



Weakly Constrained Minimization: Application to the Estimation of Images and Signals Involving Constant Regions*

MILA NIKOLOVA

CMLA UMR8536 –ENS de Cachan, 61 av. President Wilson, 94235 Cachan Cedex, France

nikolova@cmla.ens-cacahn.fr

Abstract. We focus on the question of how the shape of a cost-function determines the features manifested by its local (and hence global) minimizers. Our goal is to check the possibility that the local minimizers of an unconstrained cost-function satisfy different subsets of affine constraints dependent on the data, hence the word “weak”. A typical example is the estimation of images and signals which are constant on some regions. We provide general conditions on cost-functions which ensure that their minimizers can satisfy weak constraints when noisy data range over an open subset. These cost-functions are non-smooth at all points satisfying the weak constraints. In contrast, the local minimizers of smooth cost-functions can almost never satisfy weak constraints. These results, obtained in a general setting, are applied to analyze the minimizers of cost-functions, composed of a data-fidelity term and a regularization term. We thus consider the effect produced by non-smooth regularization, in comparison with smooth regularization. In particular, these results explain the stair-casing effect, well known in total-variation methods. Theoretical results are illustrated using analytical examples and numerical experiments.

Keywords: estimation, inverse problems, Bayesian methods, non-smooth analysis, optimization, perturbation analysis, proximal analysis, restoration, regularization, stabilization, stair-casing, total-variation, variational methods

1. Introduction

We consider the estimation of an unknown $x \in \mathcal{E}$ (an image, a signal) from data $y \in \mathcal{F}$ by minimizing with respect to x a cost-function $f: \mathcal{E} \times \mathcal{F} \rightarrow \mathbb{R}$, where \mathcal{E} and \mathcal{F} are real, affine, normed, finite-dimensional spaces. The vector spaces tangent to \mathcal{E} and \mathcal{F} will be denoted E and F , respectively. Given data $y \in \mathcal{F}$, we focus on local (and hence global) minimizers of $f(\cdot, y)$, i.e. points $\hat{x} \in \mathcal{E}$ such that

$$f(\hat{x}, y) \leq f(x, y), \quad \text{for all } x \in O, \quad (1)$$

where $O \subset \mathcal{E}$ is an open domain where $f(\cdot, y)$ has a local minimum. Cost-function f has to convey both the relationship between the data and the unknown x , and

the prior knowledge about x . The construction of cost-functions, adapted to particular estimation problems, is a critical topic in estimation theory. Typical ways to construct cost-functions are based on PDEs [10, 40, 46], or on probabilistic considerations [3, 15]. Our work suggests an alternative approach since it furnishes an insight into the relationship between the shape of a cost-function and some essential features exhibited by its local minimizers. Concretely, *this paper is dedicated to the question of how, given a set of affine operators and vectors, say $g_i: \mathcal{E} \mapsto \mathbb{R}^s$ and $\theta_i \in \mathbb{R}^s$, for $i = 1, \dots, r$, the shape of f favors, or conversely inhibits, the possibility that for some $y \in \mathcal{F}$, the function $f(\cdot, y)$ admits local minimizers $\hat{x} \in \mathcal{E}$ satisfying*

$$g_i(\hat{x}) = \theta_i, \quad \forall i \in L, \\ \text{with } L \subset \{1, \dots, r\} \text{ and } L \neq \emptyset. \quad (2)$$

For every $i = 1, \dots, r$, the linear operator corresponding with g_i reads $G_i \in \mathcal{L}(E, \mathbb{R}^s)$, where $\mathcal{L}(\cdot, \cdot)$ denotes

*This work was done while the author was with Dpt. TSI-CNRS URA820, ENST, 46 rue Barrault, 75013 Paris, France.

the set of linear operators between the relevant vector spaces. When \mathcal{E} is a vector space, $\mathcal{E} = \mathbb{E}$ and $g_i = G_i$, for all $i = 1, \dots, r$. If $\mathcal{E} = \mathbb{R}^p$, and for all $i \in \{1, \dots, p\}$, we have $\theta_i = 0$ and $[g_i(x)]_j = x_i - x_j$, for j neighbor of i , then (2) means that $f(\cdot, y)$ has minimizers \hat{x} which are constant on some regions. Such minimizers can be seen in Fig. 2(a)–(c). Let us define the function ℓ by

$$\ell(x) := \{i \in \{1, \dots, r\} : g_i(x) = \theta_i\}, \quad \forall x \in \mathcal{E}. \quad (3)$$

Reciprocally, with every $L \subset \{1, \dots, r\}$, we associate the affine subspace

$$\mathcal{K}_L := \{x \in \mathcal{E} : g_i(x) = \theta_i, \quad \forall i \in L\}. \quad (4)$$

Then (2) is equivalent to say that $\mathcal{X}(y) \in \mathcal{K}_L$ with $L \neq \emptyset$. The latter is a severe restriction since \mathcal{K}_L is a closed negligible (i.e. of zero Lebesgue measure) subset of \mathcal{E} . The possibility to satisfy (2) can be seen as a *weakly constrained minimization* since for different $y \in \mathcal{F}$, the different local minimizers \hat{x} of $f(\cdot, y)$ satisfy different subsets $\ell(\hat{x})$ of constraints. We are induced to consider the way the local minimizers of $f(\cdot, y)$ behave under variations of y . Thus we analyze local minimizer functions.

Definition 1. Let $f : \mathcal{E} \times \mathcal{F} \rightarrow \mathbb{R}$ and $N \subseteq \mathcal{F}$. We say that $\mathcal{X} : N \rightarrow \mathcal{E}$ is a local minimizer function for the family of functions $f(\cdot, N) := \{f(\cdot, y) : y \in N\}$ if for any $y \in N$, the function $f(\cdot, y)$ reaches a local minimum at $\mathcal{X}(y)$. Moreover, \mathcal{X} is said to be a strict local minimizer function if $f(\cdot, y)$ has a strict local minimum at $\mathcal{X}(y)$, for every $y \in N$.

We consider functions f which can be convex or non-convex, smooth or non-smooth. Our results address both local and global minimizers. Our main results provide general conditions on cost-functions f which ensure that either (2) is a generic property of their local minimizers, or (2) does almost never occur. The former behavior is manifested by non-smooth cost-functions, whereas the latter behavior characterizes smooth cost-functions. This work is hence an attempt to capture specific features exhibited by the local minimizers of non-smooth cost-function, in comparison with smooth cost-functions.

All results obtained in the general setting mentioned above are applied to regularized estimation. We consider regularized cost-functions f , defined on $\mathbb{R}^p \times \mathbb{R}^q$:

$$f(x, y) = \psi(x, y) + \beta \Phi(x),$$

$$\text{for } \Phi(x) = \sum_{i=1}^r \varphi_i(G_i x - \theta_i), \quad (5)$$

where $\beta > 0$ is a parameter and Φ is the regularization term. For every $i = 1, \dots, r$, the function $\varphi_i : \mathbb{R}^s \rightarrow \mathbb{R}$ is piecewise \mathcal{C}^m , with $m \geq 2$. Now $\mathcal{E} = \mathbb{R}^p$, $\mathcal{F} = \mathbb{R}^q$ and $G_i \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^s)$, for every $i = 1, \dots, r$. The function ℓ introduced in (3) now reads

$$\ell(x) := \{i \in \{1, \dots, r\} : G_i x = \theta_i\}. \quad (6)$$

In this work, the data-fidelity term $\psi : \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}$ is any explicit or implicit \mathcal{C}^m -function, with $m \geq 2$. Usually, ψ comes from a statistical modelling of the data-acquisition. The most often

$$\psi(x, y) = \|H(Ax - y)\|^2, \quad (7)$$

where $\|\cdot\|$ is the Euclidian norm on \mathbb{R}^q and $A \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^q)$ can represent a blur operator, a Radon transform in X-ray tomography, the identity in denoising problems, and many others. The expression in (7) supposes that data y are corrupted with Gaussian noise with covariance matrix $(H^T H)^{-1}$. In both emission and transmission computed tomography (ECT and TCT, respectively), data y are the observed photon counts, whereas $x \in (\mathbb{R}^p)_+$ is a map of the density of the material which is examined. Then

$$\psi(x, y) = \sum_{i=1}^q [\varrho a_i^T x - y_i \ln(a_i^T x)] \quad (\text{ECT}),$$

$$\psi(x, y) = \sum_{i=1}^q [\varrho \exp(-a_i^T x) + y_i a_i^T x] \quad (\text{TCT}), \quad (8)$$

where $a_i^T \in \mathcal{L}(\mathbb{R}^p, \mathbb{R})$, for $i = 1, \dots, q$, have non-negative entries and $\varrho > 0$ is a parameter [9, 22, 27].

The most typical form for the regularization term Φ in (5) corresponds with

$$\varphi_i(z) = \varphi(z) = \phi(\|z\|), \quad \forall i = 1, \dots, r, \quad (9)$$

where $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$ is \mathcal{C}^m for $m \geq 2$. Popular choices for ϕ are [2, 4–6, 8, 11, 20, 21, 22, 31, 35,

40, 45]

- (a) $\phi(t) = t^\alpha, 0 < \alpha \leq 1$
- (b) $\phi(t) = \alpha t / (1 + \alpha t)$
- (c) $\phi(t) = 1 - \mathbb{1}_{(t=0)}$
- (d) $\phi(t) = t^\alpha, 1 < \alpha \leq 2$
- (e) $\phi(t) = \alpha t^2 / (1 + \alpha t^2)$ (10)
- (f) $\phi(t) = 1 - \exp(-\alpha t^2)$
- (g) $\phi(t) = \sqrt{\alpha + t^2}$
- (h) $\phi(t) = t^2 \mathbb{1}_{(t \leq \alpha)} + \alpha(\alpha + 2|t - \alpha|) \mathbb{1}_{(t > \alpha)}$
- (i) $\phi(t) = \min\{1, \alpha t^2\}$

where $\mathbb{1}_{(s)} = 1$ if the expression s is true and $\mathbb{1}_{(s)} = 0$ otherwise.

