STABILITY OF MINIMIZERS OF REGULARIZED LEAST SQUARES OBJECTIVE FUNCTIONS I: STUDY OF THE LOCAL BEHAVIOR

S. DURAND* AND M. NIKOLOVA †

Abstract. Many estimation problems amount to minimizing an objective function composed of a quadratic data-fidelity term and a general regularization term. It is widely accepted that the minimizers obtained using nonsmooth and/or nonconvex regularization terms are frequently good estimates. However, very 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 nonsmooth and/or nonconvex objective functions. It consists of two parts: in this part, we focus on general local minimizers, whereas in a second part, we derive results on global minimizers. Here we demonstrate that the data domain contains an open, dense subset whose elements give rise to local and global minimizers which are necessarily strict. Moreover, we show that the relevant minimizers are stable under variations of the data.

Key words. stability analysis, regularized least-squares, non-smooth analysis, non-convex analysis, signal and image processing

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. In other words, $\hat{x} \in \mathbb{R}^p$ is a *local minimizer* of the objective function $\mathcal{E}(., y)$ since $\mathcal{E}(\hat{x}, y)$ is the minimum of $\mathcal{E}(., y)$ over O. 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, $\| \,.\, \|$ denotes the Euclidean norm and $\Phi : \mathbb{R}^p \to \mathbb{R}$ 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 $\varphi_i : \mathbb{R}^s \to \mathbb{R}$ is continuous on $\mathbb{R}^s$ and $\mathcal{C}^m$-smooth everywhere except 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 [36], 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 [6]. The operator L can represent

*CMLA UMR 8536–ENS de Cachan, 61 av. President Wilson, 94235 Cachan Cedex, France & LAMFA, Université de Picardie, 33 rue Saint-Leu, 90039 Amien Cedex, France, sdurand@cmla.ens-cachan.fr

†CNRS URA820–ENST Dpt. TSI, ENST, 46 rue Barrault, 75634 Paris Cedex 13, France & CMLA-ENS Cachan, France, nikolova@tsi.enst.fr.

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

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

where $\phi : \mathbb{R}_+ \to \mathbb{R}$ is an increasing function, often called potential function. Several functions ϕ, among the most popular, are the following [20, 5, 21, 29, 22, 33, 11, 35, 7]:

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

Objective functions as specified above are based either on PDE's [29, 33, 2, 13, 12, 37], or relay on probabilistic considerations [4, 20, 16].

Most of the potential functions cited in (5) are "irregular" in the sense that they are non-convex and/or nonsmooth. Indeed, several authors pointed out the possibility of getting signals involving jumps and images with sharp edges by using nonconvex regularization functions [26, 21, 29]. On the other hand, non-smooth regularization has been shown to avoid Gibbs artifacts and to enforce local homogeneity [19, 18, 1, 27]. In spite of this, very few facts are known about the behavior, and especially about the stability, of the local minimizers relevant to non-convex objective functions. Precisely, we study how a local minimizer $\hat{x}$ of an objective function of the form (2)-(3) behaves under variations of data y. Let us mention that the principal difficulty arises in the context of nonconvex objective functions, whereas the stability of convex objective functions is already well understood [31, 23].

Readers may associate the problem of the stability of a minimizer $\hat{x}$ with the problem of the stability of the minimum-value $\mathcal{E}(\hat{x}, y)$. Remark that the stability of a minimizer does imply the stability of the relevant minimum-value, but the inverse is false in general. Some results have been obtained on the minimum-values of nonconvex functions [32, 10, 9] but they do not have a direct relation to the problem we consider.

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 nonconvex objective function in the vicinity of some initial solution. Second, it is worth recalling that no minimization algorithm guarantees the finding of the global minimum of a general nonconvex objective function. Some algorithms allow the finding of the global minimum only with high probability, under demanding requirements (*e.g.* simulated annealing) [20, 19]. Others allow the finding of a local minimum which is expected to be close to the global minimum [8]. The practically obtained solutions are thus frequently only local minimizers, hence the importance of knowing their behavior.

Our first goal is to catch the set of all $y \in \mathbb{R}^q$ for which the relevant objective function $\mathcal{E}(.,y)$ might exhibit nonstrict minima. We shall demonstrate that all these ys are contained in a negligible subset of $\mathbb{R}^q$, provided that L is injective (one-to-one). A further question is to know whether, and in what circumstances, the strict local minimizers of $\mathcal{E}(.,y)$ give rise to a continuous local minimizer function as defined below.

DEFINITION 2.1. *A function $\mathcal{X} : O \to \mathbb{R}^p$, 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 $\mathcal{E}(.,y)$ whenever $y \in O$.*

Our second goal is therefore to show that local minimizer functions are smooth on an open, dense subset of their domains. From a practical point of view, saying that a property holds for data belonging to an open, dense subset of $\mathbb{R}^q$ means that it is systematically satisfied since it could fail only for a negligible subset of $\mathbb{R}^q$ which noisy data have no chance of coming across. So, we will quantify with respect to the Lebesgue measure on $\mathbb{R}^q$ the amount of data $y \in \mathbb{R}^q$ which *assuredly* give rise either to strict local minimizers, or to local minimizer functions $\mathcal{X}$ which remain smooth on some neighborhoods. The set given below corresponds to these properties.

DEFINITION 2.2. *Let $\mathcal{E}(.,y)$ be $\mathcal{C}^m$ (with $m \geq 1$) almost everywhere on $\mathbb{R}^p$, for every $y \in \mathbb{R}^q$. Denote*

$$\Omega := \left\{ y \in \mathbb{R}^q : \begin{array}{l} \text{if } \hat{x} \text{ is a minimizer of } \mathcal{E}(.,y) \text{ then there is} \\ \text{a } \mathcal{C}^{m-1} \text{ minimizer function } \mathcal{X} : O \to \mathbb{R}^p \\ \text{such that } y \in O \subset \mathbb{R}^q \text{ and } \hat{x} = \mathcal{X}(y) \end{array} \right\}. \tag{6}$$

The set Ω, or equivalently its complementary 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

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

It is not difficult to check that the minimizer $\hat{x}$ of $\mathcal{E}(.,y)$ 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 nonstrict minimizer.
- If $y = -1$, every $\hat{x} \in [-1,0]$ is a nonstrict minimizer.
- If $y \in (-1,1)$, the minimizer is strict and constant, $\hat{x} = 0$.

Thus we find that $\Omega^c = \{-1,\ 1\}$ which means that Ω is open and dense in $\mathbb{R}$.

