\varphi(t) \geq \gamma$ whenever $|t| \in]\,0, \eta\,[$.*

Both (H2) and (H3) originate in the practical requirement that φ be increasing on $]\,0, +\infty\,[$. Note that (H2) is satisfied by the PFs given in (1.4) and (1.5), whereas (H3) is true for the "0-1" PF given in (1.6) for any $\gamma \in]\,0, 1\,[$ and $\eta = +\infty$.

Now $\mathcal{F}_{\boldsymbol{y}}$ fails to be smooth on the union of hyperplanes $\bigcup_{k \in \mathcal{S}^\circ} [\boldsymbol{t} : t_k = 0]$.

DEFINITION 4.1. *Let $\mathcal{F}_{\boldsymbol{y}} : \mathbb{R}^M \mapsto \mathbb{R}$ be a function, $\boldsymbol{t} \in \mathbb{R}^M$ a point, and $\boldsymbol{v} \in \mathbb{I}^M$ a unit direction. We say that $\mathcal{F}_{\boldsymbol{y}}$ admits a left and a right derivative at $\boldsymbol{t}$ in the direction of $\boldsymbol{v}$, denoted by $\partial^-_{\boldsymbol{v}} \mathcal{F}_{\boldsymbol{y}}$ and $\partial^+_{\boldsymbol{v}} \mathcal{F}_{\boldsymbol{y}}$, respectively, if the limits*

$$\partial^+_{\boldsymbol{v}} \mathcal{F}_{\boldsymbol{y}}(\boldsymbol{t}) = \lim_{h \downarrow 0} \frac{\mathcal{F}_{\boldsymbol{y}}(\boldsymbol{t} + h\boldsymbol{v}) - \mathcal{F}_{\boldsymbol{y}}(\boldsymbol{t})}{h}, \qquad \partial^-_{\boldsymbol{v}} \mathcal{F}_{\boldsymbol{y}}(\boldsymbol{t}) = \lim_{h \downarrow 0} \frac{\mathcal{F}_{\boldsymbol{y}}(\boldsymbol{t} - h\boldsymbol{v}) - \mathcal{F}_{\boldsymbol{y}}(\boldsymbol{t})}{-h} \tag{4.1}$$

exist and belong to $\{-\infty\} \cup \mathbb{R} \cup \{+\infty\}$.

The down arrow in $h \downarrow 0$ means that h converges to zero by positive values. If $\mathcal{F}_{\boldsymbol{y}}$ is smooth at $\boldsymbol{t}$ in the direction $\boldsymbol{v}$, then $\partial^+_{\boldsymbol{v}} \mathcal{F}_{\boldsymbol{y}}(\boldsymbol{t}) = \partial^-_{\boldsymbol{v}} \mathcal{F}_{\boldsymbol{y}}(\boldsymbol{t})$ is its usual directional derivative; otherwise $\partial^+_{\boldsymbol{v}} \mathcal{F}_{\boldsymbol{y}}(\boldsymbol{t}) \neq \partial^-_{\boldsymbol{v}} \mathcal{F}_{\boldsymbol{y}}(\boldsymbol{t})$ and these may be infinite.

[3] By the symmetry of φ, the left derivative of φ at zero is $\varphi'_-(0) = -\varphi'_+(0) < 0$.

Conditions for minima of nonsmooth functions can be found in [26, 20, 13]. Below we give a formulation which is appropriate to our analysis.

THEOREM 4.2. *Let $\mathcal{F}_{\boldsymbol{y}} : \mathbb{R}^M \mapsto \mathbb{R}$ reach a strict minimum at $\hat{\boldsymbol{t}}$—local or global. Suppose $\mathcal{F}_{\boldsymbol{y}}$ admits side derivatives at $\hat{\boldsymbol{t}}$ in the direction of any $\boldsymbol{v} \in \mathbb{I}^M$. Then*

$$\partial^-_{\boldsymbol{v}} \mathcal{F}_{\boldsymbol{y}}(\hat{\boldsymbol{t}}) \le 0 \le \partial^+_{\boldsymbol{v}} \mathcal{F}_{\boldsymbol{y}}(\hat{\boldsymbol{t}}) \quad \textit{for all} \ \ \boldsymbol{v} \in \mathbb{I}^M. \tag{4.2}$$

The proof of this necessary condition is outlined in the appendix. Observe that if $\mathcal{F}_{\boldsymbol{y}}$ is nonsmooth at $\hat{\boldsymbol{t}}$ along $\boldsymbol{v}$, we have either $\partial^-_{\boldsymbol{v}} \mathcal{F}_{\boldsymbol{y}}(\hat{\boldsymbol{t}}) < 0 \le \partial^+_{\boldsymbol{v}} \mathcal{F}_{\boldsymbol{y}}(\hat{\boldsymbol{t}})$ or $\partial^-_{\boldsymbol{v}} \mathcal{F}_{\boldsymbol{y}}(\hat{\boldsymbol{t}}) \le 0 < \partial^+_{\boldsymbol{v}} \mathcal{F}_{\boldsymbol{y}}(\hat{\boldsymbol{t}})$, which inequalities are strict in most of the cases.

4.2. Necessary condition for a minimum of $\mathcal{F}_{\boldsymbol{y}}$. Now we specialize the necessary condition (4.2) to objective functions of the form (2.1)–(2.2). This result generalizes the conditions concerning the modulus PF provided in [2].

THEOREM 4.3. *Consider $\mathcal{F}_{\boldsymbol{y}}$ in* (2.1)–(2.2) *where φ meets* (H1) *and* (H2). *Suppose $\mathcal{F}_{\boldsymbol{y}}$ reaches a (local) minimum at $\hat{\boldsymbol{t}}$. Then $\hat{\boldsymbol{t}}$ satisfies the following system:*

$$2\left|\boldsymbol{a}_k^T(\boldsymbol{y} - A\hat{\boldsymbol{t}})\right| \le \beta\varphi'_+(0) \qquad \textit{if} \ \ k \in \hat{J}, \tag{4.3}$$

$$2\,\boldsymbol{a}_k^T(\boldsymbol{y} - A\hat{\boldsymbol{t}}) - \beta\varphi'(\hat{t}_k) = 0 \qquad \textit{if} \ \ k \in \mathcal{S}^\circ \setminus \hat{J}, \tag{4.4}$$

$$2\,\boldsymbol{a}_k^T(\boldsymbol{y} - A\hat{\boldsymbol{t}}) = 0 \qquad \textit{if} \ \ k \in \mathcal{S} \setminus \mathcal{S}^\circ, \tag{4.5}$$

where $\boldsymbol{a}_k$ are the columns of A and $\hat{J} = \mathcal{J}(\hat{\boldsymbol{t}})$; see (2.3).

The minimum reached by $\mathcal{F}_{\boldsymbol{y}}$ at $\hat{\boldsymbol{t}}$ can be local or global, strict or nonstrict.

If $\varphi'_+(0) = \infty$, then (4.3) is superfluous since it becomes $\left|\boldsymbol{a}_k^T(\boldsymbol{y} - A\hat{\boldsymbol{t}})\right| < \infty$ for all $k \in \hat{J}$. If $\hat{t}_k \ne 0$ for all $k \in \mathcal{S}^\circ$, then $\hat{J} = \emptyset$ and (4.3) is absent, while $\mathcal{F}_{\boldsymbol{y}}$ is smooth in the vicinity of $\hat{\boldsymbol{t}}$ and (4.4)–(4.5) means that the gradient of $\mathcal{F}_{\boldsymbol{y}}$ at $\hat{\boldsymbol{t}}$ is null. The only case where (4.4)–(4.5) are both absent is $\hat{J} = \mathcal{S}^\circ = \mathcal{S}$, in which case $\hat{\boldsymbol{t}} = \boldsymbol{0}$.

Proof. Let $\boldsymbol{v}$ be an arbitrary direction in $\mathbb{I}^M$. Using (4.1),

$$\partial^-_{\boldsymbol{v}} \mathcal{F}_{\boldsymbol{y}}(\boldsymbol{t}) = 2\,(A\boldsymbol{t} - \boldsymbol{y})^T A\boldsymbol{v} + \beta\partial^-_{\boldsymbol{v}}\Phi(\boldsymbol{t}), \quad \partial^+_{\boldsymbol{v}} \mathcal{F}_{\boldsymbol{y}}(\boldsymbol{t}) = 2\,(A\boldsymbol{t} - \boldsymbol{y})^T A\boldsymbol{v} + \beta\partial^+_{\boldsymbol{v}}\Phi(\boldsymbol{t}),$$

$$\partial^-_{\boldsymbol{v}}\Phi(\boldsymbol{t}) = \sum_{k\in\mathcal{S}^\circ} \partial^-_{\boldsymbol{v}}\varphi(t_k), \quad \partial^+_{\boldsymbol{v}}\Phi(\boldsymbol{t}) = \sum_{k\in\mathcal{S}^\circ} \partial^+_{\boldsymbol{v}}\varphi(t_k),$$

$$\partial^-_{\boldsymbol{v}}\varphi(t_k) = \lim_{h\downarrow 0} \frac{\varphi(t_k - hv_k) - \varphi(t_k)}{-h}, \quad \partial^+_{\boldsymbol{v}}\varphi(t_k) = \lim_{h\downarrow 0} \frac{\varphi(t_k + hv_k) - \varphi(t_k)}{h}. \tag{4.6}$$

Since $\hat{\boldsymbol{t}}$ is a minimizer of $\mathcal{F}_{\boldsymbol{y}}$, Theorem 4.2 shows that

$$\beta\partial^-_{\boldsymbol{v}}\Phi(\hat{\boldsymbol{t}}) \le 2(\boldsymbol{y} - A\hat{\boldsymbol{t}})^T A\boldsymbol{v} \le \beta\partial^+_{\boldsymbol{v}}\Phi(\hat{\boldsymbol{t}}) \quad \text{for all} \quad \boldsymbol{v} \in \mathbb{I}^M,$$

or equivalently, using the fact that $A\boldsymbol{v} = \sum_{k\in\mathcal{S}} \boldsymbol{a}_k v_k$,

$$\beta \sum_{k\in\mathcal{S}^\circ} \partial^-_{\boldsymbol{v}}\varphi(\hat{t}_k) \le 2\sum_{k\in\mathcal{S}} \boldsymbol{a}_k^T(\boldsymbol{y} - A\hat{\boldsymbol{t}})\, v_k \le \beta \sum_{k\in\mathcal{S}^\circ} \partial^+_{\boldsymbol{v}}\varphi(\hat{t}_k). \tag{4.7}$$

The sums in (4.7) are now split according to the following partition:

$$\mathcal{S} = \hat{J} \cup \left(\mathcal{S}^\circ \setminus \hat{J}\right) \cup \left(\mathcal{S} \setminus \mathcal{S}^\circ\right).$$

Below, if a term is indexed by a set that is empty, it is supposed to be zero. Since $\partial^-_{\boldsymbol{v}}\varphi(t_k) = \partial^+_{\boldsymbol{v}}\varphi(t_k) = v_k\varphi'(t_k)$ when φ is smooth at t_k, (4.7) yields

$$\beta \sum_{k\in\hat{J}} \partial^-_{\boldsymbol{v}}\varphi\left(\hat{t}_k\right) \le 2\sum_{k\in\hat{J}} \boldsymbol{a}_k^T\left(\boldsymbol{y} - A\hat{\boldsymbol{t}}\right) v_k + 2\sum_{k\in\mathcal{S}\setminus\mathcal{S}^\circ} \boldsymbol{a}_k^T\left(\boldsymbol{y} - A\hat{\boldsymbol{t}}\right) v_k$$
$$+ \sum_{k\in\mathcal{S}^\circ\setminus\hat{J}} \left[2\boldsymbol{a}_k^T(\boldsymbol{y} - A\hat{\boldsymbol{t}}) - \beta\varphi'(\hat{t}_k)\right] v_k \le \beta\sum_{k\in\hat{J}} \partial^+_{\boldsymbol{v}}\varphi\left(\hat{t}_k\right). \tag{4.8}$$

Next we compute (4.8) for each direction $\boldsymbol{e}_n$ of the canonical basis of $\mathbb{R}^M$,

$$\boldsymbol{e}_n[k] = \begin{cases} 0 & \text{if} \quad k \ne n, \\ 1 & \text{if} \quad k = n. \end{cases}$$

Recall that by (4.6) we have $\partial^\pm_{\boldsymbol{e}_n}\varphi(\hat{t}_k) = 0$ whenever $n \ne k$.

- For any $\boldsymbol{v} = \boldsymbol{e}_n$ such that $n \in \hat{J}$, (4.8) leads to

$$\beta\left[\partial^-_{\boldsymbol{e}_n}\varphi(\hat{t}_n)\right]_{\hat{t}_n=0} \le 2\boldsymbol{a}_n^T(\boldsymbol{y} - A\hat{\boldsymbol{t}}) \le \beta\left[\partial^+_{\boldsymbol{e}_n}\varphi(\hat{t}_n)\right]_{\hat{t}_n=0}.$$

  Since $\left[\partial^-_{\boldsymbol{e}_n}\varphi(t_n)\right]_{t_n=0} = -\varphi'_+(0) = -\left[\partial^+_{\boldsymbol{e}_n}\varphi(t_n)\right]_{t_n=0}$, we obtain (4.3).
  If $\hat{J} = \emptyset$, (4.8) does not involve terms of this form and (4.3) is absent.
- For any $n \in \mathcal{S}^\circ \setminus \hat{J}$, introducing $\boldsymbol{v} = \boldsymbol{e}_n$ into (4.8) yields

$$\sum_{k\in\mathcal{S}^\circ\setminus\hat{J}} \left[2\boldsymbol{a}_k^T(\boldsymbol{y} - A\hat{\boldsymbol{t}}) - \beta\varphi'(\hat{t}_k)\right] \boldsymbol{e}_n[k] = 0.$$

  Its left side becomes $2\boldsymbol{a}_n^T(\boldsymbol{y} - A\hat{\boldsymbol{t}}) - \beta\varphi'(\hat{t}_n)$, so we get (4.4).
- For any $\boldsymbol{v} = \boldsymbol{e}_n$ with $n \in \mathcal{S}\setminus\mathcal{S}^\circ$, (4.8) becomes $2\sum_{k\in\mathcal{S}\setminus\mathcal{S}^\circ} \boldsymbol{a}_k^T(\boldsymbol{y} - A\hat{\boldsymbol{t}})\,\boldsymbol{e}_n[k] = 0$, which yields $2\boldsymbol{a}_n^T(\boldsymbol{y} - A\hat{\boldsymbol{t}}) = 0$, hence (4.5).