In many applications, the minimization of $f(\cdot, y)$ is subjected to some affine constraint. Given $y \in \mathbb{R}^q$, the sought solution $\hat{x} \in \mathcal{E}$ is defined by

$$f(\hat{x}, y) \leq f(x, y),$$

for every $x \in \mathcal{O}$ such that $Cx = b$, (11)

where $C \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^{p_0})$ and $b \in \mathbb{R}^{p_0}$, are given. Unlike the problem formulated in (2), the constraint in (11) is strong since it enforces that $\hat{x} \in \mathcal{E}$ where

$$\mathcal{E} := \{x \in \mathbb{R}^p : Cx = b\}. \quad (12)$$

Now \mathcal{E} is an affine subspace of \mathbb{R}^p of dimension $p_E := p - \text{rank } C$. Noticing that for any $y \in \mathbb{R}^q$, any local solution \hat{x} to (11) is a local minimizer of $f|_{\mathcal{E}}(\cdot, y)$, the restriction of $f(\cdot, y)$ to \mathcal{E} , we will apply the results obtained for general affine spaces to constrained minimization problems of the form (11).

To our knowledge, the critical question of the way the shape of a cost-function f determines the features exhibited by its local minimizers has never been considered in a general setting. The question of implicitly, weakly constrained minimization, as presented in (2), has never been formalized previously. Several restricted questions, entirely falling into the scope of our problem, have been considered in the field of image and signal estimation. In [17], the ability of $f(x, y) = \|x - y\|^2 + \beta \sum_i |x_i|$ to recover “nearly black images,” i.e. that $\hat{x}_i = 0$ for many pixels i , is interpreted using mini-max decision theory. In an example in [20], it is discussed that if y is a one-dimensional ramp-shaped signal and if $f(x, y) = \|x - y\|^2 + \beta \sum_i \phi(|x_i - x_{i+1}|)$ with ϕ the function in (10)-(b), then \hat{x} is step-shaped—i.e. $\hat{x}_i = \hat{x}_{i+1}$ for almost all indexes except one.

Total-variation methods, pioneered in [40], amount to $f(x, y) = \|Ax - y\|^2 + \beta \sum_i \|G_i x\|$, where $\{G_i\}$ are first-order difference operators. This cost-function, in the case of one-dimensional signals, is considered in a Bayesian framework in [1]. Total-variation regularization has been observed to produce a “stair-casing effect” [13, 16] consisting in the presence of constant zones in \hat{x} , that is $G_i \hat{x} = 0$ for many indexes i . A first explanation to this phenomenon was given in [32, 34]. In [37], stair-casing for total-variation is studied in a continuous setting. Our work may also be related to stability problems which has been widely studied for the purposes of optimization [7], and especially in the framework of Moreau–Yosida regularization [24, 28].

Organization of the Paper

The notions of one-sided differentiability, used in what follows, are presented in Section 2. Conditions on cost-functions ensuring that (2) is a property of their local minimizers, are established in Section 3. In Section 4, we show that (2) can almost never occur for a smooth cost-function. Numerical illustrations are given in Section 5, with concluding remarks in Section 6. The proofs of all statements are outlined in the Appendix.

Notations

We systematically denote $S := \{x \in K : \|x\| = 1\}$ and $B(\tilde{x}, \rho) := \{x \in \mathcal{K} : \|x - \tilde{x}\| < \rho\}$, where K and \mathcal{K} are a vector space and an affine space, respectively, which are appropriate to the context. Given a set M , \overline{M} denotes its closure and $\overset{\circ}{M}$ its interior. The cardinality of a discrete set L is denoted $\#L$. If $L_1 \subset L$, the complement of L_1 in L is denoted by L_1^c . For a function $f : \mathcal{E} \rightarrow \mathbb{R}$ and a subset $N \subset \mathcal{E}$, we write down $f(N) := \{f(x) : x \in N\}$. If $f : \mathcal{E} \times \mathcal{F} \rightarrow \mathbb{R}$ is smooth, $D_i^k f$, for $i = 1, 2$, denotes its k th-order differential with respect to the i th variable. If $\mathcal{K} \subset \mathcal{E}$ is a subset, by a slight abuse of notation, $f|_{\mathcal{K}}$ will denote the restriction of $f : \mathcal{E} \times \mathcal{F} \rightarrow \mathbb{R}$ to $\mathcal{K} \times \mathcal{F}$. The orthogonal complement in a vector space E of a vector subspace $K \subset E$ is denoted by K^\perp . If A is a real-valued matrix, A^T is its transpose. The components of a vector, say $x \in \mathbb{R}^p$, are denoted either x_i , or $[x]_i$, for $i = 1, \dots, p$. When it is clear from the context, we write $\{\theta_i\}$ to address a previously introduced family $\{\theta_i, \text{ for } i = 1, \dots, r\}$.

2. One-Sided Semi-Derivatives and Derivatives

The notions introduced below are adapted to analyze non-smooth functions f leading to (2).

Definition 2. For $\tilde{y} \in \mathcal{F}$ given, the one-sided semi-derivative of a function $f(\cdot, \tilde{y}) : \mathcal{E} \rightarrow \mathbb{R}$ at a point $\hat{x} \in \mathcal{E}$ in a direction $u \in E$ is defined by

$$\delta_1 f(\hat{x}, \tilde{y})(u) := \liminf_{t \searrow 0} \frac{f(\hat{x} + tu, \tilde{y}) - f(\hat{x}, \tilde{y})}{t}, \quad (13)$$

where the index 1 in δ_1 specifies that we address derivatives with respect to the first variable of f .

The expression (13) gives rise to a *one-sided semi-derivative application*:

$$\begin{aligned} \mathcal{E} \times \mathcal{F} \times E &\rightarrow \mathbb{R}, \\ (x, y, u) &\rightarrow \delta_1 f(x, y)(u). \end{aligned} \quad (14)$$

The scalar in (13) and the application in (14) are always well-defined.

Definition 3. The function $f(\cdot, \tilde{y}) : \mathcal{E} \rightarrow \mathbb{R}$ admits at $\hat{x} \in \mathcal{E}$ a one-sided derivative in a direction $u \in E$, denoted $\delta'_1 f(\hat{x}, \tilde{y})(u)$, if the following limit exists: $\delta'_1 f(\hat{x}, \tilde{y})(u) = \lim_{t \searrow 0} (f(\hat{x} + tu, \tilde{y}) - f(\hat{x}, \tilde{y})) / t$.

The one-sided derivative of $f(\cdot, \tilde{y})$ at $\hat{x} \in \mathcal{E}$ for $u \in E$ exists if, and only if, the inferior limit in (13) can be reduced to a simple limit, in which case $\delta_1 f(\hat{x}, \tilde{y})(u) = \delta'_1 f(\hat{x}, \tilde{y})(u)$. In order to simplify the notations, we will systematically write δ in place of δ' and specify the kind of the one-sided derivative. Observe that $\delta_1 f(\hat{x}, \tilde{y})(u)$ is a *right-side* derivative and that the relevant *left-side* derivative is $-\delta_1 f(\hat{x}, \tilde{y})(-u)$. If $f(\cdot, \tilde{y})$ is differentiable at \hat{x} for u , then $\delta_1 f(\hat{x}, \tilde{y})(u) = -\delta_1 f(\hat{x}, \tilde{y})(-u) = \frac{d}{dt} f(\hat{x} + tu, \tilde{y})|_{t=0}$.

Below we give a necessary condition for a local minimum of a possibly non-smooth function [14, 24, 38].

Lemma 1. *If for $\tilde{y} \in \mathcal{F}$, the function $f(\cdot, \tilde{y})$ has a local minimum at $\hat{x} \in \mathcal{E}$, then $\delta_1 f(\hat{x}, \tilde{y})(u) \geq 0$, for every $u \in E$, where δ_1 denotes a one-sided (semi-) derivative.*

One-sided (semi-)derivatives may in particular be infinite. In this paper we restrict our attention only to

functions f for which $\delta_1 f$ is finite. The results presented in the following can be extended to functions with infinite one-sided (semi-)derivatives but this needs separate considerations. The case of infinite one-sided derivatives for a class of regularized cost-functions was considered in [34]. In that case, infinite one-sided derivatives facilitate the obtention of local minimizers which satisfy (2). This is illustrated by Example 6 given in Appendix, as well as by Fig. 2(c). In the following, we will need to know how one-sided (semi-)derivatives behave in the vicinity of a point (\hat{x}, \tilde{y}) under variations of the direction u .

Definition 4. Given an affine subspace $\mathcal{K} \subseteq \mathcal{E}$ and a vector subspace $K^\perp \subseteq E$, along with a point $(\hat{x}, \tilde{y}) \in \mathcal{K} \times \mathcal{F}$ and a direction $u \in K^\perp$, we say that the one-sided semi-derivative application $\delta_1 f$ is defined uniformly on a neighborhood of (\hat{x}, \tilde{y}, u) , included in $\mathcal{K} \times \mathcal{F} \times K^\perp$, say $(N_{\hat{x}} \times N_{\tilde{y}} \times N_u) \subset (\mathcal{K} \times \mathcal{F} \times K^\perp)$, if the function

$$\sigma \rightarrow \inf_{t \in (0, \sigma)} \frac{f(x + tu', y) - f(x, y)}{t}, \quad (15)$$

defined for $\sigma > 0$, converges towards $\delta_1 f(x, y)(u')$ as $\sigma \searrow 0$ uniformly for all $(x, y, u') \in (N_{\hat{x}} \times N_{\tilde{y}} \times N_u)$.

Notice that one-sided differentiability does not imply uniform definiteness.