EXAMPLE 2. Consider

$$\begin{aligned} &\mathcal{E} : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}, \\ &\qquad (x,y) \mapsto (x_1 - x_2 - y)^2 + \beta(x_1 - x_2)^2, \end{aligned}$$

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 $\mathcal{E}(.,y)$. Hence $\Omega^c = \mathbb{R}$.

EXAMPLE 3. Consider

$$\begin{aligned} &\mathcal{E} : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}, \\ &\qquad (x,y) \mapsto (x_1 - x_2 - y)^2 + |x_1| + |x_2|. \end{aligned}$$

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

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

Consequently, $\Omega = (-1/2, 1/2)$.

Let us 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 $L : \mathbb{R}^p \to \mathbb{R}^q$ in (2) is injective,* i.e. $\operatorname{rank} L = p$.

It is not a necessary condition to have a negligible Ω^c, but it allows us to obtain results which are strong enough.

Remark 1. We do not focus properly on the question 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 $\mathcal{E}(.,y)$ admits at least one minimizer, although the results formulated next remain trivially true in the opposite situation. This comes from an astuteness in the definition of Ω in (2.2) allowing it to contain y's for which $\mathcal{E}(.,y)$ does not admit minimizers. Practically, every objective function used for the estimation on an unknown magnitude x admits minimizers. Let us recall that $\mathcal{E}(.,y)$ is guaranteed to admit minimizers if it is coercive, *i.e.* if $\mathcal{E}(x,y) \to \infty$ along with $\|x\| \to \infty$ [15, 32]. 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 $f : \mathbb{R}^p \to \mathbb{R}$, 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 will be 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 gradient and Hessian with respect to the first variable x. By $B(x,\rho)$ we will 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 will denote 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 $\mathbb{R}^n$. Last, we denote $\mathbb{R}_+ = \{t \in \mathbb{R} : t \geq 0\}$.

The reasoning underlying our work is based upon the necessary conditions for minimum, and also upon some sufficient conditions for strict minimum. We also make a recurrent use of the implicit functions theorem [3], Sard's Theorem [3] and of several results about the minimizers of a non-smooth objective function $\mathcal{E}$ of the form (2)-(3) [28]. 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 [3].

Lemma 3.1. *Suppose $\mathcal{E} : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ is any function which is $\mathcal{C}^m$, with $m \geq 2$, with respect to both arguments. 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 $\mathcal{X}(y) = \hat{x}$.

Proof. Since $\nabla\mathcal{E}(\hat{x}, y) = 0$ and $\nabla^2\mathcal{E}(\hat{x}, y)$ is an isomorphism from $\mathbb{R}^p$ to $\mathbb{R}^p$, the Implicit functions theorem tells us that there exist $\rho > 0$ and a unique $\mathcal{C}^{m-1}$-function $\mathcal{X} : B(y, \rho) \to \mathbb{R}^p$ satisfying

$$\nabla\mathcal{E}(\mathcal{X}(y'), y') = 0 \quad \text{for all} \quad y' \in B(y, \rho).$$

In other words, each $\mathcal{X}(y')$ is a stationary point of $\mathcal{E}(x, y')$ if $y' \in B(y, \rho)$. Since $\mathcal{X}$ is continuous, for ρ sufficiently small, the constant rank theorem [3] ensures that, for every $y' \in B(y, \rho)$, we have $\operatorname{rank} \nabla^2\mathcal{E}(\mathcal{X}(y'), y') \geq \operatorname{rank} \nabla^2\mathcal{E}(\hat{x}, y) = p$, and hence $\operatorname{rank} \nabla^2\mathcal{E}(\mathcal{X}(y'), y') = p$. Then every $\mathcal{X}(y')$, relevant to $y' \in B(y, \rho)$, is a strict local minimizer of $\mathcal{E}(., y')$. $\square$

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 $m \geq 2$. If for a given $y \in \mathbb{R}^q$, a point $\hat{x} \in \mathbb{R}^p$ is a strict or non-strict local minimizer of $\mathcal{E}(., y)$, then

$$\nabla\mathcal{E}(\hat{x}, y) = 0, \tag{7}$$

$$\text{where} \quad \nabla\mathcal{E}(x, y) = 2L^T(Lx - y) + \nabla\Phi(x). \tag{8}$$

Using the fact that

$$\nabla\mathcal{E}(\hat{x}, 0) = 2L^T L\hat{x} + \nabla\Phi(\hat{x}), \tag{9}$$

the variables $\hat{x}$ and y can be separated in equation (7) which then becomes:

$$2L^T y = \nabla\mathcal{E}(\hat{x}, 0). \tag{10}$$

A point $\hat{x}$, satisfying (10), is guaranteed to be a strict minimizer of $\mathcal{E}(., y)$ if the Hessian of $\mathcal{E}(., y)$ at $\hat{x}$, namely $\nabla^2\mathcal{E}(\hat{x}, y)$, is positive definite. Furthermore, the Hessian of $\mathcal{E}(., y)$ at an arbitrary x reads

$$\nabla^2\mathcal{E}(x, y) = 2L^T L + \nabla^2\Phi(x). \tag{11}$$

We emphasize the fact that the Hessian of $\mathcal{E}(., y)$ is independent of y at any $x \in \mathbb{R}^p$, by writing $\nabla^2\mathcal{E}(x, 0)$ instead of $\nabla^2\mathcal{E}(x, y)$. Based on Lemma 3.1, we cannot guarantee that a point $\tilde{x}$ satisfying (7) is a strict minimizer of $\mathcal{E}(., y)$ if $\nabla^2\mathcal{E}(\tilde{x}, 0)$ is singular. Hence all the ys leading to a nonstrict minimizer, or to a non-continuous minimizer function, are contained in a set Ω_0^c as specified below.

LEMMA 3.2. *Suppose $\mathcal{E}$ is as in (2) where Φ is an arbitrary $\mathcal{C}^m$-function on $\mathbb{R}^p$, with $m \geq 2$. Consider the set*

$$\Omega_0 := \left\{ y \in \mathbb{R}^q : \exists \tilde{x} \in H_0 \ \textit{satisfying} \ 2L^T y \neq \nabla\mathcal{E}(\tilde{x}, 0) \right\}, \tag{12}$$

where

$$H_0 := \left\{ x \in \mathbb{R}^p : \det \nabla^2\mathcal{E}(x, 0) = 0 \right\}. \tag{13}$$

Then we have

$$\Omega_0 \subset \Omega,$$

where Ω is the set introduced in Definition 2.2.