The proof is completed. □

4.3. Smooth restriction of the objective function. Let $\hat{J} \in \mathcal{P}(\mathcal{S}^\circ)$ be nonempty and $\hat{J} \ne \mathcal{S}$. We associate to each $i \in \{1, \ldots, \#\{\mathcal{S} \setminus \hat{J}\}\}$ the integer κ_i, which is the ith element of $\mathcal{S} \setminus \hat{J}$. Thus

$$\mathcal{S} \setminus \hat{J} = \left\{\kappa_1, \kappa_2, \ldots, \kappa_{\#\{\mathcal{S}\setminus\hat{J}\}}\right\}. \tag{4.9}$$

For definiteness, assume that $\kappa_1 < \kappa_2 < \cdots < \kappa_{\#\{\mathcal{S}\setminus\hat{J}\}}$. Given $\boldsymbol{t} \in \mathbb{R}^M$, let the subscripts $\hat{J}$ and 0 mean that $\boldsymbol{t}_{\hat{J}}$ and $\boldsymbol{t}_0$ are composed of those entries of $\boldsymbol{t}$ which are indexed by $\mathcal{S} \setminus \hat{J}$ and $\hat{J}$, respectively:

$$\boldsymbol{t}_{\hat{J}}[i] := \boldsymbol{t}[\kappa_i] \quad \text{for} \quad i = 1, \ldots, \#\{\mathcal{S} \setminus \hat{J}\} \quad \text{and} \quad \boldsymbol{t}_0 = C_{\hat{J}}\boldsymbol{t}, \tag{4.10}$$

where $C_{\hat{J}}$ is the matrix defined in (3.7). Then $\boldsymbol{t}_{\hat{J}} \in \mathbb{R}^{\#\{\mathcal{S}\setminus\hat{J}\}}$ and $\boldsymbol{t}_0 \in \mathbb{R}^{\#\{\hat{J}\}}$. Similarly, let $A_{\hat{J}}$ and A_0 be composed of those columns of A that are indexed by $\mathcal{S} \setminus \hat{J}$ and by $\hat{J}$, respectively. That is,

$$A_{\hat{J}} := \left[\boldsymbol{a}_{\kappa_1}, \ldots, \boldsymbol{a}_{\kappa_{\#\{\mathcal{S}\setminus\hat{J}\}}}\right] \quad \text{and} \quad A_0 := AC_{\hat{J}}^T. \tag{4.11}$$

Thus $A\boldsymbol{t} = A_{\hat{J}}\boldsymbol{t}_{\hat{J}} + A_0\boldsymbol{t}_0$ for any $\boldsymbol{t}$.

Now focus on $\hat{\boldsymbol{t}}$, a minimizer of $\mathcal{F}_{\boldsymbol{y}}$ involving $\hat{J} = \mathcal{J}(\hat{\boldsymbol{t}})$ nonempty, $\hat{J} \neq \mathcal{S}$. Then $\hat{\boldsymbol{t}}_{\hat{\jmath}}$ is composed of both nonzero entries relevant to $\mathcal{S}^\circ \setminus \hat{J}$ and entries relevant to $\mathcal{S} \setminus \mathcal{S}^\circ$ (which are not involved in Φ and may be null). By a slight abuse of language, $\hat{\boldsymbol{t}}_{\hat{\jmath}}$ is called the *nonhomogeneous part* of $\hat{\boldsymbol{t}}$. The remaining elements of $\hat{\boldsymbol{t}}$ correspond to $\hat{J}$ and are null; that is, $\hat{\boldsymbol{t}}_0 = \mathbf{0}$. So we have

$$A\hat{\boldsymbol{t}} = \sum_{k \in \mathcal{S}\setminus \hat{J}} \boldsymbol{a}_k \hat{t}_k + \sum_{k \in \hat{J}} \boldsymbol{a}_k 0 = A_{\hat{\jmath}} \hat{\boldsymbol{t}}_{\hat{\jmath}}. \tag{4.12}$$

Take μ as defined in (3.4). The following *restricted objective function*,

$$\mathcal{F}_{\boldsymbol{y}}^{\hat{J}} : \mathbb{R}^{\#\{\mathcal{S}\setminus\hat{J}\}} \mapsto \mathbb{R},$$

$$\boldsymbol{t}_{\hat{\jmath}} \mapsto \mathcal{F}_{\boldsymbol{y}}^{\hat{J}}(\boldsymbol{t}_{\hat{\jmath}}) = \left\| A_{\hat{\jmath}} \boldsymbol{t}_{\hat{\jmath}} - \boldsymbol{y} \right\|^2 + \beta \Phi^{\hat{J}}(\boldsymbol{t}_{\hat{\jmath}}), \tag{4.13}$$

$$\text{where} \qquad \Phi^{\hat{J}}(\boldsymbol{t}_{\hat{\jmath}}) := \sum_{k \in \mathcal{S}^\circ \setminus \hat{J}} \varphi(t_k), \tag{4.14}$$

is twice differentiable on $B(\hat{\boldsymbol{t}}_{\hat{\jmath}}; \mu)$ and reaches a minimum at $\hat{\boldsymbol{t}}_{\hat{\jmath}}$; the latter fact is easily verified by using (4.12). Moreover, if $\hat{\boldsymbol{t}}$ is a strict minimizer of $\mathcal{F}_{\boldsymbol{y}}$, then $\hat{\boldsymbol{t}}_{\hat{\jmath}}$ is a strict minimizer of $\mathcal{F}_{\boldsymbol{y}}^{\hat{J}}$. If $\hat{J} = \emptyset$, then $\mathcal{F}_{\boldsymbol{y}}^{\emptyset} = \mathcal{F}_{\boldsymbol{y}}$, whereas $\Phi^{\mathcal{S}^\circ} = 0$. We will denote by $\boldsymbol{\mathcal{T}}^{\hat{J}} : U \mapsto \mathbb{R}^{\#\{\mathcal{S}\setminus\hat{J}\}}$ any minimizer function relevant to $\mathcal{F}_{\boldsymbol{y}}^{\hat{J}}$.

With these notations, (4.4)–(4.5) mean that the gradient of $\mathcal{F}_{\boldsymbol{y}}^{\hat{J}}$ at $\hat{\boldsymbol{t}}_{\hat{\jmath}}$ is null. Put

$$\theta_k := 2\, \boldsymbol{a}_k^T (\boldsymbol{y} - A_{\hat{\jmath}} \hat{\boldsymbol{t}}_{\hat{\jmath}}) \quad \text{for} \quad k \in \hat{J}. \tag{4.15}$$

Then (4.3) can be expressed as $|\theta_k| \leq \beta\varphi'_+(0)$ for all $k \in \hat{J}$. Define

$$\theta_{\max} := \max_{k \in \hat{J}} |\, \theta_k \,|; \quad \text{then} \quad \theta_{\max} \leq \beta\varphi'_+(0). \tag{4.16}$$

5. Specific properties in nonsmooth regularization.

5.1. The key feature of $\boldsymbol{\varphi}$. The following proposition states the key property of a PF which allows a minimizer of an objective function of the form (1.2)–(1.3), or equivalently (2.1)–(2.2), to involve strongly homogeneous zones.

PROPOSITION 5.1. *Let φ satisfy* (H1) *and* (H2).

(a) *If $\varphi'_+(0) > 0$ is finite, then for any $\gamma \in]\,0, 1\,[$ there exists $\eta_\gamma > 0$ such that $\varphi(t) \geq \gamma\varphi'_+(0)|t|$ whenever $|t| < \eta_\gamma$.*

(b) *If $\varphi'_+(0) = \infty$, then for any $\gamma > 0$ finite there exists $\eta_\gamma > 0$ such that $\varphi(t) \geq \gamma|t|$ whenever $|t| < \eta_\gamma$.*

Proof. Being twice differentiable on $]\,0, \infty\,[$ and null at zero, φ satisfies

$$\varphi(t) = \lim_{u \downarrow 0} \int_u^t \varphi'(s) ds \quad \text{for any} \quad t > 0.$$

Moreover, $\lim_{s\downarrow 0} \varphi'(s) = \varphi'_+(0) > 0$ shows that φ' is positive for $s > 0$ close to zero.

Case (a). Since $\varphi'_+(0)$ is finite and φ' is continuous on $]\,0, +\infty\,[$, for every $\gamma \in \;]\,0, 1\,[$ there is $\eta_\gamma > 0$ such that $\varphi'(s) \geq \gamma\varphi'_+(0)$ for all $s \in \;]\,0, \eta_\gamma\,[$. Thus $\varphi(t) \geq \int_0^t \gamma\varphi'_+(0) ds = \gamma\varphi'_+(0)\, t$ for $0 < t < \eta_\gamma$. By the symmetry of φ, we deduce (a).

Case (b). Now $\lim_{s\downarrow 0} \varphi'(s) = +\infty$. Then for any $\gamma > 0$ finite, it is possible to find $\eta_\gamma > 0$ such that $\varphi'(s) \geq \gamma$ for $0 < s < \eta_\gamma$. Hence (b). ∎

5.2. Sufficient condition for a strict minimum. We determine conditions which ensure that an arbitrary local minimizer, where $\mathcal{F}_{\boldsymbol{y}}$ is nonsmooth, is strict.

THEOREM 5.2. *Suppose $\mathcal{F}_{\boldsymbol{y}}$ is of the form* (2.1)–(2.2)*, where φ satisfies* (H1) *and* (H2) *with $\varphi'_+(0) > 0$ finite. Let $\hat{\boldsymbol{t}} \in \mathbb{R}^M$ be such that*

(a) *the inequality system* (4.3) *is strict (i.e., each one of its inequalities is strict);*

(b) *its nonhomogeneous part $\hat{\boldsymbol{t}}_{\hat{J}}$, defined as in* (4.10)*, is a strict (local) minimizer of the restricted objective function $\mathcal{F}_{\boldsymbol{y}}^{\hat{J}}$ given in* (4.13)–(4.14)*, whenever $\hat{J} \neq \mathcal{S}$.*

Then $\mathcal{F}_{\boldsymbol{y}}$ reaches a strict (local) minimum at $\hat{\boldsymbol{t}}$.

Proof. Put $\hat{J} = \mathcal{J}(\hat{\boldsymbol{t}})$. The result is trivial when $\hat{J} = \emptyset$, since $\mathcal{F}_{\boldsymbol{y}}^{\emptyset} = \mathcal{F}_{\boldsymbol{y}}$.

Suppose $\hat{J}$ is nonempty. Let $\Delta_{\boldsymbol{v}} :]\,0, +\infty\,[\ \mapsto \ \mathbb{R}$ yield the altitude increment of $\mathcal{F}_{\boldsymbol{y}}$ at $\hat{\boldsymbol{t}}$ in the direction of $\boldsymbol{v} \in \mathbb{I}^M$:

$$\text{(5.1)} \quad \begin{aligned} \Delta_{\boldsymbol{v}}(h) &:= \mathcal{F}_{\boldsymbol{y}}(\hat{\boldsymbol{t}} + h\boldsymbol{v}) - \mathcal{F}_{\boldsymbol{y}}(\hat{\boldsymbol{t}}) \\ &= h^2\|A\boldsymbol{v}\|^2 + 2h(A\hat{\boldsymbol{t}} - \boldsymbol{y})^T A\boldsymbol{v} + \beta \sum_{k\in\mathcal{S}^\circ} \varphi(\hat{t}_k + hv_k) - \beta \sum_{k\in\mathcal{S}^\circ} \varphi(\hat{t}_k). \end{aligned}$$

This proof consists in finding a radius $\rho > 0$ which ensures that $\Delta_{\boldsymbol{v}}(h) > 0$ for any $h \in]\,0, \rho\,[$, along all $\boldsymbol{v} \in \mathbb{I}^M$. For $\hat{J} \neq \mathcal{S}$, let $\boldsymbol{v}_0$ and $\boldsymbol{v}_{\hat{J}}$ be obtained from $\boldsymbol{v}$ according to (4.10). We now split $\Delta_{\boldsymbol{v}}$ into two terms:[4]

$$\text{(5.2)} \quad \begin{aligned} \Delta_{\boldsymbol{v}}(h) &= \Delta^1_{\boldsymbol{v}}(h) + \Delta^0_{\boldsymbol{v}}(h), \\ \Delta^1_{\boldsymbol{v}}(h) &:= h^2\|A_{\hat{J}}\boldsymbol{v}_{\hat{J}}\|^2 + 2h\left(A_{\hat{J}}\hat{\boldsymbol{t}}_{\hat{J}} - \boldsymbol{y}\right)^T A_{\hat{J}}\boldsymbol{v}_{\hat{J}} + \beta\Phi^{\hat{J}}(\hat{\boldsymbol{t}}_{\hat{J}} + h\boldsymbol{v}_{\hat{J}}) - \beta\Phi^{\hat{J}}(\hat{\boldsymbol{t}}_{\hat{J}}), \end{aligned}$$

$$\text{(5.3)} \quad \begin{aligned} \Delta^0_{\boldsymbol{v}}(h) &:= h^2\ \left[\ \|A_0\boldsymbol{v}_0\|^2 + 2(A_0\boldsymbol{v}_0)^T A_{\hat{J}}\boldsymbol{v}_{\hat{J}}\ \right] + 2h\left(A_{\hat{J}}\hat{\boldsymbol{t}}_{\hat{J}} - \boldsymbol{y}\right)^T A_0\boldsymbol{v}_0 \\ &\quad + \beta\sum_{k\in\hat{J}} \varphi(hv_k), \end{aligned}$$

where $\Phi^{\hat{J}}$ is defined as in (4.14), and $A_{\hat{J}}$ and A_0 are as in (4.11). Next, the terms $\Delta^1_{\boldsymbol{v}}$ and $\Delta^0_{\boldsymbol{v}}$ are analyzed separately.

