Appl Math Optim OF1–OF19 (2006)
DOI: 10.1007/s00245-005-0843-0

Applied Mathematics & Optimization

Stability of the Minimizers of Least Squares with a Non-Convex Regularization. Part II: Global Behavior

S. Durand[1,2] and M. Nikolova[2]

[1]LAMFA UMR 6140, Université de Picardie,
33 rue Saint-Leu, 90039 Amien Cedex, France

[2]CMLA UMR 8536, ENS de Cachan,
61 av. du President Wilson, 94235 Cachan Cedex, France
{sdurand,nikolova}@cmla.ens-cachan.fr

Abstract. We address estimation problems where the sought-after solution is defined as the minimizer of an objective function composed of a quadratic data-fidelity term and a regularization term. We especially focus on non-convex and possibly non-smooth regularization terms because of their ability to yield good estimates. This work is dedicated to the stability of the minimizers of such piecewise $\mathcal{C}^m$, with $m \geq 2$, non-convex objective functions. It is composed of two parts. In the previous part of this work we considered general local minimizers. In this part we derive results on global minimizers. We show that the data domain contains an open, dense subset such that for every data point therein, the objective function has a finite number of local minimizers, and a unique global minimizer. It gives rise to a global minimizer function which is $\mathcal{C}^{m-1}$ everywhere on an open and dense subset of the data domain.

Key Words. Stability analysis, Regularized least squares, Non-smooth analysis, Non-convex analysis, Signal and image processing.

AMS Classification. 26B, 49J, 68U, 94A.

1. Introduction

This is the second part of a work devoted to the stability of minimizers of regularized least squares objective functions as customarily used in signal and image reconstruction.

In the previous part [7] we considered the behavior of local minimizers, whereas now we draw conclusions about global minimizers.

Given data $y \in \mathbb{R}^q$, we consider the global minimizers $\hat{x} \in \mathbb{R}^p$ of an objective function $\mathcal{E}\colon \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ of the form

$$\mathcal{E}(x, y) := \|Lx - y\|^2 + \Phi(x), \tag{1}$$

where $L\colon \mathbb{R}^p \to \mathbb{R}^q$ is a linear operator, $\|\cdot\|$ denotes the Euclidean norm and $\Phi\colon \mathbb{R}^p \to \mathbb{R}$ is a piecewise $\mathcal{C}^m$-smooth regularization term. More precisely,

$$\Phi(x) := \sum_{i=1}^{r} \varphi_i(G_i x), \tag{2}$$

where for every $i \in \{1, \ldots, r\}$, the function $\varphi_i\colon \mathbb{R}^s \to \mathbb{R}$ is continuous on $\mathbb{R}^s$ and $\mathcal{C}^m$-smooth everywhere except possibly at a given $\theta_i \in \mathbb{R}^s$, and $G_i\colon \mathbb{R}^p \to \mathbb{R}^s$ is a linear operator. The operators G_i in the regularization term Φ usually provide the differences between neighboring samples of x. Typically, for all $i \in \{1, \ldots, r\}$, we have $\theta_i = 0$ and φ_i reads

$$\varphi_i(z) = \phi(\|z\|), \qquad \forall i \in \{1, \ldots, r\}, \tag{3}$$

where $\phi\colon \mathbb{R}_+ \to \mathbb{R}$ is an increasing function, often called the potential function. Several examples, among the most popular, are the following [8], [1], [9], [11], [10], [13], [4], [14], [2]:

$$\begin{array}{ll}
L^\alpha & \phi(t) = |t|^\alpha, \qquad 1 \le \alpha \le 2, \\
\text{Lorentzian} & \phi(t) = \alpha t^2/(1 + \alpha t^2), \\
\text{Concave} & \phi(t) = \alpha|t|/(1 + \alpha|t|), \\
\text{Gaussian} & \phi(t) = 1 - \exp(-\alpha t^2), \\
\text{Truncated quadratic} & \phi(t) = \min\{\alpha t^2, 1\}, \\
\text{Huber} & \phi(t) = \begin{cases} t^2 & \text{if} \quad |t| \le \alpha, \\ \alpha(\alpha + 2|t - \alpha|) & \text{if} \quad |t| > \alpha. \end{cases}
\end{array} \tag{4}$$

The notations in this paper are the same as in Part I. Recall that although $\mathcal{E}$ depends on two variables (x, y), $\nabla\mathcal{E}$ and $\nabla^2\mathcal{E}$ will systematically be used to denote gradient and Hessian with respect to the first variable x. By $B(x, \rho)$ we denote a ball in $\mathbb{R}^n$ with radius ρ and center x, and by S the unit sphere in $\mathbb{R}^n$ centered at the origin, for whatever dimension n appropriate to the context. For a subset $A \in \mathbb{R}^q$, its complement in $\mathbb{R}^q$ will be denoted A^{c} and its closure $\overline{A}$.

We consider minimizer functions with special attention given to those which yield the global minimum of the objective function.

Definition 1. A function $\mathcal{X}\colon O \to \mathbb{R}^p$, where O is an open domain in $\mathbb{R}^q$, is said to be a minimizer function relevant to $\mathcal{E}$ if every $\mathcal{X}(y)$ is a strict (*i.e.* isolated) local minimizer of $\mathcal{E}(\cdot, y)$ whenever $y \in O$. Moreover, $\mathcal{X}$ is called a *global minimizer function* relevant to $\mathcal{E}$ if $\mathcal{E}(\cdot, y)$ reaches its global minimum at $\mathcal{X}(y)$ for every $y \in O$.

Our goal now is to check first the uniqueness and then the regularity of the global minimizer functions relevant to $\mathcal{E}$. We make the same basic assumptions as in the previous part of this work.

H1. *The operator* $L\colon \mathbb{R}^p \to \mathbb{R}^q$ *in* (1) *is injective*, *i.e*. $\operatorname{rank} L = p$.

If Φ is $\mathcal{C}^m$-smooth, we systematically assume the following:

H2. $\nabla\Phi(tv)/t \to 0$ *uniformly with* $v \in S$ *as* $t \to \infty$.

Otherwise, for Φ piecewise $\mathcal{C}^m$ and of the form (2), the latter assumption is reformulated in the following way:

H3. *For every* $i = 1, \ldots, r$ *and for* $t \in \mathbb{R}$, *we have* $\nabla\varphi_i(tu)/t \to 0$ *uniformly with* $u \in S^s$ *when* $t \to \infty$.

The results presented in the following are meaningful if, for all $y \in \mathbb{R}^q$, the objective function $\mathcal{E}(\cdot, y)$ admits at least one minimizer. It is easy to see that by assumptions H1–H3, $\mathcal{E}(\cdot, y)$ is coercive, hence the existence of a minimizer [5], [12]. However, $\mathcal{E}(\cdot, y)$ may have several global minimizers. From a practical point of view, this means that the estimation problem is not well formulated and that there is not enough information to pick out a unique stable solution. We confine our attention to the subset of $\mathbb{R}^q$ composed of data y for which the global minimum of $\mathcal{E}(\cdot, y)$ is reached at a unique point:

$$\Gamma := \{y \in \mathbb{R}^q\colon\ \mathcal{E}(\cdot, y) \text{ has a unique global minimizer}\}.$$

We show that the interior of Γ is dense in $\mathbb{R}^q$ and that its complement Γ^c has Lebesgue measure zero in $\mathbb{R}^q$. This means that in a real-world problem there is no chance of getting data y leading to an objective function having more than one global minimizer.

On Γ, we consider the *global minimizer function* $\hat{\mathcal{X}}\colon\ \Gamma \to \mathbb{R}^p$—the function which yields $\hat{\mathcal{X}}(y)$, the unique global minimizer of $\mathcal{E}(\cdot, y)$, for every $y \in \Gamma$. Under quite general assumptions, we show that $\hat{\mathcal{X}}$ *is* $\mathcal{C}^{m-1}$*-smooth on an open subset of* Γ *which is dense in* $\mathbb{R}^q$. The global minimizer function $\hat{\mathcal{X}}$ can also be extended beyond the latter set. However, this extension may not be defined in a unique way and it can be non-smooth and even discontinuous. An interesting intermediate result says that *for all* $y \in \mathbb{R}^q$, *except those contained in a closed subset of Lebesgue measure* 0, *the objective function has a finite number of local minimizers*, *each of them corresponding to a* $\mathcal{C}^{m-1}$ *local minimizer function*.