Observe that H_0 is the set of all the critical points of $\nabla\mathcal{E}(x, y)$. Since $\nabla^2\mathcal{E}(x, 0)$ is independent of y, the set H_0 is independent of y as well.

REMARK 2. If $\nabla^2\mathcal{E}(x, 0)$ is positive definite for all $x \in \mathbb{R}^p$, the set H_0 is empty and Lemma 3.2 shows that $\Omega = \mathbb{R}^q$. It is a tautology to say that in this case, there is a unique $\mathcal{C}^{m-1}$ minimizer function $\mathcal{X}$ as stated in Definition 2.2. The above condition on $\nabla^2\mathcal{E}(x, 0)$ is satisfied whenever L is injective and Φ is convex. This is readily seen from (11) where $L^T L$ is positive definite and all the eigenvalues of the second term are non-negative.

However, H_0 is generally non-empty if Φ is non-convex. This is the reason why, in the following, we rather focus on non-convex functions Φ. More specifically, we consider functions which satisfy the following assumption.

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

This assumption is satisfied by the regularization functions used by many authors [21, 29, 22]. The theorem stated next provides the principal result of this section.

THEOREM 3.3. *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 H1 is satisfied. Then we have the following:*

(i) The set Ω^c, the complementary of Ω specified in Definition 2.2, is negligible in $\mathbb{R}^q$.

(ii) Moreover, if H2 is satisfied, $\overline{\Omega^c}$ is a negligible subset of $\mathbb{R}^q$.

REMARK 3. The results (i) and (ii) of the theorem remain true if we replace Ω by Ω_0, as defined in (12). In fact, the proof of the theorem establishes these results for Ω_0. The ultimate conclusions are obtained using $\Omega^c \subset \Omega_0^c$, according to Lemma 3.2.

REMARK 4. It is straightforward that under the conditions of Theorem 3.3, the minimum-value function $y \mapsto \mathcal{E}(\mathcal{X}(y), y)$ is $\mathcal{C}^{m-1}$ smooth. The same conclusion can be drawn also for nonsmooth objective functions as considered in Theorem 4.2.

Proof. As mentioned in Remark 3, it is sufficient to prove Theorem 3.3 for Ω_0 instead of Ω. The proof of this theorem is based on three auxiliary statements which are given below. These statements are presented in a more general form which allows their application in the context of nonsmooth regularization functions, considered later in § 4.

LEMMA 3.4. *Let L be an injective linear operator between two finite-dimensional spaces M and N. Consider an arbitrary subset $V \subset M$ which is negligible in M. Define:*

$$W := \left\{y \in N : L^T y \in V\right\}.$$

Then we have

(i) W is negligible in N;

(ii) W is closed if V is closed.

PROOF OF LEMMA 3.4. Since V is negligible, for every $\varepsilon > 0$ there is a sequence of balls $\{B_i\}$ such that

$$V \subset \bigcup_{i=1}^{\infty} B_i \quad \text{and} \quad \sum_{i=1}^{\infty} \text{measure}(B_i) < \varepsilon.$$

Then

$$W \subset \bigcup_{i=1}^{\infty} \left\{ y \in N : L^T y \in B_i \right\}.$$

However, for every i we have

$$\text{measure}\left(\left\{ y \in N : L^T y \in B_i \right\}\right) \leq \frac{1}{|\lambda|} \text{measure}(B_i), \tag{14}$$

where λ is the singular value of L which is the smallest in magnitude. By assumption H 1, we are guaranted that $\lambda \neq 0$. It follows from (14) that

$$\text{measure}(W) < \frac{1}{|\lambda|} \sum_{i=1}^{\infty} \text{measure}(B_i) < \frac{\varepsilon}{|\lambda|}.$$

The point (i) is proven.

Suppose now that V is closed. The closeness of W follows from the continuity of L^T. □

The next two results concern gradient-type functions of the form (9).

Theorem 3.5 (Sard's theorem). *Let M and N be two affine vector spaces of the same dimension. For U an open subset of M, let $\mathcal{G} : U \to N$ be a $\mathcal{C}^1$-function and H denote the set of the critical points of $\mathcal{G}$:*

$$H := \left\{ x \in U : \det \nabla \mathcal{G}(x) = 0 \right\}.$$

Then $\mathcal{G}(H) := \{\mathcal{G}(x) : x \in H\}$ is a negligible subset of N.

The proof of this statement can be found *e.g.* in [30, 25].

Lemma 3.6. *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 $\mathcal{G}(H)$ is included in a closed and negligible subset of N.

Proof of Lemma 3.6. Let $x_0 \in H$ and $C > 0$ be as required in assumption 2. Then for all $\alpha > C$,

$$\mathcal{G}(H) \cap \overline{B(\mathcal{G}(x_0), C\alpha)} \subset \mathcal{G}\left(H \cap \overline{B(x_0, \alpha)}\right) \subset \mathcal{G}(H). \tag{15}$$

The second inclusion is evident. The first one comes from the following facts. Let $y \in \mathcal{G}(H) \setminus \mathcal{G}\left(H \cap \overline{B(x_0, \alpha)}\right)$. 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, $\mathcal{G}\left(H \cap \overline{B(x_0, \alpha)}\right)$ is also compact. By the last inclusion in (15), $\mathcal{G}(H \cap \overline{B(x_0, \alpha)})$ 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 tends to infinity, we obtain the desired conclusion. □

Let us now come back to the Proof of Theorem 3.3. Note that the complementary set of Ω_0 given in (12) can also be expressed as

$$\Omega_0^c = \left\{ y \in \mathbb{R}^q : L^T y \in \nabla\mathcal{E}(H_0, 0) \right\},$$

where $\nabla\mathcal{E}(x, 0)$ is given in (9) and H_0 in (13). By applying now Theorem 3.5 with the associations $\mathcal{G} = \nabla\mathcal{E}(., 0)$, $H = H_0$ and $M = N = \mathbb{R}^p$, we find that $\nabla\mathcal{E}(H_0, 0)$ is a negligible subset of $\mathbb{R}^p$. We now apply Lemma 3.4 where we identify V with $\nabla\mathcal{E}(H_0, 0)$ and N with $\mathbb{R}^q$. Thus *(i)* is proven.

