Appl Math Optim OF1–OF24 (2006)
DOI: 10.1007/s00245-005-0842-1

Applied Mathematics & Optimization

Stability of the Minimizers of Least Squares with a Non-Convex Regularization. Part I: Local 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. Many estimation problems amount to minimizing a piecewise $\mathcal{C}^m$ objective function, with $m \geq 2$, composed of a quadratic data-fidelity term and a general regularization term. It is widely accepted that the minimizers obtained using non-convex and possibly non-smooth regularization terms are frequently good estimates. However, few facts are known on the ways to control properties of these minimizers. This work is dedicated to the stability of the minimizers of such objective functions with respect to variations of the data. It consists of two parts: first we consider all local minimizers, whereas in a second part we derive results on global minimizers. In this part we focus on data points such that every local minimizer is isolated and results from a $\mathcal{C}^{m-1}$ local minimizer function, defined on some neighborhood. We demonstrate that all data points for which this fails form a set whose closure is negligible.

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 first of two papers devoted to the stability of minimizers of regularized least squares objective functions as customarily used in signal and image reconstruction. In this part we deal with the behavior of local minimizers, whereas in the second part we draw conclusions about global minimizers.

In various inverse problems such as denoising, deblurring, segmentation or reconstruction, a sought-after object $\hat{x} \in \mathbb{R}^p$ (such as an image or a signal) is estimated from recorded data $y \in \mathbb{R}^q$ by minimizing with respect to x an objective function $\mathcal{E}$: $\mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$,

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

where $O \subset \mathbb{R}^p$ is an open domain where $\mathcal{E}(\cdot, y)$ has local minimizers. In general, $\hat{x} \in \mathbb{R}^p$ is a *local minimizer* of $\mathcal{E}(\cdot, y)$ over $\mathbb{R}^p$. This work is dedicated to objective functions of the form

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

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

$$\Phi(x) := \sum_{i=1}^{r} \varphi_i(G_i x), \tag{3}$$

where for every $i \in \{1, \dots, r\}$, the function φ_i: $\mathbb{R}^s \to \mathbb{R}$ 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: $\mathbb{R}^p \to \mathbb{R}^s$ is a linear operator. Since the publication of [45], objective functions of this form are customarily used for the restoration and the reconstruction of signals and images from noisy data y obtained at the output of a linear system L [23], [7]. The operator L can represent the blur undergone by a signal or an image, a Fourier transform on an irregular lattice in tomography, a wavelet in seismology, as well as other observation systems. The quadratic term in (2) thus accounts for the closeness of the unknown x to data y. The operators G_i in the regularization term Φ usually provide the differences between neighboring samples of x. For instance, if x is a one-dimensional signal, usually $G_i x = x_{i+1} - x_i$ or in some cases $G_i x = x_{i+1} - 2x_i + x_{i-1}$. Typically, for all $i \in \{1, \dots, r\}$, we have $\theta_i = 0$ and φ_i reads

$$\varphi_i(z) = \phi(\|z\|), \qquad \forall i \in \{1, \dots, r\}, \tag{4}$$

where ϕ: $\mathbb{R}_+ \to \mathbb{R}$ is an increasing function, often called the *potential function* (PF). Several examples among the most popular, are the following [23], [6], [24], [38], [25], [41], [10], [44], [8]:

$$\begin{array}{lll} L^\alpha & \phi(t) = |t|^\alpha, & 1 \le \alpha < 2, \\ \text{Lorentzian} & \phi(t) = \alpha t^2/(1 + \alpha t^2), & \\ \text{Concave} & \phi(t) = \alpha|t|/(1 + \alpha|t|), & \\ \text{Gaussian} & \phi(t) = 1 - \exp(-\alpha t^2), & \\ \text{Truncated quadratic} & \phi(t) = \min\{\alpha t^2, 1\}, & \\ \text{Huber} & \phi(t) = \begin{cases} t^2 & \text{if } |t| \le \alpha, \\ \alpha(\alpha + 2|t - \alpha|) & \text{if } |t| > \alpha. \end{cases} & \end{array} \tag{5}$$

Objective functions as specified above are either defined in a variational setting [38], [41], [3], [12], [11], [46] or rely on probabilistic considerations [23], [6], [7], [17].

Most of the potential functions cited in (5) are non-convex and in some cases non-smooth. Indeed, several authors pointed out the possibility of getting signals involving jumps and images with sharp edges by using non-convex regularization functions [34], [24], [38], [4]. On the other hand, non-smooth regularization has been shown to avoid Gibbs artifacts and to enforce local homogeneity [22], [19], [2], [36]. In spite of this, a few facts are known about the features of the local minimizers relevant to non-convex objective functions. We focus on the stability of the local minimizers $\hat{x}$ of objective functions $\mathcal{E}(\cdot, y)$ of the form (2)–(3) under variations of data y. Our goal is to show that if $\mathcal{E}$ is piecewise $\mathcal{C}^m$, the local and the global minimizer functions are $\mathcal{C}^{m-1}$ everywhere except on a closed negligible subset of $\mathbb{R}^q$.

Related questions have been considered in critical point theory. Conditions ensuring that the local minimizers of continuous objective functions defined on metric spaces are at the bottom of a "potential well" are considered in [27] and [16]. In such a case a local minimizer is stable under a class of homotopies of the objective function which do not introduce new critical points in a neighborhood of this local minimizer. An approach related to ours is used in studies where different concepts relevant to properties of the optimal solutions are demonstrated to hold generically. We can evoke some works on semi-definite programming [42], [1], [37], as well as on the well-posedness of some classes of optimization problems [20]. Many results have been established on the stability of the local minimizers of general smooth objective functions subject to constraints [21], [29], [28], [26]. When $\mathcal{E}$ is smooth, our results can be deduced from these references. Let us notice that in spite of the popularity of non-convex objective functions of the form (2)–(3) in image and signal restoration, and in many inverse problems—e.g. [23], [34], [6], [31], [24], [30], [7], [38], [22], [35]—the regularity of their minimizers have never been studied in a systematic way.

2. Motivation and Definitions

Studying the stability of local minimizers (rather than restricting our interest to global minimizers only) is a matter of critical importance in its own right for several reasons. In many applications, smoothing is performed by only locally minimizing a non-convex objective function in the vicinity of some initial solution. Moreover, it is worth recalling that no minimization algorithm guarantees the finding of the global minimum of a general non-convex objective function. Some algorithms allow the finding of the global minimum only with high probability, under demanding requirements (e.g. simulated annealing) [23], [22]. Others allow the finding of a local minimum which is expected to be close to the global minimum [9], [14], [35]. The practically obtained solutions are frequently only local minimizers, hence the importance of knowing their behavior.

The first question is to know whether, and in what circumstances, $\mathcal{E}(\cdot, y)$ has local minimizers which give rise to smooth local minimizer functions as defined below.

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 $y \in O$.

We focus on the subset $\Omega \subset \mathbb{R}^q$ of all data leading to minimizers which have good regularity properties as specified below:

Definition 2. Let $\mathcal{E}(\cdot, y)$ be $\mathcal{C}^m$ (with $m \geq 1$) almost everywhere on $\mathbb{R}^p$, for every $y \in \mathbb{R}^q$. Denote

$$\Omega := \left\{ y \in \mathbb{R}^q\colon \begin{array}{l} \text{if } \hat{x} \text{ is a local minimizer of } \mathcal{E}(\cdot, y) \text{ then there} \\ \text{is a } \mathcal{C}^{m-1} \text{ minimizer function } \mathcal{X}\colon\ O \to \mathbb{R}^p \\ \text{such that } y \in O \subset \mathbb{R}^q \text{ and } \hat{x} = \mathcal{X}(y) \end{array} \right\}.$$

It will be seen that data points y for which $\mathcal{E}(\cdot, y)$ has minimizers $\hat{x}$ which do not satisfy the requirements in Definition 2 are highly exceptional. We will show that the closure of Ω^c in $\mathbb{R}^q$, denoted $\overline{\Omega^c}$, is a *negligible* subset of the data domain $\mathbb{R}^q$, i.e. that it has Lebesgue measure zero. In the following, given an affine subspace $N \subset \mathbb{R}^q$, we say that *a subset $M \subset N$ is negligible in N* if it has measure zero with respect to the Lebesgue measure induced on N. The set Ω, or equivalently its complement Ω^c, can be explicitly calculated in the following examples.

Example 1. Consider the function

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

where

$$\Phi(x) = \begin{cases} 1 - (|x| - 1)^2 & \text{if } \ 0 \leq |x| \leq 1, \\ 1 & \text{if } \ |x| > 1. \end{cases}$$

It is not difficult to check that the minimizer $\hat{x}$ of $\mathcal{E}(\cdot, y)$ takes different forms according to the values of y.

- If $|y| > 1$, the minimizer is strict and reads $\hat{x} = y$.
- If $y = 1$, every $\hat{x} \in [0, 1]$ is a non-strict minimizer.
- If $y = -1$, every $\hat{x} \in [-1, 0]$ is a non-strict minimizer.
- If $y \in (-1, 1)$, the minimizer is strict and constant, $\hat{x} = 0$.

Thus we find that $\Omega^c = \{-1,\ 1\}$ which is closed and negligible in $\mathbb{R}$.

Example 2. Consider

$$\mathcal{E}\colon\ \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R},$$
$$(x, y) \mapsto (x_1 - x_2 - y)^2 + \beta(x_1 - x_2)^2,$$

where $\beta > 0$. For all $y \in \mathbb{R}$, every $\hat{x} \in \mathbb{R}^2$, such that $\hat{x}_1 - \hat{x}_2 = y/(1+\beta)$, is a minimizer of $\mathcal{E}(\cdot, y)$. Hence $\Omega^c = \mathbb{R}$.

Example 3. Consider

$$\mathcal{E}\colon\ \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R},$$
$$(x, y) \mapsto (x_1 - x_2 - y)^2 + |x_1| + |x_2|.$$

The minimizers $\hat{x}$ of $\mathcal{E}(\cdot, y)$ are obtained after a simple computation.