- Term $\Delta^1_{\boldsymbol{v}}$. If $\|\boldsymbol{v}_{\hat{J}}\| = 0$, then $\|\boldsymbol{v}_0\| = 1$, in which case $\Delta_{\boldsymbol{v}} = \Delta^0_{\boldsymbol{v}}$. So suppose $0 < \|\boldsymbol{v}_{\hat{J}}\| \le 1$. Then $\Delta^1_{\boldsymbol{v}}(h)$ gives the altitude increment of $\mathcal{F}_{\boldsymbol{y}}^{\hat{J}}$ in the vicinity of $\hat{\boldsymbol{t}}_{\hat{J}}$. By (b), there exists $\rho_1 > 0$ such that $\mathcal{F}_{\boldsymbol{y}}^{\hat{J}}\left(\hat{\boldsymbol{t}}_{\hat{J}} + h\boldsymbol{v}_{\hat{J}}/\|\boldsymbol{v}_{\hat{J}}\|\right) - \mathcal{F}_{\boldsymbol{y}}^{\hat{J}}\left(\hat{\boldsymbol{t}}_{\hat{J}}\right) > 0$ for any $h \in\ ]\,0, \rho_1[$ and for any $\boldsymbol{v}_{\hat{J}}$. It follows that

$$\Delta^1_{\boldsymbol{v}}(h) = \mathcal{F}_{\boldsymbol{y}}^{\hat{J}}\left(\hat{\boldsymbol{t}}_{\hat{J}} + h\boldsymbol{v}_{\hat{J}}\right) - \mathcal{F}_{\boldsymbol{y}}^{\hat{J}}\left(\hat{\boldsymbol{t}}_{\hat{J}}\right) > 0 \ \text{ whenever } h \in\ ]\,0, \rho_1\,[\ \text{ and }\ 0 < \|\boldsymbol{v}_{\hat{J}}\| \le 1.$$

- Term $\Delta^0_{\boldsymbol{v}}$. If $\|\boldsymbol{v}_0\| = 0$, then $\|\boldsymbol{v}_{\hat{J}}\| = 1$ and $\Delta_{\boldsymbol{v}} = \Delta^1_{\boldsymbol{v}}$.

Now suppose $0 < \|\boldsymbol{v}_0\| \le 1$, and hence $0 \le \|\boldsymbol{v}_{\hat{J}}\| < 1$. Let $\lambda > 0$ be the largest eigenvalue of A^TA; then $\|A^TA\boldsymbol{v}\| \le \lambda\|\boldsymbol{v}\|$ for any $\boldsymbol{v}$, which implies that $\|A_0^TA_{\hat{J}}\boldsymbol{v}_{\hat{J}}\| \le \lambda\|\boldsymbol{v}_{\hat{J}}\| < \lambda$, since A_0 and $A_{\hat{J}}$ are submatrices of A, and $\|\boldsymbol{v}_{\hat{J}}\| < 1$. The term multiplying h^2 in (5.3) can be bounded below as follows:

$$\text{(5.4)} \quad \begin{aligned} \|A_0\boldsymbol{v}_0\|^2 + 2\left(A_0\boldsymbol{v}_0\right)^T A_{\hat{J}}\boldsymbol{v}_{\hat{J}} &\ge -2\|\boldsymbol{v}_0\|\ \left\|A_0^TA_{\hat{J}}\boldsymbol{v}_{\hat{J}}\right\| \\ &> -2\lambda\|\boldsymbol{v}_0\| \ge -2\lambda\mathbf{1}^T|\boldsymbol{v}_0| = -2\lambda\|\boldsymbol{v}_0\|_1, \end{aligned}$$

[4]We use $\|A\boldsymbol{v}\|^2 = \|A_{\hat{J}}\boldsymbol{v}_{\hat{J}}\|^2 + \|A_0\boldsymbol{v}_0\|^2 + 2(A_0\boldsymbol{v}_0)^TA_{\hat{J}}\boldsymbol{v}_{\hat{J}}$ and $(A\hat{\boldsymbol{t}} - \boldsymbol{y})^TA\boldsymbol{v} = (A_{\hat{J}}\hat{\boldsymbol{t}}_{\hat{J}} - \boldsymbol{y})^T(A_{\hat{J}}\boldsymbol{v}_{\hat{J}} + A_0\boldsymbol{v}_0)$.

where the entries of the vector $|\boldsymbol{v}_0|$ are the *moduli* of the relevant entries of $\boldsymbol{v}_0$. Thus $\|\boldsymbol{v}_0\|_1 := \mathbf{1}^T|\boldsymbol{v}_0| = \sum_{k\in\hat{J}} |v_k|$ is the ℓ_1-norm of $\boldsymbol{v}_0$. In the last inequality in (5.4) we exploit the fact that $\|\boldsymbol{v}_0\| \leq \|\boldsymbol{v}_0\|_1$.

The term multiplying h in (5.3) can be expressed in terms of θ_k and $\theta_{\max}$, considered in (4.15)–(4.16). Then we have

$$2\left(A_{\hat{J}}\hat{\boldsymbol{t}}_{\hat{J}} - \boldsymbol{y}\right)^T A_0\boldsymbol{v}_0 = 2\sum_{k\in\hat{J}} \boldsymbol{a}_k^T\left(A_{\hat{J}}\hat{\boldsymbol{t}}_{\hat{J}} - \boldsymbol{y}\right)v_k = -\sum_{k\in\hat{J}}\theta_k v_k \geq -\theta_{\max}\|\boldsymbol{v}_0\|_1. \tag{5.5}$$

By (a), we get $|\theta_k| < \beta\varphi'_+(0)$ for any $k \in \hat{J}$, hence $\theta_{\max} < \beta\varphi'_+(0)$. Take

$$\gamma := \frac{1}{2}\left[1 + \frac{\theta_{\max}}{\beta\varphi'_+(0)}\right]; \quad \text{then} \quad \gamma \in]\,0,1\,[. \tag{5.6}$$

Proposition 5.1(a) guarantees that there exists η_γ such that

$$\varphi(hv_k) \geq \gamma\varphi'_+(0)h|v_k| \quad \text{whenever} \quad 0 \leq h|v_k| < \eta_\gamma. \tag{5.7}$$

By $|v_k| \leq 1$ we get $\{h > 0 : h|v_k| < \eta_\gamma\} \supseteq \{h > 0 : h < \eta_\gamma\}$ for any k. Hence (5.7) holds for any $k \in \hat{J}$ when $0 < h < \eta_\gamma$. The last term in (5.3) then satisfies

$$\sum_{k\in\hat{J}}\varphi(hv_k) \geq \sum_{k\in\hat{J}}\gamma\varphi'_+(0)h|v_k| = \gamma\varphi'_+(0)h\,\|\boldsymbol{v}_0\|_1 \quad \text{if} \quad 0 \leq h < \eta_\gamma.$$

This result, combined with (5.4), (5.5), and (5.6), yields

$$\begin{aligned}\Delta^0_{\boldsymbol{v}}(h) &> -2\lambda h^2\|\boldsymbol{v}_0\|_1 - \theta_{\max}h\|\boldsymbol{v}_0\|_1 + \frac{\beta}{2}\left[1 + \frac{\theta_{\max}}{\beta\varphi'_+(0)}\right]\varphi'_+(0)\,h\,\|\boldsymbol{v}_0\|_1 \\ &= h\left[-2\lambda\,h + \frac{\beta\varphi'_+(0)}{2} - \frac{\theta_{\max}}{2}\right]\|\boldsymbol{v}_0\|_1 \quad \text{if} \quad 0 < h < \eta_\gamma.\end{aligned} \tag{5.8}$$

The right side above is positive for $0 < h < \rho_2$, where

$$\rho_2 = \frac{\beta\varphi'_+(0) - \theta_{\max}}{4\lambda}.$$

If $\hat{J} = \mathcal{S}$, we have $\hat{\boldsymbol{t}} = \mathbf{0}$ in (5.1). Now, $\boldsymbol{v}_0 = \boldsymbol{v}$ and $\Delta^1_{\boldsymbol{v}}$ is absent, so take $\rho_1 = 1$. By (4.3) and (4.16), $-2\boldsymbol{y}^T A\boldsymbol{v} \geq -\theta_{\max}\|\boldsymbol{v}\|_1$. Writing down $\|A\boldsymbol{v}\| > -2\lambda\|\boldsymbol{v}\|_1$, and taking γ as in (5.6), leads to (5.8).

In conclusion, we see that $\Delta_{\boldsymbol{v}}(h) > 0$ along any $\boldsymbol{v} \in \mathbb{I}^M$, if $h \in]\,0,\rho\,[$ with $\rho = \min\{\rho_1,\ \eta_\gamma,\ \rho_2\}$. □

The possibility of checking whether a point in $\mathbb{R}^M$ is a strict minimizer of $\mathcal{F}_{\boldsymbol{y}}$, by splitting $\mathcal{F}_{\boldsymbol{y}}$ into two parts that are analyzed for strictness *separately*, exists only because $\mathcal{F}_{\boldsymbol{y}}$ is nonsmooth at this point.

It is worth noting that neither Theorem 4.3 nor Theorem 5.2 permits conditions for a strict minimum, which are simultaneously necessary and sufficient, to be derived. Indeed, there may exist strict minima for which (4.3) involves equalities, but such minima do not fall into the scope of Theorem 5.2. The next theorem focuses on PFs which either have an infinite right derivative at zero or are discontinuous at zero.

THEOREM 5.3. *Consider $\mathcal{F}_{\boldsymbol{y}}$ of the form* (2.1)–(2.2), *where φ satisfies* (H1). *In addition, suppose that φ satisfies either* (H2) *with $\varphi'_+(0) = +\infty$, or* (H3). *Let $\hat{\boldsymbol{t}} \in \mathbb{R}^M$ be such that the assumption* (b) *of Theorem* 5.2 *holds.*

Then $\mathcal{F}_{\boldsymbol{y}}$ has a strict (local) minimum at $\hat{\boldsymbol{t}}$.

It turns out that the sufficient conditions for a strict minimum are simpler when $\varphi'_+(0) = \infty$ than $\varphi'_+(0)$ is finite. The proof of this theorem is given in the appendix.

6. Strong homogeneity.

6.1. The general situation. Next we show that taking φ nonsmooth at zero ensures that the minimizer functions relevant to $\mathcal{F}_{\boldsymbol{y}}$ are locally strongly homogeneous in quite general conditions. This is the main contribution of our work.

THEOREM 6.1. *Suppose φ satisfies* (H1) *and* (H2) *with $\varphi'_+(0) > 0$ finite, and $\mathcal{F}_{\boldsymbol{y}}$ is as in* (2.1)–(2.2). *Let $\hat{\boldsymbol{t}}$ be a strict (local) minimizer of $\mathcal{F}_{\boldsymbol{y}}$ such that $\hat{J} = \mathcal{J}(\hat{\boldsymbol{t}})$ is nonempty. Suppose that*

(a) *the inequality system* (4.3) *is strict;*

(b) *the restricted objective function $\mathcal{F}^{\hat{J}}_{\boldsymbol{y}}$, given in* (4.13) *for $\hat{J} \neq \mathcal{S}$, admits a (local) minimizer function $\boldsymbol{\mathcal{T}}^{\hat{J}}$ which is defined on a neighborhood of $\boldsymbol{y}$ and is continuous at $\boldsymbol{y}$, where it yields $\hat{\boldsymbol{t}}_{\hat{J}} = \boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y})$—the nonhomogeneous part of $\hat{\boldsymbol{t}}$; see* (4.10).

Then there exists $\boldsymbol{\mathcal{T}}$, a (local) minimizer function relevant to $\mathcal{F}_{\boldsymbol{y}}$, which is continuous at $\boldsymbol{y}$ and yields $\hat{\boldsymbol{t}} = \boldsymbol{\mathcal{T}}(\boldsymbol{y})$, and which has the following remarkable property:

$$\textit{there exists } \xi > 0 \textit{ such that} \quad (\mathcal{J} \circ \boldsymbol{\mathcal{T}})(\boldsymbol{y}') = \hat{J} \quad \textit{for all} \quad \boldsymbol{y}' \in B(\boldsymbol{y}; \xi).$$

Typically, there is a neighborhood of $\hat{\boldsymbol{t}}$, say, V, and a neighborhood of $\boldsymbol{y}$, say, U, such that each $\mathcal{F}_{\boldsymbol{y}'}$, corresponding to any $\boldsymbol{y}' \in U$, has a unique local minimizer over V. In such a situation, we can assert that all the minimizers $\hat{\boldsymbol{t}}' = \boldsymbol{\mathcal{T}}(\boldsymbol{y}') \in V$, obtained from $\boldsymbol{y}' \in B(\boldsymbol{y}; \xi) \subseteq U$, have the same set of strong homogeneity $\hat{J}$. If $\mathcal{F}_{\boldsymbol{y}}$ is strictly convex, this assertion holds with $U = \mathbb{R}^N$ and $V = \mathbb{R}^M$. As to (b): recall that the relation between $\hat{\boldsymbol{t}}$ and $\hat{\boldsymbol{t}}_{\hat{J}}$ is determined both by (4.10) and by $\hat{\boldsymbol{t}}[k] = 0$ for $k \in \hat{J}$.