We complete our reasoning in order to prove *(ii)*. By the continuity of $\nabla\mathcal{E}(., 0)$, the set H_0 is closed. Let us check whether assumption 2 of Lemma 3.6 is true for $\mathcal{G} = \nabla\mathcal{E}(., 0)$ 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, assumption H2 means that there is $C > 0$ such that $\|x\| > C$ leads to $\|\nabla\Phi(x)\| \leq \lambda^2\|x\|$. Hence, the assumption 2 of Lemma 3.6 is true for $x_0 = 0$, which fact allows us to deduce that $\nabla\mathcal{E}(H_0, 0)$ is contained in a closed, negligible subset of $\mathbb{R}^p$. The statement *(ii)* is obtained by applying Lemma 3.4 again. □

Remark 5. Even if the chances of getting a point y, yielding a nonstrict minimizer, is “almost null”, it is legitimate to ask what is the shape of H_0, as defined in (13), since it contains all the non-strict minimizers of $\mathcal{E}(., y)$, for all y. A key point in Theorem 3.3 is that $\nabla\mathcal{E}(H_0, 0)$ is negligible although H_0 itself may be of positive measure. However, we observe that for the most important classes of functions Φ, the set H_0 is negligible as well. 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 H_0 is of positive measure. Being closed, H_0 contains an open p-cell. As $\nabla^2\mathcal{E}(., 0)$ 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

$$\nabla^2\mathcal{E}(x_0, 0) = 2L^T L + \nabla^2\Phi(x_0) \tag{16}$$

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 [21, 24, 29] where the typical choices for ϕ read

$$\phi(t) = \frac{t^2}{t^2 + \alpha}, \tag{17}$$

$$\phi(t) = 1 - e^{-\alpha t^2}, \tag{18}$$

where $\alpha > 0$ is a parameter.

Let us come back to the expression of $\nabla^2 \mathcal{E}(x_0, 0)$ in (16). Observe that if there is a point x_0 such that $\nabla^2 \Phi(x_0)$ is positive definite, the set H_0 is negligible independently of the injectivity of L.

4. Objective function involving nonsmooth regularization. We shall now consider regularization terms Φ as introduced in (3), namely

$$\Phi(x) = \sum_{i=1}^{r} \varphi_i(G_i x), \tag{19}$$

where $G_i : \mathbb{R}^p \to \mathbb{R}^s$ are linear operators, for all $i = 1, \ldots, r$. We will 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 $\mathbb{R}^s \setminus \{\theta_i\}$, with $m \geq 2$, and continuous on $\mathbb{R}^s$. Typically, φ_i is nonsmooth at θ_i. Potential functions which are non-smooth at more than one point, say θ_i, can be seen as a combination of several φ_i, which are nonsmooth at θ_i, applied to the same $G_i x$. Notice that the regularization function studied in §3 can be seen as a special case of (19) corresponding to $r = 1$, $G_1 = I$ and $\varphi_1 \in \mathcal{C}^m(\mathbb{R}^s)$. In the context of piecewise smooth potential functions, the assumption H2 is specified as it follows:

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

We restrict our attention to potential functions φ_i which admit at θ_i directional derivatives for every direction $u \in \mathbb{R}^s$.

Definition 4.1. *Consider a function $f : M \to \mathbb{R}$ with M a finite-dimensional real affine space. For $x \in M$ and u in the relevant vector space, f is said to admit a one-sided directional derivative at x in the direction of u, denoted by $d^+ f(x)(u)$, if the difference quotient $t \to [f(x + tu) - f(x)]/t$ for $t \in \mathbb{R}$ has a limit when $t \searrow 0$:*

$$d^+ f(x)(u) := \lim_{t \searrow 0} \frac{f(x + tu) - f(x)}{t}.$$

In order to simplify the notations, we introduce the normalization application $\mathcal{N}$ which, with each vector v, associates its projection on the relevant unit sphere, *i.e.*

$$\mathcal{N}(v) = \frac{v}{\|v\|}. \tag{20}$$

Whenever φ_i is nonsmooth at θ_i, the directional derivative $d^+ \varphi_i(\theta_i)(u)$ is a nonlinear function of the direction u. We will 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 *directional gradient*. More rigourously, we will focus on functions φ_i which satisfy 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

$$\nabla^+ \varphi_i(\theta_i) \left(\lim_{h \to 0} \mathcal{N}(h) \right) := \lim_{h \to 0} \nabla \varphi_i(\theta_i + h). \tag{21}$$

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

$$\nabla^+ \varphi_i(\theta_i)(u) = \begin{cases} \nabla^+ \varphi_i(\theta_i) \left(\mathcal{N}(u) \right) & \text{if} \quad u \neq 0, \\ 0 & \text{if} \quad u = 0. \end{cases} \tag{22}$$

The vector $\nabla^+\varphi_i(\theta_i)(u)$ is the directional gradient of φ_i at θ_i for u, as it was mentioned above. In particular, if φ_i is smooth at θ_i, for every $u \neq 0$, we have $\nabla^+\varphi_i(\theta_i)(u) = \nabla\varphi_i(\theta_i)$, *i.e.* we get the gradient of φ_i at θ_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 $u \neq 0$. When the directional gradient $\nabla^+\varphi_i(\theta_i)$ exists, 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

$$
\begin{aligned}
d^+\varphi_i(z)(u) &= \lim_{t \searrow 0} \frac{\varphi_i(z+tu) - \varphi_i(z)}{t} \\
&= \lim_{t \searrow 0} u^T \nabla\varphi_i(z + \kappa_t tu) \quad \text{for} \quad \kappa_t \in (0,1) \qquad \text{[by the mean value theorem]} \\
&= u^T \nabla^+\varphi_i(z)(u). \qquad (23)
\end{aligned}
$$

We will also use two other assumptions which are given below.

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

Remark 6. Under assumption H5, the relation reached in (23) shows that the application $u \to d^+\varphi_i(\theta_i)(u)$ is Lipschitz on $\mathbb{R}^s$.

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

Definition 4.1 and the last assumptions are illustrated in the context of the most typical potential functions as mentioned in (4).

Example 4. Consider

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

where $\phi \in \mathcal{C}^m(\mathbb{R}_+)$, $m \geq 2$, and $\phi'(0) > 0$. The latter inequality implies that φ_i is nonsmooth at θ_i. By applying (21)-(22), it is easily obtained that

$$
\begin{cases}
\nabla\varphi_i(z) = \phi'(\|z - \theta_i\|) \dfrac{z - \theta_i}{\|z - \theta_i\|} & \text{if } z \neq \theta_i, \\[2ex]
\nabla^+\varphi_i(\theta_i)(u) = \phi'(0^+) \dfrac{u}{\|u\|} & \text{if } z = \theta_i.
\end{cases}
$$

Both assumptions H5 and H6 are clearly satisfied. The 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 (23), the directional derivative of φ_i at θ_i for u reads

$$d^+\varphi_i(\theta_i)(u) = u^T \nabla^+\varphi_i(\theta_i)(u) = \phi'(0^+)\|u\|. \qquad (24)$$

Below we extend Theorem 3.3 to objective functions involving nonsmooth regularization terms.