- If $y > \frac{1}{2}$, every $\hat{x} = (\alpha, \alpha - y + \frac{1}{2})$ for $\alpha \in [0, y - \frac{1}{2}]$ is a non-strict minimizer.
- If $y \in (-\frac{1}{2}, \frac{1}{2})$, the only minimizer is $\hat{x} = (0, 0)$.
- If $y < -\frac{1}{2}$, every $\hat{x} = (\alpha, \alpha - y - \frac{1}{2})$ for $\alpha \in [y + \frac{1}{2}, 0]$ is a non-strict minimizer.

Consequently, $\Omega = (-\frac{1}{2}, \frac{1}{2})$.

We remark that L is injective in Example 1 whereas it is non-injective in Examples 2 and 3. We can construct many other examples of objective functions $\mathcal{E}$ involving L non-injective for which Ω^c is non-negligible. This suggests we make the following assumption:

H1. The operator $L\colon \mathbb{R}^p \to \mathbb{R}^q$ in (2) is injective, i.e. rank $L = p$.

Notice that Ω^c may be negligible even if H1 fails. However, this assumption allows strong results to be obtained.

Remark 1. We do not focus properly on the question of whether or not $\mathcal{E}$ admits minimizers when y ranges over $\mathbb{R}^q$. 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, although they remain trivially true in the opposite situation. Practically, objective functions are defined in such a way that they do have minimizers. We recall that $\mathcal{E}(\cdot, y)$ admits minimizers if it is coercive, i.e. if $\mathcal{E}(x, y) \to \infty$ along with $\|x\| \to \infty$ [15], [40]. For instance, this situation occurs, for all $y \in \mathbb{R}^q$ when L is injective and Φ does not decrease faster or equally as fast as $-\|Lx\|^2$ as $\|x\| \to \infty$. This is trivially satisfied in practice where Φ is bounded below.

For any function $f\colon \mathbb{R}^p \to \mathbb{R}$, we denote by $\nabla f(x) \in \mathbb{R}^p$ the gradient of f at a point $x \in \mathbb{R}^p$ and by $\nabla^2 f(x) \in \mathbb{R}^p \times \mathbb{R}^p$ the Hessian matrix of f at x. Although $\mathcal{E}$ depends on two variables (x, y), we are concerned only with its derivatives with respect to x. For simplicity, $\nabla\mathcal{E}$ and $\nabla^2\mathcal{E}$ will systematically be used to denote the gradient and the Hessian of $\mathcal{E}$ 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, for whatever dimension n appropriate to the context. Furthermore, the letter S denotes the unit sphere in $\mathbb{R}^n$ centered at the origin. When necessary, the superscript n is used to specify that S^n is the unit sphere in $\mathbb{R}^n$. We denote $\mathbb{R}_+ = \{t \in \mathbb{R}\colon t \geq 0\}$. For a subset $A \in \mathbb{R}^q$, its complement in $\mathbb{R}^q$ will be denoted A^c and its closure $\overline{A}$.

The subsequent considerations are split into two parts according to the differentiability of Φ.

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

The characterization of Ω, developed in this section, is based on the next lemma which constitutes a straightforward extension of the Implicit Functions Theorem [5]. Its proof can be found, e.g. in [21, Theorem 6, p. 34] or in [28, Lemma 6.1.1, p. 268].

Lemma 1. *Suppose $\mathcal{E}$: $\mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ is any function which is $\mathcal{C}^m$, with $m \geq 2$. Fix $y \in \mathbb{R}^q$. Let $\hat{x}$ be such that $\nabla \mathcal{E}(\hat{x}, y) = 0$ and $\nabla^2 \mathcal{E}(\hat{x}, y)$ is positive definite. Then there exist $\rho > 0$ and a unique $\mathcal{C}^{m-1}$-minimizer function $\mathcal{X}$: $B(y, \rho) \to \mathbb{R}^p$ such that $\mathcal{X}(y) = \hat{x}$.*

In the following we focus on objective functions $\mathcal{E}$ of the form (2) where Φ is any $\mathcal{C}^m$ function on $\mathbb{R}^p$, with $m \geq 2$. Given $y \in \mathbb{R}^q$, if $\hat{x} \in \mathbb{R}^p$ is a strict or non-strict local minimizer of $\mathcal{E}(\cdot, y)$, then

$$\nabla \mathcal{E}(\hat{x}, y) = 0, \tag{6}$$

where

$$\nabla \mathcal{E}(x, y) = 2L^T(Lx - y) + \nabla \Phi(x).$$

Using the fact that

$$\nabla \mathcal{E}(\hat{x}, 0) = 2L^T L\hat{x} + \nabla \Phi(\hat{x}), \tag{7}$$

the variables $\hat{x}$ and y can be separated in (6) which then becomes

$$2L^T y = \nabla \mathcal{E}(\hat{x}, 0). \tag{8}$$

A point $\hat{x}$, satisfying (8), is a strict minimizer of $\mathcal{E}(\cdot, y)$ if the Hessian of $\mathcal{E}(\cdot, y)$ at $\hat{x}$, namely $\nabla^2 \mathcal{E}(\hat{x}, y)$, is positive definite. For an arbitrary x, we have

$$\nabla^2 \mathcal{E}(x, y) = 2L^T L + \nabla^2 \Phi(x). \tag{9}$$

We emphasize the fact that the Hessian of $\mathcal{E}(\cdot, y)$ is independent of y at any $x \in \mathbb{R}^p$, by writing $\nabla^2 \mathcal{E}(x, 0)$ instead of $\nabla^2 \mathcal{E}(x, y)$. Let H_0 be the set of all the critical points of $\nabla \mathcal{E}(x, y)$:

$$H_0 := \{x \in \mathbb{R}^p\colon\ \det \nabla^2 \mathcal{E}(x, 0) = 0\}. \tag{10}$$

The set H_0 is independent of y as well. We cannot guarantee that a point $\tilde{x}$ satisfying (6) is a strict minimizer of $\mathcal{E}(\cdot, y)$ if $\nabla^2 \mathcal{E}(\tilde{x}, 0)$ is singular. Define the set

$$\Omega_0 := \{y \in \mathbb{R}^q\colon\ \exists \tilde{x} \in H_0 \text{ satisfying } 2L^T y \neq \nabla \mathcal{E}(\tilde{x}, 0)\}. \tag{11}$$

All the y's leading to a non-strict minimizer, or to a non-continuous minimizer function, are contained in Ω_0^c which reads

$$\Omega_0^c = \{y \in \mathbb{R}^q\colon\ L^T y \in \nabla \mathcal{E}(H_0, 0)\}. \tag{12}$$

Then we have

$$\Omega_0 \subset \Omega. \tag{13}$$

If $\nabla^2 \mathcal{E}(x, 0)$ is positive definite for all $x \in \mathbb{R}^p$, the set H_0 is empty and (13) shows that $\Omega = \mathbb{R}^q$, and there is a unique $\mathcal{C}^{m-1}$ minimizer function $\mathcal{X}$ as stated in Definition 2. If L is injective and Φ is convex, $\nabla^2 \mathcal{E}(x, 0)$ is clearly positive definite for all x. However, if Φ is non-convex, H_0 is non-empty. In the following, we focus on non-convex functions Φ which satisfy the assumption below.

H2. As $t \to \infty$, we have $\nabla\Phi(tv)/t \to 0$ uniformly with $v \in S$.

This assumption is satisfied by the regularization functions used by many authors [24], [38], [25].

Theorem 1. *Suppose $\mathcal{E}$ is as in* (2) *where Φ is an arbitrary $\mathcal{C}^m$ function on $\mathbb{R}^p$, with $m \geq 2$. Suppose that H*1 *is satisfied. Then we have the following*:

(i) *The set Ω^c, the complement of Ω specified in Definition* 2, *has Lebesgue measure zero in $\mathbb{R}^q$.*
(ii) *Moreover, if H*2 *is satisfied, $\overline{\Omega^c}$ has Lebesgue measure zero in $\mathbb{R}^q$.*

Remark 2. The proof of the theorem establishes (i) and (ii) for Ω_0, as given in (11), instead of Ω. The ultimate conclusions are obtained by noticing that (13) entails that $\Omega^c \subset \Omega_0^c$ where Ω_0^c is given in (12).

Proof. Since $\nabla\mathcal{E}(\cdot, 0)$ is $\mathcal{C}^1$ and H_0 is the set of its critical points, Sard's theorem [39], [33] shows that $\nabla\mathcal{E}(H_0, 0)$ is a negligible subset of $\mathbb{R}^p$. Since L is injective, Ω_0^c is negligible, hence (i). Statement (ii) is proven in the Appendix.

Remark 3. Now we ask what is the shape of H_0, as defined in (10), since it contains all the non-strict minimizers of $\mathcal{E}(\cdot, y)$, for all y. A key point in Theorem 1 is that $\nabla\mathcal{E}(H_0, 0)$ is negligible although H_0 itself may be of positive measure. We observe that for the most important classes of functions Φ, the set H_0 has an empty interior. For instance, such is the case if L is injective, Φ is analytic and there is $x_0 \in \mathbb{R}^p$ for which the Hessian matrix $\nabla^2\Phi(x_0)$ has all its eigenvalues non-negative. Indeed, assume that the interior of H_0 is non-empty. As $\nabla^2\mathcal{E}(\cdot, 0)$ is analytic on $\mathbb{R}^p$, it follows that $\det \nabla^2\mathcal{E}(x, 0) = 0$ for all $x \in \mathbb{R}^p$. However, the latter is impossible because by assumption there is x_0 such that

$$\nabla^2\mathcal{E}(x_0, 0) = 2L^T L + \nabla^2\Phi(x_0) \tag{14}$$

is positive definite, as being the sum of a positive definite and of a semi-positive definite matrix.

More specifically, the assumption about the positive definiteness of $\nabla^2\Phi(x_0)$ holds for $x_0 = 0$ whenever Φ is of the form of (3)–(4) with ϕ analytic and symmetric, and $\phi''(0) \geq 0$—this comes from the fact that $\nabla^2\Phi(0) = \phi''(0)\sum_{i=1}^r G_i^T G_i$. These requirements are satisfied by the objective functions used in [24], [30] and [38] where the typical choices for ϕ are the Lorentzian and the Gaussian potential functions given in (5). We return to the expression of $\nabla^2\mathcal{E}(x_0, 0)$ in (14). Observe that if there is a point x_0 such that $\nabla^2\Phi(x_0)$ is positive definite, then H_0 has a non-empty interior independently of the injectivity of L.