Definition 5. The one-sided derivative application $(x, y, u) \rightarrow \delta_1 f(x, y)(u)$ is defined uniformly on a neighborhood $(N_{\hat{x}} \times N_{\tilde{y}} \times N_u) \subset (\mathcal{K} \times \mathcal{F} \times K^\perp)$ of (\hat{x}, \tilde{y}, u) , if the function

$$t \rightarrow \frac{f(x + tu', y) - f(x, y)}{t}, \quad (16)$$

defined for $t > 0$, converges towards $\delta_1 f(x, y)(u')$ as $t \searrow 0$ uniformly for all $(x, y, u') \in (N_{\hat{x}} \times N_{\tilde{y}} \times N_u)$.

Our theory does not address functions for which the one-sided (semi-)derivative application is not uniformly defined on any neighborhood of (\hat{x}, \tilde{y}, u) . Observe that Definitions 4 and 5 address only functions f for which $\delta_1 f$ is finite on $(N_{\hat{x}} \times N_{\tilde{y}} \times N_u) \subset (\mathcal{K} \times \mathcal{F} \times K^\perp)$. A function admitting uniformly defined one-sided semi-derivatives, but not admitting one-sided derivatives, is considered below.

Example 1. Let $f : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$ be given by

$$\begin{aligned} f(x, y) &= \psi(x, y) + \varphi(x_1 - x_2), \\ \psi(x, y) &= \frac{1}{2}(x_1 + x_2 - y_1)^2 + \frac{1}{2}(x_1 - x_2 - y_2)^2, \\ \varphi(t) &= \begin{cases} t \left(2 + \sin \frac{1}{t} \right) & \text{if } t > 0, \\ -3t - (t - 1)^2 + 1 & \text{if } -1/2 \leq t \leq 0, \\ 1/4 & \text{if } t < -1/2. \end{cases} \end{aligned}$$

Notice that φ is C^∞ on $(-1/2, 0) \cup (0, +\infty)$ and that it is continuous at zero where $\varphi(0) = 0$. In this case, $G_1 = [1, -1]$. We focus on $L = \{1\}$. By (4), we have $\mathcal{K}_{\{1\}} = K_{\{1\}} = \{u \in \mathbb{R}^2 : u_1 = u_2\}$ and $K_{\{1\}}^\perp = \{u \in \mathbb{R}^2 : u_1 = -u_2\}$. Consider some $x \in \mathcal{K}_{\{1\}}$. The function f does not admit at x a one-sided derivative for any $u \in K_{\{1\}}^\perp$ [Definition 3]. Its one-sided semi-derivative $\delta_1 f$ [Definition 3] reads (the details are given in the Appendix)

$$\delta_1 f(x, y)(u) = -y_2(u_1 - u_2) + |u_1 - u_2|. \quad (17)$$

The application $\delta_1 f$ is continuous. Moreover, for every $u \in K_{\{1\}}^\perp$,

$$\begin{aligned} & \left| \delta_1 f(x, y)(u) - \inf_{t \in (0, \sigma)} \frac{f(x + tu, y) - f(x, y)}{t} \right| \\ &= \begin{cases} 0 & \text{if } u_1 > u_2, \\ \sigma |u_1 u_2| & \text{if } u_1 \leq u_2. \end{cases} \end{aligned}$$

The above residual goes to zero when $\sigma \searrow 0$ uniformly on $\mathcal{K}_{\{1\}} \times \mathbb{R}^2 \times (K_{\{1\}}^\perp \cap S)$, hence $\delta_1 f$ is defined uniformly on this set.

Related notions of one-sided differentiability have been considered by many authors [14, 24, 28, 39, 42]. For instance, the semi-differential used in [39] amounts to $\liminf_{t \searrow 0, u' \rightarrow u} (f(\hat{x} + tu', \tilde{y}) - f(\hat{x}, \tilde{y}))/t$, so the behavior of the difference quotient is tested for all half-lines $\hat{x} + tu'$ converging to $\hat{x} + tu$. In comparison, for one-sided (semi-)derivatives, it is tested only for the half-line $\hat{x} + tu$. The definitions we adopt reflect the ambition to use quite weak assumptions still allowing us to exhibit the possibility to have property (2).

3. Minimizers Involving Weak Constraints

3.1. General Cost-Functions

We start by considering cost-functions which do not necessarily admit one-sided derivatives. The theorem below contains the main contribution of this paper.

Theorem 1. Consider an $f : \mathcal{E} \times \mathcal{F} \rightarrow \mathbb{R}$, along with a set of affine functions $g_i : \mathcal{E} \rightarrow \mathbb{R}^s$ and vectors $\theta_i \in \mathbb{R}^s$, for $i = 1, \dots, r$. Let $(\hat{x}, \tilde{y}) \in \mathcal{E} \times \mathcal{F}$ and put $L := \ell(\hat{x})$, where ℓ is given in (3). Let \mathcal{K}_L be defined by (4) and K_L be its tangent. Suppose that the semi-derivative application [Definition 4]

$$\begin{aligned} & \mathcal{K}_L \times \mathcal{F} \times (K_L^\perp \cap S) \rightarrow \mathbb{R}, \\ & (x, y, u) \rightarrow \delta_1 f(x, y)(u), \end{aligned}$$

1. is lower semi-continuous on $\{\hat{x}\} \times \{\tilde{y}\} \times (K_L^\perp \cap S)$;
2. is defined uniformly on $N_{\hat{x}} \times N \times (K_L^\perp \cap S)$ where $(N_{\hat{x}} \times N) \subset \mathcal{K}_L \times \mathcal{F}$ is a neighborhood of (\hat{x}, \tilde{y}) .

Suppose also that

- (a) $\delta_1 f(\hat{x}, \tilde{y})(u) > 0$ for every $u \in K_L^\perp \cap S$;
- (b) \tilde{y} is contained in a neighborhood $N \subset \mathcal{F}$ such that $f|_{\mathcal{K}_L(\cdot, N)}$ admits a local minimizer function $\mathcal{X}_L : N \rightarrow \mathcal{K}_L$ [Definition 1] which is continuous at \tilde{y} and satisfies $\hat{x} = \mathcal{X}_L(\tilde{y})$.

Then there is a neighborhood V_L of \tilde{y} such that $f(\cdot, V_L)$ admits a local minimizer function $\mathcal{X} : V_L \rightarrow \mathcal{E}$ which satisfies $\mathcal{X}(\tilde{y}) = \hat{x}$ and

$$y \in V_L \Rightarrow g_i(\mathcal{X}(y)) = \theta_i, \quad \text{for all } i \in L. \quad (18)$$

More precisely, $\mathcal{X}(y) = \mathcal{X}_L(y) \in \mathcal{K}_L$, for every $y \in V_L$. In addition, if \mathcal{X}_L is a strict local minimizer function for $f|_{\mathcal{K}_L(\cdot, N)}$, then \mathcal{X} is a strict local minimizer function for $f(\cdot, V_L)$.

Notice that the tangent of \mathcal{K}_L reads $K_L = \{u \in E : G_i u = 0, \forall i \in L\}$ where $\{G_i\}$ are the linear operators associated with $\{g_i\}$. Assumptions 1 and 2 concern the regularity of f on a neighborhood of (\hat{x}, \tilde{y}) while (a) and (b) state a proper sufficient condition for local minimum. We see that once a data point \tilde{y} yields a local minimizer \hat{x} of $f(\cdot, \tilde{y})$ with $L := \ell(\hat{x})$ non-empty (such a \tilde{y} can usually be determined), there is a neighborhood

V_L of \tilde{y} such that if $y \in V_L$, then $g_i(\mathcal{X}(y)) = \theta_i$, for all $i \in L$. It is striking to observe that in spite of the fact that y ranges over an open subset of \mathcal{F} , the minimizer function \mathcal{X} evolves in an affine subspace strictly included in \mathcal{E} .

To illustrate Theorem 1, we calculate a minimizer of the function considered in Example 1.

Example 2. By (17), Assumptions 1 and 2 of Theorem 1 are valid. By (17), condition (a) reads $y_2(u_1 - u_2) < |u_1 - u_2|$, $\forall u \in K_{\{1\}}^\perp$, which is equivalent to $|y_2| < 1$. The restriction of $f(\cdot, y)$ to $\mathcal{K}_{\{1\}}$ reads $f|_{\mathcal{K}_{\{1\}}}(x, y) = \frac{1}{2}(x_1 + x_2 - y_1)^2$, where $x_1 = x_2$. This function is strictly convex and yields

$$\mathcal{X}_{\{1\}}(y) = \left(\frac{y_1}{2}, \frac{y_1}{2} \right), \quad \forall y \in \mathbb{R}^2.$$

Hence, condition (b) is satisfied for every $y \in \mathbb{R}^2$. According to Theorem 1,

$$y \in V_{\{1\}} := \{y \in \mathbb{R}^2 : |y_2| < 1\} \Rightarrow \mathcal{X}(y) = \left(\frac{y_1}{2}, \frac{y_1}{2} \right) \in \mathcal{K}_{\{1\}}. \quad (19)$$

Notice that \mathcal{X} is a strict global minimizer function and that $f(\cdot, y)$ has an infinity of local minimizers.

Next we focus on functions which admit one-sided derivatives in the sense of Definition 3.

Proposition 1. *Consider an $f : \mathcal{E} \times \mathcal{F} \rightarrow \mathbb{R}$ and $(\hat{x}, \tilde{y}) \in \mathcal{E} \times \mathcal{F}$. Let $\{g_i\}, \{\theta_i\}, L, \mathcal{K}_L$ and K_L be as in Theorem 1. Let (\hat{x}, \tilde{y}) be contained in a neighborhood $(N_{\hat{x}} \times N) \subset (\mathcal{K}_L \times \mathcal{F})$ such that f is continuous and has a uniformly defined one-sided derivative application $\delta_1 f$ on $N_{\hat{x}} \times N \times (K_L^\perp \cap S)$ [Definition 3]. Then $\delta_1 f$ is continuous on $\{\hat{x}\} \times \{\tilde{y}\} \times (K_L^\perp \cap S)$.*

If there is a neighborhood of (\hat{x}, \tilde{y}) where f is continuous and admits a uniformly defined one-sided derivative application $\delta_1 f$, and if conditions (a) and (b) of Theorem 1 are satisfied, then (18) holds.