Theorem 4.2. *Suppose $\mathcal{E}$ is as in (2)-(3). For all $i \in \{1, \ldots, r\}$, let φ_i be $\mathcal{C}^m$ on $\mathbb{R} \setminus \{\theta_i\}$ with $m \geq 2$ and continuous at θ_i where the assumptions H4, H5 and H6 hold. Suppose that H1 is satisfied. Then we have the following:*

(i) The set Ω^c, the complementary of Ω specified in Definition 2.2, is negligible in $\mathbb{R}^q$.

(ii) Moreover, if H3 is satisfied, $\overline{\Omega^c}$ is a negligible subset in $\mathbb{R}^q$.

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

$$F := \bigcup_{i=1}^{r} \{x \in \mathbb{R}^p : G_i x = \theta_i\}. \tag{25}$$

If $G_i \neq 0$, $\forall i \in \{1,\ldots,r\}$, it is obvious that F is both closed and negligible in $\mathbb{R}^p$. Then it is legitimate to ask what is the chance of a minimizer of $\mathcal{E}(.,y)$, for some $y \in \mathbb{R}^q$, coming across to F. It has been shown in [27] that if the φ_i are $\mathcal{C}^2$ on $\mathbb{R}^s \setminus \{\theta_i\}$ and such that

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

the minimizers $\hat{x}$ of $\mathcal{E}(.,y)$ 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 [17, 14].

The conditions for a point $\hat{x} \in F$ to be a minimizer of $\mathcal{E}(.,y)$ are now more tricky than in the case when $\mathcal{E}(.,y)$ is smooth in the vicinity of $\hat{x}$. For every $\hat{x} \in F$, and for every $i \in J$, where

$$J := \{i \in \{1,\ldots,r\} : G_i\hat{x} = \theta_i\}, \tag{26}$$

the function $\varphi_i(G_i\,.)$ is nonsmooth at $\hat{x}$. Otherwise, for $i \in J^c = \{i \in \{1,\ldots,r\} : i \notin J\}$, 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,

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

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

$$\Theta_J := \left\{ x \in \mathbb{R}^p : \left[\begin{array}{lll} G_i x = \theta_i & \text{for all} & i \in J, \\ G_i x \neq \theta_i & \text{for all} & i \in J^c \end{array} \right. \right\}. \tag{27}$$

By the way, $\overline{\Theta}_J$ 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 satisfies

$$T_J = \bigcap_{i \in J} \operatorname{Ker} G_i. \tag{28}$$

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

$$\bigcup_{J \in \mathcal{P}(\{1,\cdots,r\})} \{y \in \mathbb{R}^q : \exists \hat{x} \in \Theta_J \text{ minimizer of } \mathcal{E}(.,y)\}$$

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

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

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

Let J be associated with $\hat{x}$ according to (26). For any $x \in \Theta_J$, and for any $v \in \mathbb{R}^p$, we have

$$d^+\mathcal{E}(x,y)(v) = v^T \nabla \mathcal{E}_J(x,y) + \sum_{i\in J} v^T G_i^T \nabla^+ \varphi_i(\theta_i)(G_i v), \tag{29}$$

with

$$\nabla \mathcal{E}_J(x,y) = 2L^T(Lx-y) + \sum_{i \in J^c} G_i^T \nabla \varphi_i(G_i x). \tag{30}$$

Below we shall evoke $\mathcal{E}|_{\Theta_J}(.,y)$—the restriction of $\mathcal{E}(.,y)$ to the manifold Θ_J. Note that

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

and consequently $\mathcal{E}|_{\Theta_J}(.,y)$ is $\mathcal{C}^m$ on Θ_J. Based on these expressions, we formulate a result which extends Lemma 3.1. The proofs of all statements given in what follows are detailed in the appendix.

PROPOSITION 4.3. *Consider $\mathcal{E}$ defined as in (2)-(3) and $y \in \mathbb{R}^q$. For all $i \in \{1,\dots,r\}$, let φ_i be $\mathcal{C}^m$ on $\mathbb{R} \setminus \{\theta_i\}$ with $m \geq 2$ and continuous at θ_i where the assumptions H4, H5 and H6 hold. Focus on $\hat{x} \in \mathbb{R}^p$ and let J be defined as in (26). Suppose $\hat{x}$ is a local minimizer of $\mathcal{E}|_{\Theta_J}(.,y)$ such that:*

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

(B) *if J is nonempty,*

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

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

All data points $y \in \mathbb{R}^q$ for which all local minimizers of $\mathcal{E}(.,y)$ satisfy the conditions of Proposition 4.3 clearly belong to Ω. Reciprocally, its complementary Ω^c is included in the set of those data points y for which the conditions of Proposition 4.3 are liable to fail. As previously, we will try to confine the latter set to a closed negligible subset of $\mathbb{R}^q$.

COROLLARY 4.4. *Let $\mathcal{E}$ be as in Proposition 4.3. For $J \in \mathcal{P}(\{1,\dots,r\})$, define*

$$H_0^J := \{x \in \Theta_J : \ \det \nabla^2(\mathcal{E}|_{\Theta_J})(x,0) = 0\}, \tag{31}$$

$$W_J := \left\{ w \in T_J^\perp : \ v^T w \leq \sum_{i \in J} v^T G_i^T \nabla^+ \varphi_i(\theta_i)(G_i v), \ \forall v \in T_J^\perp \right\}. \tag{32}$$

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

$$A_J := \left\{ y \in \mathbb{R}^q : 2\Pi_{T_J} L^T y \in \nabla\left(\mathcal{E}|_{\Theta_J}\right)(H_0^J, 0)\right\}, \tag{33}$$

$$B_J := \left\{ y \in \mathbb{R}^q : 2L^T y \in \nabla \mathcal{E}_J(\Theta_J, 0) + \partial_{T_J^\perp} W_J \right\}, \tag{34}$$

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 2.2, satisfies

$$\Omega^c \subseteq \bigcup_{J\in\mathcal{P}(\{1,\ldots,r\})} (A_J \cup B_J)\,. \tag{35}$$

The reasoning underlying Corollary 4.4 can be summarized in the following way. The set A_J in (33) contains all the $y \in \mathbb{R}^q$ which lead to a stationary point of $\mathcal{E}|_{\Theta_J}(.,y)$ belonging to Θ_J where the Hessian of $\mathcal{E}|_{\Theta_J}(.,y)$ is singular, i.e. for which the condition (A) of Proposition 4.3 is not valid. For any J nonempty, B_J contains all y for which $\mathcal{E}(.,y)$ can exhibit minimizers for which the condition (B) of Proposition 4.3 fails. It remains to consider the extent of the sets A_J and B_J. The set A_J is addressed next.