Remark 4 (Extension to Mumford–Shah-like Regularization). The most frequently used non-convex regularization is certainly the Mumford–Shah functional which in the discrete setting amounts to inserting the truncated quadratic potential function, see (5), into (3)–(4). This is a special form of non-smooth regularization which does not fall into the scope of Section 4, but can be assimilated to the smooth case considered above. More generally, consider that there is a constant $\tau > 0$ such that

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

and that ϕ is $\mathcal{C}^m$, $m \geq 2$ on $\mathbb{R}_+\backslash\{\tau\}$, with $\phi'(0) = 0$. For the truncated quadratic PF, $\tau = 1/\sqrt{\alpha}$ where $\varphi'(\tau^-) = 2\sqrt{\alpha}$ and $\varphi'(\tau^+) = 0$. We define

$$M = \bigcup_{i=1}^{r} \{x \in \mathbb{R}^p\colon\ \|G_i x\| = \tau\}.$$

Obviously, for every $y \in \mathbb{R}^q$, the function $\mathcal{E}(\cdot, y)$ is $\mathcal{C}^m$ on $\mathbb{R}^p\backslash M$. An important observation is that if $\mathcal{E}(\cdot, y)$ has a (local) minimum at $\hat{x} \in \mathbb{R}^p$, then

$$\hat{x} \notin M.$$

The proof can be found in [36]. We now define

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

Notice that M is a finite union of $\mathcal{C}^\infty$ manifolds of dimension $\leq p - 1$ and that $\mathbb{R}^p\backslash M$ is composed of a finite number of open subsets of $\mathbb{R}^p$, say M_i^c, $i = 1, \ldots, n$. Using Sard's theorem, $\nabla\mathcal{E}(H_0 \cap M_i^c, 0)$ is negligible in $\mathbb{R}^p$, for every $i = 1, \ldots, n$, and thus $\nabla\mathcal{E}(H_0, 0)$ is negligible in $\mathbb{R}^p$. The injectivity of L yet again leads to statement (i) of Theorem 1. Statement (ii) can be shown by extending the arguments developed in the Appendix.

For the truncated quadratic PF in particular, using that L is injective and $\nabla^2\Phi(x) \geq 0$, for all $x \in \mathbb{R}^p\backslash M$, we get $H_0 = \emptyset$ and hence $\Omega = \mathbb{R}^q$.

4. Objective Function Involving Non-smooth Regularization

We now consider regularization terms Φ as introduced in (3), namely

$$\Phi(x) = \sum_{i=1}^{r} \varphi_i(G_i x), \tag{15}$$

where $G_i\colon \mathbb{R}^p \to \mathbb{R}^s$ are linear operators, for all $i = 1, \ldots, r$. We assume that for each $i = 1, \ldots, r$, there is a constant $\theta_i \in \mathbb{R}^s$ such that φ_i is $\mathcal{C}^m$ on $\mathbb{R}^s\backslash\{\theta_i\}$, with $m \geq 2$, and continuous on $\mathbb{R}^s$. Typically, φ_i is non-smooth at θ_i. Potential functions which are non-smooth at more than one point, say θ_i, can be expressed as the sum of several φ_i, which are non-smooth at θ_i, applied to the same $G_i x$. Notice that the regularization function studied in Section 3 can be seen as a special case of (15) corresponding to $r = 1$, $G_1 = I$ and $\varphi_1 \in \mathcal{C}^m(\mathbb{R}^s)$. Assumption H2 is adapted to the context of piecewise smooth potential functions as it follows:

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

We assume that every φ_i has at θ_i a one-sided directional derivative $d^+\varphi_i(\theta_i)(u)$, for every direction $u \in \mathbb{R}^s$ [40]. When φ_i is non-smooth at θ_i, the function $u \mapsto d^+\varphi_i(\theta_i)(u)$ is non-linear. We focus on functions φ_i for which $d^+\varphi_i(\theta_i)(u)$ can be expressed as the scalar product of the direction u and a direction-dependent vector, that we call a *directional gradient*. Let $\mathcal{N}$ be the normalization mapping which for every vector v gives its projection on the unit sphere,

$$\mathcal{N}(v) = \frac{v}{\|v\|}. \tag{16}$$

We assume that every φ_i satisfies the following property:

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 $\lim_{h\to 0} \mathcal{N}(h)$.

We put

$$\nabla^+\varphi_i(\theta_i)\left(\lim_{h\to 0} \mathcal{N}(h)\right) := \lim_{h\to 0} \nabla\varphi_i(\theta_i + h). \tag{17}$$

By a slight abuse of notation, we extend this definition to every $u \in \mathbb{R}^s$ in the following way:

$$\nabla^+\varphi_i(\theta_i)(u) = \begin{cases} \nabla^+\varphi_i(\theta_i)\,(\mathcal{N}(u)) & \text{if} \quad u \neq 0, \\ 0 & \text{if} \quad u = 0. \end{cases} \tag{18}$$

The vector $\nabla^+\varphi_i(\theta_i)(u)$ is the directional gradient of φ_i at θ_i for u. In particular, if φ_i is smooth at θ_i, for every $u \neq 0$, we have $\nabla^+\varphi_i(\theta_i)(u) = \nabla\varphi_i(\theta_i)$. This fact suggests we extend the definition of $\nabla^+\varphi_i$ on $\mathbb{R}^s$ by taking $\nabla^+\varphi_i(z)(u) = \nabla\varphi_i(z)$ for every $z \neq \theta_i$ and for every $u \neq 0$. If the directional gradient $\nabla^+\varphi_i(\theta_i)$ exists, then φ_i is semi-smooth [32] since $\lim_{h\to 0} \mathcal{N}(h)^T \nabla\varphi_i(\theta_i + h)$ exists provided that $\lim_{h\to 0} \mathcal{N}(h)$ exists. In such a case, the one-sided directional derivative $d^+\varphi_i(\theta_i)$ is well defined and, more generally, for any $z \in \mathbb{R}^s$ and for any $u \in \mathbb{R}^s$, we have

$$d^+\varphi_i(z)(u) = \lim_{t\searrow 0} \frac{\varphi_i(z + tu) - \varphi_i(z)}{t} = u^T \nabla^+\varphi_i(z)(u). \tag{19}$$

We also use two other assumptions which are given below.

H5. For every $i \in \{1, \ldots, r\}$, the mapping $u \to \nabla^+\varphi_i(\theta_i)(u)$ is Lipschitz on S^s.

Assumption H5 and (19) show that the mapping $u \to d^+\varphi_i(\theta_i)(u)$ is Lipschitz on $\mathbb{R}^s$.

H6. For every $i \in \{1, \ldots, r\}$, the mapping $u \mapsto \nabla\varphi_i(\theta_i + hu)$ converges to $\nabla^+\varphi_i(\theta_i)$ as $h \searrow 0$, uniformly on S^s.

These assumptions are illustrated in the context of typical potential functions as mentioned in (4).

Example 4. 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 $\phi'(0) > 0$. Using (17)–(18), it is easily obtained that

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

Both assumptions H5 and H6 are clearly satisfied. Assumption H3 amounts to saying that $\phi'(t)/t \to 0$ when $t \to \infty$. This is satisfied by all the functions cited in (5). By (19), the one-sided directional derivative of φ_i at θ_i for u reads

$$d^+\varphi_i(\theta_i)(u) = u^T\nabla^+\varphi_i(\theta_i)(u) = \phi'(0^+)\|u\|. \tag{20}$$

Below we extend Theorem 1 to objective functions involving non-smooth regularization terms.

Theorem 2. *Suppose $\mathcal{E}$ is as in* (2)–(3) *and that* H1 *is satisfied. 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 where assumptions* H4–H6 *hold. Then we have the following*:

(i) *The set Ω^c, the complement of Ω specified in Definition* 2, *has Lebesgue measure zero in $\mathbb{R}^q$.*
(ii) *Moreover, if* H3 *is satisfied, $\overline{\Omega^c}$ has Lebesgue measure zero in $\mathbb{R}^q$.*

The proof of this theorem relies on several propositions and lemmas. Before we present them, we first exhibit some basic facts entailed by the non-smoothness of Φ. Let $\hat{x}$ be a minimizer of $\mathcal{E}(\cdot, y)$. If $G_i\hat{x} \neq \theta_i$ for all $i = 1, \dots, r$, then $(\hat{x}, y)$ is contained in a neighborhood where $\mathcal{E}$ is $\mathcal{C}^m$. So every minimizer $\hat{x}'$ of $\mathcal{E}(\cdot, y')$ satisfies $\nabla\mathcal{E}(\hat{x}', y') = 0$ and the second differential $\nabla^2\mathcal{E}(\cdot, y')$ is well defined on this neighborhood. For all $(\hat{x}', y')$ in the neighborhood, we can apply the results on smooth regularization presented in Section 3. Otherwise, all minimizers $\hat{x}$ of $\mathcal{E}(\cdot, y)$, involving at least one index i for which $G_i\hat{x} = \theta_i$, belong to the following set

$$F := \bigcup_{i=1}^{r}\{x \in \mathbb{R}^p \colon G_i x = \theta_i\}. \tag{21}$$

If $G_i \neq 0$, $\forall i \in \{1, \dots, r\}$, it is obvious that F is closed and negligible in $\mathbb{R}^p$. It is legitimate to ask what is the chance of a minimizer of $\mathcal{E}(\cdot, y)$, for some $y \in \mathbb{R}^q$, coming across to F. It has been shown in [36] that if the φ_i are $\mathcal{C}^2$ on $\mathbb{R}^s\backslash\{\theta_i\}$ and such that

$$d^+\varphi_i(\theta_i)(v) > -d^+\varphi_i(\theta_i)(-v), \qquad \forall v \in \mathbb{R}^s\backslash\{0\},$$

the minimizers $\hat{x}$ of $\mathcal{E}(\cdot, y)$ involve numerous indices i for which $G_i\hat{x} = 0$, that is $\hat{x} \in F$. When $\{G_i\}$ yield the first-order differences between adjacent neighbors, this amounts to the *stair-casing effect* which has been experimentally observed by many authors [18], [13].