Example 1. Consider the function $\mathcal{E}\colon \mathbb{R} \to \mathbb{R}$,

$$\mathcal{E}(x, y) = (x - y)^2 + \phi(x), \qquad \text{for} \quad \phi(x) = \min\{x^2, 1\}.$$

The function ϕ above was already given in (4). For every $y \in \mathbb{R}$, $\mathcal{E}(\cdot, y)$ admits either one or two local minimizer functions. These are

$$\begin{aligned} \mathcal{X}_1(y) &= y/2 &&\text{if} \quad |y| \le 2,\\ \mathcal{X}_2(y) &= y &&\text{if} \quad |y| \ge 1. \end{aligned}$$

For $1 \le |y| \le 2$, $\mathcal{E}(\cdot, y)$ has two local minimizers. The set Γ reads

$$\Gamma = (-\infty, -\sqrt{2}) \cup (-\sqrt{2}, \sqrt{2}) \cup (\sqrt{2}, +\infty).$$

Clearly, Γ^c is closed and has Lebesgue measure zero in $\mathbb{R}$. The global minimizer function $\mathcal{X}$ is well defined on Γ and reads

$$\mathcal{X}(y) = \begin{cases} \mathcal{X}_1(y) & \text{if} \quad |y| < \sqrt{2}, \\ \mathcal{X}_2(y) & \text{if} \quad |y| > \sqrt{2}. \end{cases}$$

For $y \in \Gamma^c = \{-\sqrt{2}, \sqrt{2}\}$, there are two global minimizers, $\hat{x}_1 = \text{sign}(y)/\sqrt{2}$ and $\hat{x}_2 = \text{sign}(y)\sqrt{2}$. Although $\mathcal{X}$ is $\mathcal{C}^\infty$ on Γ, and $\mathcal{X}_1$ and $\mathcal{X}_2$ are $\mathcal{C}^\infty$ on a neighborhood of Γ^c, there is no continuous extension of $\mathcal{X}$ on $\overline{\Gamma} = \mathbb{R}$.

The regularity of the local minimizer functions does not generally imply that there is a regular global minimizer function, as required in Definition 1. Another example is the function

$$\mathcal{E}(x, y) = \big(|x-1| + \mathbb{1}(y \in \mathbb{Q})\big)\mathbb{1}(x \geq \tfrac{1}{2}) + \big(|x| + 1 - \mathbb{1}(y \in \mathbb{Q})\big)\mathbb{1}(x \leq \tfrac{1}{2}).$$

It has two local minimizer functions, $\mathcal{X}_1$ and $\mathcal{X}_2$, which are $\mathcal{C}^\infty$ on $\mathbb{R}$:

$$\mathcal{X}_1(y) = 1 \quad \text{and} \quad \mathcal{X}_2(y) = 0, \qquad \forall y \in \mathbb{R}.$$

There is a global minimizer function $\mathcal{X}$ defined on $\mathbb{R}$ but it is nowhere continuous since

$$\mathcal{X}(y) = \begin{cases} 1 & \text{if} \quad y \in \mathbb{Q}, \\ 0 & \text{if} \quad y \notin \mathbb{Q}. \end{cases}$$

An overview of the results on global minimizers for several classes of objective functions can be found in [6]. For some classes, existence and uniqueness of the global minimizers are shown to be a generic property—using our terminology, they hold for all data except those included in a countable union of rare sets. The stability of the global minimizer is studied in several ways, including the continuity of the optimal solution. The setting being quite abstract, the results presented there are difficult to use as a starting point for our work.

2. $\mathcal{C}^m$-Smooth Objective Function

The context of smooth objective functions allows us to see easily the main reasons yielding the result alluded to above.

Theorem 1. *Suppose $\mathcal{E}$ is of the form* (1) *where Φ is an arbitrary $\mathcal{C}^m$-function on $\mathbb{R}^p$, with $m \geq 2$. Let assumptions* H1 *and* H2 *be true. Then we have the following statements*:

(i) *Γ^c has Lebesgue measure zero in $\mathbb{R}^q$ and the interior of Γ is dense in $\mathbb{R}^q$.*
(ii) *The global minimizer function $\hat{\mathcal{X}}$: $\Gamma \to \mathbb{R}^p$ is $\mathcal{C}^{m-1}$ on an open subset of Γ which is dense in $\mathbb{R}^q$.*

The first part of statement (i) is a straightforward consequence of several well-known facts. Under assumption H2 or H3, it is easy to see that the minimum-value function

$y \to v(y) = \min_{x \in \mathbb{R}^p} \mathcal{E}(x, y)$ is locally Lipschitz. By the Rademacher theorem (e.g. see p. 403 of [12]), v is differentiable on a subset $\Gamma' \subset \mathbb{R}^q$ such that the Lebesgue measure of $\mathbb{R}^q \backslash \Gamma'$ is zero. Under H1, the Danskin theorem (see p. 275 of [3]) shows that v is differentiable at a point y if, and only if, $\mathcal{E}(\cdot, y)$ has a unique global minimizer. It follows that $\mathcal{E}(\cdot, y)$ has a unique global minimizer for every $y \in \Gamma'$.

The proof of the other statements uses two auxiliary propositions given below.

Proposition 1. *Suppose also that* Φ *is* $\mathcal{C}^m$ *and that assumptions* H1 *and* H2 *are true. Then there exists* Ω_0*—with* $\overline{\Omega_0^c}$ *of Lebesgue measure zero in* $\mathbb{R}^q$*—such that every* $y \in \Omega_0$ *is contained in a neighborhood* $N \in \mathbb{R}^q$, *associated with an integer* $n > 0$, *so that for every* $y' \in N$, *the relevant objective function* $\mathcal{E}(\cdot, y')$ *admits at most n local minimizers.*

Proof. The set Ω_0 evoked in the proposition can be taken as defined in (12) in Part I [7], namely

$$\Omega_0 = \{y \in \mathbb{R}^q \colon 2L^T y \notin \nabla\mathcal{E}(H_0, 0)\} \subset \Omega, \tag{5}$$

where

$$H_0 = \{x \in \mathbb{R}^p \colon \det \nabla^2 \mathcal{E}(x, 0) = 0\}. \tag{6}$$

As seen from the proof of Theorem 1 in Part I [7], the set $\overline{\Omega_0^c}$ has Lebesgue measure zero in $\mathbb{R}^q$. The proof of Proposition 1 relies on the following lemma.

Lemma 1. *Let the assumptions of Proposition* 1 *hold. Then for every bounded subset* $N \subset \mathbb{R}^q$, *there exists a compact set* $C \subset \mathbb{R}^p$ *such that for every* $y \in N$, *every local minimizer* $\hat{x}$ *of* $\mathcal{E}(\cdot, y)$ *satisfies* $\hat{x} \in C$.

Proof. All minimizers of all functions $\mathcal{E}(\cdot, y)$ when y ranges over N, are contained in the set

$$\{x \in \mathbb{R}^p \colon \nabla\mathcal{E}(x, 0) \in 2L^T N\}. \tag{7}$$

The set $2L^T N$ is clearly bounded. Moreover, by assumptions H1 and H2 we have $\nabla\mathcal{E}(x, 0) \sim 2L^T Lx$ as $\|x\| \to \infty$, where $2L^T L$ is invertible. Hence the set given in (7) is bounded as well. □

We will show that if for some $y \in \mathbb{R}^q$ the property stated in Proposition 1 is not satisfied, then $y \in \Omega_0^c$. So consider $y \in \mathbb{R}^q$ and suppose that for every integer $n > 0$, there exists $y_n \in B(y, 1/n)$ such that $\mathcal{E}(\cdot, y_n)$ admits at least n different local minimizers. This gives rise to a sequence, indexed by n, every element of which is a set of n minimizers among all the minimizers of $\mathcal{E}(\cdot, y_n)$. For every n, let d_n denote the smallest distance between two minimizers of $\mathcal{E}(\cdot, y_n)$ belonging to the selected set of n minimizers. The set being finite, the distance d_n is reached for a pair of minimizers, say $\hat{x}_n$ and $\hat{x}'_n$. Any such two minimizers satisfy

$$\nabla\mathcal{E}(\hat{x}'_n, y_n) = 0 = \nabla\mathcal{E}(\hat{x}_n, y_n). \tag{8}$$

By the mean-value theorem, there is $\tilde{x}_n \in \{t\hat{x}'_n + (1-t)\hat{x}_n\colon\ 0 < t < 1\}$ for which

$$\det \nabla^2 \mathcal{E}(\tilde{x}_n, y_n) = 0.$$

Since $\nabla^2 \mathcal{E}(\tilde{x}_n, y_n) = \nabla^2 \mathcal{E}(\tilde{x}_n, 0)$, we deduce that $\tilde{x}_n \in H_0$.