PROPOSITION 4.5. *Let $\mathcal{E}$ be as in Proposition 4.3. We have the following statements.*

(i) The set A_J, defined in (33), 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 the sets B_J.

PROPOSITION 4.6. *Let $\mathcal{E}$ be as in Proposition 4.3. We have the following statements.*

(i) The set B_J, defined in (34), 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 4.2 is a straightforward consequence of Corollary 4.4 and Propositions 4.5 and 4.6.

5. Appendix.

Proof of Proposition 4.3. As a first stage we will consider the consequences of the assumption (A). The point $\hat{x}$ satisfies

$$\nabla\left(\mathcal{E}|_{\Theta_J}\right)(\hat{x}, y) = 0, \tag{36}$$

$$\nabla^2\left(\mathcal{E}|_{\Theta_J}\right)(\hat{x}, 0) \ \text{ is positive definite.} \tag{37}$$

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

$$\nabla\left(\mathcal{E}|_{\Theta_J}\right)(\mathcal{X}_J(y'), y') = 0 \ \text{ when } y' \in B(y,\rho). \tag{38}$$

In addition, by (37) and the fact that $\nabla^2\left(\mathcal{E}|_{\Theta_J}\right)$ is continuous, there is $\eta > 0$ such that $\nabla^2\left(\mathcal{E}|_{\Theta_J}\right)(x, 0)$ is positive definite whenever $x \in B(\hat{x},\eta)$. Since $\mathcal{X}_J$ is continuous, for ρ small enough, we have $\mathcal{X}_J\left(B(y,\rho)\right) \subseteq B(\hat{x},\eta)$. In other words, $\nabla^2\left(\mathcal{E}|_{\Theta_J}\right)(\mathcal{X}_J(y'), y')$ is positive definite on $B(y,\rho)$. This fact, combined with (38) shows that $\mathcal{X}_J$ is a local minimizer function on $B(y,\rho)$, relevant to the restricted objective function $\mathcal{E}|_{\Theta_J}$.

By taking into account also 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 $\mathcal{E}(.,y')$. To this end, we analyze the growth of $\mathcal{E}(.,y')$ near to $\hat{x}'$ along arbitrary directions $v \in \mathbb{R}^p$. Since any $v \in \mathbb{R}^p$ is decomposed in a unique way into

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

we can write

$$\begin{aligned} \text{(39)} \qquad \mathcal{E}(\hat{x}'+v,y')-\mathcal{E}(\hat{x}',y') &= [\mathcal{E}(\hat{x}'+v_J+v_J^{\perp},y')-\mathcal{E}(\hat{x}'+v_J,y')] \\ &\quad + [\mathcal{E}(\hat{x}'+v_J,y')-\mathcal{E}(\hat{x}',y')]. \end{aligned}$$

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 $\mathcal{E}|_{\Theta_J}(.,y')$, for any $y' \in B(y,\rho)$ there exists $\nu_1 > 0$ such that

$$\text{(40)} \qquad \mathcal{E}|_{\Theta_J}(\hat{x}'+v_J,y')-\mathcal{E}|_{\Theta_J}(\hat{x}',y') > 0 \quad \text{if} \quad 0 < \|v_J\| < \nu_1.$$

Now we focus on the first term on the right side of (39) 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 $\mathcal{E}(.,y')$ reaches a strict minimum at such a 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 the 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

$$\text{(41)} \qquad d^+\mathcal{E}(\hat{x},y)(v_J^{\perp}) > \frac{\eta}{2}\|v_J^{\perp}\|, \quad \forall v_J^{\perp} \in T_J^{\perp} \setminus \{0\}.$$

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

$$\text{(42)} \qquad \left|\mathcal{E}(x'+v_J^{\perp},y')-\mathcal{E}(x',y')-d^+\mathcal{E}(\hat{x},y)(v_J^{\perp})\right| < \frac{\eta}{2}\|v_J^{\perp}\|.$$

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

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

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

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

The one-sided derivative of $\mathcal{E}$ given in (29), is written

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

Based on the last two equations,

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

The expression in (43)-(44) is bounded by

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

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 (43)-(44) 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 (45) can be bounded above by $(\eta/6)\|v_J^\perp\|$. Last, by hypothesis H6, the expression (46) can be bounded by $(\eta/6)\|v_J^\perp\|$ as well. We thus obtain that the expression in (42) is smaller than $\eta\|v_J^\perp\|$. Hence the conclusion.

Proof of Corollary 4.4. Let $y\in\Omega^c$, then $\mathcal{E}(.,y)$ admits at least one minimizer $\hat{x}\in\mathbb{R}^p$ such that the conclusion of Proposition 4.3 fails. Let J be calculated according to (26), then $\hat{x}\in\Theta_J$. Clearly $\hat{x}$ is also a stationary point of $\mathcal{E}|_{\Theta_J}(.,y)$, which means that

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

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

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

We deduce

$$2\Pi_{T_J}L^Ty=\Pi_{T_J}\nabla\mathcal{E}_J(\hat{x},0)=\nabla\left(\mathcal{E}|_{\Theta_J}\right)(\hat{x},0) \tag{47}$$

Since y is in Ω^c, at least one of the conditions (A) or (B) of Proposition 4.3 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 (47) 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 $J=\emptyset$, (32) shows that $W_\emptyset=\emptyset$, since $T_\emptyset=\mathbb{R}^p$. Consequently, $B_\emptyset=\emptyset$ as well. Let us now consider the case when J is nonempty. The fact that $\hat{x}$ is a minimizer of $\mathcal{E}(.,y)$ implies that

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

which expression comes from (29). Using the definition of W_J in (32), 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 down

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

Since $\hat{x}$ minimizes $\mathcal{E}(.,y)$, (47) is true. Adding it to the expression above yields

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

Hence $y \in B_J$.