3.2. Application to Non-Smooth Regularization

We will apply the results obtained in Section 3.1 to possibly constrained minimization problems of the form (11) for cost-functions $f : \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}$ as given in (5). For any $i = 1, \dots, r$, we suppose that

H1 the function $\varphi_i : \mathbb{R}^s \rightarrow \mathbb{R}$ is \mathcal{C}^m on $\mathbb{R}^s \setminus \{0\}$, for $m \geq 2$;

H2 at zero, φ_i admits a one-sided derivative application $u \rightarrow \delta\varphi_i(0)(u)$ [Definition 3] which is defined uniformly on S .

In some cases, ψ may assume the form $\psi(x, y) = \tilde{\psi}(x, y) + \sum_{i=r+1}^{r_0} \tilde{\varphi}_i(G_i x - \theta_i)$, where $\tilde{\psi}$ is a data-fidelity term and $\tilde{\varphi}_i$, for $i = r + 1, \dots, r_0$, are \mathcal{C}^m -functions, for $m \geq 2$. Thus the formulation in (5) allows us to address cost-functions combining both smooth and non-smooth regularization terms.

We will consider the subspace \mathcal{E} defined in (12). Similarly to (4), for any $L \subset \{1, \dots, r\}$ we define

$$\mathcal{K}_L = \{x \in \mathcal{E} : G_i x = \theta_i, \forall i \in L\}. \quad (20)$$

It makes sense to consider only subsets L such that \mathcal{K}_L is of dimension ≥ 1 and is strictly included in \mathcal{E} . The tangents of \mathcal{E} and \mathcal{K}_L now read $E = \{x \in \mathbb{R}^p : Cx = 0\}$ and $K_L = \{x \in E : G_i x = 0, \forall i \in L\}$. The next proposition concerns the regularity of f .

Proposition 2. *Let f be as given in (5), where all φ_i , for $i = 1, \dots, r$, satisfy H1 and H2. Let \mathcal{E} be as defined in (12). Suppose that there is an open subset $(N_x \times N_y) \subset (\mathcal{E} \times \mathbb{R}^q)$ where ψ is \mathcal{C}^m with $m \geq 2$. Then for every $\hat{x} \in \mathcal{K}_L \cap N_x$, such that $G_i \hat{x} \neq \theta_i$ for all $i \in L^c$, and for every $\tilde{y} \in N_y$, there is a neighborhood $(N_{\hat{x}} \times N) \subset (\mathcal{K}_L \times \mathbb{R}^q)$ containing (\hat{x}, \tilde{y}) , so that f has a uniformly defined one-sided derivative application $\delta_1 f$ on $N_{\hat{x}} \times N \times (K_L^\perp \cap S)$ [Definition 5].*

Given $L \subset \{1, \dots, r\}$, we decompose f into $f(x, y) = f_0(x, y) + f_L(x)$, where

$$\begin{aligned} f_0(x, y) &= \psi(x, y) + \beta \sum_{i \in L^c} \varphi_i(G_i x - \theta_i), \\ f_L(x) &= \beta \sum_{i \in L} \varphi_i(G_i x - \theta_i). \end{aligned} \quad (21)$$

If $L = \ell(\hat{x})$, for ℓ as defined in (6), then $f_0(\cdot, y)$ is \mathcal{C}^m on a neighborhood of \hat{x} included in \mathcal{E} , whereas f_L is non-smooth at \hat{x} for all directions in K_L^\perp . Clearly, $\delta_1 f(\hat{x}, y)(u) = D_1 f_0(\hat{x}, y)u + \delta f_L(\hat{x})(u)$ for all $u \in E$.

Theorem 2. *Consider problem (11) with f of the form (5), where every φ_i , for $i = 1, \dots, r$, satisfies H1 and H2. For $\tilde{y} \in \mathbb{R}^q$ given, let $\hat{x} \in \mathcal{E}$ be a solution to (11) where \mathcal{E} reads as in (12). Put $L := \ell(\hat{x})$*

for ℓ as given in (6). Let \mathcal{K}_L be as in (20) and K_L be its tangent. Assume that (\hat{x}, \tilde{y}) is contained in an open neighborhood $(N_{\hat{x}} \times N) \subset (\mathcal{E} \times \mathbb{R}^q)$ where ψ is C^m with $m \geq 2$. Suppose also that

- (a) $D_1 f_0(\hat{x}, \tilde{y})u > -\beta \sum_{i \in L} \delta \varphi_i(0)(G_i u)$, for every $u \in K_L^\perp \cap S$;
- (b) $f_0|_{\mathcal{K}_L}(\cdot, N)$ has a local minimizer function $\mathcal{X}_L : N \rightarrow \mathcal{K}_L$ which is continuous at \tilde{y} and $\hat{x} = \mathcal{X}_L(\tilde{y})$,

where f_0 is as given in (21). Then there is a neighborhood V_L of \tilde{y} such that $f(\cdot, V_L)$ admits a local minimizer function $\mathcal{X} : V_L \rightarrow \mathcal{E}$ which satisfies $\mathcal{X}(\tilde{y}) = \hat{x}$ and

$$y \in V_L \Rightarrow G_i \mathcal{X}(y) = \theta_i, \quad \text{for all } i \in L.$$

More precisely, $\mathcal{X}(y) = \mathcal{X}_L(y) \in \mathcal{K}_L$ for every $y \in V_L$. In addition, if \mathcal{X}_L is a strict local minimizer function for $f|_{\mathcal{K}_L}(\cdot, N)$, then \mathcal{X} is a strict local minimizer function for $f(\cdot, V_L)$.

Next we check H1 and H2 for $\{\varphi_i\}$ of the form (9).

Proposition 3. Let $\varphi : \mathbb{R}^s \rightarrow \mathbb{R}$ read

$$\varphi(u) = \phi(\|u\|), \quad (22)$$

where $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$ is continuous, C^m for $m \geq 2$ on $(0, +\infty)$ with $\phi'(0^+) > 0$ finite. Then the function $u \rightarrow \varphi(u)$ satisfies H1 and H2, and we have $\delta \varphi(0)(v) > 0$, for all $v \in \mathbb{R}^s$.

The fact that $\phi'(0^+) > 0$ implies that φ is non-smooth at the origin. Among the functions given in (10), we have $\phi'(0^+) = 1$ for (a) with $\alpha = 1$ and $\phi'(0^+) = \alpha$ for (b).

3.3. Commentary

Assumptions 1 and 2 of Theorem 1 are general enough. By Proposition 2, they are satisfied by general regularized cost-functions. They hold also for the function in Example 1.

3.3.1. On Condition (a). In Theorems 1 and 2, $f(\cdot, \tilde{y})$ has a local minimum at \hat{x} , hence $\delta_1 f(\hat{x}, \tilde{y})(u) \geq 0$, for all $u \in E$, according to Lemma 1. In comparison, (a) requires only that this inequality be strict for every $u \in K_L^\perp \subset E$. So, (a) is not a strong requirement. If

f is a regularized cost-function of the form (5)–(7) and satisfies some technical assumptions, the analysis in [18] shows that for almost every y (except those contained in a negligible subset of \mathbb{R}^q), every local minimizer \hat{x} of $f(\cdot, y)$ satisfies (a).

Let us come back to Example 1. If for some y , $f(\cdot, y)$ has a minimizer $\hat{x} \in \mathcal{K}_{\{1\}}$, then (17) shows that $|y_2| \leq 1$. For all $y \in \mathbb{R}^2$ with $|y_2| > 1$, if $f(\cdot, y)$ has a local minimizer \hat{x} , then $\hat{x} \in \mathbb{R}^2 \setminus \mathcal{K}_{\{1\}}$. Joining this with (20) shows that the only points \tilde{y} for which $f(\cdot, \tilde{y})$ may have a minimizer $\hat{x} \in \mathcal{K}_{\{1\}}$, without satisfying (a), correspond to $\tilde{y} \in \{y \in \mathbb{R}^2 : |y_2| = 1\}$; the latter is a closed, negligible subset of \mathbb{R}^2 .

The remark below gives indications how to construct cost-functions whose minimizers satisfy some weak constraints of the form (2).

Remark 1 (f is non-smooth on all $\mathcal{K}_{\{i\}}$). By (a), $f(\cdot, \tilde{y})$ is non-smooth at \hat{x} for every non-zero direction $u \in K_L^\perp$, since $-\delta_1 f(\hat{x}, \tilde{y})(-u) < 0 < \delta_1 f(\hat{x}, \tilde{y})(u)$. Moreover, there is a neighborhood $N_{\hat{x}}$ of \hat{x} , included in \mathcal{K}_L , and a neighborhood N of \tilde{y} , so that for every $y \in N$, the function $f(\cdot, y)$ is non-smooth at each $x \in N_{\hat{x}} \subset \mathcal{K}_L$, for every $u \in K_L^\perp$. More generally, if we wish that for different y the minimizers \hat{x} of $f(\cdot, y)$ satisfy (2) for different subsets L , we will ensure that for every $y \in \mathcal{F}$, the function $f(\cdot, y)$ is non-smooth on $\bigcup_{i=1}^r \mathcal{K}_{\{i\}}$ in such a way that its one-sided (semi-)derivatives are ≥ 0 . If f is a regularized cost-functions of the form (5), the latter condition is satisfied if for every $i = 1, \dots, r$, the function φ_i is such that $-\delta \varphi_i(0)(-v) < \delta \varphi_i(0)(v)$, for all $v \in \mathbb{R}^s$, i.e. if its left-side derivatives are smaller than its right-side derivatives. In particular, this holds if $\delta \varphi_i(0)(v) > 0$, for all $v \in \mathbb{R}^s$.

3.3.2. On Condition (b). Notice that the requirement that \mathcal{X}_L is C^{m-1} , for $m \geq 1$, does not need $f|_{\mathcal{K}_L}$ to be smooth. The next lemma, which can be found in [19], is important to check (b).