On the other hand, Lemma 1 tells us that all the minimizers of $\mathcal{E}(\cdot, y_n)$, for every n, are contained in the same compact set, whose convex hull is also compact and will be denoted C. Then $\tilde{x}_n \in C$ as well. By the compactness of C, the sequence $\{\tilde{x}_n\}$ admits a subsequence which converges to a point $\tilde{x}$ as long as $n \to \infty$. Moreover, C contains an increasing number (equal or larger than n) of minimizers when $n \to \infty$, so d_n goes to zero when $n \to \infty$. Hence, $\hat{x}_n \to \tilde{x}$ when $n \to \infty$. At the same time, $y_n \to y$ by construction. Since $(x, y) \to \nabla\mathcal{E}(x, y)$ is continuous, at the limit when $n \to \infty$, (8) yields

$$\nabla\mathcal{E}(\tilde{x}, y) = 0. \tag{9}$$

Moreover, since H_0 is closed, $\tilde{x} \in H_0$. Combining this with (9) shows that $y \in \Omega_0^{\mathrm{c}}$. □

Remark 1. Extending the arguments of this proof, we can see that local minimizer functions never cross on Ω_0. We consider two minimizer functions $\mathcal{X}_1$ and $\mathcal{X}_2$ defined on an open and connected domain $O \subset \Omega_0$. We claim that either $\mathcal{X}_1 \equiv \mathcal{X}_2$ on O, or

$$\mathcal{X}_1(y) \neq \mathcal{X}_2(y), \qquad \forall y \in O. \tag{10}$$

The reason is the following. Consider the set $\tilde{O} := \{y \in O\colon\ \mathcal{X}_1(y) = \mathcal{X}_2(y)\}$ and suppose that $\tilde{O}$ is non-empty and different from O. By the continuity of $\mathcal{X}_i,\ i = 1, 2$, the set $\tilde{O}$ is closed in O. Focus on y belonging to the boundary of $\tilde{O}$ in O. Then there is a sequence $\{y_n\}$ with $y_n \in O\backslash\tilde{O}$, converging to y as $n \to \infty$, such that $\mathcal{X}_1(y_n) \neq \mathcal{X}_2(y_n)$. Since $\mathcal{X}_1$ and $\mathcal{X}_2$ are continuous, the points $\hat{x}_n := \mathcal{X}_1(y_n)$ and $\hat{x}'_n := \mathcal{X}_2(y_n)$ come arbitrarily close to each other as long as $n \to \infty$. By applying the reasoning developed next to (8), we deduce that $\det \nabla^2\mathcal{E}(\mathcal{X}_i(y), y) = 0$, for $i = 1, 2$, which contradicts the fact that $y \in \Omega_0$. Hence the boundary of $\tilde{O}$ in O is empty. Since O is connected and open, this entails that either $\tilde{O} = O$ or $\tilde{O}$ is empty.

The proposition below reinforces this observation.

Proposition 2. *Let the assumptions of Proposition* 1 *be true. Every open set of* $\mathbb{R}^q$ *contains an open subset* O *on which* $\mathcal{E}$ *admits exactly* n *minimizer functions* $\mathcal{X}_i\colon O \to \mathbb{R}^p$, $i = 1, \ldots, n$, *which are* $\mathcal{C}^{m-1}$ *and are such that for all* $y \in O$, *all the local minimizers of* $\mathcal{E}(\cdot, y)$ *read*

$$\mathcal{X}_i(y), \qquad i = 1, \ldots, n, \tag{11}$$

and satisfy

$$\mathcal{E}(\mathcal{X}_i(y), y) \neq \mathcal{E}(\mathcal{X}_j(y), y), \qquad \forall i, j \in \{1, \ldots, n\} \quad \textit{with} \quad i \neq j. \tag{12}$$

Proof. Since Ω_0 is open and dense in $\mathbb{R}^q$, we can take our open set in Ω_0. Let y belong to this set. By Proposition 1, y has a neighborhood N composed of elements y' for which

$\mathcal{E}(\cdot, y')$ has at most n local minimizers, where $n > 0$ is the smallest integer for which this property holds. Even if it means interchanging two elements of N, we can assume that $\mathcal{E}(\cdot, y)$ has exactly n local minimizers $\hat{x}_i$, $i = 1, \ldots, n$. By $y \in \Omega_0 \subset \Omega$, each minimizer $\hat{x}_i$, $i = 1, \ldots, n$, results from the application of a $\mathcal{C}^m$ minimizer function $\mathcal{X}_i$, i.e. $\hat{x}_i = \mathcal{X}_i(y)$. Each $\mathcal{X}_i$ being defined on an open domain containing y, we can additionally restrict N in such a way that it is connected and included in the intersection of these domains.

Statement (11) holds for any $O \subset N$, because of the following two arguments. On the one hand, every $\mathcal{E}(\cdot, y')$, for $y' \in N$, has at most n minimizers. On the other hand, by Remark 1, for every $y' \in N$ and i, j with $i \neq j$, we have $\mathcal{X}_i(y') \neq \mathcal{X}_j(y')$.

The proof of (12) relies on the following lemma.

Lemma 2. *Let $\mathcal{X}_1$ and $\mathcal{X}_2$ be two differentiable (local) minimizer functions relevant to $\mathcal{E}$, defined on the same open domain $\tilde{O} \subset \Omega$. Suppose we have*

$$\mathcal{E}(\mathcal{X}_1(y), y) = \mathcal{E}(\mathcal{X}_2(y), y), \qquad \forall y \in \tilde{O}. \tag{13}$$

Then

$$\mathcal{X}_1(y) = \mathcal{X}_2(y), \qquad \forall y \in \tilde{O}.$$

Proof. By differentiating both sides of (13) with respect to y, we obtain

$$\begin{aligned} &D_1\mathcal{E}(\mathcal{X}_1(y), y)\, D\mathcal{X}_1(y) + D_2\mathcal{E}(\mathcal{X}_1(y), y) \\ &\quad = D_1\mathcal{E}(\mathcal{X}_2(y), y)\, D\mathcal{X}_2(y) + D_2\mathcal{E}(\mathcal{X}_2(y), y), \end{aligned} \tag{14}$$

where $D_i\mathcal{E}$ denotes the differential of $\mathcal{E}$ with respect to its ith argument—thus $D_1\mathcal{E} = (\nabla\mathcal{E})^T$—and $D\mathcal{X}_i$ is the Jacobian matrix of $\mathcal{X}_i$. Since, for $i \in \{1, 2\}$, $\mathcal{X}_i$ is a minimizer function,

$$D_1\mathcal{E}(\mathcal{X}_i(y), y) = 0, \qquad \forall y \in \tilde{O}.$$

On the other hand, differentiating $\mathcal{E}(x, y)$ in (1) with respect to y leads to

$$D_2\mathcal{E}(x, y) = 2Lx - 2y. \tag{15}$$

Introducing these last two expressions in (14), shows that

$$L\mathcal{X}_1(y) = L\mathcal{X}_2(y), \qquad \forall y \in \tilde{O}.$$

The conclusion follows from the injectivity of L. □

We now pursue the proof of (12). For all $i, j \in \{1, \ldots, n\}$ with $i \neq j$ we consider

$$N_{i,j} := \{y \in N\colon\ \mathcal{E}(\mathcal{X}_i(y), y) = \mathcal{E}(\mathcal{X}_j(y), y)\}.$$

Then introduce the subset

$$O := N\setminus\left(\bigcup_{i,j\in\{1,\dots,n\}} N_{i,j}\right).$$

Equivalently,

$$O = \{y \in N\colon\ \mathcal{E}(\mathcal{X}_i(y), y) \neq \mathcal{E}(\mathcal{X}_j(y), y),\ \forall i, j \in \{1, \dots, n\} \text{ with } i \neq j\}.$$

Since, for every $i = 1, \dots, n$, the function $y \to \mathcal{E}(\mathcal{X}_i(y), y)$ is continuous on N, every $N_{i,j}$ is closed in N. By the same argument, O is open.

Suppose O is empty. Since N is open, its interior is non-empty. Hence, there exist $i, j \in \{1, \dots, n\}$ for which the interior of $N_{i,j}$ is also non-empty. Associating $\mathcal{X}_1$, $\mathcal{X}_2$ and $\tilde{O}$ of Lemma 2 with $\mathcal{X}_i$, $\mathcal{X}_j$ and the interior of $N_{i,j}$, respectively, we obtain that $\mathcal{X}_i = \mathcal{X}_j$ on this interior. This contradicts the fact that $\mathcal{X}_i(y) \neq \mathcal{X}_j(y)$, for all $y \in N$. It follows that O is non-empty. The second statement of the proposition is proven. □

Proof of Theorem 1. The statement on the measure of Γ^c was considered earlier. The proof of all other statements follows directly from Proposition 2. Actually, we show a stronger result, namely that these statements remain true if we replace Γ by

$$\Gamma_0 := \left\{ y \in \Omega_0\colon \begin{array}{l}\text{every local minimum of } \mathcal{E}(\cdot, y) \text{ is}\\ \text{reached for a unique local minimizer}\end{array} \right\} \subset \Gamma.$$

So, Γ_0 is the set of all $y \in \Omega_0$ for which $\mathcal{E}(\cdot, y)$ reaches a different value $\mathcal{E}(\hat{x}_i, y)$ at each local minimizer $\hat{x}_i$. Hence the uniqueness of the global minimizer, i.e. $\Gamma_0 \subset \Gamma$.

Let $y \in \mathbb{R}^q$ and consider a neighborhood of y in $\mathbb{R}^q$. By Proposition 2, it contains an open set O on which the conclusion of the proposition holds. Clearly, O belongs to the interior of Γ_0. Since O can be arbitrarily close to y, we have proved that the interior of Γ_0 is dense in $\mathbb{R}^q$.