Proof of Proposition 4.5. By applying Sard's Theorem (see Theorem 3.5) to $M = \overline{\Theta}_J$, $N = T_J$, $U = \Theta_J$ and $\mathcal{G} = \nabla\left(\mathcal{E}|_{\Theta_J}\right)(.,0)$, the set $\nabla\left(\mathcal{E}|_{\Theta_J}\right)(H_0^J,0)$, with H_0^J as in (31), is negligible in T_J. Next, we notice $\operatorname{rank}\Pi_{T_J}L^T = \dim T_J$. By identifying $\Pi_{T_J}L^T$ with the operator L^T of Lemma 3.4, and $\nabla\left(\mathcal{E}|_{\Theta_J}\right)(H_0^J,0)$ with V, we obtain *(i)*. Similarly to Theorem 3.3, the assumptions H1 and H3 shows that $\mathcal{G}$ satisfies the condition 2 of Lemma 3.6. The same Lemma then implies that $\overline{\nabla\left(\mathcal{E}|_{\Theta_J}\right)(H_0^J,0)}$ is negligible in T_J. Applying again Lemma 3.4 along with $V = \overline{\nabla\left(\mathcal{E}|_{\Theta_J}\right)(H_0^J,0)}$ yields *(ii)*.

Proof of Proposition 4.6. The proof of this proposition relies on the following theorem.

THEOREM 5.1. *Let U be an open subset of $\mathbb{R}^n$ and $f : U \to \mathbb{R}^n$ a locally Lipschitz function. If W is a negligible subset of U, then $f(W)$ is a negligible subset of $\mathbb{R}^n$.*

The proof of this theorem can be found for instance in [34].

Proof of Proposition 4.6.. 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 4.6 shows that $\nabla\mathcal{E}_J(\Theta_J,0) + \partial_{T_J^\perp}W_J$ is also negligible in $\mathbb{R}^p$. Then Lemma 3.4 applied to $V = \nabla\mathcal{E}_J(\Theta_J,0) + \partial_{T_J^\perp}W_J$ leads to *(i)*.

In order to prove *(ii)*, we show that under the assumption H3, $\overline{B_J}$ is also negligible in $\mathbb{R}^q$. Based on Lemma 3.4 again, 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 development below is dedicated to show the latter statement. The term $\overline{\nabla\mathcal{E}_J(\Theta_J,0)}$ reads

$$\begin{aligned}&\overline{\nabla\mathcal{E}_J(\Theta_J,0)}\\ (48)\quad &= \left\{\lim_{n\to\infty}\nabla\mathcal{E}_J(x_n,0) : x_n \in \Theta_J,\ \forall n \in \mathbf{N} \text{ and } \lim_{n\to\infty}\nabla\mathcal{E}_J(x_n,0) \text{ exists}\right\}.\end{aligned}$$

Assumption H3, joined to the fact that $\nabla\mathcal{E}_J(x_n,0)$ is bounded when $n \to \infty$, implies that $\{x_n\}_{n\in\mathbf{N}}$ is also bounded. Consequently, $\{x_n\}_{n\in\mathbf{N}}$ admits a subsequence which converges in $\overline{\Theta_J}$; by a slight abuse of notation, the latter will be denoted by $\{x_n\}_{n\in\mathbf{N}}$ 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 set $I_0 \subset J^c$ such that $\overline{x} \in \Theta_{J\cup I_0}$.

If $I_0 = \emptyset$, we can write down 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 more intricate and are developed in several stages. Starting with I_0, for every $k = 1, 2, \cdots$ we define recursively

$$u_k := \lim_{n\to\infty} \mathcal{N}\left(\Pi_{T_J\cap(\cap_{i\in I_{k-1}}\text{Ker } G_i)^\perp}(x_n - \overline{x})\right), \tag{49}$$

$$I_k := \{i \in I_{k-1} : G_i u_k = 0\}. \tag{50}$$

The limit in (49) 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 $I_{k-1} \neq \emptyset$. The definitions in (49) and (50) are considered in the following intermediate statements:

Lemma 5.2. *There exists K, $1 \leq K \leq r$, such that the sequence $\{I_k\}_{k\in\{0,\cdots,K\}}$ is strictly decreasing with respect to the inclusion relation, and $I_K = \emptyset$.*

Proof of Lemma 5.2. For k small enough, the definition of u_k shows that $u_k \notin \cap_{I_{k-1}}\text{Ker } G_i$, hence there exists $i \in I_{k-1}$ for which $G_i u_k \neq 0$. Consequently $\{I_k\}_{k\in\mathbf{N}}$ is strictly decreasing whenever I_k is nonempty. The existence of K is straightforward. □

Lemma 5.3. *For every $k \in \{1, \cdots, K\}$ we have $u_k \in U_k$ where*

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

Proof of Lemma 5.3. By the definitions of u_k and of I_k,

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

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

$$\left(\bigcap_{i\in I_{k-1}} \text{Ker } G_i\right)^\perp = \left(\left(\bigcap_{i\in I_{k-1}\setminus I_k} \text{Ker } G_i\right) \cap \left(\bigcap_{i\in I_k} \text{Ker } G_i\right)\right)^\perp$$

$$= \left(\bigcap_{i \in I_{k-1} \setminus I_k} \text{Ker } G_i \right)^{\perp} + \left(\bigcap_{i \in I_k} \text{Ker } G_i \right)^{\perp}$$

we find that

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

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

Lemma 5.4. *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 of Lemma 5.4. From the 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. Let us 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

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

Hence,

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

Letting $n \to \infty$, we obtain

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

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

We now come back to the proof of the proposition. Given $I \subset \{1, \cdots, r\}$, let us introduce the function

$$F_I : \mathbb{R}^p \setminus \{\cup_{i \in I} \text{Ker } G_i\} \to \mathbb{R}^p, \quad u \to F_I(u) := \sum_{i \in I} G_i^T \nabla^+ \varphi_i(\theta_i)(\mathcal{N}(G_i u)) \tag{51}$$

By the definition of I_k in (50), $u_k \notin \cup_{i \in I_{k-1} \setminus I_k} \text{Ker } G_i$. Then, according to lemma 5.4,