The conditions for a point $\hat{x} \in F$ to be a minimizer of $\mathcal{E}(\cdot, y)$ are now more tricky than in the case when $\mathcal{E}(\cdot, y)$ is smooth in the vicinity of $\hat{x}$. Given $\hat{x} \in F$, define

$$J := \{i \in \{1, \ldots, r\}\colon\ G_i\hat{x} = \theta_i\}. \tag{22}$$

For $i \in J^{\mathrm{c}} = \{i \in \{1, \ldots, r\}\colon\ i \notin J\}$, the function $x \to \varphi_i(G_i x)$ is differentiable on a neighborhood of $\hat{x}$ in the usual sense. This suggests we introduce the following partial objective function,

$$\mathcal{E}_J(x, y) = \|Lx - y\|^2 + \sum_{i \in J^{\mathrm{c}}} \varphi_i(G_i x),$$

which is $\mathcal{C}^m$ on a neighborhood of $\hat{x}$. Moreover, for every $y \in \mathbb{R}^q$, we see that $\mathcal{E}_J(\cdot, y)$ is $\mathcal{C}^m$ at any x belonging to the set

$$\Theta_J := \left\{ x \in \mathbb{R}^p\colon \begin{bmatrix} G_i x = \theta_i \text{ for all } i \in J \\ G_i x \neq \theta_i \text{ for all } i \in J^{\mathrm{c}} \end{bmatrix} \right\}. \tag{23}$$

Incidentally, $\overline{\Theta}_J$ is an affine space and Θ_J is a differentiable manifold. The relevant tangent space at any point of Θ_J is denoted T_J and reads

$$T_J = \bigcap_{i \in J} \operatorname{Ker} G_i. \tag{24}$$

Notice that the family of all Θ_J, when J ranges over $\mathcal{P}(\{1, \ldots, r\})$, forms a partition of $\mathbb{R}^p$. We can notice also that

$$\bigcup_{J \in \mathcal{P}(\{1,\ldots,r\})} \left\{ y \in \mathbb{R}^q\colon\ \exists \hat{x} \in \Theta_J \text{ minimizer of } \mathcal{E}(\cdot, y) \right\}$$

is a covering of $\mathbb{R}^q$ provided that for every y, $\mathcal{E}(\cdot, y)$ admits at least one minimizer. In particular, this is a partition of $\mathbb{R}^q$ if $\mathcal{E}(\cdot, y)$ admits a unique strict minimizer for every y.

Any minimizer $\hat{x}$ of $\mathcal{E}(\cdot, y)$ satisfies

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

Let J be associated with $\hat{x}$ according to (22). For any $x \in \Theta_J$ and $v \in \mathbb{R}^p$ we have

$$d^+\mathcal{E}(x, y)(v) = v^T \nabla \mathcal{E}_J(x, y) + \sum_{i \in J} (G_i v)^T \nabla^+ \varphi_i(\theta_i)(G_i v), \tag{25}$$

with

$$\nabla \mathcal{E}_J(x, y) = 2L^T(Lx - y) + \sum_{i \in J^{\mathrm{c}}} G_i^T \nabla \varphi_i(G_i x). \tag{26}$$

The restriction of $\mathcal{E}(\cdot, y)$ to the manifold Θ_J satisfies

$$\mathcal{E}|_{\Theta_J}(\cdot, y) = \mathcal{E}_J|_{\Theta_J}(\cdot, y) + K \qquad \text{where} \quad K = \sum_{i \in J} \varphi_i(\theta_i),$$

and consequently $\mathcal{E}|_{\Theta_J}(\cdot, y)$ is $\mathcal{C}^m$ on Θ_J. Based on these expressions, we formulate a result which extends Lemma 1. The proofs of all statements given in what follows are detailed in the Appendix.

Proposition 1. *Consider $\mathcal{E}$ defined as in* (2)–(3) *and $y \in \mathbb{R}^q$. For all $i \in \{1, \ldots, r\}$, let φ_i be $\mathcal{C}^m$ on $\mathbb{R}\backslash\{\theta_i\}$ with $m \geq 2$ and continuous at θ_i where assumptions* H4–H6 *hold. Given $\hat{x} \in \mathbb{R}^p$, let J be defined as in* (22). *Suppose $\hat{x}$ is a local minimizer of $\mathcal{E}|_{\Theta_J}(\cdot, y)$ such that*:

(A) $\nabla^2 \left(\mathcal{E}|_{\Theta_J}\right)(\hat{x}, y)$ *is positive definite*;
(B) *if J is non-empty,*

$$d^+\mathcal{E}(\hat{x}, y)(v) > 0, \qquad \forall v \in T_J^\perp \cap S.$$

Then there exist $\rho > 0$ and a unique $\mathcal{C}^{m-1}$ local minimizer function $\mathcal{X}\colon B(y, \rho) \to \mathbb{R}^p$ such that $\hat{x} = \mathcal{X}(y)$. Moreover, $\mathcal{X}(y') \in \Theta_J$ for all $y' \in B(y, \rho)$.

Remark 5. For $y' \in B(y, \rho)$, since $\mathcal{X}(y') \in \Theta_J$, we can say that minimizing $\mathcal{E}(\cdot, y')$ on a neighborhood N of $\hat{x}$, such that $G_i x \neq \theta_i$, for all $i \in J^c$, is equivalent to solving the problem: minimize $\mathcal{E}_J(\cdot, y')$ on N under the constraint $G_i x = \theta_i$, for all $i \in J$. Notice that $\mathcal{E}_J(\cdot, y')$ is smooth on N and the latter problem is classically solved using Lagrange multipliers. Our conditions can be reformulated in such a way that using Theorem 6 on p. 34 of [21], or Theorem 6.1.1 on p. 275 of [28], we can deduce that $\mathcal{X} \in \Theta_J$—the function exhibited in our Proposition 1—solves the latter constrained optimization problem. This could be used as a starting point in the proof of Proposition 1, in which case we must also show that $\mathcal{X}$ solves our unconstrained minimization problem as well. Notice that when y' is far enough from $B(y, \rho)$, the relevant local minimizer $\hat{x}'$ of $\mathcal{E}(\cdot, y')$ will satisfy another set of constraints $J' \neq J$.

All data points $y \in \mathbb{R}^q$ for which all local minimizers of $\mathcal{E}(\cdot, y)$ satisfy the conditions of Proposition 1 clearly belong to Ω. Reciprocally, its complement Ω^c is included in the set of those data points y for which the conditions of Proposition 1 are liable to fail. As previously, we will try to confine the latter set to a closed negligible subset of $\mathbb{R}^q$.

Corollary 1. *Let $\mathcal{E}$ be as in Proposition* 1. *For $J \in \mathcal{P}(\{1, \ldots, r\})$, define*

$$H_0^J := \{x \in \Theta_J\colon \det \nabla^2(\mathcal{E}|_{\Theta_J})(x, 0) = 0\}, \tag{27}$$

$$W_J := \left\{ w \in T_J^\perp\colon v^T w \leq \sum_{i \in J} (G_i v)^T \nabla^+ \varphi_i(\theta_i)(G_i v),\ \forall v \in T_J^\perp \right\}. \tag{28}$$

Let Π_{T_J} be the orthogonal projection onto T_J. Put

$$A_J := \{y \in \mathbb{R}^q\colon 2\Pi_{T_J} L^T y \in \nabla(\mathcal{E}|_{\Theta_J})(H_0^J, 0)\}, \tag{29}$$

$$B_J := \{y \in \mathbb{R}^q\colon 2L^T y \in \nabla\mathcal{E}_J(\Theta_J, 0) + \partial_{T_J^\perp} W_J\}, \tag{30}$$

where $\partial_{T_J^\perp} W_J$ is the boundary of W_J considered in $T_J^\perp$. Then Ω^c, the complement of Ω in $\mathbb{R}^q$ introduced in Definition 2, *satisfies*

$$\Omega^c \subseteq \bigcup_{J \in \mathcal{P}(\{1,\ldots,r\})} (A_J \cup B_J). \tag{31}$$

The reasoning underlying Corollary 1 can be summarized in the following way. The set A_J in (29) contains all the $y \in \mathbb{R}^q$ which lead to a stationary point of $\mathcal{E}|_{\Theta_J}(\cdot, y)$ belonging to Θ_J where the Hessian of $\mathcal{E}|_{\Theta_J}(\cdot, y)$ is singular, i.e. for which condition (A) of Proposition 1 is not valid. For any J non-empty, B_J contains all y for which $\mathcal{E}(\cdot, y)$ can exhibit minimizers for which condition (B) of Proposition 1 fails. It remains to consider the extent of the sets A_J and B_J. The set A_J is addressed next.

Proposition 2. *Let $\mathcal{E}$ be as in Proposition* 1. *Then*:

(i) *The set A_J, defined in* (29), *is negligible in $\mathbb{R}^q$.*
(ii) *If all φ_i satisfy* H3, *the closure of A_J is a negligible subset of $\mathbb{R}^q$.*

Although the proof is totally different, we have a similar statement for B_J.

Proposition 3. *Let $\mathcal{E}$ be as in Proposition* 1. *Then*:

(i) *The set B_J, defined in* (30), *is negligible in $\mathbb{R}^q$.*
(ii) *If the assumption* H3 *is true*, *the closure of B_J is a negligible subset of $\mathbb{R}^q$.*

The proof of Theorem 2 is a straightforward consequence of Corollary 1 and Propositions 2 and 3.

Appendix

The lemma below is used in the proof of Theorem 1 and Proposition 2.

Lemma 2. *Let M and N be two real vector spaces of the same finite dimension. Consider a closed subset H of M. Let $\mathcal{G}$ be a continuous function from H to N such that*

1. *$\mathcal{G}(H)$ is a negligible subset of N*;
2. *there is a point $x_0 \in H$ as well as a positive constant C, such that for all $x \in H$ satisfying $\|x - x_0\| > C$ we have $\|\mathcal{G}(x) - \mathcal{G}(x_0)\| \geq C\,\|x - x_0\|$.*

Then $\overline{\mathcal{G}(H)}$ is a negligible subset of N.