We now consider an arbitrary $y' \in O$. By (12), there is an index $i \in \{1, \dots, n\}$ for which

$$\mathcal{E}(\mathcal{X}_i(y'), y') < \mathcal{E}(\mathcal{X}_j(y'), y'), \qquad \forall j \in \{1, \dots, n\}\setminus\{i\}.$$

As the functions $y'' \to \mathcal{E}(\mathcal{X}_j(y''), y'')$ are continuous on O, there is a neighborhood $N \subset O$ of y' such that

$$\mathcal{E}(\mathcal{X}_i(y''), y'') < \mathcal{E}(\mathcal{X}_j(y''), y''), \qquad \forall y'' \in N, \quad \forall j \in \{1, \dots, n\}\setminus\{i\}.$$

Therefore $\hat{\mathcal{X}} = \mathcal{X}_i$ on N which implies that N belongs to the interior of $\{y'' \in \Gamma\colon\ \hat{\mathcal{X}} \text{ is } \mathcal{C}^{m-1} \text{ at } y''\}$. Noticing that N can be arbitrarily close to y yields the conclusion. □

Remark 2 (Extension to Mumford–Shah-like Regularization). In order to pursue the ideas presented in Remark 3 in [7], we focus again on objective functions involving non-smooth regularization of the kind of the Mumford–Shah functional. Let us recall that these correspond to (2)–(3) where ϕ is such that

$$\phi'(\tau^-) > \phi'(\tau^+),$$

for some constant $\tau > 0$, and ϕ is $\mathcal{C}^m$, $m \geq 2$, on $\mathbb{R}_+ \backslash \{\tau\}$, with $\phi'(0) = 0$. It can be seen that the statements of the theorem remain true in this context as well. The key point is that minimizers are in $\mathbb{R}^p \backslash M$, where $M = \bigcup_{i=1}^r \{x \in \mathbb{R}^p\colon \|G_i x\| = \tau\}$ is a union of manifolds of dimension $\leq p-1$, and that $\mathcal{E}(\cdot, y)$ is $\mathcal{C}^m$ on $\mathbb{R}^p \backslash M$. One can see that Proposition 1, Remark 1 and Lemma 2 can be extended to this case. The proof of Proposition 2 will remain the same.

For the truncated quadratic function—see (4)—the arguments are much simpler. It can be shown that each (local) minimizer functions $\mathcal{X}_i$ corresponds with a subset $J_i \subset \{1, \ldots, p\}$ such that

$$\mathcal{X}_i(y) \left(2L^T L + \sum_{k \in J_i} G_k^T G_k\right)^{-1} 2L^T y.$$

Each minimizer function is hence linear with respect to y whereas the minimum value $\mathcal{E}(\mathcal{X}_i(y), y)$ is a second degree polynomial with respect to y.

3. Objective Function Involving Non-Smooth Regularization

We now consider regularization terms of the form (2) where for every $i \in \{1, \ldots, r\}$, the function $\varphi_i\colon \mathbb{R}^s \to \mathbb{R}$ is continuous on $\mathbb{R}^s$ and $\mathcal{C}^m$ on $\mathbb{R}^s \backslash \{\theta_i\}$ for a given $\theta_i \in \mathbb{R}^s$, and $G_i\colon \mathbb{R}^p \to \mathbb{R}^s$ is a linear operator. For every $i \in \{1, \ldots, r\}$, the function φ_i is supposed to satisfy the same conditions as in Part I [7]:

H4. *For every net* $h \in \mathbb{R}^s$ *converging to* 0 *and such that* $\lim_{h\to 0} \mathcal{N}(h)$ *exists, the limit* $\lim_{h\to 0} \nabla\varphi_i(\theta_i + h)$ *exists and depends only on* $\lim_{h\to 0} \mathcal{N}(h)$.

In the expression above, $\mathcal{N}$ denotes the normalization mapping defined by $\mathcal{N}(v) = v/\|v\|$, for every vector v. We put again

$$\nabla^+\varphi_i(\theta_i)\left(\lim_{h\to 0} \mathcal{N}(h)\right) := \lim_{h\to 0} \nabla\varphi_i(\theta_i + h), \tag{16}$$

and then extend this definition to every $u \in \mathbb{R}^s$,

$$\nabla^+\varphi_i(\theta_i)(u) = \begin{cases} \nabla^+\varphi_i(\theta_i)\,(\mathcal{N}(u)) & \text{if} \quad u \neq 0, \\ 0 & \text{if} \quad u = 0. \end{cases} \tag{17}$$

Recall that for every $i \in \{1, \ldots, r\}$, we also have the assumptions:

H5. $u \to \nabla^+\varphi_i(\theta_i)(u)$ *is Lipschitz on* S^s.

H6. $u \mapsto \nabla\varphi_i(\theta_i + hu)$ *converges to* $\nabla^+\varphi_i(\theta_i)$ *as* $h \searrow 0$, *uniformly on* S^s.

We need two additional assumptions which are usually satisfied in practice. For all $i \in \{1, \dots, r\}$, we assume that

H7. $\liminf_{z\to\theta_i} \inf_{v\in S^s} v^T \nabla^2 \varphi_i(z) v > -\infty$

and

H8. $u^T \nabla^+ \varphi_i(\theta_i)(u) \geq u^T \nabla^+ \varphi_i(\theta_i)(v)$, $\forall u \in S^s$ *and* $\forall v \in S^s$.

Observe that by the definition of $\nabla^+\varphi_i$ in (17), the inequality in assumption H8 can be extended to all u and v in $\mathbb{R}^s$.

Example 2. To illustrate the two last assumptions, consider

$$\varphi_i(z) = \phi(\|z-\theta_i\|) \qquad \text{for} \quad z \in \mathbb{R}^s,$$

where $\phi \in \mathcal{C}^m(\mathbb{R}_+)$, $m \geq 2$, and $\phi'(0) > 0$. By applying (16)–(17), it becomes

$$\begin{cases} \nabla\varphi_i(z) = \phi'(\|z-\theta_i\|)\dfrac{z-\theta_i}{\|z-\theta_i\|} & \text{if} \quad z \neq \theta_i, \\ \nabla^+\varphi_i(\theta_i)(u) = \phi'(0^+)\dfrac{u}{\|u\|} & \text{if} \quad z = \theta_i. \end{cases}$$

Differentiating $\nabla\varphi_i$ for $z \neq \theta_i$, we obtain

$$\nabla^2\varphi_i(z) = \frac{\phi'(\|z-\theta_i\|)}{\|z-\theta_i\|}\, I + \left(\phi''(\|z-\theta_i\|) - \frac{\phi'(\|z-\theta_i\|)}{\|z-\theta_i\|}\right)\mathcal{N}(z-\theta_i)(\mathcal{N}(z-\theta_i))^T.$$

For any $v \in S^s$, we have

$$\begin{aligned} v^T\nabla^2\varphi_i(z)v &= \frac{\phi'(\|z-\theta_i\|)}{\|z-\theta_i\|} + \left(\phi''(\|z-\theta_i\|) - \frac{\phi'(\|z-\theta_i\|)}{\|z-\theta_i\|}\right)(v^T\mathcal{N}(z-\theta_i))^2 \\ &= \frac{\phi'(\|z-\theta_i\|)}{\|z-\theta_i\|}(1-(v^T\mathcal{N}(z-\theta_i))^2) + \phi''(\|z-\theta_i\|)(v^T\mathcal{N}(z-\theta_i))^2. \end{aligned}$$

The first term is always positive. So assumption H7 amounts to saying that

$$\liminf_{t\searrow 0} \phi''(t) > -\infty.$$

For instance, for the concave function in (4), we find that $\phi''(0) = -2\alpha$. For $\alpha = 1$, the L^α-function is non-smooth at zero and we have $\phi'(0) = 1$ and $\phi''(0) = 0$.

Furthermore, the inequality required in assumption H8 reads

$$\phi'(0^+) \geq \phi'(0^+)\, u^T v, \qquad \forall u, v \in S^s,$$

which amounts to the Schwarz inequality.

The developments in the case of piecewise $\mathcal{C}^m$ regularization follow the same lines as those relevant to $\mathcal{C}^m$-functions and some details can therefore be skipped. The next theorem is an extension of Theorem 1 and gives the main result of this paper.

Theorem 2. *Consider $\mathcal{E}$ represented by* (1) *where Φ has the form* (2). *For all $i \in \{1, \dots, r\}$, let φ_i be $\mathcal{C}^m$ on $\mathbb{R}\backslash\{\theta_i\}$ with $m \geq 2$ and continuous at θ_i and let assumptions* H3–H8 *be true. Suppose that* H1 *is satisfied. Then*:

- (i) Γ^c *has Lebesgue measure zero in $\mathbb{R}^q$ and the interior of Γ is dense in $\mathbb{R}^q$.*
- (ii) *The global minimizer function $\hat{\mathcal{X}}\colon \Gamma \to \mathbb{R}^p$ is $\mathcal{C}^{m-1}$ on an open subset of Γ which is dense in $\mathbb{R}^q$.*

The proof of the statement on the measure of Γ^c is the same as in the smooth case considered in Theorem 1. The proof of the other statements relies on the two propositions given below.