Proof. Let first $\hat{J} \neq \mathcal{S}$. By (b), there exists $\omega > 0$ such that (b) holds on $B(\boldsymbol{y}; \omega)$ and, whenever $\hat{J} \neq \mathcal{S}^\circ$,

$$\text{(6.1)} \quad \boldsymbol{y}' \in B(\boldsymbol{y}; \omega) \;\text{ leads to }\; \left| \mathcal{T}^{\hat{J}}_i(\boldsymbol{y}') - \hat{\boldsymbol{t}}_{\hat{J}}[i] \right| < \mu \;\text{ for any } i \text{ such that } \kappa_i \in \mathcal{S}^\circ \setminus \hat{J},$$

with μ the bound defined in (3.4) and κ_i defined as in (4.9). Then ω is similar to η in (3.5). From (6.1), we see that $|\, \mathcal{T}^{\hat{J}}_i(\boldsymbol{y}') \,| > \left| \hat{\boldsymbol{t}}_{\hat{J}}[i] \right| - \mu \geq 0$ for any i such that $\kappa_i \in \mathcal{S}^\circ \setminus \hat{J}$ and for all $\boldsymbol{y}' \in B(\boldsymbol{y}; \omega)$.

Define $\tilde{\boldsymbol{\mathcal{T}}} : B(\boldsymbol{y}; \omega) \mapsto \mathbb{R}^M$ as follows:

$$\text{(6.2)} \qquad \begin{aligned} \tilde{\mathcal{T}}_{\kappa_i}(\boldsymbol{y}') &= \mathcal{T}^{\hat{J}}_i(\boldsymbol{y}') \quad \text{if} \quad i = 1, \ldots, \#\{\mathcal{S} \setminus \hat{J}\}, \\ \tilde{\mathcal{T}}_k(\boldsymbol{y}') &= 0 \qquad\qquad \text{if} \quad k \in \hat{J}. \end{aligned}$$

By construction, $\tilde{\boldsymbol{\mathcal{T}}}$ is continuous at $\boldsymbol{y}$. Moreover, (6.1) and (6.2) together show that

$$\mathcal{J}\left(\tilde{\boldsymbol{\mathcal{T}}}(\boldsymbol{y}') \right) = \hat{J} \quad \text{for all} \quad y' \in B(\boldsymbol{y}, \omega).$$

However, it is not clear whether or not $\tilde{\boldsymbol{\mathcal{T}}}$ is a minimizer function relevant to $\mathcal{F}_{\boldsymbol{y}}$ on a neighborhood of $\boldsymbol{y}$. From Definition 2.2, we have to check whether or not $\tilde{\boldsymbol{t}}' = \tilde{\boldsymbol{\mathcal{T}}}(\boldsymbol{y}')$

are *strict* minimizers of $\mathcal{F}_{\boldsymbol{y}'}$ when $\boldsymbol{y}'$ is in the vicinity of $\boldsymbol{y}$. Our considerations are based on Theorem 5.2. First, the assumption (b) above says that each $\tilde{\boldsymbol{t}}' = \tilde{\boldsymbol{\mathcal{T}}}(\boldsymbol{y}')$, obtained from $\boldsymbol{y}' \in B(\boldsymbol{y};\omega)$, satisfies the condition (b) of Theorem 5.2. So, it remains to find a radius $\xi \leq \omega$ such that $\tilde{\boldsymbol{t}}' = \tilde{\boldsymbol{\mathcal{T}}}(\boldsymbol{y}')$ satisfies the condition (a) of Theorem 5.2 for all $\boldsymbol{y}' \in B(\boldsymbol{y};\xi)$. That is, we seek $\xi \in]\,0,\omega\,]$, which ensures that any $\boldsymbol{y}' = \boldsymbol{y} + h\boldsymbol{u}$ for $0 < h < \xi$ and $\boldsymbol{u} \in \mathbb{I}^N$ yields $\tilde{\boldsymbol{t}}' = \tilde{\boldsymbol{\mathcal{T}}}(\boldsymbol{y}+h\boldsymbol{u})$, which strictly satisfies (4.3).

For any $k \in \hat{J}$, let $\mathcal{L}_k$ give the left side of (4.3), relevant to $\tilde{\boldsymbol{t}}' = \tilde{\boldsymbol{\mathcal{T}}}(\boldsymbol{y}+h\boldsymbol{u})$:

$$\mathcal{L}_k(h;\boldsymbol{u}) := 2\left|\boldsymbol{a}_k^T\left[(\boldsymbol{y}+h\boldsymbol{u}) - A\tilde{\boldsymbol{\mathcal{T}}}(\boldsymbol{y}+h\boldsymbol{u})\right]\right|. \tag{6.3}$$

Next, we determine a convenient upper bound of $\mathcal{L}_k$. Put

$$a := \max\left\{\max_{k\in\hat{J}}\|\boldsymbol{a}_k\|,\ 1\right\}. \tag{6.4}$$

By (6.2) and (4.11) we see that $A\tilde{\boldsymbol{\mathcal{T}}}(\boldsymbol{y}+h\boldsymbol{u}) = A_{\hat{J}}\boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y}+h\boldsymbol{u})$. Hence

$$\begin{aligned}\mathcal{L}_k(h;\boldsymbol{u}) &= 2\left|\boldsymbol{a}_k^T\left[\boldsymbol{y} - A_{\hat{J}}\boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y})\right] + h\boldsymbol{a}_k^T\boldsymbol{u} - \boldsymbol{a}_k^T A_{\hat{J}}\left[\boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y}+h\boldsymbol{u}) - \boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y})\right]\right| \\ &= \left|\theta_k + 2\,\boldsymbol{a}_k^T\left[h\boldsymbol{u} - A_{\hat{J}}\left(\boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y}+h\boldsymbol{u}) - \boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y})\right)\right]\right| \quad \text{for any } k\in\hat{J},\end{aligned} \tag{6.5}$$

where θ_k are defined as in (4.15). Let λ be as in (5.4), then $\|\boldsymbol{a}_k^T A_{\hat{J}}\| \leq \sqrt{\lambda}\|\boldsymbol{a}_k\|$. Also using $\theta_{\max}$ as given in (4.16) and the fact that $\|\boldsymbol{u}\| = 1$, we get

$$\begin{aligned}\mathcal{L}_k(h;\boldsymbol{u}) &\leq |\theta_k| + 2h\,\|\boldsymbol{a}_k\| + 2\|\boldsymbol{a}_k^T A_{\hat{J}}\|\,\left\|\boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y}+h\boldsymbol{u}) - \boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y})\right\| \\ &\leq \theta_{\max} + 2ha + 2a\sqrt{\lambda}\left\|\boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y}+h\boldsymbol{u}) - \boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y})\right\| \quad \text{for any } k\in\hat{J}.\end{aligned} \tag{6.6}$$

From (b), there exists $\sigma \in]\,0,\omega\,]$, which ensures that for all $\boldsymbol{u} \in \mathbb{I}^N$ we have that

$$h \in]\,0,\sigma\,[\ \text{ leads to } \ \left\|\boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y}+h\boldsymbol{u}) - \boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y})\right\| < \frac{\beta\varphi'_+(0) - \theta_{\max}}{4\,a\sqrt{\lambda}}. \tag{6.7}$$

The upper bound above is positive by (a). Introducing (6.7) into (6.6) yields

$$\begin{aligned}&\mathcal{L}_k(h;\boldsymbol{u}) < \overline{\mathcal{L}}(h) \ \text{ for all } \ h\in]\,0,\sigma\,[, \text{ for all } \boldsymbol{u}\in\mathbb{I}^N \text{ and for any } k\in\hat{J}, \\ &\text{where} \quad \overline{\mathcal{L}}(h) := \theta_{\max} + 2ha + \frac{\beta\varphi'_+(0) - \theta_{\max}}{2}.\end{aligned} \tag{6.8}$$

This $\overline{\mathcal{L}}$ is the sought upper bound of $\mathcal{L}_k$ for any k. Now put

$$\xi := \min\left\{\sigma,\ \frac{\beta\varphi'_+(0) - \theta_{\max}}{4a}\right\}. \tag{6.9}$$

Using that $a > 0$ and that $\xi \leq \left(\beta\varphi'_+(0) - \theta_{\max}\right)/(4a)$, we deduce that

$$\overline{\mathcal{L}}(h) < \theta_{\max} + 2a\xi + \frac{\beta\varphi'_+(0) - \theta_{\max}}{2} \leq \beta\varphi'_+(0) \ \text{ whenever } \ h\in[0,\xi[. \tag{6.10}$$

This inequality, combined with (6.8), shows that

$$h\in]0,\xi[\text{ leads to } \ \mathcal{L}_k(h;\boldsymbol{u}) < \beta\varphi'_+(0) \ \text{ for all } \boldsymbol{u}\in\mathbb{I}^N, \quad \text{for any } k\in\hat{J}. \tag{6.11}$$

In other words, $\tilde{\boldsymbol{t}}' = \tilde{\boldsymbol{\mathcal{T}}}(\boldsymbol{y} + h\boldsymbol{u})$ satisfies (4.3) in its strict form.

If $\hat{J} = \mathcal{S}$, put $\tilde{\boldsymbol{\mathcal{T}}} = \mathbf{0}$ and $\omega = 1$ in (6.2). Taking $\sigma = 1$ and ξ as in (6.9), we get (6.11).

By Theorem 5.2, every $\tilde{\boldsymbol{t}}' = \tilde{\boldsymbol{\mathcal{T}}}(\boldsymbol{y}')$ obtained from $\boldsymbol{y}' \in B(\boldsymbol{y}; \xi)$ is a strict minimizer of $\mathcal{F}_{\boldsymbol{y}'}$. Thus $\boldsymbol{\mathcal{T}} := \tilde{\boldsymbol{\mathcal{T}}}$ is a *strict* minimizer function on $\boldsymbol{y}' \in B(\boldsymbol{y}; \xi)$. ◻

The next theorem gives a similar result for continuous PFs with $\varphi'_+(0) = \infty$ and for discontinuous-at-zero PFs.

THEOREM 6.2. *Consider $\mathcal{F}_{\boldsymbol{y}}$ in* (2.1)–(2.2) *with φ satisfying* (H1). *Suppose also that φ satisfies either* (H2) *with $\varphi'_+(0) = \infty$, or* (H3). *Let $\hat{\boldsymbol{t}}$ be a strict (local) minimizer of $\mathcal{F}_{\boldsymbol{y}}$ such that $\hat{J} = \mathcal{J}(\hat{\boldsymbol{t}})$ is nonempty and that the assumption* (b) *of Theorem* 6.1 *holds.*

Then there exists $\boldsymbol{\mathcal{T}}$ a (local) minimizer function relevant to $\mathcal{F}_{\boldsymbol{y}}$, which is continuous at $\boldsymbol{y}$ and yields $\hat{\boldsymbol{t}} = \boldsymbol{\mathcal{T}}(\boldsymbol{y})$, and which exhibits the following particularity:

$$\text{there exists } \xi > 0 \text{ such that} \quad (\mathcal{J} \circ \boldsymbol{\mathcal{T}})(\boldsymbol{y}') = \hat{J} \quad \text{for all} \ \ \boldsymbol{y}' \in B(\boldsymbol{y}; \xi).$$

Proof. Take $\tilde{\boldsymbol{\mathcal{T}}}$ as in the proof of Theorem 6.1, then each $\tilde{\boldsymbol{\mathcal{T}}}(\boldsymbol{y}')$ with $\boldsymbol{y}' \in B(\boldsymbol{y}; \omega)$ satisfies the conditions of Theorem 5.3. Thus $\tilde{\boldsymbol{\mathcal{T}}}$ is a strict minimizer function of $\mathcal{F}_{\boldsymbol{y}}$, and therefore $\boldsymbol{\mathcal{T}} = \tilde{\boldsymbol{\mathcal{T}}}$ on $B(\boldsymbol{y}; \omega)$. So take $\xi = \omega$. ◻

Note that Theorems 6.1 and 6.2 require that *only* $\boldsymbol{\mathcal{T}}^{\hat{J}}$—the minimizer function of $\mathcal{F}^{\hat{J}}_{\boldsymbol{y}}$—be locally defined and continuous. In particular, this holds if $\mathcal{F}^{\hat{J}}_{\boldsymbol{y}}$ is $\mathcal{C}^2$ in the vicinity of $\hat{\boldsymbol{t}}_{\hat{J}}$ and its Hessian at $\hat{\boldsymbol{t}}_{\hat{J}}$ is positive definite.

Both Theorems 6.1 and 6.2 reveal a particular form of *local resistance* to noise, proper to the strongly homogeneous zones of a local minimizer where the objective function is nonsmooth: in the presence of small data variations, the nonhomogeneous part of the minimizer function evolves continuously, while its zero part remains constant. Moreover, the proofs of these theorems show that $\xi \le \omega$: this local resistance is stronger when $\varphi'_+(0)$ is infinite.

6.2. Boundary situations. Theorem 6.1 establishes the strong homogeneity of a local minimizer function relevant to $\mathcal{F}_{\boldsymbol{y}}$ when all the inequalities in (4.3) are strict. Now we focus on a minimizer involving one, or several, equalities in (4.3). Note that such a situation cannot occur unless $\varphi'_+(0)$ is *finite*.