Lemma 2. Let $f : \mathcal{E} \times \mathcal{F} \rightarrow \mathbb{R}$ be C^m , with $m \geq 2$, on a neighborhood of $(\hat{x}, \tilde{y}) \in \mathcal{E} \times \mathcal{F}$. Suppose that $f(\cdot, \tilde{y})$ reaches at \hat{x} a local minimum such that $D_1^2 f(\hat{x}, \tilde{y})$ is positive definite (i.e. its eigenvalues are > 0). Then there are a neighborhood $N \subset \mathcal{F}$ containing \tilde{y} and a unique C^{m-1} strict local minimizer function $\mathcal{X} : N \rightarrow \mathcal{E}$, such that $D_1^2 f(\mathcal{X}(y), y)$ is positive definite for every $y \in N$ and $\mathcal{X}(\tilde{y}) = \hat{x}$.

Thus, condition (b) holds if $f|_{\mathcal{K}_L}$ is \mathcal{C}^m on a neighborhood of (\hat{x}, \tilde{y}) belonging to $\mathcal{K}_L \times \mathcal{F}$, and if

$$D_1(f|_{\mathcal{K}_L})(\hat{x}, \tilde{y}) = 0, \quad \text{and} \\ D_1^2(f|_{\mathcal{K}_L})(\hat{x}, \tilde{y}) \text{ is positive definite.} \quad (23)$$

In many applications f is convex and $D_1^2(f|_{\mathcal{K}_L})(x, y)$ is positive definite for all $x \in \mathcal{E}$ with $L = \ell(x)$, and for all $y \in \mathbb{R}^q$. Analyzing the possibility to have (b) when f is non-convex is more intricate. In [18] it is shown that if f is of the form (5) with HA one-to-one, and $\{\varphi_i\}$ possibly non-convex and satisfying some non-restrictive assumptions, then for almost every $\tilde{y} \in \mathbb{R}^q$ (except those contained in a negligible subset of \mathbb{R}^q), for every local minimizer \hat{x} of $f(\cdot, \tilde{y})$, there is a neighborhood $N \subset \mathbb{R}^q$ and a \mathcal{C}^{m-1} local minimizer function $\mathcal{X} : N \rightarrow \mathcal{K}_L$, with $\hat{x} = \mathcal{X}(\tilde{y})$ and $L = \ell(\hat{x})$. Then $\mathcal{X}_L = \mathcal{X}$ is as required in (b).

3.3.3. Significance of the Results. The conclusion of Theorems 1 and 2 can be reformulated as

$$y \in V_L \Rightarrow \ell(\mathcal{X}(y)) \supseteq L, \quad (24)$$

where $L := \ell(\mathcal{X}(\tilde{y}))$. In particular, V_L can contain points y yielding $g_i(\mathcal{X}(y)) = \theta_i$ for some $i \in L^c$. However, by the continuity of \mathcal{X} at \tilde{y} , there is a neighborhood $\tilde{V}_L \subset V_L$ containing \tilde{y} such that

$$y \in \tilde{V}_L \Rightarrow \begin{cases} g_i(\mathcal{X}(y)) = \theta_i & \text{if } i \in L, \\ g_i(\mathcal{X}(y)) \neq \theta_i & \text{if } i \in L^c. \end{cases} \quad (25)$$

In other words, $\ell(\mathcal{X}(y)) = L$ for all $y \in \tilde{V}_L$.

Focus on a minimizer function $\mathcal{X} : N \rightarrow \mathcal{E}$ for $f(\cdot, N)$ and put $L = \ell(\mathcal{X}(\tilde{y}))$ for some $\tilde{y} \in N$. By Theorems 1 and 2, and by (25), the sets V_L and \tilde{V}_L ,

$$V_L := \{y \in N : \ell(\mathcal{X}(y)) \supseteq L\} \\ = \{y \in N : \mathcal{X}(y) \in \mathcal{K}_L\}, \quad (26)$$

$$\tilde{V}_L := \{y \in N : \ell(\mathcal{X}(y)) = L\}, \quad (27)$$

contain *open* subsets of N , hence V_L and \tilde{V}_L are of *positive measure* in \mathcal{F} . The chance that random points y (e.g. noisy data) come across V_L , or \tilde{V}_L , is positive. When data y range over N , the set-valued function $(\ell \circ \mathcal{X})$ generally takes several distinct values, say $\{L_j\}$. Thus, with a minimizer function \mathcal{X} , defined on an open set N , there is associated a family of subsets $\{\tilde{V}_{L_j}\}$ which form a covering of N . When $y \in \tilde{V}_{L_j}$,

we find a minimum $\hat{x} = \mathcal{X}(y)$ satisfying $\ell(\hat{x}) = L_j$. This is the reason why non-smooth cost-functions, as those considered here, exhibit local minimizers which generically satisfy weak constraints of the form (2). For a regularized cost-function of the form (5), with $\{\varphi_i\}$ as in (9), $\{G_i\}$ first-order difference operators and $\theta_i = 0$, for all i , minimizers \hat{x} are typically constant on many regions. This explains in particular the *stair-casing effect* observed in total-variation methods [13, 16, 40]. In the example below we derive the sets \tilde{V}_L , for every $L \subset \{1, \dots, r\}$, in the context of a particular f .

Example 3 (Total variation). Let $f : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$ be given by

$$f(x, y) = \|Ax - y\|^2 + \beta \sum_{i=1}^{p-1} |x_i - x_{i+1}|, \quad (28)$$

where $A \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^p)$ is invertible and $\beta > 0$. It is easy to see that there is a unique minimizer function \mathcal{X} for $f(\cdot, \mathbb{R}^p)$. We exhibit two striking phenomena characterizing non-smooth regularization:

1. for every point $\hat{x} \in \mathbb{R}^p$, there is a polyhedron $W_{\hat{x}} \subset \mathbb{R}^p$ of dimension $\#\ell(\hat{x})$, such that for every $y \in W_{\hat{x}}$, the same point $\mathcal{X}(y) = \hat{x}$ is the unique minimizer of $f(\cdot, y)$;
2. for every $L \subset \{1, \dots, p-1\}$, there is a subset $\tilde{V}_L \subset \mathbb{R}^p$, composed of $2^{p-\#\ell(L)}$ unbounded polyhedra of \mathbb{R}^p , such that for every $y \in \tilde{V}_L$, the minimizer \hat{x} of $f(\cdot, y)$ satisfies $\hat{x}_i = \hat{x}_{i+1}$ for all $i \in L$ and $\hat{x}_i \neq \hat{x}_{i+1}$ for all $i \in L^c$.

Remark 2 (Classification rule). The function ℓ naturally provides a classification rule if for any $x \in \mathcal{E}$ we consider the two classes $L = \ell(x)$ and L^c , its complement. By (24) and (25), the function $y \rightarrow \ell(\mathcal{X}(y))$ is constantly equal to $L = \ell(\mathcal{X}(\tilde{y}))$ on \tilde{V}_L and $\ell(\mathcal{X}(y)) \supseteq \ell(\mathcal{X}(\tilde{y}))$ for all $y \in V_L$. The resultant classification is hence *stable* with respect to small variations of the data (e.g. due to noise perturbations).

Remark 3. By Theorems 1 and 2, we have $\mathcal{X}(V_L) \subset \mathcal{K}_L$, and hence $\text{rank } D\mathcal{X}(y) \leq \dim \mathcal{K}_L$, for all $y \in V_L$.

4. Locally Smooth Cost-Function

4.1. General Case

In this section we consider the possibility that a smooth cost-function yields local minimizers which satisfy

weak constraints of the form (2). In what follows, \mathcal{E} and \mathcal{F} are of dimensions p and q , respectively.

Theorem 3. Consider an $f : \mathcal{E} \times \mathcal{F} \rightarrow \mathbb{R}$ and let $\mathcal{X} : N \rightarrow \mathcal{E}$ be a differentiable local minimizer function for $f(\cdot, N)$, where $N \subset \mathcal{F}$ is open. Suppose f is twice differentiable on $N_x \times N$ where $N_x \subset \mathcal{E}$ is open and $\mathcal{X}(N) \subset N_x$. Let $\theta \in \mathbb{R}^s$ and let $g : \mathcal{E} \rightarrow \mathbb{R}^s$ be an affine function such that

$$\begin{aligned} \text{rank } Dg(x) &> p - \text{rank } D_{12}f(x, y), \\ \forall(x, y) &\in N_x \times N. \end{aligned} \quad (29)$$

For any $N' \subseteq N$ with $N' \neq \emptyset$, define the subset $V_{N'}$ by

$$V_{N'} := \{y \in N' : g(\mathcal{X}(y)) = \theta\}. \quad (30)$$

Then

- (i) the interior of V_N is empty;
- (ii) if \mathcal{X} is \mathcal{C}^1 on N , for any compact $N' \subset N$, the relevant $V_{N'}$ is included in a closed, negligible subset of \mathcal{F} .

The proof of this theorem uses the following lemma which relates the ranks of $D\mathcal{X}$ and $D_{12}f$ on N .

Lemma 3. Let f , N and \mathcal{X} be as in Theorem 3. Then $\text{rank } D\mathcal{X}(y) \geq \text{rank } D_{12}f(\mathcal{X}(y), y)$, for all $y \in N$.

We next focus on functions f which at some points $(x, y) \in \mathcal{E} \times \mathcal{F}$ are non-smooth with respect to x in such a way that $-\delta_1 f(x, y)(-u) > \delta_1 f(x, y)(u)$ for some direction $u \in E$. A popular function of this kind is considered in Example 4. Let M denote the set of all such points:

$$M := \{(x, y) \in \mathcal{E} \times \mathcal{F} : \exists u \in E \text{ with } -\delta_1 f(x, y)(-u) > \delta_1 f(x, y)(u)\}, \quad (31)$$

where δ_1 denotes a one-sided (semi-)derivative [Definition 2 or 3]. Notice that from Rademacher's theorem [24, 36], when f is Lipschitz, M is included in a negligible subset of $\mathcal{E} \times \mathcal{F}$.