Proof. Let $x_0 \in H$ and $C > 0$ be as required in assumption 2. Then for all $\alpha > C$,

$$\mathcal{G}(H) \cap \overline{B(\mathcal{G}(x_0), C\alpha)} \subset \mathcal{G}(H \cap \overline{B(x_0, \alpha)}) \subset \mathcal{G}(H). \tag{32}$$

The second inclusion is evident. The first one comes from the following facts. Let $y \in \mathcal{G}(H)\backslash\mathcal{G}(H \cap \overline{B(x_0, \alpha)})$. Hence there is x such that $y = \mathcal{G}(x)$ and $\|x - x_0\| > \alpha > C$. By assumption 2, we get $\|\mathcal{G}(x) - \mathcal{G}(x_0)\| \geq C\alpha$. This means that $y \notin \mathcal{G}(H) \cap \overline{B(\mathcal{G}(x_0), C\alpha)}$.

Furthermore, as $\mathcal{G}$ is continuous and $H \cap \overline{B(x_0, \alpha)}$ is compact, $\mathcal{G}(H \cap \overline{B(x_0, \alpha)})$ is also compact. By the last inclusion in (32), $\mathcal{G}(H \cap \overline{B(x_0, \alpha)})$ is a negligible subset of N. Then, by the first inclusion, $\mathcal{G}(H) \cap \overline{B(x_0, C\alpha)}$ is included in a negligible compact set. This is true for all α, so by making α tend to infinity, we obtain the conclusion. ☐

Proof of Theorem 1(ii)

By the continuity of $\nabla\mathcal{E}(\cdot, 0)$, the set H_0 is closed. Let us check whether assumption 2 of Lemma 2 is true for $\mathcal{G} = \nabla\mathcal{E}(\cdot, 0)$ and $H = H_0$. We have

$$\|\mathcal{G}(x)\| \geq 2\|L^T Lx\| - \|\nabla\Phi(x)\|.$$

Moreover, $\|L^T Lx\| \geq \lambda^2\|x\|$ for any $x \in \mathbb{R}^p$, where λ^2 is the least eigenvalue of $L^T L$; since L is injective, $\lambda^2 > 0$. Next, by assumption H2, there is $C > 0$ such that $\|x\| > C$ leads to $\|\nabla\Phi(x)\| \leq \lambda^2\|x\|$. Hence, assumption 2 of Lemma 2 is true for $x_0 = 0$, a fact which allows us to deduce that $\overline{\nabla\mathcal{E}(H_0, 0)}$ is a negligible subset of $\mathbb{R}^p$. Using that L is injective, we see that $\overline{\Omega_0^c}$ is negligible, which proves (ii). □

Proof of Proposition 1

As a first stage we consider the consequences of assumption (A). Point $\hat{x}$ satisfies

$$\nabla(\mathcal{E}|_{\Theta_J})(\hat{x}, y) = 0, \tag{33}$$

$$\nabla^2(\mathcal{E}|_{\Theta_J})(\hat{x}, 0) \text{ is positive definite.} \tag{34}$$

By the Implicit Functions Theorem, there are $\rho_1 > 0$ and a unique $\mathcal{C}^{m-1}$-function $\mathcal{X}_J\colon B(y, \rho_1) \to \Theta_J$ such that

$$\nabla\left(\mathcal{E}|_{\Theta_J}\right)(\mathcal{X}_J(y'), y') = 0 \qquad \text{when} \quad y' \in B(y, \rho_1). \tag{35}$$

In addition, by (34) and the fact that $\nabla^2(\mathcal{E}|_{\Theta_J})$ is continuous, there is $\nu_1 > 0$ such that $\nabla^2(\mathcal{E}|_{\Theta_J})(x, 0)$ is positive definite whenever $x \in B(\hat{x}, \nu_1)$. Since $\mathcal{X}_J$ is continuous, there is $\rho_2 \in (0, \rho_1)$ such that $\mathcal{X}_J\,(B(y, \rho_2)) \subseteq B(\hat{x}, \nu_1)$. In other words, $\nabla^2(\mathcal{E}|_{\Theta_J})(\mathcal{X}_J(y'), y')$ is positive definite if $y' \in B(y, \rho_2)$. This fact, combined with (35) shows that $\mathcal{X}_J$ is a local minimizer function on $B(y, \rho_2)$, relevant to $\mathcal{E}|_{\Theta_J}$.

By also taking into account the consequences of assumption (B), we will show that for every y' belonging to a neighborhood of y, the point $\hat{x}' := \mathcal{X}_J(y') \in \Theta_J$ is a strict minimizer of the relevant non-restricted objective function $\mathcal{E}(\cdot, y')$. To this end, we analyze the growth of $\mathcal{E}(\cdot, y')$ near to an $\hat{x}' \in B(\hat{x}, \nu_1/2)$ along arbitrary directions $v \in \mathbb{R}^p$. Since any $v \in \mathbb{R}^p$ is decomposed in a unique way into

$$v = v_J + v_J^{\perp} \qquad \text{with} \quad v_J \in T_J \quad \text{and} \quad v_J^{\perp} \in T_J^{\perp},$$

we can write

$$\begin{aligned}\mathcal{E}(\hat{x}' + v, y') - \mathcal{E}(\hat{x}', y') &= [\mathcal{E}(\hat{x}' + v_J + v_J^{\perp}, y') - \mathcal{E}(\hat{x}' + v_J, y')] \\ &\quad + [\mathcal{E}(\hat{x}' + v_J, y') - \mathcal{E}(\hat{x}', y')].\end{aligned} \tag{36}$$

The sign of the two terms between the brackets will be checked separately. The fact that $G_i\hat{x}' \neq \theta_i$ for all $i \in J^c$ entails that there is $\nu_2 \in (0, \nu_1)$ such that $G_i(\hat{x}' + v) \neq \theta_i$ for all $i \in J^c$, if $\|v\| < \nu_2$. In such a case, $\hat{x}' + v_J \in \Theta_J$, so we have

$$\mathcal{E}(\hat{x}' + v_J, y') - \mathcal{E}(\hat{x}', y') = \mathcal{E}|_{\Theta_J}(\hat{x}' + v_J, y') - \mathcal{E}|_{\Theta_J}(\hat{x}', y').$$

Because, by construction, $\hat{x}'$ is a minimizer of $\mathcal{E}|_{\Theta_J}(\cdot, y')$, for any $y' \in B(y, \rho_2)$ there exists $\nu_3 \in (0, \nu_2)$ such that

$$\mathcal{E}|_{\Theta_J}(\hat{x}' + v_J, y') - \mathcal{E}|_{\Theta_J}(\hat{x}', y') > 0 \qquad \text{if} \quad 0 < \|v_J\| < \nu_3. \tag{37}$$

Now we focus on the first term on the right side of (36) which will be shown to be positive when $\|v\|$ is small enough. Instead of $\hat{x}' + v_J \in \Theta_J$, we consider any $x' \in \Theta_J$ in a neighborhood of $\hat{x}$. We show that for any y' near y, the function $\mathcal{E}(\cdot, y')$ reaches a strict minimum at such an x' in the direction of $T_J^{\perp}$.

Since, by H5, $u \mapsto d^+\varphi_i(\theta_i)(u)$ is lower semi-continuous on S^s, we see that $u \mapsto d^+\mathcal{E}(\hat{x}, y)(u)$ is lower semi-continuous on S^p. Then assumption (B) implies that

$$\eta := \inf_{u \in T_J^{\perp} \cap S} d^+\mathcal{E}(\hat{x}, y)(u) > 0,$$

where the positivity of η is due to the compactness of $T_J^{\perp} \cap S$. It follows that

$$d^+\mathcal{E}(\hat{x}, y)(v_J^{\perp}) > \frac{\eta}{2}\|v_J^{\perp}\|, \qquad \forall v_J^{\perp} \in T_J^{\perp}\backslash\{0\}. \tag{38}$$

Then $\mathcal{E}(x' + v_J^{\perp}, y') - \mathcal{E}(x', y')$ will be positive for $(x', y', v_J^{\perp})$ on a neighborhood of $(\hat{x}, y, 0)$ if

$$\left|\mathcal{E}(x' + v_J^{\perp}, y') - \mathcal{E}(x', y') - d^+\mathcal{E}(\hat{x}, y)(v_J^{\perp})\right| < \frac{\eta}{2}\|v_J^{\perp}\|. \tag{39}$$

In order to show this statement, for $v_J^{\perp} \in T_J^{\perp}$, we define

$$I := \{i \in \{1, \dots, r\}\colon\ G_i v_J^{\perp} = 0\}.$$

Then for $x' \in \Theta_J$ near $\hat{x}$, we have

$$\begin{aligned}
&\mathcal{E}(x' + v_J^{\perp}, y') - \mathcal{E}(x', y') \\
&\quad = 2(Lv_J^{\perp})^T(Lx' - y') + \|Lv_J^{\perp}\|^2 + \sum_{i \in I^c}[\varphi_i(G_i x' + G_i v_J^{\perp}) - \varphi_i(G_i x')].
\end{aligned}$$

The one-sided derivative of $\mathcal{E}$ given in (25), is written

$$\begin{aligned}
d^+\mathcal{E}(\hat{x}, y)(v_J^{\perp}) &= 2(Lv_J^{\perp})^T(L\hat{x} - y) + \sum_{i \in J^c \cap I^c}(G_i v_J^{\perp})^T \nabla\varphi_i(G_i\hat{x}) \\
&\quad + \sum_{i \in J \cap I^c}(G_i v_J^{\perp})^T \nabla^+\varphi_i(\theta_i)(G_i v_J^{\perp}).
\end{aligned}$$

Based on the last two equations,

$$\begin{aligned}
&\left|\mathcal{E}(x' + v_J^{\perp}, y') - \mathcal{E}(x', y') - d^+\mathcal{E}(\hat{x}, y)(v_J^{\perp})\right| \\
&\quad \le \Bigg|(v_J^{\perp})^T\Bigg(2L^TL(x' - \hat{x}) - 2L^T(y' - y) + L^TLv_J^{\perp} && (40)\\
&\qquad\qquad - \sum_{i \in J^c \cap I^c} G_i^T(\nabla\varphi_i(G_i\hat{x}) - \nabla\varphi_i(G_i x'))\Bigg)\Bigg| && (41)\\
&\qquad + \sum_{i \in J^c \cap I^c}\left|\varphi_i(G_i x' + G_i v_J^{\perp}) - \varphi_i(G_i x') - (v_J^{\perp})^T G_i^T \nabla\varphi_i(G_i x')\right| && (42)\\
&\qquad + \sum_{i \in J \cap I^c}\left|\varphi_i(\theta_i + G_i v_J^{\perp}) - \varphi_i(\theta_i) - d^+\varphi_i(\theta_i)(G_i v_J^{\perp})\right|. && (43)
\end{aligned}$$