PROPOSITION 6.3. *Let φ satisfy* (H1) and (H2) *and $\varphi'_+(0) > 0$ be finite. Consider $\mathcal{F}_{\boldsymbol{y}}$ as given by* (2.1)–(2.2). *Let $\hat{\boldsymbol{t}}$ be a strict (local) minimizer of $\mathcal{F}_{\boldsymbol{y}}$ with $\hat{J} = \mathcal{J}(\hat{\boldsymbol{t}}) \neq \emptyset$. Suppose*

(a) *the set $\hat{J}^0 := \{k \in \hat{J} : 2\left|\boldsymbol{a}_k^T(\boldsymbol{y} - A\hat{\boldsymbol{t}})\right| = \beta\varphi'_+(0)\}$ is nonempty; i.e.,* (4.3) *involves several equalities;*

(b) *the restricted objective function $\mathcal{F}^{\hat{J}}_{\boldsymbol{y}}$, given in* (4.13)–(4.14) *for $\hat{J} \neq \mathcal{S}$, admits a minimizer function $\boldsymbol{\mathcal{T}}^{\hat{J}}$ such that $\boldsymbol{\mathcal{T}}^{\hat{J}}$ is differentiable on a neighborhood of $\boldsymbol{y}$, $\mathrm{D}\boldsymbol{\mathcal{T}}^{\hat{J}}$ is continuous at $\boldsymbol{y}$, and $\boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y}) = \hat{\boldsymbol{t}}_{\hat{J}}$ is the nonhomogeneous part of $\hat{\boldsymbol{t}}$; see* (4.10)*;*

(c) *for any $k \in \hat{J}^0$, $\boldsymbol{\tau}_k^T := \boldsymbol{a}_k^T - \boldsymbol{a}_k^T A_{\hat{J}} \mathrm{D}\boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y}) \neq \mathbf{0}$ whenever $\hat{J} \neq \mathcal{S}$. If $\hat{J} = \mathcal{S}$, put $\boldsymbol{\tau}_k := \boldsymbol{a}_k$.*

By using θ_k as given in (4.15)*, introduce the set*

$$\mathcal{Q} := \bigcap_{k \in \hat{J}_0} \mathcal{Q}_k \quad \text{where} \quad \mathcal{Q}_k := \left\{ \boldsymbol{u} \in \mathbb{R}^N : \theta_k \boldsymbol{\tau}_k^T \boldsymbol{u} < 0 \right\}. \tag{6.12}$$

Then there exists $\boldsymbol{\mathcal{T}}$ *a (local) minimizer function relevant to* $\mathcal{F}_{\boldsymbol{y}}$*, characterized by the following: with any* $\boldsymbol{u} \in \mathcal{Q}$ *there is associated a* $\zeta(\boldsymbol{u}) > 0$ *such that*

$$\text{any } \ \boldsymbol{y}' = \boldsymbol{y} + h\boldsymbol{u} \text{ with } 0 < h < \zeta(\boldsymbol{u}) \ \text{ leads to } \ (\mathcal{J} \circ \boldsymbol{\mathcal{T}})(\boldsymbol{y}') = \hat{J},$$

and $\boldsymbol{\mathcal{T}}$ *is continuous at* $\boldsymbol{y}'$.

The situations where (c) fails to hold are unlikely since they may occur only at special data points. The proof of this proposition is given in the appendix.

7. Data sets and strong homogeneity.

7.1. Sets of data yielding the same estimate. Given an arbitrary $\tilde{\boldsymbol{t}} \in \mathbb{R}^M$, the next theorem shows the possibility of finding connected sets of data $\boldsymbol{y}$ such that all relevant objective functions $\mathcal{F}_{\boldsymbol{y}}$ reach a strict minimum at this $\tilde{\boldsymbol{t}}$.

PROPOSITION 7.1. *Suppose* φ *satisfies* (H1) and (H2). *Let* $\tilde{\boldsymbol{t}} \in \mathbb{R}^M$ *be an arbitrary point involving* $\tilde{J} = \mathcal{J}(\tilde{\boldsymbol{t}})$ *nonempty. In addition, suppose*

(a) *A is invertible, i.e.,* Rank $(A) = M = N$*;*

(b) *the matrix* $2A_{\tilde{J}}^T A_{\tilde{J}} + \beta \mathrm{D}^2 \Phi^{\tilde{J}}(\tilde{\boldsymbol{t}}_{\tilde{J}})$ *is positive definite, where* $\tilde{\boldsymbol{t}}_{\tilde{J}}$ *is the nonhomogeneous part of* $\tilde{\boldsymbol{t}}$*, defined as in* (4.10)*, and* $\Phi^{\tilde{J}}$ *and* $A_{\tilde{J}}$ *are defined according to* (4.14) *and* (4.11)*, respectively.*

Then there exists a polyhedron $\mathcal{V}_{\tilde{\boldsymbol{t}}} \subset \mathbb{R}^N$ *of dimension* $\#\{\tilde{J}\}$ *such that for any* $\boldsymbol{y} \in \mathcal{V}_{\tilde{\boldsymbol{t}}}$*, the relevant* $\mathcal{F}_{\boldsymbol{y}}$ *reaches a strict local minimum at* $\tilde{\boldsymbol{t}}$.

The assumption (b) holds in numerous important situations; for instance, take φ and $\tilde{\boldsymbol{t}}$ such that $\varphi''(\tilde{t}_k) > 0$ for any $k \in \mathcal{S}^\circ \setminus \mathcal{J}$.

Proof. We reformulate the strict form of (4.3)–(4.5) as follows:

$$-\beta\varphi'_+(0)\mathbf{1} + 2A_0^T A_{\tilde{J}} \tilde{\boldsymbol{t}}_{\tilde{J}} < 2A_0^T \boldsymbol{y} < \beta\varphi'_+(0)\mathbf{1} + 2A_0^T A_{\tilde{J}} \tilde{\boldsymbol{t}}_{\tilde{J}}, \tag{7.1}$$

$$2A_{\tilde{J}}^T \boldsymbol{y} \;=\; 2A_{\tilde{J}}^T A_{\tilde{J}} \tilde{\boldsymbol{t}}_{\tilde{J}} + \beta \mathrm{D}\Phi^{\hat{J}}(\tilde{\boldsymbol{t}}_{\tilde{J}}), \tag{7.2}$$

where the operations $=$ and $<$ are performed element by element. If we wish for $\tilde{\boldsymbol{t}}$ to satisfy the conditions of Theorem 5.2 with respect to $\mathcal{F}_{\boldsymbol{y}}$, then $\boldsymbol{y}$ must meet (7.1)–(7.2) with respect to $\tilde{\boldsymbol{t}}_{\tilde{J}}$. Since A is invertible, Rank $(A_{\tilde{J}}^T) = M - \#\{\tilde{J}\}$, then the assumption (a) shows that

$$\mathcal{G}_{\tilde{\boldsymbol{t}}} := \left\{ \boldsymbol{y} \in \mathbb{R}^N : (7.2) \text{ holds} \right\}$$

is an affine subspace with $\dim(\mathcal{G}_{\tilde{\boldsymbol{t}}}) = N - \text{Rank}\,(A_{\tilde{J}}^T) = \#\{\tilde{J}\} \geq 1$.

Next, $\mathcal{H}_{\tilde{\boldsymbol{t}}} := \left\{ \boldsymbol{y} \in \mathbb{R}^N : (7.1) \text{ holds} \right\}$ is a polyhedron of $\mathbb{R}^N$. Notice that $\mathcal{H}_{\tilde{\boldsymbol{t}}} = \mathbb{R}^N$ if $\varphi'_+(0) = \infty$. By (a), we see that $\mathcal{V}_{\tilde{\boldsymbol{t}}} := \mathcal{G}_{\tilde{\boldsymbol{t}}} \bigcap \mathcal{H}_{\tilde{\boldsymbol{t}}}$ is nonempty. Moreover, $\mathcal{V}_{\tilde{\boldsymbol{t}}}$ contains cells of dimension $\#\{\tilde{J}\}$. The condition (a) of Theorem 5.2 is valid at $\tilde{\boldsymbol{t}}$ for any $\mathcal{F}_{\boldsymbol{y}}$ corresponding to $\boldsymbol{y} \in \mathcal{V}_{\tilde{\boldsymbol{t}}}$.

Combining (7.2) and (b) above, that is,

$$\begin{matrix} \mathrm{D}\mathcal{F}_{\boldsymbol{y}}^{\tilde{J}}(\tilde{\boldsymbol{t}}_{\tilde{J}}) = \mathbf{0}^T \\ \mathrm{D}^2\mathcal{F}_{\boldsymbol{y}}^{\tilde{J}}(\tilde{\boldsymbol{t}}_{\tilde{J}}) \ \text{ positive definite} \end{matrix} \quad \text{for all } \ \boldsymbol{y} \in \mathcal{V}_{\tilde{\boldsymbol{t}}}, \tag{7.3}$$

shows that $\tilde{\boldsymbol{t}}_{\tilde{J}}$ is a strict (local) minimizer of any $\mathcal{F}_{\boldsymbol{y}}^{\tilde{J}}$ corresponding to $\boldsymbol{y} \in \mathcal{V}_{\tilde{\boldsymbol{t}}}$. So $\tilde{\boldsymbol{t}}$ satisfies the condition (b) of Theorem 5.2 with respect to $\mathcal{F}_{\boldsymbol{y}}$ whenever $\boldsymbol{y} \in \mathcal{V}_{\tilde{\boldsymbol{t}}}$. Hence the result. ◻

Remark. If φ is $\mathcal{C}^2$ on $\mathbb{R}\setminus\{0\}$, there is a minimizer function $\boldsymbol{\mathcal{T}}$, such that $\boldsymbol{\mathcal{T}}(\boldsymbol{y}) = \tilde{\boldsymbol{t}}$ on $\mathcal{V}_{\tilde{\boldsymbol{t}}}$, and which is continuous on $\mathcal{V}_{\tilde{\boldsymbol{t}}}$. To see the latter, choose $\boldsymbol{y} \in \mathcal{V}_{\tilde{\boldsymbol{t}}}$ arbitrarily. By (b) and (3.4), there exists $\varrho \in]\,0, \mu\,]$ such that $\mathrm{D}^2\mathcal{F}^{\tilde{J}}_{\boldsymbol{y}}(\boldsymbol{t}_{\tilde{J}})$ is positive definite for any $\boldsymbol{t}_{\tilde{J}} \in B(\tilde{\boldsymbol{t}}_{\tilde{J}}; \varrho)$. By the implicit function theorem [4], the equation $\mathrm{D}\mathcal{F}^{\tilde{J}}_{\boldsymbol{y}}(\boldsymbol{t}_{\tilde{J}}) = \boldsymbol{0}^T$ defines a unique, implicit, differentiable function $\boldsymbol{\mathcal{T}}^{\hat{J}}$ on a neighborhood of $\boldsymbol{y}$, such that $\boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y}') \in B(\tilde{\boldsymbol{t}}_{\tilde{J}}; \varrho)$ and $\mathrm{D}\mathcal{F}^{\tilde{J}}_{\boldsymbol{y}'}[\boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y}')] = \boldsymbol{0}^T$ whenever $\boldsymbol{y}'$ is in this neighborhood. Hence the condition (b) of Theorem 6.1 holds for any $\boldsymbol{y} \in \mathcal{V}_{\tilde{\boldsymbol{t}}}$. Using Theorem 6.1 if $\varphi'_+(0)$ is finite, or Theorem 6.2 if $\varphi'_+(0) = \infty$, we deduce that $\boldsymbol{\mathcal{T}}$ is continuous at any $\boldsymbol{y} \in \mathcal{V}_{\tilde{\boldsymbol{t}}}$.

7.2. Volumes of data preserving ($\mathcal{J} \circ \boldsymbol{\mathcal{T}}$) constant. We now turn to the organization of the data space $\mathbb{R}^N$, induced by the distinct sets of strong homogeneity yielded by the strict minimizers of an objective function. Theorems 6.1 and 6.2 mean that $\mathbb{R}^N$ contains open *volumes* composed of data that lead to local minimizers having the same set of strong homogeneity. Consider a minimizer function $\boldsymbol{\mathcal{T}}$, and let $\boldsymbol{y} \in U$ give rise to $\hat{J} = (\mathcal{J} \circ \boldsymbol{\mathcal{T}})(\boldsymbol{y})$ nonempty. Such a volume reads

$$\mathcal{W}^{\mathcal{T}}_{\hat{J}} = \left\{ \boldsymbol{y}' \in U : (\mathcal{J} \circ \boldsymbol{\mathcal{T}})(\boldsymbol{y}') = \hat{J} \right\}.$$

Note that $\mathcal{W}^{\mathcal{T}}_{\hat{J}} \supset B(\boldsymbol{y}, \xi)$ with ξ the radius found in Theorems 6.1 and 6.2, but $\mathcal{W}^{\mathcal{T}}_{\hat{J}}$ are not connected in general. The proofs of these theorems and of Proposition 6.3 show that the extent of a connected component of such a volume depends on both the domain of validity of (4.3) and the domain of definition of $\boldsymbol{\mathcal{T}}^{\hat{J}}$. These domains depend on the forms of A and of φ, and on the model parameters (β and those involved in φ). *It is crucial that the probability that noisy data belong to a volume $\mathcal{W}^{\mathcal{T}}_{\hat{J}}$, with $\hat{J}$ nonempty, be strictly positive.*

When $\boldsymbol{y}$ ranges over U, the set-valued function $(\mathcal{J} \circ \boldsymbol{\mathcal{T}})(\boldsymbol{y})$ generally takes several distinct values $\hat{J}_n \in \mathcal{P}(\mathcal{S}^\circ)$, $n \in \mathcal{I}_{\mathcal{T}}$ where $\mathcal{I}_{\mathcal{T}}$ is a set of indices relevant to the minimizer function $\boldsymbol{\mathcal{T}}$. Each $\hat{J}_n$, in its turn, gives rise to a volume $\mathcal{W}^{\mathcal{T}}_{\hat{J}_n}$. Thus, with a minimizer function $\boldsymbol{\mathcal{T}}$ there is associated a set of volumes $\{\mathcal{W}^{\mathcal{T}}_{\hat{J}_n},\ n \in \mathcal{I}_{\mathcal{T}}\}$: each time $\boldsymbol{y} \in \mathcal{W}^{\mathcal{T}}_{\hat{J}_n}$, we get a minimizer $\boldsymbol{\mathcal{T}}(\boldsymbol{y})$ with $\mathcal{J}(\boldsymbol{\mathcal{T}}(\boldsymbol{y})) = \hat{J}_n$. Reciprocally, the domain of $\boldsymbol{\mathcal{T}}$ equivalently reads $U = \cup_{n \in \mathcal{I}_{\mathcal{T}}} \mathcal{W}^{\mathcal{T}}_{\hat{J}_n}$. Speaking more loosely, the volumes corresponding to the different *nonempty* sets of strong homogeneity $\hat{J}_n$ are placed side by side. For a "reasonable" choice of the model parameters these volumes are large enough and the noisy data come across their union. This is the reason why the minimizers of an objective function, involving a nonsmooth-at-zero PF, exhibit large strongly homogeneous zones.