Proposition 3. *Let Φ have the form* (2) *and let assumptions* H1, H3, H4, H7 *and* H8 *be true. Then there exists Ω_0 whose interior is dense in $\mathbb{R}^q$ such that every $y \in \Omega_0$ is contained in a neighborhood $N \in \mathbb{R}^q$, associated with an integer $n > 0$, so that for every $y' \in N$, the relevant objective function $\mathcal{E}(\cdot, y')$ admits at most n local minimizers.*

Proof. Given $J \in \mathcal{P}(\{1, \dots, r\})$, let Θ_J be the manifold

$$\Theta_J := \left\{ x \in \mathbb{R}^p \colon \begin{bmatrix} G_i x = \theta_i \text{ for all } i \in J \\ G_i x \neq \theta_i \text{ for all } i \in J^c \end{bmatrix} \right\},$$

and let T_J be its tangent. Let Π_{T_J} be the orthogonal projection onto T_J. Similarly to [7], we define

$$H_0^J := \{x \in \Theta_J \colon \det \nabla^2(\mathcal{E}|_{\Theta_J})(x, 0) = 0\}, \tag{18}$$

$$W_J := \left\{ w \in T_J^{\perp} \colon v^T w \leq \sum_{i \in J} v^T G_i^T \nabla^+ \varphi_i(\theta_i)(G_i v),\ \forall v \in T_J^{\perp} \right\}. \tag{19}$$

The set Ω_0 is now constructed in close relation with Corollary 1 in Part I [7]:

$$\Omega_0 := \bigcap_{J \subset \mathcal{P}(\{1,\dots,r\})} (A_J^c \cap B_J^c) \subset \Omega, \tag{20}$$

where we recall that

$$A_J := \{y \in \mathbb{R}^q \colon 2\Pi_{T_J} L^T y \in \nabla(\mathcal{E}|_{\Theta_J})(H_0^J, 0)\}, \tag{21}$$

$$B_J := \{y \in \mathbb{R}^q \colon 2L^T y \in \nabla\mathcal{E}_J(\Theta_J, 0) + \partial_{T_J^{\perp}} W_J\}, \tag{22}$$

and $\partial_{T_J^{\perp}} W_J$ is the boundary of W_J considered in $T_J^{\perp}$. As seen from Propositions 2 and 3 in Part I [7], the interiors of the sets A_J^c and B_J^c are dense in $\mathbb{R}^q$. Hence the interior of Ω_0 is dense in $\mathbb{R}^q$ as well. Next we need a lemma which generalizes Lemma 1 in Section 2.

Lemma 3. *Let Φ be as in* (2) *and let assumptions* H1 *and* H3 *hold. Then for every open and bounded set $N \subset \mathbb{R}^q$ there exists a compact set $C \subset \mathbb{R}^p$ such that for every $y \in N$, every local minimizer $\hat{x}$ of $\mathcal{E}(\cdot, y)$ satisfies $\hat{x} \in C$.*

Proof. Let $\hat{x} \in \Theta_J$ be a minimizer of $\mathcal{E}(\cdot, y)$. Then we can write that

$$\nabla(\mathcal{E}|_{\Theta_J})(\hat{x}, y) = 0.$$

Equivalently,

$$\nabla(\mathcal{E}|_{\Theta_J})(\hat{x}, 0) = 2\Pi_{T_J} L^T y.$$

Then all minimizers of all functions $\mathcal{E}(\cdot, y)$ corresponding to $y \in N$ are contained in the set

$$\bigcup_{J \in \mathcal{P}(\{1,\dots,n\})} \left\{ x \in \Theta_J \colon \nabla(\mathcal{E}|_{\Theta_J})(x, 0) \in 2\Pi_{T_J} L^T N \right\}.$$

Each one of the sets composing this union is bounded because $2L^T N$ is bounded and $x \to \|\nabla(\mathcal{E}|_{\Theta_J})(x, 0)\|$ is coercive due to H1 and H3. Hence their union is bounded as well. □

Below we develop the proof of Proposition 3. Similarly to Proposition 1, we shall show that if $y \in \mathbb{R}^q$ does not satisfy the conclusion, then $y \notin \Omega_0$. Consider therefore a point $y \in \mathbb{R}^q$ such that for every integer $n > 0$, there is a point $y_n \in B(y, 1/n)$ for which $\mathcal{E}(\cdot, y_n)$ has at least n different local minimizers. This gives rise to a sequence, indexed by n, every element of which is a set of n minimizers among all the minimizers of $\mathcal{E}(\cdot, y_n)$. Notice that for every J, the set Θ_J is composed of a finite number of convex subsets. For instance, we can consider the following decomposition:

$$\begin{aligned}
\Theta_J &= \left\{ x \in \mathbb{R}^p \colon G_i x = \begin{bmatrix} \theta_i, \ \forall i \in J \\ G_k x \neq \theta_k, \ \forall k \in J^c \end{bmatrix} \right\} \\
&= \left\{ x \in \mathbb{R}^p \colon G_i x = \begin{bmatrix} \theta_i, \ \forall i \in J \\ \forall k \in J^c, \ \exists j_k \in \{1, \dots, s\} \text{ such that } [G_k x]_{j_k} \neq [\theta_k]_{j_k} \end{bmatrix} \right\} \\
&= \bigcup_{\{j_k\} \in \{1,\dots,s\}^{J^c}} \ \bigcup_{\lambda \in \{-1,1\}} \left\{ x \in \mathbb{R}^p \colon G_i x = \begin{bmatrix} \theta_i, \ \forall i \in J \\ \lambda [G_k x - \theta_k]_{j_k} > 0, \ \forall k \in J^c \end{bmatrix} \right\},
\end{aligned}$$

where for a vector z, $[z]_k$ denotes its kth entry. Using also the fact that $\mathcal{P}(\{1, \dots, n\})$ is finite, it is easy to see that there exist a set J of indexes and a subsequence of $\{y_n\}$, denoted by $\{y_n\}$ again, such that for every integer $n > 0$, the function $\mathcal{E}(\cdot, y_n)$ has at least n local minimizers belonging to the same convex subset $\tilde{\Theta}_J$ of Θ_J. Using the same arguments as in the proof of Proposition 1, we see that there are two convergent subsequences of local minimizers of $\mathcal{E}(\cdot, y_n)$ in $\tilde{\Theta}_J$, say $\{\hat{x}_n\}$ and $\{\hat{x}'_n\}$ such that the distance between them $\|\hat{x}_n - \hat{x}'_n\|$ goes to zero as long as $n \to \infty$. Similarly, the convexity of $\tilde{\Theta}_J$ allows the mean-value theorem to be applied. Then we see that there exists $\tilde{x}_n \in \{t\hat{x}'_n + (1-t)\hat{x}_n\colon\ 0 < t < 1\}$ for which

$$(\hat{x}_n - \hat{x}'_n)^T \nabla^2(\mathcal{E}|_{\Theta_J})(\tilde{x}_n, y_n)(\hat{x}_n - \hat{x}'_n) = 0. \tag{23}$$

As $\|\hat{x}_n - \hat{x}'_n\| \to 0$ when $n \to \infty$, all the three sequences, $\{\hat{x}_n\}$, $\{\hat{x}'_n\}$ and $\{\tilde{x}_n\}$, converge to the same point $\tilde{x}$ whereas $y_n \to y$ by construction. Now, two situations can occur according to the position of $\tilde{x}$. These are considered in Lemmas 4 and 5 below.

Lemma 4. *Suppose that $\tilde{x} \in \Theta_J$. Then $y \in A_J \subset \Omega_0^{\mathrm{c}}$.*

Proof. Returning to the definitions of A_J and H_0^J, we have to show that $\nabla(\mathcal{E}|_{\Theta_J})(\tilde{x}, y) = 0$ and $\det \nabla^2(\mathcal{E}|_{\Theta_J})(\tilde{x}, y) = 0$. As to the gradient, the continuity of the function

$$(x, y) \to \nabla(\mathcal{E}|_{\Theta_J})(x, y) = \Pi_{T_J}\left(2L^T(Lx - y) + \sum_{i \in J^{\mathrm{c}}} G_i^T \nabla\varphi_i(G_i x)\right)$$

on $\Theta_J \times \mathbb{R}^q$ entails that

$$\nabla(\mathcal{E}|_{\Theta_J})(\tilde{x}, y) = \lim_{n\to\infty} \nabla(\mathcal{E}|_{\Theta_J})(\hat{x}_n, y_n) = 0.$$

We now check that $\nabla^2(\mathcal{E}|_{\Theta_J})(\tilde{x}, y)$ is semi-positive definite. Since every $\hat{x}_n$ is a local minimizer of $\mathcal{E}(\cdot, y_n)$ and $\hat{x}_n \in \Theta_J$, it is a local minimizer of $\left(\mathcal{E}|_{\Theta_J}\right)(\cdot, y_n)$. Then

$$v^T \nabla^2(\mathcal{E}|_{\Theta_J})(\hat{x}_n, y_n)\, v \geq 0, \qquad \forall v \in T_J.$$