Proposition 4. Let f be twice differentiable on $\mathcal{E} \times \mathcal{F} \setminus M$ where M is the set defined in (31). Then for every

$\tilde{y} \in \mathcal{F}$ for which $f(\cdot, \tilde{y})$ admits some local minimizers, and for every local minimizer \hat{x} of $f(\cdot, \tilde{y})$,

$$(\hat{x}, \tilde{y}) \notin M.$$

Suppose there is an open set $N \subset \mathcal{F}$ such that $f(\cdot, N)$ admits a differentiable local minimizer function $\mathcal{X} : N \rightarrow \mathcal{E}$. Suppose M is closed and negligible in $\mathcal{E} \times \mathcal{F}$. For $g : \mathcal{E} \rightarrow \mathbb{R}^s$ and $\theta \in \mathbb{R}^s$ as in Theorem 3, let (29) hold for all $(x, y) \in (\mathcal{E} \times \mathcal{F}) \setminus M$. Then the conclusions of Theorem 3 are valid.

4.2. Application to Smooth Regularization

We focus on problem (11) and (12) where f is of the form (5), and is smooth in the vicinity of its local minimizers. We suppose that for every $i = 1, \dots, r$, there is a closed subset $T_i \subset \mathbb{R}^s$, so that

H3 the function $\varphi_i : \mathbb{R}^s \rightarrow \mathbb{R}$ is \mathcal{C}^m -smooth on $\mathbb{R}^s \setminus \{T_i\}$, for $m \geq 2$;

H4 if $T_i \neq \emptyset$, then T_i is a \mathcal{C}^1 -manifold with $\dim T_i = s - 1$, and for every $z \in T_i$ there is $\tilde{v} \in \mathbb{R}^s$ such that $-\delta\varphi_i(z)(-\tilde{v}) > \delta\varphi_i(z)(\tilde{v})$, whereas $-\delta\varphi_i(z)(-v) \geq \delta\varphi_i(z)(v)$ for every $v \in \mathbb{R}^s$.

T_i is nonempty for φ_i is of the form (9) with ϕ as in (10)-(f). It is empty for (10)-(d), (e), (g), (h).

Example 4 (Truncated quadratic function). Consider $\varphi_i : \mathbb{R}^s \rightarrow \mathbb{R}$ of the form (9) where

$$\phi(t) = \min\{1, \alpha t^2\} \quad \text{with } \alpha > 0.$$

This function was initially introduced in [21] for the restoration of images involving sharp edges, and in the sequel this ϕ was considered by many authors [6, 12, 31]. It is non-smooth at $1/\sqrt{\alpha}$ and \mathcal{C}^∞ elsewhere. The set T_i , evoked in H4, reads

$$T_i = \left\{ z \in \mathbb{R}^s : \|z\| = \frac{1}{\sqrt{\alpha}} \right\}.$$

Indeed, T_i is a \mathcal{C}^∞ -manifold of dimension $s - 1$. By Definition 3, for any $z \in T_i$,

$$\delta\varphi_i(z)(u) = \begin{cases} 0 & \text{if } z^T u \geq 0, \\ -2\alpha |z^T u| & \text{if } z^T u < 0. \end{cases}$$

Hence H4 is satisfied in its full form. If $s = 1$, $\varphi_i(z) = \phi(|z|)$ for $z \in \mathbb{R}$, then $T_i = \{-1/\sqrt{\alpha}, 1/\sqrt{\alpha}\}$ is a manifold of dimension 0, and $\delta\varphi_i(1/\sqrt{\alpha})(1) = 0$ and $-\delta\varphi_i(1/\sqrt{\alpha})(-1) = 2\sqrt{\alpha}$. Focus next on $\varphi_i(G_i x - \theta_i) = \phi(\|G_i x\|)$, with $G_i \neq 0$, for every $i = 1, \dots, r$. The set M introduced in (31) now reads

$$\begin{aligned} M &= \bigcup_{i=1}^r M_i \times \mathbb{R}^q, \text{ where } M_i = \{x \in \mathbb{R}^p : G_i x \in T_i\} \\ &= \left\{ x \in \mathbb{R}^p : \|G_i x\| = \frac{1}{\sqrt{\alpha}} \right\}. \end{aligned}$$

Every M_i , for $i = 1, \dots, r$, is a C^∞ -manifold of dimension $p - 1$, hence M is a closed, negligible subset of $\mathbb{R}^p \times \mathbb{R}^q$.

Theorem 4. *Consider problem (11) where $f : \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}$ is of the form (5), and H3 and H4 are satisfied. For \mathcal{E} as defined in (12), let $\mathcal{X} : N \rightarrow \mathcal{E}$ be a differentiable local minimizer function for $f|_{\mathcal{E}(\cdot, N)}$ where $N \subset \mathbb{R}^q$ is an open set. Suppose that ψ is twice differentiable on $(N_x \times N)$ where N_x is open and $\mathcal{X}(N) \subset N_x$. For some $L = \{j_1, j_2, \dots, j_l\} \subset \{1, \dots, r\}$, put*

$$G := \begin{bmatrix} G_{j_1} \\ G_{j_2} \\ \dots \\ G_{j_l} \end{bmatrix} \quad \text{and} \quad \theta := \begin{bmatrix} \theta_{j_1} \\ \theta_{j_2} \\ \dots \\ \theta_{j_l} \end{bmatrix}. \quad (32)$$

Let $\Pi_E : \mathbb{R}^p \rightarrow E$ denote the orthogonal projection and $p_E := \dim E$. Suppose that

$$\begin{aligned} \text{rank}(G\Pi_E) &> p_E - \text{rank } D_{12}\psi|_{\mathcal{E}(x, y)}, \\ \forall (x, y) &\in N_x \times N. \end{aligned} \quad (33)$$

For every $i \in \{1, \dots, r\}$ such that $T_i \neq \emptyset$, assume that $\text{rank}(G_i\Pi_E) = s$. For any $N' \subseteq N$, define

$$V_{L, N'} := \{y \in N' : G\mathcal{X}(y) = \theta\}. \quad (34)$$

Then

- (i) the interior of $V_{L, N}$ is empty;
- (ii) if \mathcal{X} is C^1 on N , then for any compact $N' \subset N$, the relevant $V_{L, N'}$ is included in a closed, negligible subset of \mathbb{R}^q .

This theorem is illustrated below using a very classical cost-function.

Example 5 (Regularized least-squares). Consider the function $f : \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}$,

$$f(x, y) = \|Ax - y\|^2 + \beta\|Gx - \theta\|^2, \quad (35)$$

where $\beta > 0$ and $G \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^s)$. Since [44], cost-functions of this form are among the most widely used tools in signal and image estimation [15, 26, 29, 43]. Under the assumption $\ker(A^T A) \cap \ker(G^T G) = \{0\}$, for every $y \in \mathbb{R}^q$, the function $f(\cdot, y)$ is strictly convex and its unique minimizer \hat{x} satisfies $D_1 f(\hat{x}, y) = 0$, where $D_1 f(\hat{x}, y) = 2(A\hat{x} - y)^T A + 2\beta(G\hat{x} - \theta)^T G$. The relevant minimizer function $\mathcal{X} : \mathbb{R}^q \rightarrow \mathbb{R}^p$ reads

$$\mathcal{X}(y) = (A^T A + \beta G^T G)^{-1}(A^T y + \beta G^T \theta). \quad (36)$$

Let the rows of G be denoted $G_i \in \mathcal{L}(\mathbb{R}^p, \mathbb{R})$ for $i = 1, \dots, r$. For a given index $i \in \{1, \dots, r\}$, we will determine the set $V_{\{i\}}$ of all data points $y \in \mathbb{R}^q$ for which satisfy exactly the equation $G_i \mathcal{X}(y) = \theta_i$,

$$V_{\{i\}} := \{y \in \mathbb{R}^q : G_i \mathcal{X}(y) = \theta_i\}.$$

Using (36), $V_{\{i\}}$ equivalently reads

$$\begin{aligned} V_{\{i\}} &= \{y \in \mathbb{R}^q : y^T p_i(\beta) = c_i(\beta)\}, \\ \text{where } p_i(\beta) &= A(A^T A + \beta G^T G)^{-1} G_i^T, \\ c_i(\beta) &= \theta_i - \beta \theta^T G(A^T A + \beta G^T G)^{-1} G_i^T. \end{aligned}$$

We can have $p_i(\beta) = 0$ only if $\text{rank } A < p$ and β is such that $G_i^T \in \ker A(A^T A + \beta G^T G)^{-1}$: the latter is a system of r polynomials of degree p . If there are $\beta > 0$ satisfying this system, they form a finite, discrete set of values. However, β in (35) will almost never belong to such a set, so in general, $p_i(\beta) \neq 0$. Then $V_{\{i\}} \subset \mathbb{R}^q$ is an affine subspace of dimension $q - 1$. More generally, we have the implication

$$\begin{aligned} \exists i \in \{1, \dots, r\} \quad \text{such that } G_i \mathcal{X}_i(y) &= \theta_i \\ \Rightarrow y &\in \bigcup_{i=1}^r V_{\{i\}}. \end{aligned}$$

The union on the right side is composed of r subspaces of \mathbb{R}^q of dimension $q - 1$. The chance that noisy data come across this union is null. Hence, the chance that noisy data y yield a minimizer $\mathcal{X}(y)$ which satisfies $G_i \mathcal{X}(y) = \theta_i$, even for one index i , is null.