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

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

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

Based on (48) and Lemma 5.3, we can write

$$\overline{\nabla \mathcal{E}_J(\Theta_J, 0)} \subset \nabla \mathcal{E}_J(\Theta_J, 0) \cup \left(\bigcup_{K=1}^{r} \bigcup_{\{I_k\}_{k=1}^K \subset \mathcal{I}_K} \left(\nabla \mathcal{E}_{J \cup I_0}(\Theta_{J \cup I_0}, 0) + \sum_{k=1}^{K} F_{I_{k-1} \setminus I_k} \left(U_k \right) \right) \right), \tag{52}$$

where

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

Lemma 5.5. *Let $\{I_k\}_{k=0}^K$ be a strictly decreasing sequence (with respect to the inclusion relation) and $\{U_k\}_{k=0}^K$ are defined as in Lemma 5.3. Then we have $\overline{U_k} \perp \overline{U_l}$ for every $k \neq l$ and*

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

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

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

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

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

Proof of Lemma 5.5. This proof is based on the following identity:

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

Consequently

$$T_J \cap \left(\bigcap_{i\in I_k} \text{Ker } G_i\right) = \left[T_J \cap \left(\bigcap_{i\in I_{k-1}} \text{Ker } G_i\right)\right] \oplus U_k. \tag{53}$$

By using recursively the obtained identity we get

$$\begin{aligned}
T_J &= T_J \cap \left[\left(\bigcap_{i\in I_{K-1}} \text{Ker } G_i\right) \oplus \left(\bigcap_{i\in I_{K-1}} \text{Ker } G_i\right)^{\perp}\right] \\
&= \left[T_J \cap \left(\bigcap_{i\in I_{K-1}} \text{Ker } G_i\right)\right] \oplus \overline{U_K} \\
&= \left[T_J \cap \left(\bigcap_{i\in I_{K-2}} \text{Ker } G_i\right)\right] \oplus \overline{U_{K-1}} \oplus \overline{U_K} \qquad [\text{by (53)}] \\
&= \qquad \cdots \\
&= \left[T_J \cap \left(\bigcap_{i\in I_0} \text{Ker } G_i\right)\right] \oplus \left(\bigoplus_{k=1}^{K} \overline{U_k}\right) \\
&= T_{J\cup J_0} \oplus \left(\bigoplus_{k=1}^{K} \overline{U_k}\right)
\end{aligned}$$

The proof is complete. □

We can now complete the proof of the proposition. By Lemma 5.5, we have the following inclusion

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

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 5.5, any $x \in \mathbb{R}^p$ can be decomposed in a unique way in the form

$$\begin{aligned} & x = \tilde{x} + x_{J\cup I_0} + x_1 + \cdots + x_K + x_J^\perp \\ \text{where} \quad & x_{J\cup I_0} \in T_{J\cup I_0} \\ & x_k \in \overline{U_k}, \quad \forall k \in \{1, \cdots, K\} \\ & x_J^\perp \in T_J^\perp \end{aligned}$$

Based on this decomposition and using F_I defined in (51), 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 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 (52), $\overline{\nabla \mathcal{E}_J(\Theta_J, 0)} + \partial_{T_J^\perp} W_J$ is a negligible subset of $\mathbb{R}^p$, as being a finite union of negligible subsets. The proof is complete.

REFERENCES

[1] S. Alliney, *Digital filters as absolute norm regularizers*, IEEE Transactions on Medical Imaging, MI-12 (1993), pp. 173–181.

[2] 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), pp. 845–866.

[3] A. Avez, *Calcul différentiel*, Masson, 1991.

[4] J. E. Besag, *Statistical analysis of non-lattice data*, The Statistician, 24 (1975), pp. 179–195.

[5] ———, *On the statistical analysis of dirty pictures (with discussion)*, Journal of the Royal Statistical Society B, 48 (1986), pp. 259–302.

[6] ———, *Digital image processing : Towards Bayesian image analysis*, Journal of Applied Statistics, 16 (1989), pp. 395–407.

[7] M. Black, G. Sapiro, D. Marimont, and D. Heeger, *Robust anisotropic diffusion*, IEEE Transactions on Image Processing, 7 (1998), pp. 421–432.

[8] 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), pp. 2–12.

[9] J. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer-Verlag, New York, 2000.

[10] J. F. Bonnans and A. Shapiro, *Optimization problems with perturbations: A guided tour*, SIAM Review, 40 (1998), pp. 228–264.

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

[12] V. Caselles, J.-M. Morel, and C. Sbert, *An axiomatic approach to image interpolation*, IEEE Transactions on Image Processing, 7 (1998), pp. 59–83.

[13] 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), pp. 182–193.

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

[15] P. G. Ciarlet, *Introduction à l'analyse numérique matricielle et à l'optimisation*, Collection mathématiques appliquées pour la maîtrise, Masson, Paris, 1990.

[16] G. Demoment, *Image reconstruction and restoration : Overview of common estimation structure and problems*, IEEE Transactions on Acoustics Speech and Signal Processing, ASSP-37 (1989), pp. 2024–2036.

[17] D. Dobson and F. Santosa, *Recovery of blocky images from noisy and blurred data*, SIAM Journal on Applied Mathematics, 56 (1996), pp. 1181–1199.

[18] 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), pp. 41–81.

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

[20] 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), pp. 721–741.

[21] S. Geman and D. McClure, *Statistical methods for tomographic image reconstruction*, in Proc. of the 46-th Session of the ISI, Bulletin of the ISI, vol. 52, 1987, pp. 22–26.

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

[23] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and Minimization Algorithms, vol. I and II*, Springer-Verlag, Berlin, 1996.

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

[25] J. Milnor, *Topology from the differential point of view*, The University Press of Virginia, Charlottesville, 1965.

[26] D. Mumford and J. Shah, *Boundary detection by minimizing functionals*, in Proceedings of IEEE ICASSP, 1985, pp. 22–26.

[27] M. Nikolova, *Estimées localement fortement homogènes*, Comptes-Rendus de l'Académie des Sciences, t. 325, série 1 (1997), pp. 665–670.

[28] ———, *Reconstruction of locally homogeneous images*, tech. report, TSI–ENST, Paris, France, 1999.

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

[30] G. d. Rham, *Variétés différentiables*, Hermann, Paris, 1955.

[31] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.

[32] R. T. Rockafellar and J. B. Wets, *Variational analysis*, Springer-Verlag, New York, 1997.

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

[34] M. A. Spivak, *A comprehensive introduction to differential geometry*, vol. 1, Brandeis University, Waltham, Mass., 1970.

[35] 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), pp. 387–397.

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

[37] J. Weickert, *Anisotropic Diffusion in Image Processing*, B.G. Teubner, Stuttgart, 1998.