The continuity of the function

$$(x, y) \to \nabla^2(\mathcal{E}|_{\Theta_J})(x, y) = \Pi_{T_J}\left(2L^T L + \sum_{i \in J^{\mathrm{c}}} G_i^T \nabla^2\varphi_i(G_i x) G_i\right)\Pi_{T_J}^T$$

shows that at the limit when $n \to \infty$,

$$v^T \nabla^2(\mathcal{E}|_{\Theta_J})(\tilde{x}, y)\, v \geq 0, \qquad \forall v \in T_J.$$

Yet consider subsequences of $\{\hat{x}_n\}$ and $\{\hat{x}'_n\}$ such that $\{\mathcal{N}(\hat{x}_n - \hat{x}'_n)\}$ converges, and denote

$$u := \lim_{n\to\infty} \mathcal{N}(\hat{x}_n - \hat{x}'_n).$$

The facts that $\hat{x}_n$ and $\hat{x}'_n$ are in T_J, for every n, shows that $u \in T_J$. Next we divide (23) by $\|\hat{x}_n - \hat{x}'_n\|^2 \neq 0$ and take the limit when $n \to \infty$. This yields

$$u^T \nabla^2(\mathcal{E}|_{\Theta_J})(\tilde{x}, y)\, u = 0.$$

It follows that $\det \nabla^2\left(\mathcal{E}|_{\Theta_J}\right)(\tilde{x}, y) = 0$. Hence the result. □

The other possibility is that $\tilde{x}$ belongs to the boundary of Θ_J in $\overline{\Theta_J}$, which means that $\tilde{x} \in \Theta_{\tilde{J}}$ with $\tilde{J} \supset J$, $\tilde{J} \neq J$.

Lemma 5. *Suppose that $\tilde{x} \in \Theta_{\tilde{J}}$. Then $\nabla(\mathcal{E}|_{\Theta_{\tilde{J}}})(\tilde{x}, y) = 0$.*

Proof. By $\tilde{J} \supset J$, we have $T_{\tilde{J}} \subset T_J$, and hence $\Pi_{T_{\tilde{J}}} \circ \Pi_{T_J} = \Pi_{T_{\tilde{J}}}$. This allows us to write

$$\begin{aligned}
&\Pi_{T_{\tilde{J}}} \nabla \left(\mathcal{E}|_{\Theta_J}\right)(\hat{x}_n, y_n) \\
&\quad = \Pi_{T_{\tilde{J}}} \circ \Pi_{T_J} \left(2L^T(L\hat{x}_n - y_n) + \sum_{i \in J^c} G_i^T \nabla \varphi_i(G_i \hat{x}_n) \right) \\
&\quad = \Pi_{T_{\tilde{J}}} \left(2L^T(L\hat{x}_n - y_n) + \sum_{i \in \tilde{J}^c} G_i^T \nabla \varphi_i(G_i \hat{x}_n) \right) \\
&\qquad + \sum_{i \in \tilde{J} \setminus J} \Pi_{T_{\tilde{J}}} G_i^T \nabla \varphi_i(G_i \hat{x}_n).
\end{aligned}$$

Since $\Pi_{T_{\tilde{J}}} G_i^T = 0$, $\forall i \in \tilde{J}$, the last term above vanishes, hence

$$\Pi_{T_{\tilde{J}}} \nabla \left(\mathcal{E}|_{\Theta_J}\right)(\hat{x}_n, y_n) = \Pi_{T_{\tilde{J}}} \left(2L^T(L\hat{x}_n - y_n) + \sum_{i \in \tilde{J}^c} G_i^T \nabla \varphi_i(G_i \hat{x}_n) \right).$$

The obtained function is continuous with respect to $(\hat{x}_n, y_n) \in \{x \in \mathbb{R}^p \colon G_i x \neq \theta_i,\ \forall i \in \tilde{J}^c\} \times \mathbb{R}^q$. Since $\nabla\left(\mathcal{E}|_{\Theta_J}\right)(\hat{x}_n, y_n) = 0$, $\forall n$, at the limit when $n \to \infty$ we get

$$\Pi_{T_{\tilde{J}}} \left(2L^T(L\tilde{x} - y) + \sum_{i \in \tilde{J}^c} G_i^T \nabla \varphi_i(G_i \tilde{x}) \right) = 0.$$

This completes the proof. □

Hence, $\tilde{x}$ satisfies the necessary condition for a local minimum of $\mathcal{E}|_{\Theta_{\tilde{J}}}$. Next we exhibit a direction $u \in T_J$ which shows that either $y \in A_{\tilde{J}}$ or $y \in B_{\tilde{J}}$. As above, we take a convergent subsequence of $\mathcal{N}(\hat{x}_n - \hat{x}'_n)$ and consider

$$u := \lim_{n \to \infty} \mathcal{N}(\hat{x}_n - \hat{x}'_n). \tag{24}$$

Since $\hat{x}_n \in \Theta_J$ and $\hat{x}'_n \in \Theta_J$, $\forall n$, we see that $u \in T_J$. Two cases now arise which are considered in the two following lemmas.

Lemma 6. *Suppose that $\tilde{x} \in \Theta_{\tilde{J}}$ and $u \in T_{\tilde{J}}$. Then $y \in A_{\tilde{J}} \subset \Omega_0^c$.*

Proof. By developing (23) and dividing by $\|\hat{x}_n - \hat{x}'_n\|^2 \neq 0$, we obtain

$$\begin{aligned}
&\mathcal{N}(\hat{x}_n - \hat{x}'_n)^T 2L^T L \mathcal{N}(\hat{x}_n - \hat{x}'_n) \\
&\quad + \sum_{i \in J^c} \mathcal{N}(\hat{x}_n - \hat{x}'_n)^T G_i^T \nabla^2 \varphi_i(G_i \tilde{x}_n) G_i \mathcal{N}(\hat{x}_n - \hat{x}'_n) = 0.
\end{aligned}$$

Noticing that $J^c = \tilde{J}^c \cup (\tilde{J}\setminus J)$, we put the last equation into the form

$$\mathcal{N}(\hat{x}_n - \hat{x}'_n)^T 2L^T L \mathcal{N}(\hat{x}_n - \hat{x}'_n) \tag{25}$$

$$+ \sum_{i\in\tilde{J}^c} \mathcal{N}(\hat{x}_n - \hat{x}'_n)^T G_i^T \nabla^2 \varphi_i(G_i\tilde{x}_n) G_i \mathcal{N}(\hat{x}_n - \hat{x}'_n) \tag{26}$$

$$= - \sum_{i\in\tilde{J}\setminus J} \mathcal{N}(\hat{x}_n - \hat{x}'_n)^T G_i^T \nabla^2 \varphi_i(G_i\tilde{x}_n) G_i \mathcal{N}(\hat{x}_n - \hat{x}'_n). \tag{27}$$

We consider separately the limit of (25)–(26) and (27) when $n \to \infty$. Noticing that $G_i\mathcal{N}(\hat{x}_n - \hat{x}'_n) \to G_i u$, (25)–(26) becomes

$$u^T 2L^T L u + \sum_{i\in\tilde{J}^c} u^T G_i^T \nabla^2 \varphi_i(G_i\tilde{x}) G_i u = u^T \nabla^2 \left(\mathcal{E}|_{\Theta_{\tilde{J}}}\right)(\tilde{x}, 0)\, u.$$

Notice that for every n, the point $\hat{x}_n \in \Theta_{\tilde{J}}$ is a local minimizer of $\mathcal{E}(\cdot, y_n)$, and so it is a minimizer of $\mathcal{E}|_{\Theta_{\tilde{J}}}(\cdot, y_n)$ as well. Consequently,

$$v^T \nabla^2 \left(\mathcal{E}|_{\Theta_{\tilde{J}}}\right)(\hat{x}_n, 0)\, v = v^T \nabla^2 \left(\mathcal{E}|_{\Theta_{\tilde{J}}}\right)(\hat{x}_n, y_n)\, v \geq 0, \qquad \forall n,\ \forall v \in T_{\tilde{J}}.$$

As $n \to \infty$,

$$v^T \nabla^2 \left(\mathcal{E}|_{\Theta_{\tilde{J}}}\right)(\tilde{x}, 0)\, v \geq 0, \qquad \forall v \in T_{\tilde{J}}. \tag{28}$$

In particular, for $v = u$ we deduce that (25)–(26) has a positive limit which is

$$u^T \nabla^2 \left(\mathcal{E}|_{\Theta_{\tilde{J}}}\right)(\tilde{x}, 0)\, u \geq 0. \tag{29}$$