Theorems 6.1 and 6.2 show how the estimation using nonsmooth regularization reduces the "richness" of $\mathbb{R}^M$—the domain of the original signal $\boldsymbol{x}$—to a finite family of solutions in $\mathbb{R}^M$. This kind of *soft classification* is the way in which the prior, expressed through a regularizer Φ involving a nonsmooth-at-zero PF, is *effectively* taken into account by the estimator $\boldsymbol{\mathcal{T}}$, or equivalently by $\boldsymbol{\mathcal{X}}$.

7.3. Example. Consider the following strictly convex objective function:

$$\mathcal{E}_{\boldsymbol{y}}(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{y}\|^2 + \beta \sum_{k \in \mathcal{S}} |x_k|; \quad \text{then} \quad \hat{J} = \{k \in \mathcal{S} : \hat{x}_k = 0\}.$$

Its global minimizer $\hat{\boldsymbol{x}}$ is the unique point satisfying (4.3)–(4.5). Assume for the moment that $0 < \#\{\hat{J}\} < M$. Since A is the identity and $\varphi'_+(0) = 1$, the system

(4.3)–(4.5) becomes

$$
\begin{aligned}
&2(y_k - \hat{x}_k) - \beta \mathrm{sign}\,(\hat{x}_k) = 0 \quad \text{for} \quad k \in \mathcal{S} \setminus \hat{J}, \\
&2|y_k| \le \beta \quad \text{for} \quad k \in \hat{J}.
\end{aligned} \tag{7.4}
$$

The minimizer $\hat{\boldsymbol{x}} = \boldsymbol{\mathcal{X}}(\boldsymbol{y})$ admits an explicit expression:

$$
\begin{aligned}
&\hat{x}_k = \mathcal{X}_k(\boldsymbol{y}) = y_k - \frac{\beta}{2}\mathrm{sign}\,(y_k) \quad \text{if} \quad |y_k| > \frac{\beta}{2}, \\
&\hat{x}_k = \mathcal{X}_k(\boldsymbol{y}) = 0 \qquad\qquad\qquad\quad \text{if} \quad |y_k| \le \frac{\beta}{2};
\end{aligned}
$$

hence

$$
(\mathcal{J} \circ \mathcal{X})\,(\boldsymbol{y}) = \left\{ k \in \mathcal{S} : |y_k| \le \frac{\beta}{2} \right\}.
$$

This $\boldsymbol{\mathcal{X}}$ is a nonlinear filter which is used in [18, 16] to perform *soft thresholding*.

Provided that (7.4) is strict, i.e., that $|y_k| \ne \frac{1}{2}\beta$ for all $k \in \hat{J}$, a radius ξ, as the one exhibited in Theorem 6.1, reads

$$
\xi = \min_{k \in \mathcal{S}} \left| \frac{\beta}{2} - |y_k| \right|.
$$

Now let (7.4) involve one nonstrict inequality, say, $\hat{J}^0 = \{m\}$ with $y_m = \beta/2$. According to Proposition 6.3 and (4.15), we have $\theta_m = 2y_m = \beta$ and $\boldsymbol{\tau}_m = \boldsymbol{e}_m$, hence

$$
\mathcal{Q}_m = \left\{ \boldsymbol{u} \in \mathbb{I}^N : u_m < 0 \right\} \quad \text{and} \quad \zeta(\boldsymbol{u}) = \min \left\{ \beta, \ \left| \frac{\beta}{2} - |y_k| \right| \quad \text{for} \quad k \in \mathcal{S} \setminus \{m\} \right\}.
$$

On the other hand, $(\mathcal{J} \circ \mathcal{X})\,(\boldsymbol{y} + h\boldsymbol{u}) \ne \hat{J}$ whenever $u_m > 0$.

Conversely, fix $\hat{\boldsymbol{x}} \in \mathbb{R}^M$ with $\hat{J} = \mathcal{J}(\hat{\boldsymbol{x}})$ nonempty. As in Proposition 7.1, we seek $\mathcal{V}_{\hat{\boldsymbol{x}}}$, the set of all $\boldsymbol{y}$ such that $\hat{\boldsymbol{x}}$ minimizes $\mathcal{E}_{\boldsymbol{y}}$:

$$
\mathcal{V}_{\hat{\boldsymbol{x}}} = \left\{ \boldsymbol{y} : |y_k| \le \frac{\beta}{2} \quad \text{for} \quad k \in \hat{J} \quad \text{and} \quad y_k = \hat{x}_k + \frac{\beta}{2}\mathrm{sign}\,(\hat{x}_k) \quad \text{for} \quad k \in \mathcal{S} \setminus \hat{J} \right\}.
$$

Now fix $\hat{J} \ne \emptyset$. As in section 7.2, the set $\mathcal{W}_{\hat{J}}$ of all $\boldsymbol{y}$ yielding $(\mathcal{J} \circ \mathcal{X})\,(\boldsymbol{y}) = \hat{J}$ reads

$$
\mathcal{W}_{\hat{J}} = \left\{ \boldsymbol{y} : \ |y_k| \le \frac{\beta}{2} \text{ for } k \in \hat{J} \ \text{ and } \ |y_k| > \frac{\beta}{2} \text{ for } k \in \mathcal{S} \setminus \hat{J} \right\}.
$$

Such a $\mathcal{W}_{\hat{J}}$ is composed of a finite number of connected sets. In particular, the set $\mathcal{W}_{\mathcal{S}} = \{\boldsymbol{y} : |y_k| \le \frac{\beta}{2} \text{ for } k \in \mathcal{S}\}$ gives rise to $\hat{\boldsymbol{x}} = \boldsymbol{0}$, and hence $\hat{J} = \mathcal{S}$. On the contrary, $\mathcal{W}_{\emptyset} = \{\boldsymbol{y} : |y_k| > \frac{\beta}{2} \quad \text{for any} \quad k \in \mathcal{S}\}$ corresponds to minimizers $\hat{\boldsymbol{x}}$ with $\mathcal{J}(\hat{\boldsymbol{x}}) = \emptyset$.

Finally, $\{\mathcal{W}_{\hat{J}}, \ \hat{J} \in \mathcal{P}(\mathcal{S})\}$ form a *partition* of the data space $\mathbb{R}^N$.

8. Numerical illustration. By way of illustration, we consider the estimation of a signal $\boldsymbol{x}$ from noisy data $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n}$. In order to evaluate the ability of different PFs to recover, and to conserve, the strongly homogeneous zones yielded by minimizing the relevant $\mathcal{E}_{\boldsymbol{y}}$, we process in the same numerical conditions two data sets, contaminated by two very different noise realizations; see Figure 8.1. In all subsequent figures, the

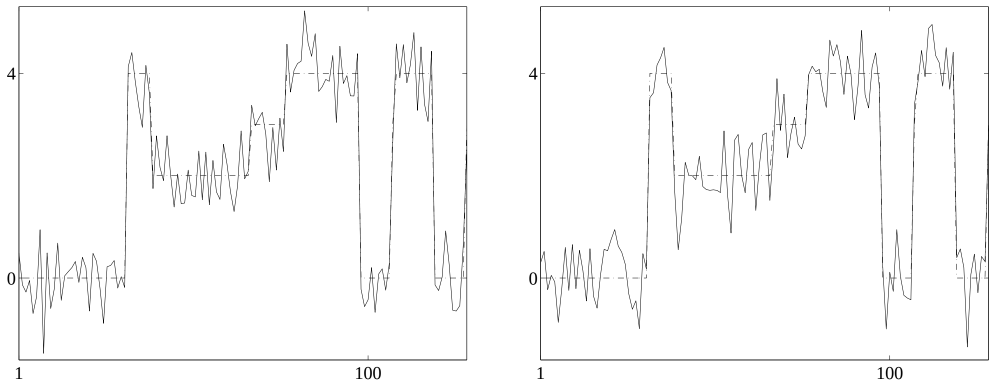

FIG. 8.1. Data $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n}$ *(—), corresponding to the* original $\boldsymbol{x}$ *(-.-.), contaminated with two different noise samples* $\boldsymbol{n}$*. The two data sets are shown in the left and in the right.*

estimates are presented on the left and on the right, respectively, while the shape of the PF is plotted in the middle.

Figure 8.2 shows an estimation using a *Huber* PF[5], which is quadratic near the origin, $\varphi(t) = \alpha\, t^2/2$ if $|t| < 1/\alpha$, and affine beyond it, $\varphi(t) = (\,|t| - 1/2\alpha\,)$ if $|t| \geq 1/\alpha$. This PF smooths the small differences while it adds a bias to the large differences. The obtained solutions do not contain strongly homogeneous zones.

Figure 8.3 illustrates the effect of a *dislocated quadratic* PF, which is nonsmooth at 0 and *quadratic* elsewhere, $\varphi(t) = (t - \alpha)^2$ if $t < 0$ and $\varphi(t) = (t + \alpha)^2$ if $t > 0$, with $\varphi(0) = 0$. Because of the quadratic shape of φ beyond 0, the large differences are noticeably underestimated. However, large strongly homogeneous zones are recovered under both noise samples. This experiment nicely corroborates our assertion that the estimation of strongly homogeneous zones is related only to the differentiability of φ at zero.

Figure 8.4 shows an estimation using the *modulus* PF in (1.4). This PF differs from the Huber PF only on $]-1/\alpha, 1/\alpha\,[$. However, the obtained solutions are essentially different: the strongly homogeneous zones are now well retrieved and they are globally the same for the two data sets. Note that now the amplitude of the signal is better estimated than with the dislocated quadratic PF.

Figure 8.5 presents an estimation using the *concave* PF given in (1.5). The same set of strong homogeneity is found under both noise samples. As expected, the large differences—the jumps—are slightly different.

9. Conclusion. Signals involving strongly homogeneous zones arise in various practical situations. The ability of a regularized estimator—the minimizer of an objective function—both to yield a solution containing large strongly homogeneous zones and to conserve them under small variations of the data has been formalized mathematically. It has been shown that an estimator involving a PF which is smooth at zero cannot recover nor conserve such zones. Instead, it has been demonstrated that nonsmoothness of the PF at zero ensures that the strict local minimizers of the

[5]This PF is not twice differentiable at $\pm 1/\alpha$. Nevertheless, it can be shown that the data $\boldsymbol{y} \in \mathbb{R}^N$, yielding a minimizer which involves one or several differences $\hat{t}_k$ equal to $\pm 1/\alpha$, belong to a closed, negligible set in $\mathbb{R}^N \ldots$. Correspondingly, $\mathcal{F}_{\boldsymbol{y}}$ is twice differentiable in the vicinity of its minimizer for almost any $\boldsymbol{y} \in \mathbb{R}^N$.