4.3. Commentary

4.3.1. On Conditions (a) and (b). For the data-fidelity term given in (7),

$$D_{12}\psi(x, y) = -2H^T HA, \quad \forall (x, y) \in \mathbb{R}^p \times \mathbb{R}^q,$$

where $\text{rank } H^T H = q$. For the ECT and TCT applications, given in (8), we have

$$\begin{aligned} D_{12}\psi(x, y) &= (\text{diag}(a_1^T x, \dots, a_q^T x))^{-1} A \quad (\text{ECT}), \\ D_{12}\psi(x, y) &= A \quad (\text{TCT}), \end{aligned}$$

where a_i^T , for $i = 1, \dots, q$, are the rows of A . The diagonal matrix in the expression for ECT is always positive definite, since $x \in (\mathbb{R}_+^*)^p$ and the entries of A are non-negative. In all these cases,

$$\text{rank } D_{12}f(x, y) = \text{rank } D_{12}\psi(x, y) = \text{rank } A.$$

If the constraint in (11) is absent, $E = \mathbb{R}^p$, $\Pi_E = I$, $f|_E = f$ and $\psi|_E = \psi$, and $p_E = p$. The condition in (33) reads $\text{rank } G > p - \text{rank } A$ and it vanishes if $\text{rank } A = p$. If the constraint in (11) is applied, $\text{rank } D_{12}\psi|_E(x, y) = \text{rank}(A\Pi_E)$. If $\ker A \cap E = \emptyset$, we find $\text{rank } D_{12}\psi|_E(x, y) = p_E$. For a good choice of the operators $\{G_i\}$ in (5), we have $\text{rank}(G_i\Pi_E) = \text{rank } G_i$, for all i , and then $\text{rank}(G\Pi_E) = \text{rank } G$.

Observe that the conclusion of Example 5 is independent of $\text{rank } A$, provided that $\ker(A^T A) \cap \ker(G^T G) = \{0\}$. This fact suggests that (29) and (33) are quite strong sufficient conditions, and that the conclusions of Theorems 3 and 4 hold even if they fail. The latter is corroborated by the numerical experiments: the images in Fig. 3, restored using smooth regularization, satisfy $G_i\hat{x} \approx 0$, for all i .

4.3.2. On the Minimizer Function \mathcal{X} . By Lemma 2 it is seen that if for every $y \in \mathcal{F}$, $f(\cdot, y)$ is convex and $D_1^2(\cdot, y)$ is positive definite on \mathcal{E} , there is a unique \mathcal{C}^{m-1} minimizer function $\mathcal{X} : \mathcal{F} \rightarrow \mathcal{E}$ for $f(\cdot, \mathcal{F})$. The latter conditions are satisfied if f is of the form (5)–(7) where HA is one-to-one, and $\{\varphi_i\}$ are \mathcal{C}^m and convex. Non-convex functions f of the form (5)–(7), where HA is one-to-one and Φ is \mathcal{C}^m and satisfies some loose requirements, are considered in [18]. The analysis there shows that for almost every $y \in \mathbb{R}^q$ (except those contained in a closed, negligible subset of \mathbb{R}^q), every local minimizer \hat{x} of $f(\cdot, y)$ results from a \mathcal{C}^{m-1} local minimizer function \mathcal{X} .

4.3.3. Significance of the Results. Theorems 3 and 4 and Proposition 4 reveal that the minimizers of smooth cost-functions, corresponding to noisy data, can almost never satisfy weak constraints as those given in (2). The reason is that *all* data points \tilde{y} , which may give rise to $g(\mathcal{X}(\tilde{y})) = \theta$ for some local minimizer function \mathcal{X} , are contained in closed negligible subsets of \mathcal{F} . The chance that noisy data come across such sets is null. In particular, no smooth cost-function dependant on noisy data yields minimizers which are constant on some regions—this is nicely demonstrated by Fig. 3 (a), (b) and (c).

In some applications it is important to know whether the minimizers \hat{x} of a cost-function f can avoid a given constraint $g(\hat{x}) = \theta$, where g is a general function. *E.g.*, in some signal processing applications, one is seeking a non-singular covariance matrix $\hat{x} \in \mathbb{R}^{p \times p}$. Although the set $\{x \in \mathbb{R}^{p \times p} : \det(x) = 0\}$ is closed and negligible in $\mathbb{R}^{p \times p}$ [41], the question whether the set $\{y \in \mathcal{F} : \det(\mathcal{X}(y)) = 0\}$ is negligible in \mathcal{F} when f is smooth, remains. Generalization to non-affine weak constraints is a problem to consider.

Incidentally, Remark 3 and Lemma 3 provide an instructive comparison of the ranks of $D\mathcal{X}$ corresponding to a non-smooth, and to a smooth cost-function, respectively.

5. Restoration of an Image

By way of illustration, we present the restoration of a blurred, noisy 128×128 synthetic image using both non-smooth and smooth regularized cost-functions. The original image in Fig. 1(a) presents smoothly varying regions, constant regions and sharp edges. Data in Fig. 1(b) correspond to $y = a * x + n$, where a is a blur operator with entries $a_{i,j} = \exp(-(i^2 + j^2)/12.5)$ for $-4 \leq i, j \leq 4$, and n is white Gaussian noise yielding 20 dB of SNR. The amplitudes of the original image are in the range of $[0, 1.32]$ and those of the data in $[-5, 50]$. All restorations are calculated by minimizing

$$f(x, y) = \|a * x - y\|^2 + \beta \sum_{(i,j)} \sum_{(k,l) \in U_{i,j}} \phi(|x_{i,j} - x_{k,l}|),$$

$$\text{where } U_{i,j} = \{(i-1, j), (i, j-1)\},$$

for $1 \leq i, j \leq 128$, and boundary condition $x_{i,0} = x_{i,1}$ and $x_{1,j} = x_{128,j}$, as well as $x_{129,j} = x_{128,j}$ and $x_{j,129} = x_{j,128}$. Thus, with every pixel (i, j) , we

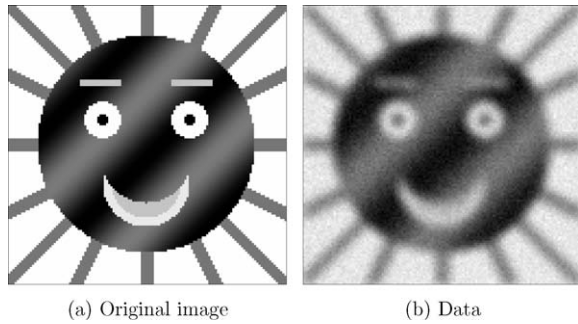


Figure 1. The original image contains weakly varying and constant regions, separated by edges. Data are a blurred and noisy version of the original image.

associated $g_{i,j}^1(x) = x_{i,j} - x_{i-1,j}$ and $g_{i,j}^2(x) = x_{i,j} - x_{i,j-1}$, along with $\theta_{i,j}^1 = \theta_{i,j}^2 = 0$. All restored images below correspond with different functions ϕ among those given in (10), and α and β are set experimentally. (Notice that the properties we illustrate are independent of the exact value of α and β .) In all figures, the obtained minimizers are displayed on the top, while below we give three sections of the restored images, corresponding to rows 35, 54 and 90.

In the experiments in Fig. 2, ϕ is non-smooth at zero. Figure 2(a) illustrates a minimizer of the total-variation cost-function, where $\phi(t) = t$ and $\beta = 80$. This image is constant on many regions. The minimizer in Fig. 2(b) is obtained using $\phi(t) = \alpha t / (1 + \alpha t)$, and ($\alpha = 20, \beta = 100$). It is composed of mainly constant regions. A “0-1” function, $\phi(t) = 1 - \mathbb{1}_{(t=0)}$, along with $\beta = 25$, is applied in Fig. 2(c). The corresponding ϕ has infinite one-sided derivatives at zero. Although such functions are not considered in this paper, the relevant minimizers behave similarly. In all these images, constant regions are naturally encircled by closed contours which justifies the use of non-smooth regularization for the segmentation images.

The minimizers in Fig. 3 correspond to functions ϕ which are smooth at zero. The image in Fig. 3(a) corresponds to $\phi(t) = t^\alpha$ and ($\alpha = 1.4, \beta = 7$). It presents a smoothly varying surface. The minimizer in Fig. 3(b) is defined using $\phi(t) = \alpha t^2 / (1 + \alpha t^2)$, and ($\alpha = 25, \beta = 35$). It exhibits both smoothly varying regions and sharp edges, but contours are not closed. The restoration in Fig. 3(c) is obtained using $\phi(t) = \min\{1, \alpha t^2\}$, for ($\alpha = 60, \beta = 10$). The resultant image corroborates Theorem 4, since it exhibits weakly varying zones.

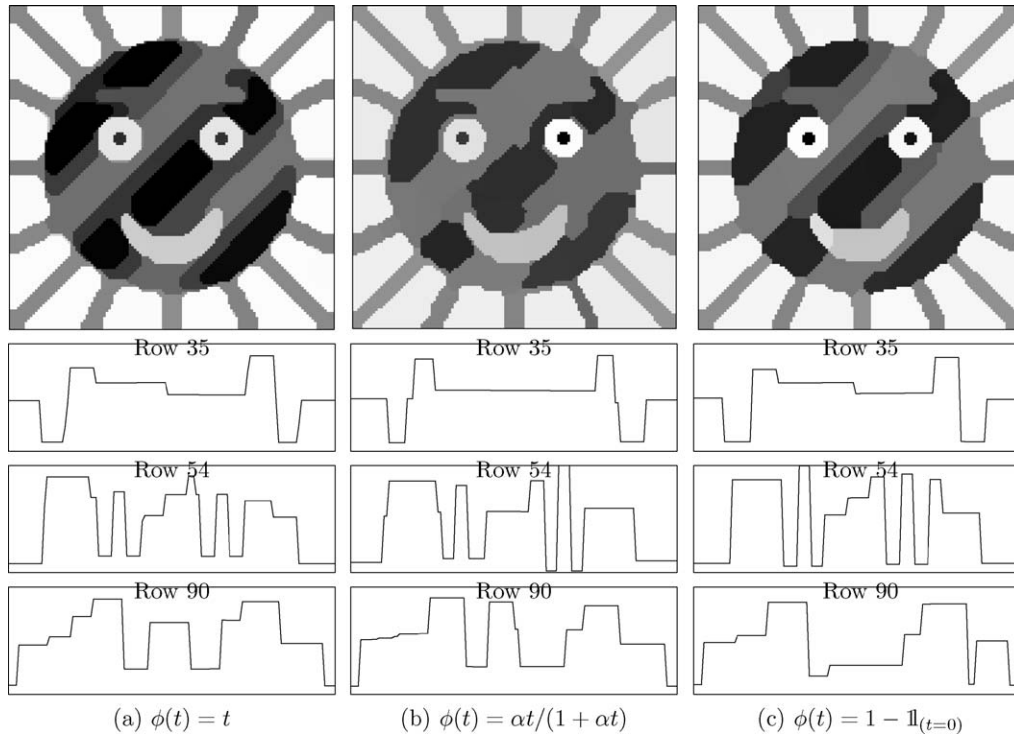


Figure 2. Restoration using non-smooth regularization.