We now examine the upper bound of (27) as $n \to \infty$. Using the identity $G_i\mathcal{N}(\hat{x}_n - \hat{x}'_n) = \|G_i\mathcal{N}(\hat{x}_n - \hat{x}'_n)\|\, \mathcal{N}\left(G_i(\hat{x}_n - \hat{x}'_n)\right)$, we obtain

$$\begin{aligned} &\mathcal{N}(\hat{x}_n - \hat{x}'_n)^T G_i^T \nabla^2 \varphi_i(G_i\tilde{x}_n) G_i \mathcal{N}(\hat{x}_n - \hat{x}'_n) \\ &\quad = \|G_i\mathcal{N}(\hat{x}_n - \hat{x}'_n)\|^2\, \mathcal{N}\left(G_i(\hat{x}_n - \hat{x}'_n)\right)^T \varphi_i(G_i\tilde{x}_n) \mathcal{N}\left(G_i(\hat{x}_n - \hat{x}'_n)\right) \\ &\quad \geq \|G_i\mathcal{N}(\hat{x}_n - \hat{x}'_n)\|^2 \inf_{v\in S^s} v^T \varphi_i(G_i\tilde{x}_n) v. \end{aligned}$$

Furthermore, for every $i \in \tilde{J}\setminus J$ we have $G_i\mathcal{N}(\hat{x}_n - \hat{x}'_n) \to 0$ and $G_i\tilde{x}_n \to \theta_i$ as long as $n \to \infty$. At this point, assumption H7 shows that

$$\liminf_{n\to\infty} \mathcal{N}(\hat{x}_n - \hat{x}'_n)^T G_i^T \nabla^2 \varphi_i(G_i\tilde{x}_n) G_i \mathcal{N}(\hat{x}_n - \hat{x}'_n) \geq 0.$$

It follows that the limit of (27) satisfies

$$\begin{aligned} &\limsup_{n\to\infty} - \sum_{i\in\tilde{J}\setminus J} \mathcal{N}(\hat{x}_n - \hat{x}'_n)^T G_i^T \nabla^2 \varphi_i(G_i\tilde{x}_n) G_i \mathcal{N}(\hat{x}_n - \hat{x}'_n) \\ &\quad = - \sum_{i\in\tilde{J}\setminus J} \liminf_{n\to\infty} \mathcal{N}(\hat{x}_n - \hat{x}'_n)^T G_i^T \nabla^2 \varphi_i(G_i\tilde{x}_n) G_i \mathcal{N}(\hat{x}_n - \hat{x}'_n) \leq 0. \end{aligned}$$

As (25)–(26) and (27) have the same limit when $n \to \infty$, the latter result, combined with (29), shows that

$$u^T \nabla^2 \left(\mathcal{E}|_{\Theta_{\tilde{J}}}\right)(\tilde{x}, 0)\, u = 0. \tag{30}$$

Joining (28) to (30) and the fact that $\nabla^2 \mathcal{E}|_{\Theta_J}(\tilde{x}, 0)$ is symmetric, we see that $\tilde{x} \in H_0^{\tilde{J}}$ where $H_0^{\tilde{J}}$ was defined in (18). The latter, combined with Lemma 5 shows that $y \in A_{\tilde{J}}$. By (20), $y \in \Omega_0^{\mathrm{c}}$. □

Lemma 7. *Suppose that $\tilde{x} \in \Theta_{\tilde{J}}$ and $u \in T_J \backslash T_{\tilde{J}}$. Then $y \in B_{\tilde{J}} \subset \Omega_0^{\mathrm{c}}$.*

Proof. Being a minimizer of $\mathcal{E}(\cdot, y_n)$, for every n, the point $\hat{x}_n$ satisfies

$$d^+\mathcal{E}(\hat{x}_n, y_n)(v) \geq 0, \qquad \forall v \in \mathbb{R}^p. \tag{31}$$

We now expand this side-derivative.

$$d^+\mathcal{E}(\hat{x}_n, y_n)(v) = 2v^T L^T (L\hat{x}_n - y_n) + \sum_{i \in \tilde{J}^{\mathrm{c}}} v^T G_i^T \nabla\varphi_i(G_i \hat{x}_n) + \sum_{i \in \tilde{J} \backslash J} v^T G_i^T \nabla\varphi_i(G_i \hat{x}_n) + K,$$

where $K = \sum_{i \in J} v^T G_i^T \nabla^+ \varphi_i(\theta_i)(G_i v)$ is independent of n. Take a subsequence $\{\hat{x}_n\}$ for which $\mathcal{N}(G_i \hat{x}_n - \theta_i)$ converges for every $i \in \tilde{J} \backslash J$. When $n \to \infty$, we have

$$\lim_{n\to\infty} d^+\mathcal{E}(\hat{x}_n, y_n)(v) = 2v^T L^T (L\tilde{x} - y) + \sum_{i \in \tilde{J}^{\mathrm{c}}} v^T G_i^T \nabla\varphi_i(G_i \tilde{x}) + \sum_{i \in \tilde{J} \backslash J} v^T G_i^T \nabla^+\varphi_i(\theta_i) \left(\lim_{n\to\infty} \mathcal{N}(G_i \hat{x}_n - \theta_i)\right) + K.$$

Using assumption H8, the last term can be upper-bounded:

$$v^T G_i^T \nabla^+\varphi_i(\theta_i) \left(\lim_{n\to\infty} \mathcal{N}(G_i \hat{x}_n - \theta_i)\right) \leq v^T G_i^T \nabla^+\varphi_i(\theta_i)(G_i v).$$

It follows that $d^+\mathcal{E}(\tilde{x}, y)(v) \geq \lim_{n\to\infty} d^+\mathcal{E}(\hat{x}_n, y_n)(v)$. Putting this together with (31) we see that

$$d^+\mathcal{E}(\tilde{x}, y)(v) \geq 0, \qquad \forall v \in \mathbb{R}^p. \tag{32}$$

In other words, $\tilde{x}$ satisfies the necessary condition for a local minimum.

Consider convergent subsequences of $\{\mathcal{N}(\hat{x}_n - \tilde{x})\}$ and of $\{\mathcal{N}(\hat{x}'_n - \tilde{x})\}$. Since $u \notin T_{\tilde{J}}$, at least one of the following limits, $v := \lim_{n\to\infty} \mathcal{N}(\hat{x}_n - \tilde{x})$ and $v' := \lim_{n\to\infty} \mathcal{N}(\hat{x}'_n - \tilde{x})$, does not belong to $T_{\tilde{J}}$. For definiteness, suppose $v \notin T_{\tilde{J}}$. By the latter, the projection of v onto $T_{\tilde{J}}^{\perp}$ is non-null. Put $w := \mathcal{N}(\Pi_{T_{\tilde{J}}^{\perp}} v)$ and notice that $w \in T_J$, because $v \in T_J$ and $T_{\tilde{J}} \subset T_J$. Since $\nabla \left(\mathcal{E}|_{\Theta_J}\right)(\hat{x}_n, y_n) = 0$, we deduce that

$$d^+\mathcal{E}(\hat{x}_n, y_n)(w) = 0, \qquad \forall n. \tag{33}$$

Moreover, noticing that $G_i w = 0$ for every $i \in J$, we have

$$\sum_{i \in J} w^T G_i^T \nabla^+ \varphi_i(\theta_i)(G_i w) = 0.$$

Thus we obtain

$$d^+\mathcal{E}(\hat{x}_n, y_n)(w) = 2w^T L^T (L\hat{x}_n - y_n) + \sum_{i \in \tilde{J}^c} w^T G_i^T \nabla \varphi_i(G_i \hat{x}_n) + \sum_{i \in \tilde{J} \setminus J} w^T G_i^T \nabla \varphi_i(G_i \hat{x}_n).$$

We will calculate the limit of all the terms in $d^+\mathcal{E}(\hat{x}_n, y_n)(w)$ when $n \to \infty$. The limit of the first two terms on the right side of the equation given above is easily obtained by continuity. We now focus on the limit of $\nabla \varphi_i(G_i \hat{x}_n)$ for $i \in \tilde{J} \setminus J$. We start by considering the case when $G_i w \neq 0$. We have

$$\begin{aligned} \lim_{n \to \infty} \mathcal{N}(G_i \hat{x}_n - \theta_i) &= \lim_{n \to \infty} \mathcal{N}(G_i \mathcal{N}(\hat{x}_n - \tilde{x})) \\ &= \mathcal{N}\left(G_i \lim_{n \to \infty} \mathcal{N}(\hat{x}_n - \tilde{x})\right) \\ &= \mathcal{N}(G_i v) = \mathcal{N}(G_i w). \end{aligned}$$