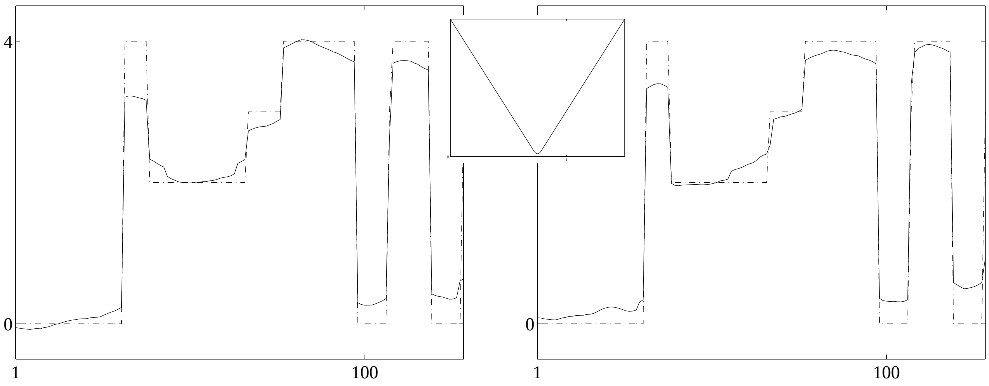

FIG. 8.2. Huber PF: $\varphi(t) = \alpha t^2/2$ *if* $|t| < 1/\alpha$ *and* $\varphi(t) = (\,|t| - 1/2\alpha\,)$ *if* $|t| \geq 1/\alpha$, *where* $\alpha = 30$ *and* $\beta = 4$. *Original (-.-.); estimate (—). This PF is smooth at zero and the obtained solutions do not contain strongly homogeneous zones.*

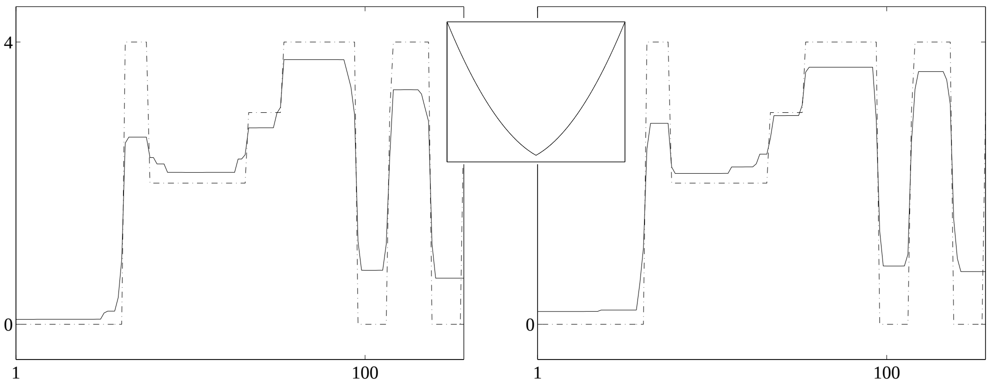

FIG. 8.3. Dislocated quadratic PF: $\varphi(t) = (t - \alpha)^2$ *if* $t < 0$ *and* $\varphi(t) = (t + \alpha)^2$ *if* $t > 0$, *where* $\alpha = 3$ *and* $\beta = 1$. *Original (-.-.); estimate (—). The minimizer has large strongly homogeneous zones. The large differences are underestimated.*

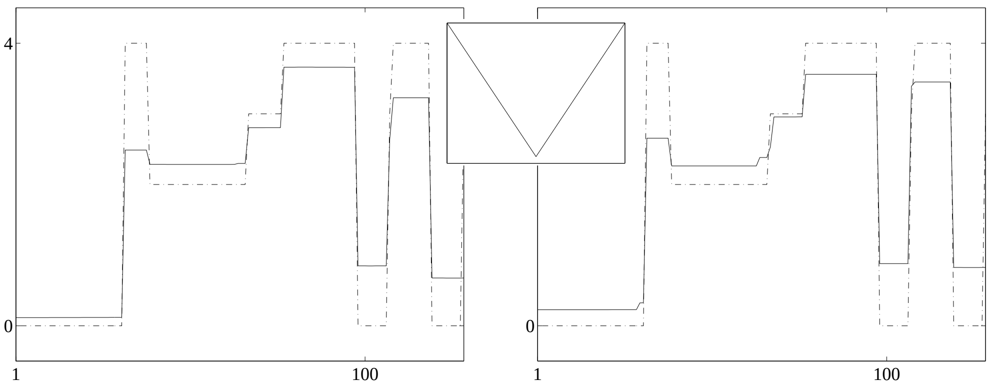

FIG. 8.4. Modulus PF: $\varphi(t) = |t|$ *with* $\beta = 9$. *Original (-.-.); estimate (—). The location of the strongly homogeneous zones, corresponding to both data sets, differ only at a few points. The large differences are better estimated than in Figure* 8.3.

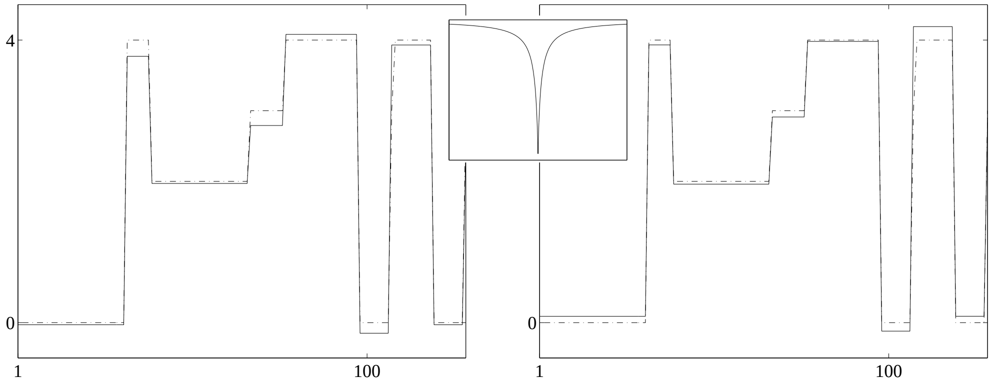

FIG. 8.5. Concave PF: $\varphi(t) = \alpha|t|/(1+\alpha|t|)$, *with* $\alpha = 10$ *and* $\beta = 3$. *Original (-.-.); estimate (—). Both data sets yield minimizers that are strongly homogeneous over the same zones. This PF is nonconvex and large differences are detected in these estimates.*

objective function conserve their strong homogeneity zones under small variations of the data. More precisely, any strict minimizer involving strongly homogeneous zones arises from a set of data which, in turn, give rise to strict minimizers having the *same* strongly homogeneous zones. Conversely, an estimator, comprising a PF that is nonsmooth at zero, recovers solutions that involve large strongly homogeneous zones. Two analytical examples and a set of numerical illustrations corroborate the relevance of our mathematical results.

10. Appendix.

Proof of Theorem 4.2. Function $\mathcal{F}_{\boldsymbol{y}}$ has a strict local minimum at $\hat{\boldsymbol{t}}$ whenever there exists $\rho > 0$ such that for any $\boldsymbol{t} \in B(\hat{\boldsymbol{t}};\rho) \setminus \{\hat{\boldsymbol{t}}\}$ we have $\mathcal{F}_{\boldsymbol{y}}(\boldsymbol{t}) > \mathcal{F}_{\boldsymbol{y}}(\hat{\boldsymbol{t}})$. Equivalently, for any $\boldsymbol{v} \in \mathbb{I}^M$ and for any $h \in]\,0,\rho\,[$:

$$\mathcal{F}_{\boldsymbol{y}}(\hat{\boldsymbol{t}} + h\boldsymbol{v}) > \mathcal{F}_{\boldsymbol{y}}(\hat{\boldsymbol{t}}) \quad \text{and} \quad \mathcal{F}_{\boldsymbol{y}}(\hat{\boldsymbol{t}} - h\boldsymbol{v}) > \mathcal{F}_{\boldsymbol{y}}(\hat{\boldsymbol{t}}).$$

The latter is equivalent to

$$\frac{\mathcal{F}_{\boldsymbol{y}}(\hat{\boldsymbol{t}} - h\boldsymbol{v}) - \mathcal{F}_{\boldsymbol{y}}(\hat{\boldsymbol{t}})}{-h} < 0 < \frac{\mathcal{F}_{\boldsymbol{y}}(\hat{\boldsymbol{t}} + h\boldsymbol{v}) - \mathcal{F}_{\boldsymbol{y}}(\hat{\boldsymbol{t}})}{h}.$$

Taking these inequalities in the limit $h \downarrow 0$ leads to (4.2). □

Proof of Theorem 5.3. As in Theorem 5.2, we examine the altitude increment $\Delta_{\boldsymbol{v}}$ given in (5.1) with $h \geq 0$. For $\hat{J} \neq \mathcal{S}$, the latter is split into two terms, as in (5.2)–(5.3).

• Term $\Delta^1_{\boldsymbol{v}}$. The condition ensuring that $\Delta^1_{\boldsymbol{v}}(h) > 0$ comes from the arguments presented in the proof of Theorem 5.2 and it has the form $0 < h < \rho_1$.

• Term $\Delta^0_{\boldsymbol{v}}$. Two situations arise, according to the continuity of φ at 0.

—φ satisfies (H2) with $\varphi'_+(0) = +\infty$. Choose

$$\gamma := \frac{2\theta_{\max}}{\beta}.$$

By Proposition 5.1 (b), there exists η_γ such that

$$\varphi(hv_k) \geq \gamma h|v_k| \quad \text{if} \quad 0 \leq h|v_k| < \eta_\gamma \quad \text{and in particular if} \quad 0 \leq h < \eta_\gamma,$$

since $0 \leq |v_k| \leq 1$. Then

$$\sum_{k\in\hat{J}} \varphi(hv_k) \geq \sum_{k\in\hat{J}} \gamma h|v_k| = \gamma h\,\mathbf{1}^T|\boldsymbol{v}_0| = h\frac{2\theta_{\max}}{\beta}\,\|\boldsymbol{v}_0\|_1 \quad \text{if} \quad 0 \leq h < \eta_\gamma.$$

This result, combined with (5.4) and (5.5), is introduced in (5.3):

$$\Delta^0_{\boldsymbol{v}}(h) > -2\lambda h^2\|\boldsymbol{v}_0\|_1 - \theta_{\max}h\|\boldsymbol{v}_0\|_1 + 2\theta_{\max}h\|\boldsymbol{v}_0\|_1 = h\,(-2\lambda h + \theta_{\max})\|\boldsymbol{v}_0\|_1.$$

The last term is positive if $0 < h < \theta_{\max}/2\lambda$.

When φ is continuous at 0, we conclude that $\Delta_{\boldsymbol{v}}(h) > 0$ along any $\boldsymbol{v} \in \mathbb{I}^M$ if

$$0 < h < \rho \quad \text{with} \quad \rho = \min\left\{\rho_1,\ \eta_\gamma,\ \frac{\theta_{\max}}{2\lambda}\right\}.$$

—φ satisfies (H3). The fact that $|v_k| \leq 1$ allows us to write that $\varphi(hv_k) \geq \gamma\,|v_k|$ when $h \in]\,0, \eta\,[$. In consequence,

$$\sum_{k\in\hat{J}} \varphi(hv_k) \geq \gamma\,\mathbf{1}^T|\boldsymbol{v}_0| = \gamma\|\boldsymbol{v}_0\|_1 \quad \text{if} \quad 0 < h < \eta.$$

The following lower bound on $\Delta^0_{\boldsymbol{v}}$ can then be obtained:

$$\Delta^0_{\boldsymbol{v}}(h) > -2\lambda h^2\|\boldsymbol{v}_0\|_1 - \theta_{\max}h\|\boldsymbol{v}_0\|_1 + \beta\gamma\|\boldsymbol{v}_0\|_1 = (-2\lambda h^2 - \theta_{\max}\,h + \beta\gamma)\,\|\boldsymbol{v}_0\|_1,$$

which holds if $0 < h < \eta$. The quadratic term in parentheses is concave and has a positive and a negative real root; it is positive between these roots.

For φ discontinuous at 0, we see that $\Delta_{\boldsymbol{v}}(h) > 0$ for any $\boldsymbol{v} \in \mathbb{I}^M$ if

$$0 < h < \rho \quad \text{where} \quad \rho = \min\left\{\rho_1,\ \eta,\ \frac{-\theta_{\max} + \sqrt{\theta^2_{\max} + 8\lambda\beta\gamma}}{4\lambda}\right\}.$$

If $\hat{J} = \mathcal{S}$, we have the same arguments as in the proof of Theorem 5.2. □

Proof of Proposition 6.3. This proof extends the reasoning underlying the proof of Theorem 6.1. The considerations about $\hat{\boldsymbol{t}}_{\hat{J}} = \boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y})$ for $\hat{J} \neq \mathcal{S}$, presented in the beginning of the latter proof, are still valid. So, consider ω—the radius in (6.1)—and $\tilde{\boldsymbol{\mathcal{T}}}$ as defined in (6.2). Our goal now is to find an open domain, contained in $B(\boldsymbol{y};\omega)$, whose elements $\boldsymbol{y}'$ are such that $\tilde{\boldsymbol{t}}' = \tilde{\boldsymbol{\mathcal{T}}}(\boldsymbol{y}')$ are assuredly strict minimizers of the relevant $\mathcal{F}_{\boldsymbol{y}'}$. To this end, we will seek a family of directions, say $\mathcal{Q} \subset \mathbb{I}^N$, in connection with a family of bounds $\{\zeta(\boldsymbol{u}) \in]\,0, \omega\,] : \boldsymbol{u} \in \mathcal{Q}\}$, such that $\tilde{\boldsymbol{t}}' = \tilde{\boldsymbol{\mathcal{T}}}(\boldsymbol{y}+h\boldsymbol{u})$ satisfies[6] the conditions of Theorem 5.2 whenever $\boldsymbol{u} \in \mathcal{Q}$ and $h \in]\,0, \zeta(\boldsymbol{u})\,[$. Observe that the condition (b) of Theorem 5.2 is actually satisfied by any $\tilde{\boldsymbol{\mathcal{T}}}(\boldsymbol{y}')$ with $\boldsymbol{y}' \in B(\boldsymbol{y};\omega)$.

Now we determine a domain where the condition (a) of Theorem 5.2 holds as well, i.e., that (4.3) holds and is strict. By (a), each $k \in \hat{J}^0$ gives rise to the alternative:

$$\text{(10.1)} \quad \text{either} \quad -\beta\varphi'_+(0) = 2\boldsymbol{a}_k^T(\boldsymbol{y} - A_{\hat{J}}\hat{\boldsymbol{t}}_{\hat{J}}) < \beta\varphi'_+(0), \quad \text{i.e.,} \quad \theta_k = -\beta\varphi'_+(0),$$

$$\text{(10.2)} \quad \text{or} \quad -\beta\varphi'_+(0) < 2\boldsymbol{a}_k^T(\boldsymbol{y} - A_{\hat{J}}\hat{\boldsymbol{t}}_{\hat{J}}) = \beta\varphi'_+(0), \quad \text{i.e.,} \quad \theta_k = \beta\varphi'_+(0).$$

[6]Notice that $\boldsymbol{y}$ belongs not to the domain determined by $\mathcal{Q}$ and ζ but to its closure.