The expression in (40)–(41) is bounded by

$$\|v_J^\perp\|\Bigg(2\|L^TL\|\,\|x'-\hat{x}\|+2\|L\|\,\|y'-y\|+\|L^TL\|\,\|v_J^\perp\|$$
$$+\sum_{i\in J^c\cap I^c}\|G_i\|\,\|\nabla\varphi_i(G_i\hat{x})-\nabla\varphi_i(G_ix')\|\Bigg).$$

The term between the parentheses will be smaller than $\eta/6$ if $(x', y', v_J^\perp)$ is close enough to $(\hat{x}, y, 0)$. Hence the term in (40)–(41) is upper bounded by $(\eta/6)\|v_J^\perp\|$. As the functions φ_i are at least $\mathcal{C}^1$ in a neighborhood of G_ix' when $i\in J^c\cap I^c$, the expression in (42) can be bounded above by $(\eta/6)\|v_J^\perp\|$. Last, by hypothesis H6, the expression (43) can be bounded by $(\eta/6)\|v_J^\perp\|$ as well. We thus obtain that the expression in (39) is smaller than $\eta\|v_J^\perp\|$. Hence the conclusion. □

Proof of Corollary 1

Let $y\in\Omega^c$, then $\mathcal{E}(\cdot, y)$ admits at least one minimizer $\hat{x}\in\mathbb{R}^p$ such that the conclusion of Proposition 1 fails. Let J be calculated according to (22), then $\hat{x}\in\Theta_J$. Clearly, $\hat{x}$ is also a stationary point of $\mathcal{E}|_{\Theta_J}(\cdot, y)$, which means that

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

By noticing that for every direction $v\in T_J$ we have $\mathcal{E}(\hat{x}+v, y)=\mathcal{E}_J(\hat{x}+v, y)+K$, where $K=\sum_{i\in J}\varphi_i(\theta_i)$ is independent of v, we see that

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

We deduce

$$2\Pi_{T_J}L^Ty=\Pi_{T_J}\nabla\mathcal{E}_J(\hat{x}, 0)=\nabla(\mathcal{E}|_{\Theta_J})(\hat{x}, 0). \tag{44}$$

Since y is in Ω^c, at least one of the conditions (A) or (B) of Proposition 1 is not satisfied. If (A) fails, we have

$$\det\nabla^2\left(\mathcal{E}|_{\Theta_J}(\hat{x}, y)\right)=0,$$

which means that $\hat{x}\in H_0^J$. Since $\hat{x}$ satisfies (44) as well, it follows that $y\in A_J$. It is easy to see that these considerations are trivially satisfied if $J=\emptyset$.

Next, we focus on the case when (B) fails. In the particular case when $J=\emptyset$, (28) shows that $W_\emptyset=\emptyset$, since $T_\emptyset=\mathbb{R}^p$. Consequently, $B_\emptyset=\emptyset$ as well. We now consider the case when J is non-empty. Since $\hat{x}$ is a minimizer of $\mathcal{E}(\cdot, y)$,

$$d^+\mathcal{E}(\hat{x}, y)(v)=v^T\nabla\mathcal{E}_J(\hat{x}, 0)-2v^TL^Ty$$
$$+\sum_{i\in J}v^TG_i^T\nabla^+\varphi_i(\theta_i)(G_iv)\geq 0,\qquad \forall v\in T_J^\perp,$$

which expression comes from (25). Using the definition of W_J in (28), the latter expression is equivalent to

$$2\Pi_{T_J^\perp}L^Ty-\Pi_{T_J^\perp}\nabla\mathcal{E}_J(\hat{x}, 0)\in W_J.$$

Saying that (B) fails means that $\exists v \in T_J^\perp$, $v \neq 0$ such that $d^+\mathcal{E}(\hat{x}, y)(v) = 0$. Hence we can write

$$2\Pi_{T_J^\perp} L^T y - \Pi_{T_J^\perp} \nabla\mathcal{E}_J(\hat{x}, 0) \in \partial_{T_J^\perp} W_J.$$

Since $\hat{x}$ minimizes $\mathcal{E}(\cdot, y)$, (44) is true. Adding it to the expression above yields

$$2L^T y \in \nabla\mathcal{E}_J(\hat{x}, 0) + \partial_{T_J^\perp} W_J.$$

Hence $y \in B_J$. □

Proof of Proposition 2

By (27), H_0^J is the set of the critical points of the $\mathcal{C}^1$-mapping $\nabla\left(\mathcal{E}|_{\Theta_J}\right)(\cdot, 0)\colon \Theta_J \to T_J$. Then Sard's theorem [39], [33] shows that the set $\nabla(\mathcal{E}|_{\Theta_J})(H_0^J, 0)$ is negligible in T_J. Noticing that $\Pi_{T_J} L^T$ is surjective shows that A_J, given in (29), is negligible in $\mathbb{R}^q$, hence (i). Similarly to Theorem 1, assumptions H1 and H3 show that $\overline{\nabla\left(\mathcal{E}|_{\Theta_J}\right)(\cdot, 0)}$ satisfies condition 2 of Lemma 2. The same lemma then implies that $\overline{\nabla\left(\mathcal{E}|_{\Theta_J}\right)(H_0^J, 0)}$ is negligible in T_J. Using the surjectivity of $\Pi_{T_J} L^T$ again yields (ii).

Proof of Proposition 3

The proof of this proposition relies on the following theorem whose proof can be found for instance in [43].

Theorem 3. *Let U be an open subset of $\mathbb{R}^n$ and let $f\colon U \to \mathbb{R}^n$ be a locally Lipschitz function. If W is a negligible subset of U, then $f(W)$ is a negligible subset of $\mathbb{R}^n$.*

As $B_\emptyset = \emptyset$, we just have to prove the proposition for $J \neq \emptyset$. Since W_J is convex, $\partial_{T_J^\perp} W_J$ is negligible in $T_J^\perp$, hence the set $\Theta_J + \partial_{T_J^\perp} W_J$ is negligible in $\mathbb{R}^p$. By noticing that the function $x + \tilde{x} \mapsto \nabla\mathcal{E}_J(x, 0) + \tilde{x}$ is $\mathcal{C}^1$ on $\Theta_J + T_J^\perp = \mathbb{R}^p$, Theorem 3 shows that $\nabla\mathcal{E}_J(\Theta_J, 0) + \partial_{T_J^\perp} W_J$ is also negligible in $\mathbb{R}^p$. Since L is injective, the set B_J in (30) is negligible in $\mathbb{R}^q$.

In order to prove (ii), we show that under assumption H3, $\overline{B_J}$ is also negligible in $\mathbb{R}^q$. Since L is injective, this is true provided that $\overline{\nabla\mathcal{E}_J(\Theta_J, 0)} + \partial_{T_J^\perp} W_J$ is negligible in $\mathbb{R}^p$. The latter statement is shown below. The term $\overline{\nabla\mathcal{E}_J(\Theta_J, 0)}$ reads

$$\begin{aligned}&\overline{\nabla\mathcal{E}_J(\Theta_J, 0)}\\ &\qquad = \left\{\lim_{n\to\infty} \nabla\mathcal{E}_J(x_n, 0)\colon\ x_n \in \Theta_J,\ \forall n \in \mathbb{N} \text{ and } \lim_{n\to\infty} \nabla\mathcal{E}_J(x_n, 0) \text{ exists}\right\}. \end{aligned} \tag{45}$$

Assumption H3, joined to the fact that $\nabla\mathcal{E}_J(x_n, 0)$ is bounded when $n \to \infty$, implies that $\{x_n\}_{n\in\mathbb{N}}$ is also bounded. Consequently, $\{x_n\}_{n\in\mathbb{N}}$ admits a subsequence which converges in $\overline{\Theta_J}$; by a slight abuse of notation, the latter will be denoted by $\{x_n\}_{n\in\mathbb{N}}$ again. Let $\bar{x} := \lim_{n\to\infty} x_n$. Then $\overline{x} \in \overline{\Theta_J}$ where

$$\overline{\Theta_J} = \bigcup_{I \subset J^c} \Theta_{J\cup I}.$$

Since all the sets $\Theta_{J\cup I}$ in the above union are disjoint, there is a unique $I_0 \subset J^{\mathrm{c}}$ such that $\overline{x} \in \Theta_{J\cup I_0}$.

If $I_0 = \emptyset$, we can write that

$$\lim_{n\to\infty} \nabla\mathcal{E}_J(x_n, 0) = \nabla\mathcal{E}_J(\overline{x}, 0).$$

For $I_0 \neq \emptyset$, the considerations are developed in several stages. Starting with I_0, for every $k = 1, 2, \ldots$, we define recursively

$$u_k := \lim_{n\to\infty} \mathcal{N}\big(\Pi_{T_J\cap(\cap_{i\in I_{k-1}}\mathrm{Ker}\, G_i)^{\perp}}(x_n - \overline{x})\big), \tag{46}$$

$$I_k := \{i \in I_{k-1}\colon\ G_i u_k = 0\}. \tag{47}$$

The limit in (46) is taken over an arbitrary convergent subsequence. More precisely, for every k, we recursively extract a subsequence of $\{x_n\}$ that is denoted $\{x_n\}$ again, and which ensures the existence of the limit. Clearly, u_k is well defined only when $I_{k-1} \neq \emptyset$. For k small enough, the definition of u_k shows that $u_k \notin \bigcap_{I_{k-1}} \mathrm{Ker}\, G_i$, hence there exists $i \in I_{k-1}$ for which $G_i u_k \neq 0$. Consequently $\{I_k\}_{k\in\mathbb{N}}$ is strictly decreasing whenever I_k is non-empty. It follows that there is an integer K, with $1 \le K \le r$, such that the sequence $\{I_k\}_{k\in\{0,\ldots,K\}}$ is strictly decreasing with respect to the inclusion relation, and $I_K = \emptyset$. The expressions in (46) and (47) are considered in the lemmas presented below.