The last equality comes from the fact that for $i \in \tilde{J}$ we have $G_i v = G_i \Pi_{T_{\tilde{J}}^\perp} v + G_i \Pi_{T_{\tilde{J}}} v = G_i \Pi_{T_{\tilde{J}}^\perp} v = G_i w \|\Pi_{T_{\tilde{J}}^\perp} v\|$, since $\Pi_{T_{\tilde{J}}} v \in T_{\tilde{J}}$ and hence $G_i \Pi_{T_{\tilde{J}}} v = 0$. Thus, for $i \in \tilde{J} \setminus J$ and $G_i w \neq 0$, we find that $w^T G_i^T \nabla \varphi_i(G_i \hat{x}_n) \to w^T G_i^T \nabla^+ \varphi_i(\theta_i)(G_i w)$. Otherwise, if $G_i w = 0$ for some $i \in \tilde{J} \setminus J$, obviously $(G_i w)^T \nabla \varphi_i(G_i \hat{x}_n) = 0 = (G_i w)^T \nabla^+ \varphi_i(\theta_i)(G_i w)$. Consequently,

$$\begin{aligned} \lim_{n \to \infty} d^+\mathcal{E}(\hat{x}_n, y_n)(w) &= 2w^T L^T (L\tilde{x} - y) + \sum_{i \in \tilde{J}^c} (G_i w)^T \nabla \varphi_i(G_i \tilde{x}) \\ &\quad + \sum_{i \in \tilde{J} \setminus J} (G_i w)^T \nabla^+ \varphi_i(\theta_i)(G_i w) \\ &= d^+\mathcal{E}(\tilde{x}, y)(w). \end{aligned}$$

Using (33), at the limit we get $d^+\mathcal{E}(\tilde{x}, y)(w) = 0$. However, $w \in T_{\tilde{J}}^\perp$ which shows that $\tilde{x}$, although being a local minimizer of $\mathcal{E}|_{\Theta_{\tilde{J}}}(\cdot, y)$, does not satisfy condition (B) of Proposition 1 in the previous part [7]. Then $y \in B_{\tilde{J}}$ as given in (22). Using (20) we see that $y \in \Omega_0^c$. □

These lemmas show that if the claim of Proposition 3 fails to hold then y belongs to Ω_0^c.

This completes the proof of Proposition 3. □

We can now extend Remark 1 to the class of objective functions considered in this section.

Remark 3. Consider two minimizer functions $\mathcal{X}_1$ and $\mathcal{X}_2$ defined on an open and connected domain $O \subset \Omega_0$. Then we have either $\mathcal{X}_1 \equiv \mathcal{X}_2$ on O, or

$$\mathcal{X}_1(y) \neq \mathcal{X}_2(y), \qquad \forall y \in O.$$

The arguments are similar to those given in Remark 1. Put $\tilde{O} := \{y \in O\colon \mathcal{X}_1(y) = \mathcal{X}_2(y)\}$ and suppose that $\tilde{O} \neq \emptyset$ and $\tilde{O} \neq O$. Clearly, $\tilde{O}$ is closed in O. Focus on y belonging to the boundary of $\tilde{O}$ in O. Then there is a sequence $\{y_n\}$ with $y_n \in O \backslash \tilde{O}$ and $y_n \to y$ when $n \to \infty$, such that $\mathcal{X}_1(y_n) \neq \mathcal{X}_2(y_n)$. Since $\mathcal{X}_1$ and $\mathcal{X}_2$ are continuous, the points $\hat{x}_n := \mathcal{X}_1(y_n)$ and $\hat{x}'_n := \mathcal{X}_2(y_n)$ come arbitrarily close to each other as long as $n \to \infty$. Then we apply the same reasoning developed after (23) and deduce that $y \in \Omega_0^{\mathrm{c}}$. This contradicts the fact that $O \subset \Omega_0$.

Proposition 4. *Let the assumptions of Proposition* 3 *hold. Then every open set of* $\mathbb{R}^q$ *contains an open subset* O *on which* $\mathcal{E}$ *admits* n *minimizer functions* $\mathcal{X}_i\colon O \to \mathbb{R}^p$, $i = 1, \ldots, n$, *which are* $\mathcal{C}^{m-1}$ *and such that for all* $y \in O$, *all the local minimizers of* $\mathcal{E}(\cdot, y)$ *read*

$$\mathcal{X}_i(y), \qquad i = 1, \ldots, n, \tag{34}$$

and satisfy

$$\mathcal{E}(\mathcal{X}_i(y), y) \neq \mathcal{E}(\mathcal{X}_j(y), y), \qquad \forall i, j \in \{1, \ldots, n\} \quad \textit{with} \quad i \neq j. \tag{35}$$

Proof. We take into consideration that the smoothness of Φ is not exploited in the proof of Proposition 2, but is in the proofs of Proposition 1, Remark 1 and Lemma 2. The generalization of these statements to the conditions of Proposition 4 is then sufficient to prove this proposition. The first two statements have been generalized in Proposition 3 and Remark 3, the last one is given in Lemma 8 below.

Lemma 8. *Let* $\mathcal{X}_1$ *and* $\mathcal{X}_2$ *be two differentiable* (*local*) *minimizer functions relevant to* $\mathcal{E}$ *and defined on the same open domain* $\tilde{O} \subset \Omega$. *Suppose*, *we have*

$$\mathcal{E}(\mathcal{X}_1(y), y) = \mathcal{E}(\mathcal{X}_2(y), y), \qquad \forall y \in \tilde{O}. \tag{36}$$

Then

$$\mathcal{X}_1(y) = \mathcal{X}_2(y), \qquad \forall y \in \tilde{O}.$$

Proof. Let us consider $y \in \tilde{O}$. Then there are two sets of indexes J_1 and J_2 such that we have $\mathcal{X}_1(y) \in Q_{J_1}$ and $\mathcal{X}_2(y) \in Q_{J_2}$. By Proposition 1 of the previous part [7], y is contained in a neighborhood $N \subset \tilde{O}$ such that for all $y' \in N$ we have in addition $\mathcal{X}_1(y') \in Q_{J_1}$ and $\mathcal{X}_2(y') \in Q_{J_2}$. On this neighborhood, (36) can equivalently be written

$$\mathcal{E}|_{Q_{J_1}}(\mathcal{X}_1(y'), y') = \mathcal{E}|_{Q_{J_2}}(\mathcal{X}_2(y'), y'), \qquad \forall y' \in N. \tag{37}$$

By differentiating both sides of (37) with respect to y', we obtain

$$\begin{aligned} &D_1(\mathcal{E}|_{Q_{J_1}})(\mathcal{X}_1(y'), y')\, D\mathcal{X}_1(y') + D_2\mathcal{E}(\mathcal{X}_1(y'), y') \\ &\quad = D_1(\mathcal{E}|_{Q_{J_2}})(\mathcal{X}_2(y'), y')\, D\mathcal{X}_2(y') + D_2\mathcal{E}(\mathcal{X}_2(y'), y'). \end{aligned} \tag{38}$$

Since, for $i \in \{1, 2\}$, $\mathcal{X}_i$ is a minimizer function relevant to $\mathcal{E}|_{Q_{J_i}}$,

$$D_1(\mathcal{E}|_{Q_{J_i}})(\mathcal{X}_i(y'), y') = 0, \qquad \forall y' \in N.$$

By also using the expression of $D_2\mathcal{E}$ given in (15), equation (38) yields

$$L\mathcal{X}_1(y') = L\mathcal{X}_2(y'), \qquad \forall y' \in N.$$

The conclusion follows from the injectivity of L. □

References

1. J. E. Besag, On the statistical analysis of dirty pictures (with discussion), Journal of the Royal Statistical Society B, 48 (1986), 259–302.
2. M. Black, G. Sapiro, D. Marimont, and D. Heeger, Robust anisotropic diffusion, IEEE Transactions on Image Processing, 7 (1998), 421–432.
3. J. Bonnans and A. Shapiro, Perturbation Analysis of Optimization Problems, Springer-Verlag, New York, 2000.
4. C. Bouman and K. Sauer, A generalized Gaussian image model for edge-preserving map estimation, IEEE Transactions on Image Processing, 2 (1993), 296–310.
5. P. G. Ciarlet, Introduction à l'analyse numérique matricielle et à l'optimisation, Collection mathématiques appliquées pour la maîtrise, Masson, Paris, 1990.
6. A. L. Dontchev and T. Zollezi, Well-Posed Optimization Problems, Springer-Verlag, New York, 1993.
7. S. Durand and M. Nikolova, Stability of the minimizers of least squares with a non-convex regularization. Part I: local behavior, Applied Mathematics & Optimization, DOI: 10.1007/s00245-005-0842-1 (to appear).
8. S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6 (1984), 721–741.
9. S. Geman and D. E. McClure, Statistical methods for tomographic image reconstruction, in Proceedings of the 46th Session of the ISI, Bulletin of the ISI, vol. 52, 1987, pp. 22–26.
10. P. J. Green, Bayesian reconstructions from emission tomography data using a modified em algorithm, IEEE Transactions on Medical Imaging, MI-9 (1990), 84–93.
11. P. Perona and J. Malik, Scale-space and edge detection using anisotropic diffusion, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-12 (1990), 629–639.
12. R. T. Rockafellar and J. B. Wets, Variational Analysis, Springer-Verlag, New York, 1997.
13. L. Rudin, S. Osher, and C. Fatemi, Nonlinear total variation based noise removal algorithm, Physica, 60 D (1992), 259–268.
14. S. Teboul, L. Blanc-Féraud, G. Aubert, and M. Barlaud, Variational approach for edge-preserving regularization using coupled pde's, IEEE Transactions on Image Processing, 7 (1998), 387–397.

Accepted 13 *August* 2005. *Online publication* 17 *February* 2006.