Requiring that (4.3) be strict for $\boldsymbol{y}' = \boldsymbol{y} + h\boldsymbol{u}$ is equivalent to the following:

find $\boldsymbol{u} \in \mathbb{I}^N$ and $\zeta(\boldsymbol{u}) > 0$ such that $h \in \,]\,0, \zeta(\boldsymbol{u})\,[$ leads to

$$\mathcal{L}_k(h;\boldsymbol{u}) < \beta\varphi'_+(0) \quad \text{for any} \quad k \in \hat{J}^0, \tag{10.3}$$

$$\mathcal{L}_k(h;\boldsymbol{u}) < \beta\varphi'_+(0) \quad \text{for any} \quad k \in \hat{J} \setminus \hat{J}^0, \tag{10.4}$$

where $\mathcal{L}_k$ are defined as in (6.3). Below, we determine conditions on h and on $\boldsymbol{u}$ which ensure that (10.3) and (10.4) are true.

• Condition (10.4). Following (4.16), we now introduce

$$\theta^0_{\max} := \max\left\{ |\,\theta_k\,| \quad \text{for} \quad k \in \hat{J} \setminus \hat{J}^0 \right\};$$

then $\theta^0_{\max} < \beta\varphi'_+(0)$ strictly. By (b), there exists $\sigma_0 \in]\,0, \omega\,[$ such that

$$h \in]\,0, \sigma_0\,[\quad \text{yields} \quad \left\| \boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y} + h\boldsymbol{u}) - \boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y}) \right\| < \frac{\beta\varphi'_+(0) - \theta^0_{\max}}{4\,a\sqrt{\lambda}} \quad \text{for all } \boldsymbol{u} \in \mathbb{I}^N, \tag{10.5}$$

where a and λ are as in (6.4) and (5.4), respectively. Similarly to (6.9), define

$$\xi_0 = \min\left\{ \sigma_0, \; \frac{\beta\varphi'_+(0) - \theta^0_{\max}}{4a} \right\}.$$

The implication (10.5) is similar to (6.7), whereas the arguments that yielded (6.10) can now be applied to (10.4). Thus we see that (10.4) holds for all $\boldsymbol{u} \in \mathbb{I}^N$ whenever $0 < h < \xi_0$.

• Condition (10.3). These inequalities are more difficult to satisfy. Define

$$\mathcal{R}_k(h;\boldsymbol{u}) := h\boldsymbol{a}_k^T\boldsymbol{u} - \boldsymbol{a}_k^T A_{\hat{J}}\left[\boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y} + h\boldsymbol{u}) - \boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y}) \right]. \tag{10.6}$$

Using (6.5) and (10.1)–(10.2), any $\mathcal{L}_k$ corresponding to $k \in \hat{J}^0$ can be reformulated as

$$\mathcal{L}_k(h;\boldsymbol{u}) = \begin{cases} \left| -\beta\varphi'_+(0) + 2\mathcal{R}_k(h;\boldsymbol{u}) \right| & \text{if} \quad \theta_k = -\beta\varphi'_+(0), \\ \left| \beta\varphi'_+(0) + 2\mathcal{R}_k(h;\boldsymbol{u}) \right| & \text{if} \quad \theta_k = \beta\varphi'_+(0). \end{cases}$$

Then (10.3) is equivalent to requiring that for all $k \in \hat{J}^0$ we have

$$\begin{cases} 0 < \mathcal{R}_k(h;\boldsymbol{u}) < \beta\varphi'_+(0) & \text{if} \quad \theta_k = -\beta\varphi'_+(0), \\ 0 < -\mathcal{R}_k(h;\boldsymbol{u}) < \beta\varphi'_+(0) & \text{if} \quad \theta_k = \beta\varphi'_+(0). \end{cases} \tag{10.7}$$

For each $k \in \hat{J}^0$ we next determine conditions ensuring that (10.7) is true.

• Upper inequalities in (10.7). From the definition of $\mathcal{R}_k$ in (10.6) we get

$$\begin{aligned} |\mathcal{R}_k(h;\boldsymbol{u})| &\le h\|\boldsymbol{a}_k\| + \|\boldsymbol{a}_k^T A_{\hat{J}}\| \left\| \boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y} + h\boldsymbol{u}) - \boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y}) \right\| \\ &\le ha + a\sqrt{\lambda} \left\| \boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y} + h\boldsymbol{u}) - \boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y}) \right\|. \end{aligned}$$

From (10.5) we see that $h \in]\,0, \sigma_0\,[$ guarantees that[7]

$$\left\| \boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y} + h\boldsymbol{u}) - \boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y}) \right\| < \frac{\beta\varphi'_+(0)}{2\,a\sqrt{\lambda}},$$

[7]We use the fact that $\dfrac{\beta\varphi'_+(0) - \theta^0_{\max}}{4\,a\sqrt{\lambda}} < \dfrac{\beta\varphi'_+(0)}{4\,a\sqrt{\lambda}} < \dfrac{\beta\varphi'_+(0)}{2\,a\sqrt{\lambda}}$.

and consequently

$$|\mathcal{R}_k(h;\boldsymbol{u})| < ha + \frac{\beta\varphi'_+(0)}{2}.$$

Now define

$$\zeta_1 := \min\left\{\frac{\beta\varphi'_+(0)}{2a},\ \sigma_0\right\}.$$

Since $\zeta_1 \le \left(\beta\varphi'_+(0)\right)/(2a)$ we see that

$$0 < h < \zeta_1 \ \text{ leads to } \ |\mathcal{R}_k(h;\boldsymbol{u})| < \zeta_1 a + \frac{\beta\varphi'_+(0)}{2} \le \beta\varphi'_+(0).$$

Hence, both upper inequalities (10.7) (relevant to $\theta_k < 0$ and to $\theta_k > 0$) are satisfied for any $\boldsymbol{u} \in \mathbb{I}^N$, provided that $0 < h < \zeta_1$.

• Lower inequalities in (10.7). By (b), the function $h \mapsto \boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y}+h\boldsymbol{u})$ allows a local Taylor expansion:

$$\boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y}+h\boldsymbol{u}) - \boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y}) = h\mathrm{D}\boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y})\boldsymbol{u} + \mathcal{P}_{\boldsymbol{u}}(h) \quad \text{if} \quad 0 \le h < \chi_{\boldsymbol{u}},$$

where $\chi_{\boldsymbol{u}} \in]\,0,\omega\,]$ is a bound and $\mathcal{P}_{\boldsymbol{u}}$ is the remainder. Then for each $k \in \hat{J}^0$ we have

$$\mathcal{R}_k(h;\boldsymbol{u}) = h\boldsymbol{a}_k^T\boldsymbol{u} - \boldsymbol{a}_k^T A_{\hat{J}}\left[h\mathrm{D}\boldsymbol{\mathcal{T}}^{\hat{J}}(\boldsymbol{y})\boldsymbol{u} + \mathcal{P}_{\boldsymbol{u}}(h)\right] = h\boldsymbol{\tau}_k^T\boldsymbol{u} - \boldsymbol{a}_k^T A_{\hat{J}}\mathcal{P}_{\boldsymbol{u}}(h),$$

where $\boldsymbol{\tau}_k$ was introduced in the assumption (c).

Fix $k \in \hat{J}^0$ and suppose $\boldsymbol{u} \in \mathcal{Q}_k$ with $\mathcal{Q}_k$ as given in (6.12). Two cases arise now.

—Case $\theta_k = -\beta\varphi'_+(0)$. By the assumption (b), the function $h \mapsto \mathcal{P}_{\boldsymbol{u}}(h)/h$ is continuous on $[\,0,\chi_{\boldsymbol{u}}\,[$ and $\lim_{h\downarrow 0}\mathcal{P}_{\boldsymbol{u}}(h)/h = \boldsymbol{0}$. By $\boldsymbol{u} \in \mathcal{Q}_k$ we get $\boldsymbol{\tau}_k^T\boldsymbol{u} > 0$; then there exists $\nu_k(\boldsymbol{u}) \in]\,0,\chi_{\boldsymbol{u}}]$ such that

$$\boldsymbol{\tau}_k^T\boldsymbol{u} > \boldsymbol{a}_k^T A_{\hat{J}}\frac{\mathcal{P}_{\boldsymbol{u}}(h)}{h} \ \text{whenever}\ 0 < h < \nu_k(\boldsymbol{u}).$$

For any such h we have $\mathcal{R}_k(h;\boldsymbol{u}) > 0$.

—Case $\theta_k = \beta\varphi'_+(0)$. Now $\boldsymbol{u} \in \mathcal{Q}_k$ means that $\boldsymbol{\tau}_k^T\boldsymbol{u} < 0$. The same considerations as previously yield a bound $\nu_k(\boldsymbol{u}) \in]\,0,\chi_{\boldsymbol{u}}\,]$ such that

$$\boldsymbol{\tau}_k^T\boldsymbol{u} < \boldsymbol{a}_k^T A_{\hat{J}}\frac{\mathcal{P}_{\boldsymbol{u}}(h)}{h} \ \text{ if } \ 0 < h < \nu_k(\boldsymbol{u}).$$

Hence $\mathcal{R}_k(h;\boldsymbol{u}) > 0$ for any $h \in]\,0,\nu_k(\boldsymbol{u})\,[$.

Now put

$$\nu(\boldsymbol{u}) := \min\left\{\nu_k(\boldsymbol{u}) \ \text{ for } \ k \in \hat{J}^0\right\}.$$

Hence, the inequality system (4.3) is strict for any $\boldsymbol{y}' = \boldsymbol{y} + h\boldsymbol{u}$ such that

$$\boldsymbol{u} \in \bigcap_{k\in\hat{J}^0}\mathcal{Q}_k \quad \text{and} \quad 0 < h < \zeta(\boldsymbol{u}),\ \text{with}\ \ \zeta(\boldsymbol{u}) = \min\{\xi_0,\ \zeta_1,\ \nu(\boldsymbol{u})\}.$$

All conditions of Theorem 5.2 are satisfied for the relevant $\tilde{\boldsymbol{t}}' = \tilde{\boldsymbol{\mathcal{T}}}(\boldsymbol{y}')$; hence these are strict minimizers of $\mathcal{F}_{\boldsymbol{y}'}$ at which $\boldsymbol{\mathcal{T}}$ is continuous.

The case $\hat{J} = \mathcal{S}$ is easily analyzed by using the arguments evoked in the proof of Theorem 6.1. □

Acknowledgments. The author wishes to express her gratitude to Sylvain Durand and Francois Malgouyres for carefully examining this paper. She also wishes to thank Alfred Hero for proposing other ramifications of this work, Marc Lavielle for his remarks, and Stéphane Chrétien for suggesting interesting references on nonsmooth optimization.

REFERENCES

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

[2] S. Alliney and S. A. Ruzinsky, *An algorithm for the minimization of mixed l_1 and l_2 norms with application to Bayesian estimation*, IEEE Trans. Signal Processing, SP-42 (1994), pp. 618–627.

[3] L. Alvarez, P.-L. Lions, and J.-M. Morel, *Image selective smoothing and edge detection by nonlinear diffusion.* II. SIAM J. Numer. Anal., 29 (1992), pp. 845–866.

[4] A. Avez, *Calcul différentiel*, Masson, Paris, 1991.

[5] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, New York, 1985.

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

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

[8] F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll, *Image selective smoothing and edge detection by nonlinear diffusion,* SIAM J. Numer. Anal., 29 (1992), pp. 182–193.

[9] A. Chambolle, *Image segmentation by variational methods: Mumford and Shah functional and the discrete approximation*, SIAM J. Appl. Math., 55 (1997), pp. 827–863.

[10] A. Chambolle and P.-L. Lions, *Image recovery via total variation minimization and related problems*, Numer. Math., 76 (1997), pp. 167–188.

[11] T. F. Chan and C. K. Wong, *Total variation blind deconvolution*, IEEE Trans. Image Processing, 7 (1998), pp. 370–375.

[12] S. Chretien and A. O. Hero, *Generalized Proximal Point Agorithms and Bundle Implementations*, Research Report CSPL 3139, Dept. of EECS, University of Michigan, Ann Arbor, MI, 1998.

[13] F. Clarke, *Optimization and Nonsmooth Analysis*, 2nd ed., Classics Appl. Math. 5, SIAM, Philadelphia, 1990.

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

[15] D. C. Dobson and F. Santosa, *Recovery of blocky images from noisy and blurred data*, SIAM J. Appl. Math., 56 (1996), pp. 1181–1199.

[16] D. Donoho, *De-noising by soft-thresholding*, IEEE Trans. Inform. Theory, 41 (1995), pp. 613–627.

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

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

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

[20] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms,* Vols. I and II, Springer-Verlag, Berlin, 1996.

[21] R. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.

[22] P. J. Huber, *Robust Statistics*, Wiley, New York, 1981.

[23] Y. Leclerc, *Constructing simple stable description for image partitioning*, Internat. J. Computer Vision, 3 (1989), pp. 73–102.

[24] C. Lemaréchal and C. Sagastizábal, *Practical aspects of the Moreau–Yosida regularization: Theoretical properties*, SIAM J. Optim., 7 (1997), pp. 367–385.

[25] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., 14 (1976), pp. 877–898.

[26] R. T. Rockafellar, *The Theory of Subgradients and Its Application to Problems of Optimization: Convex and Nonconvex Functions*, Helderman-Verlag, Berlin, 1981.

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

[28] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*, Springer-Verlag, New York, 1985.

[29] D. M. Titterington, *Common structure of smoothing techniques in statistics*, International Statistical Review, 53 (1985), pp. 141–170.

[30] C. R. Vogel, *Nonsmooth regularization*, in Inverse Problems in Geophysical Applications, H. W. Engl., A. K. Louis, and W. Rundell, eds., SIAM, Philadelphia, 1997, pp. 1–11.