Lemma 3. *For every $k \in \{1, \ldots, K\}$ we have $u_k \in U_k$ where*

$$U_k := \begin{cases} \left(T_J \cap \left(\bigcap_{i\in I_{k-1}\setminus I_k} \mathrm{Ker}\, G_i\right)^{\perp} \cap \left(\bigcap_{i\in I_k} \mathrm{Ker}\, G_i\right)\right) \\ \qquad\qquad \setminus \left(\bigcup_{i\in I_{k-1}\setminus I_k} \mathrm{Ker}\, G_i\right) & \text{if} \quad k < K, \\ \left(T_J \cap \left(\bigcap_{i\in I_{K-1}} \mathrm{Ker}\, G_i\right)^{\perp}\right) \setminus \left(\bigcup_{i\in I_{k-1}\setminus I_k} \mathrm{Ker}\, G_i\right) & \text{if} \quad k = K. \end{cases}$$

Proof. By the definitions of u_k and I_k,

$$u_k \in T_J \cap \left(\bigcap_{i\in I_{k-1}} \mathrm{Ker}\, G_i\right)^{\perp} \quad \text{and} \quad u_k \in \bigcap_{i\in I_k} \mathrm{Ker}\, G_i,$$

respectively. Hence, u_k belongs to the intersection of the above sets. By using the following trivial decomposition when $k < K$,

$$\begin{aligned} \left(\bigcap_{i\in I_{k-1}} \mathrm{Ker}\, G_i\right)^{\perp} &= \left(\left(\bigcap_{i\in I_{k-1}\setminus I_k} \mathrm{Ker}\, G_i\right) \cap \left(\bigcap_{i\in I_k} \mathrm{Ker}\, G_i\right)\right)^{\perp} \\ &= \left(\bigcap_{i\in I_{k-1}\setminus I_k} \mathrm{Ker}\, G_i\right)^{\perp} + \left(\bigcap_{i\in I_k} \mathrm{Ker}\, G_i\right)^{\perp} \end{aligned}$$

we find that

$$\begin{aligned}
u_k &\in T_J \cap \left(\left(\left(\bigcap_{i\in I_{k-1}\setminus I_k} \operatorname{Ker} G_i \right)^{\perp} + \left(\bigcap_{i\in I_k} \operatorname{Ker} G_i \right)^{\perp} \right) \cap \left(\bigcap_{i\in I_k} \operatorname{Ker} G_i \right) \right.\\
&= \left(T_J \cap \left(\bigcap_{i\in I_{k-1}\setminus I_k} \operatorname{Ker} G_i \right)^{\perp} \cap \left(\bigcap_{i\in I_k} \operatorname{Ker} G_i \right) \right)\\
&\quad + \left(T_J \cap \left(\bigcap_{i\in I_k} \operatorname{Ker} G_i \right)^{\perp} \cap \left(\bigcap_{i\in I_k} \operatorname{Ker} G_i \right) \right)\\
&= T_J \cap \left(\bigcap_{i\in I_{k-1}\setminus I_k} \operatorname{Ker} G_i \right)^{\perp} \cap \left(\bigcap_{i\in I_k} \operatorname{Ker} G_i \right).
\end{aligned}$$

We obtain the result relevant to $k = K$ likewise. □

Lemma 4. *If* $i \in I_{k-1}\setminus I_k$,

$$\lim_{n\to\infty} \nabla\varphi_i(G_i x_n) = \nabla^+\varphi_i(\theta_i)(\mathcal{N}(G_i u_k)).$$

Proof. From hypothesis H4, we have

$$\begin{aligned}
\lim_{n\to\infty} \nabla\varphi_i(G_i x_n) &= \lim_{n\to\infty} \nabla\varphi_i(\theta_i + G_i(x_n - \overline{x}))\\
&= \nabla^+\varphi_i(\theta_i) \left(\lim_{n\to\infty} \mathcal{N}\left(G_i(x_n - \overline{x})\right) \right)
\end{aligned}$$

provided that the limit between the parentheses is well defined. We examine the latter question. The fact that x_n and $\overline{x}$ are elements of $\overline{\Theta_J}$ implies that $x_n - \overline{x} \in T_J$ and moreover

$$\begin{aligned}
G_i(x_n - \overline{x}) &= G_i\, \Pi_{T_J}(x_n - \overline{x})\\
&= G_i\, \Pi_{T_J\cap(\cap_{j\in I_{k-1}} \operatorname{Ker} G_j)^{\perp}}(x_n - \overline{x}) + G_i\, \Pi_{T_J\cap(\cap_{j\in I_{k-1}} \operatorname{Ker} G_j)}(x_n - \overline{x})\\
&= G_i\, \Pi_{T_J\cap(\cap_{j\in I_{k-1}} \operatorname{Ker} G_j)^{\perp}}(x_n - \overline{x}).
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathcal{N}(G_i(x_n - \overline{x})) &= \mathcal{N}\big(G_i\, \Pi_{T_J\cap(\cap_{j\in I_{k-1}} \operatorname{Ker} G_j)^{\perp}}(x_n - \overline{x})\big)\\
&= \mathcal{N}\big(G_i\, \mathcal{N}\big(\Pi_{T_J\cap(\cap_{j\in I_{k-1}} \operatorname{Ker} G_j)^{\perp}}(x_n - \overline{x})\big)\big).
\end{aligned}$$

Letting $n \to \infty$, we obtain

$$\begin{aligned}
\lim_{n\to\infty} \mathcal{N}(G_i(x_n - \overline{x})) &= \mathcal{N}\left(G_i \lim_{n\to\infty} \mathcal{N}\big(\Pi_{T_J\cap(\cap_{j\in I_{k-1}} \operatorname{Ker} G_j)^{\perp}}(x_n - \overline{x})\big)\right)\\
&= \mathcal{N}(G_i u_k).
\end{aligned}$$

The last expression is well defined since $i \notin I_k$ ensures $G_i u_k \neq 0$. □

We now return to the proof of the proposition. Given $I \subset \{1, \ldots, r\}$, we introduce the function

$$F_I\colon\ \mathbb{R}^p \backslash \left\{ \bigcup_{i \in I} \operatorname{Ker} G_i \right\} \to \mathbb{R}^p,$$
$$u \mapsto F_I(u) := \sum_{i \in I} G_i^T \nabla^+ \varphi_i(\theta_i)(\mathcal{N}(G_i u)). \tag{48}$$

By the definition of I_k in (47), $u_k \notin \bigcup_{i \in I_{k-1} \backslash I_k} \operatorname{Ker} G_i$. Then, according to Lemma 4,

$$\lim_{n \to \infty} \sum_{i \in I_{k-1} \backslash I_k} G_i^T \nabla \varphi_i(G_i x_n) = F_{I_{k-1} \backslash I_k}(u_k).$$

Hence, from the definition of $\mathcal{E}_J$, we have

$$\lim_{n \to \infty} \nabla \mathcal{E}_J(x_n, 0) = \nabla \mathcal{E}_{J \cup I_0}(\overline{x}, 0) + \sum_{k=1}^{K} F_{I_{k-1} \backslash I_k}(u_k).$$

Based on (45) and Lemma 3, we can write

$$\overline{\nabla \mathcal{E}_J(\Theta_J, 0)} \subset \nabla \mathcal{E}_J(\Theta_J, 0)$$
$$\cup \left(\bigcup_{K=1}^{r} \quad \bigcup_{\{I_k\}_{k=1}^K \subset \mathcal{I}_K} \left(\nabla \mathcal{E}_{J \cup I_0}(\Theta_{J \cup I_0}, 0) + \sum_{k=1}^{K} F_{I_{k-1} \backslash I_k}\, (U_k) \right) \right), \tag{49}$$

where

$$\mathcal{I}_K := \left\{ \{I_k\}_{k=1}^K \subset (\mathcal{P}(\{1, \ldots, r\}))^K\colon\ \{I_k\}_{k=1}^K \text{ is strictly decreasing and } I_K = \emptyset \right\}.$$

Lemma 5. *Let $\{I_k\}_{k=0}^K$ be a strictly decreasing sequence* (*with respect to the inclusion relation*) *and let $\{U_k\}_{k=0}^K$ be defined as in Lemma* 3. *Then we have $\overline{U_k} \perp \overline{U_l}$ for every $k \neq l$ and*

$$T_J = T_{J \cup I_0} \oplus \left(\bigoplus_{k=1}^{K} \overline{U_k} \right).$$

Note that $U_k \neq \{0\}$ since $u_k \in U_k$ and $u_k \neq 0$. It follows that

$$\dim \left(\bigcup_{i \in I_{k-1} \backslash I_k} \operatorname{Ker} G_i \right) < \dim \left(T_J \cap \left(\bigcap_{i \in I_{k-1} \backslash I_k} \operatorname{Ker} G_i \right)^{\perp} \cap \left(\bigcap_{i \in I_k} \operatorname{Ker} G_i \right) \right).$$

Then $\overline{U_k}$ is a vector space which reads

$$\overline{U_k} = T_J \cap \left(\bigcap_{i \in I_{k-1} \backslash I_k} \operatorname{Ker} G_i \right)^{\perp} \cap \left(\bigcap_{i \in I_k} \operatorname{Ker} G_i \right).$$

Proof. This proof is based on the following identity:

$$\begin{aligned}\bigcap_{i\in I_k}\operatorname{Ker} G_i &= \left[\left(\bigcap_{i\in I_{k-1}\setminus I_k}\operatorname{Ker} G_i\right)\cap\left(\bigcap_{i\in I_k}\operatorname{Ker} G_i\right)\right]\\ &\quad\oplus\left[\left(\bigcap_{i\in I_{k-1}\setminus I_k}\operatorname{Ker} G_i\right)^{\perp}\cap\left(\bigcap_{i\in I_k}\operatorname{Ker} G_i\right)\right]\\ &=\left(\bigcap_{i\in I_{k-1}}\operatorname{Ker} G_i\right)\\ &\quad\oplus\left[\left(\bigcap_{i\in I_{k-1}\setminus I_k}\operatorname{Ker} G_i\right)^{\perp}\cap\left(\bigcap_{i\in I_k}\operatorname{Ker} G_i\right)\right]\qquad(\text{by } I_k\subset I_{k-1}).\end{aligned}$$

Consequently

$$T_J\cap\left(\bigcap_{i\in I_k}\operatorname{Ker} G_i\right)=\left[T_J\cap\left(\bigcap_{i\in I_{k-1}}\operatorname{Ker} G_i\right)\right]\oplus U_k. \tag{50}$$

By using recursively the obtained identity we get

$$\begin{aligned}T_J &= T_J\cap\left[\left(\bigcap_{i\in I_{K-1}}\operatorname{Ker} G_i\right)\oplus\left(\bigcap_{i\in I_{K-1}}\operatorname{Ker} G_i\right)^{\perp}\right]\\ &=\left[T_J\cap\left(\bigcap_{i\in I_{K-1}}\operatorname{Ker} G_i\right)\right]\oplus\overline{U_K}\\ &=\left[T_J\cap\left(\bigcap_{i\in I_{K-2}}\operatorname{Ker} G_i\right)\right]\oplus\overline{U_{K-1}}\oplus\overline{U_K}\qquad(\text{by (50)})\\ &=\cdots\\ &=\left[T_J\cap\left(\bigcap_{i\in I_0}\operatorname{Ker} G_i\right)\right]\oplus\left(\bigoplus_{k=1}^{K}\overline{U_k}\right)\\ &=T_{J\cup J_0}\oplus\left(\bigoplus_{k=1}^{K}\overline{U_k}\right).\end{aligned}$$

The proof is complete. □

We can now complete the proof of the proposition. By Lemma 5, we have the following inclusion:

$$\left(\Theta_{J\cup I_0}+\sum_{k=1}^{K}U_k+\partial_{T_J^{\perp}}W_J\right)\subset(\overline{\Theta}_J+\partial_{T_J^{\perp}}W_J).$$

Since $\partial_{T_J^\perp} W_J$ is negligible in $T_J^\perp$, the expression in the right side above determines a set which is negligible in $\mathbb{R}^p$. Hence the term in the left side is negligible as well.

Let $\tilde{x} \in \overline{\Theta_{J\cup I_0}}$ be given, then $\overline{\Theta_{J\cup I_0}} = \{\tilde{x}\} + T_{J\cup I_0}$. By Lemma 5, any $x \in \mathbb{R}^p$ can be decomposed in a unique way in the form

$$x = \tilde{x} + x_{J\cup I_0} + x_1 + \cdots + x_K + x_J^\perp,$$

where

$$x_{J\cup I_0} \in T_{J\cup I_0},$$
$$x_k \in \overline{U_k}, \quad \forall k \in \{1, \ldots, K\},$$
$$x_J^\perp \in T_J^\perp.$$

Based on this decomposition and using F_I defined in (48), the function

$$\Theta_{J\cup I_0} + \sum_{k=1}^{K} U_k + T_J^\perp \to \mathbb{R}^p,$$

$$\tilde{x} + x_{J\cup I_0} + x_1 + \cdots + x_K + x_J^\perp \mapsto \nabla\mathcal{E}_{J\cup I_0}(\tilde{x} + x_{J\cup I_0}, 0) + \sum_{k=1}^{K} F_{I_{k-1}\setminus I_k}(x_k) + x_J^\perp,$$

is locally Lipschitz since $\nabla\mathcal{E}_{J\cup I_0}$ is $\mathcal{C}^1$ and $F_{I_{k-1}\setminus I_k}$ is Lipschitz by H5. Its image when x ranges over $\Theta_{J\cup I_0} + \sum_{k=1}^{K} U_k + \partial_{T_J^\perp} W_J$, that is

$$\nabla\mathcal{E}_{J\cup I_0}(\Theta_{J\cup I_0}, 0) + \sum_{k=1}^{K} F_{I_{k-1}\setminus I_k}(U_k) + \partial_{T_J^\perp} W_J,$$

is consequently negligible in $\mathbb{R}^p$.

We prove in the same way that $\nabla\mathcal{E}_J(\Theta_J, 0) + \partial_{T_J^\perp} W_J$ is negligible in $\mathbb{R}^p$. Thus, according to (49), $\overline{\nabla\mathcal{E}_J(\Theta_J, 0)} + \partial_{T_J^\perp} W_J$ is a negligible subset of $\mathbb{R}^p$, being a finite union of negligible subsets. The proof is complete. □

References

1. F. Alizadeh, J.-P. Haeberly, and M. Overton, Complementarity and non-degeracy in semi-definite programming, Mathematical Programming, Series B, 77 (1997), 111–128.
2. S. Alliney, Digital filters as absolute norm regularizers, IEEE Transactions on Medical Imaging, MI-12 (1993), 173–181.
3. L. Alvarez, P. L. Lions, and J. M. Morel, Image selective smoothing and edge detection by nonlinear diffusion (II), SIAM Journal on Numerical Analysis, 29 (1992), 845–866.
4. G. Aubert and P. Kornprobst, Mathematical Problems in Images Processing, Springer-Verlag, Berlin, 2002.
5. A. Avez, Calcul différentiel, Masson, Paris, 1991.
6. J. E. Besag, On the statistical analysis of dirty pictures (with discussion), Journal of the Royal Statistical Society B, 48 (1986), 259–302.
7. J. E. Besag, Digital image processing: towards Bayesian image analysis, Journal of Applied Statistics, 16 (1989), 395–407.

8. M. Black, G. Sapiro, D. Marimont, and D. Heeger, Robust anisotropic diffusion, IEEE Transactions on Image Processing, 7 (1998), 421–432.
9. A. Blake, Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-11 (1989), 2–12.
10. C. Bouman and K. Sauer, A generalized Gaussian image model for edge-preserving map estimation, IEEE Transactions on Image Processing, 2 (1993), 296–310.
11. V. Caselles, J.-M. Morel, and C. Sbert, An axiomatic approach to image interpolation, IEEE Transactions on Image Processing, 7 (1998), 59–83.
12. F. Catte, T. Coll, P. L. Lions, and J. M. Morel, Image selective smoothing and edge detection by nonlinear diffusion (I), SIAM Journal on Numerical Analysis, 29 (1992), 182–193.
13. T. F. Chan and C. K. Wong, Total variation blind deconvolution, IEEE Transactions on Image Processing, 7 (1998), 370–375.
14. P. Charbonnier, L. Blanc-Féraud, and M. Barlaud, An adaptive reconstruction method involving discontinuities, in Proceedings of IEEE ICASSP, vol. V, 1993, 491–494.
15. P. G. Ciarlet, Introduction à l'analyse numérique matricielle et à l'optimisation, Collection mathématiques appliquées pour la maîtrise, Masson, Paris, 1990.
16. J.-N. Corvellec and A. Hantoute, Homotopical stability of isomated critical points of continuous functionals, Set-Valued Analysis, 10(2) (2002), 143–164.
17. G. Demoment, Image reconstruction and restoration: overview of common estimation structure and problems, IEEE Transactions on Acoustics Speech and Signal Processing, ASSP-37 (1989), 2024–2036.
18. D. Dobson and F. Santosa, Recovery of blocky images from noisy and blurred data, SIAM Journal on Applied Mathematics, 56 (1996), 1181–1199.
19. D. Donoho, I. Johnstone, J. Hoch, and A. Stern, Maximum entropy and the nearly black object, Journal of the Royal Statistical Society B, 54 (1992), 41–81.
20. A. L. Dontchev and T. Zollezi, Well-Posed Optimization Problems, Springer-Verlag, New York, 1993.
21. A. Fiacco and G. McCormic, Nonlinear Programming, Classics in Applied Mathematics, SIAM, Philadelphia, PA, 1990.
22. D. Geman and G. Reynolds, Constrained restoration and recovery of discontinuities, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-14 (1992), 367–383.
23. 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.
24. 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.
25. P. J. Green, Bayesian reconstructions from emission tomography data using a modified em algorithm, IEEE Transactions on Medical Imaging, MI-9 (1990), 84–93.
26. M. A. Hasan, Bifurcation analysis for perturbations of the first-order conditions of optimality, Journal of Mathematical Analysis and Applications, 188(3) (1994), 901–927.
27. A. Ioffe and E. Schwartzman, Metric critical point theory 1, morse regularity and homotopic stability of a minimum, Journal de Mathématiques Pures et Appliquées, 75 (1996), 125–153.
28. H. T. Jongen, P. Jonker, and T. F., Nonlinear Optimization in Finite Dimensions, 1st edn., Kluwer, Dordrecht, 1986.
29. D. Klatte and B. Kummer, Nonsmooth Equations in Optimization, Kluwer, Dordrecht, London, 2002.
30. Y. Leclerc, Constructing simple stable description for image partitioning, International Journal of Computer Vision, 3 (1989), 73–102.
31. J. Marroquin, S. Mitter, and T. Poggio, Probabilistic solution of ill-posed problems in computational vision, Journal of the American Statistical Association, 82 (1987), 76–89.
32. R. Mifflin, Semismooth and semiconvex functions in constrained optimization, SIAM Journal of Control and Optimization, 15 (1977), 959–972.
33. J. Milnor, Topology from the Differential Point of View, The University Press of Virginia, Charlottesville, VA, 1965.
34. D. Mumford and J. Shah, Boundary detection by minimizing functionals, in Proceedings of IEEE ICASSP, 1985, pp. 22–26.
35. M. Nikolova, Markovian reconstruction using a GNC approach, IEEE Transactions on Image Processing, 8 (1999), 1204–1220.

36. M. Nikolova, Weakly constrained minimization. Application to the estimation of images and signals involving constant regions, Journal of Mathematical Imaging and Vision, 21 (2004), 155–175.
37. G. Pataki, On the rank of extreme matrices in semi-definite programs and the multiplicity of optimal eigenvalues, Mathematics of Operations Research, 23(2) (1998), 339-358.
38. 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.
39. G. d. Rham, Variétés différentiables, Hermann, Paris, 1955.
40. R. T. Rockafellar and J. B. Wets, Variational Analysis, Springer-Verlag, New York, 1997.
41. L. Rudin, S. Osher, and C. Fatemi, Nonlinear total variation based noise removal algorithm, Physica, 60 D (1992), 259–268.
42. A. Shapiro and M. K. H. Fan, On eigenvalue optimization, SIAM Journal on Optimization, 3 (1995), 552–568.
43. M. A. Spivak, A Comprehensive Introduction to Differential Geometry, vol. 1, Brandeis University, Waltham, MA, 1970.
44. 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.
45. A. Tikhonov and V. Arsenin, Solutions of Ill-Posed Problems, Winston, Washington, DC, 1977.
46. J. Weickert, Anisotropic Diffusion in Image Processing, B.G. Teubner, Stuttgart, 1998.

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