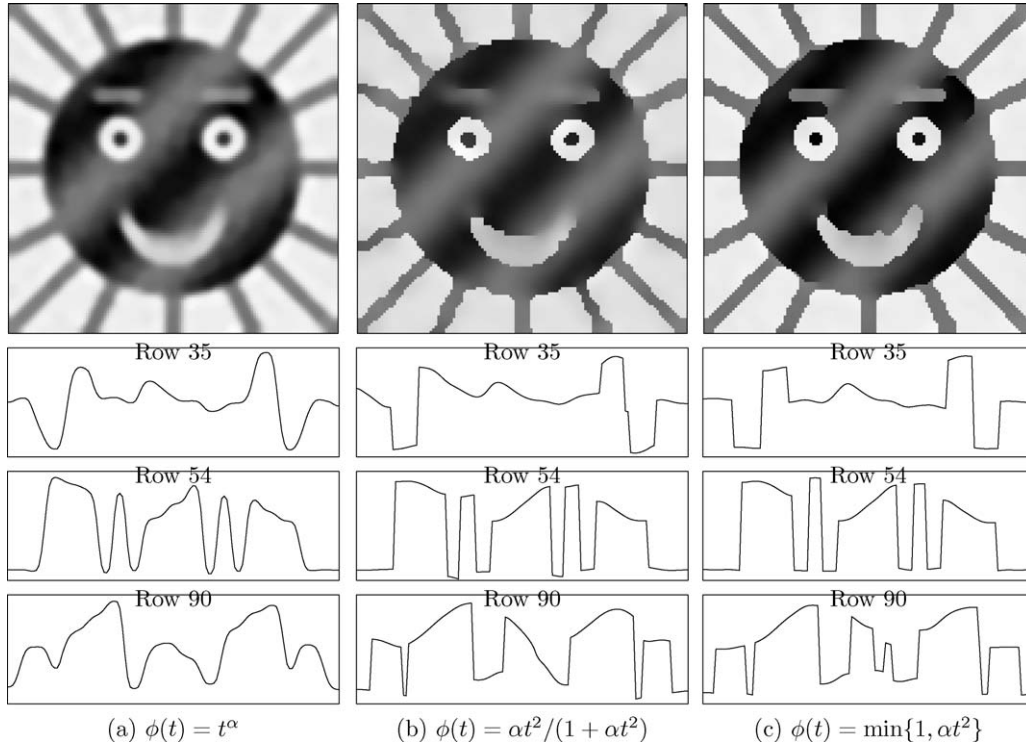


Figure 3. Restoration using smooth regularization.

All minimizers presented here are calculated using a generalized graduated non-convexity method presented in [33]. The images for which ϕ is non-convex may correspond to local minimizers which are not global. Recall that our theory holds for any local minimizers.

6. Conclusion

We introduced the notion of weakly constrained minimization as the possibility that the local minimizers of a cost-function satisfy a varying subset of constraints. We focused on affine constraints. We provided sufficient conditions ensuring that the local minimizers of cost-functions satisfy a set of weak constraints when data ranges over open subsets. We saw that the relevant cost-functions are non-smooth at all points satisfying these constraints. We proved that in contrast, the minimizers of smooth cost-functions can almost never satisfy weak constraints. All these results were applied to possibly constrained regularized cost-functions. Assumptions were justified and analytical examples were

presented. A numerical experiment illustrates our theoretical results.

Appendix

Proof of Lemma 1: Since $f(., \tilde{y})$ has a minimum at \hat{x} , there is $\sigma > 0$ such that for every $t \in (0, \sigma)$ and $u \in E \cap S$, we have $f(\hat{x} + tu, \tilde{y}) - f(\hat{x}, \tilde{y}) \geq 0$, and hence $(f(\hat{x} + tu, \tilde{y}) - f(\hat{x}, \tilde{y}))/t \geq 0$. Then

$$\inf_{t \in (0, \sigma)} \frac{f(\hat{x} + tu, \tilde{y}) - f(\hat{x}, \tilde{y})}{t} \geq 0, \quad \forall u \in E \cap S.$$

At the limit when $\sigma \searrow 0$, we get the result. \square

Example 6 (Infinite one-sided derivative). Consider the function $f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$f(x, y) = (x - y)^2 + \beta\varphi(x),$$

$$\text{where } \varphi(0) = 0 \quad \text{and} \quad \varphi(x) = 1, \quad \forall x \neq 0.$$

Using Definition 3 we find $\delta\varphi(0)(1) = \varphi'(0^+) = \infty$. We have $g(x) = x, \theta = 0$, and $\ell(0) = \{1\}$ and $\ell(x) = \emptyset$

if $x \neq 0$. There are two local minimizer functions on \mathbb{R} , namely $\mathcal{X}_1(y) = 0$ and $\mathcal{X}_2(y) = y$, which merge at $y = 0$. There is a global minimum at $\mathcal{X}_1(y)$ if $|y| \leq \sqrt{\beta}$ and at $\mathcal{X}_2(y)$ if $|y| \geq \sqrt{\beta}$.

Example 1 (Detail). For any $x \in \mathcal{K}_{\{1\}}$ and $u \in K_{\{1\}}^\perp$, we calculate $\delta_1 f$ using Definition 2. We have

$$\begin{aligned} \frac{f(x + tu, y) - f(x, y)}{t} &= D_1 \psi(x, y)u + t(u_1^2 + u_2^2) \\ &+ \begin{cases} (u_1 - u_2) \left(2 + \sin \frac{1}{t(u_1 - u_2)} \right) & \text{if } u_1 > u_2, \\ -(u_1 - u_2) - t(u_1 - u_2)^2 & \text{if } u_1 \leq u_2, \end{cases} \end{aligned}$$

where $D_1 \psi(x, y) = (2x_1 - y_1 - y_2, 2x_2 - y_1 + y_2)$. If $\sigma > 0$ and $u \in K_{\{1\}}^\perp$ with $u \neq 0$, then

$$\begin{aligned} \inf_{t \in (0, \sigma)} \frac{f(x + tu, y) - f(x, y)}{t} &= D_1 \psi(x, y)u \\ &+ \begin{cases} |u_1 - u_2| & \text{if } u_1 > u_2, \\ -(u_1 - u_2) + \sigma u_1 u_2 & \text{if } u_1 \leq u_2, \end{cases} \end{aligned}$$

where we use the facts that $u_1 u_2 \leq 0$ for $u \in K_{\{1\}}^\perp$, and that for every $\sigma > 0$, and for every $\kappa \geq 0$, $\inf_{t \in (0, \sigma)} (\kappa t + 2 + \sin(1/t)) = 1$. Taking the limit when $\sigma \searrow 0$ shows that for any $x \in \mathcal{K}_{\{1\}}$,

$$\delta_1 f(x, y)(u) = D_1 \psi(x, y)u + |u_1 - u_2|, \quad \forall u \in K_{\{1\}}^\perp.$$

Since $x \in \mathcal{K}_{\{1\}}$ and $u \in K_{\{1\}}^\perp$ entail that $x_1 u_1 + x_2 u_2 = 0$ and $y_1(u_1 + u_2) = 0$, we get (17).

Proof of Theorem 1: The result is trivial if $L = \emptyset$, since $\mathcal{K}_\emptyset = \mathcal{E}$ and $K_\emptyset^\perp = \{0\}$. So consider that L is nonempty. From Assumption 1, the application $u \rightarrow \delta_1 f(\hat{x}, \tilde{y})(u)$ is lower semi-continuous on $K_L^\perp \cap S$. Combining this with (a), and with the compactness of $K_L^\perp \cap S$, shows that there exists $\kappa > 0$ such that

$$\delta_1 f(\hat{x}, \tilde{y})(u) \geq \kappa, \quad \forall u \in K_L^\perp \cap S. \quad (37)$$

By Assumption 1 yet again, with every $u \in K_L^\perp \cap S$ there is associated $\rho_u > 0$, with $(B(\hat{x}, \rho_u) \cap \mathcal{K}_L) \subset N_{\hat{x}}$ and $B(\tilde{y}, \rho_u) \subset N$, such that for every $(x, y) \in (B(\hat{x}, \rho_u) \cap \mathcal{K}_L) \times B(\tilde{y}, \rho_u)$,

$$\begin{aligned} \delta_1 f(x, y)(u') &\geq \delta_1 f(\hat{x}, \tilde{y})(u) - \frac{\kappa}{2} \geq \frac{\kappa}{2}, \\ \forall u' &\in B(u, \rho_u) \cap K_L^\perp \cap S, \end{aligned} \quad (38)$$

where the last inequality comes from (37). The set $\{B(u, \rho_u) \cap K_L^\perp \cap S : u \in K_L^\perp \cap S\}$ is an open covering of $K_L^\perp \cap S$. Since $K_L^\perp \cap S$ is compact, we can extract a finite sub-covering, say

$$\begin{aligned} \{B(u^i, \rho_{u^i}) \cap K_L^\perp \cap S : u^i \in K_L^\perp \cap S \\ \text{for } i = 1, \dots, m\}. \end{aligned} \quad (39)$$

Put $\rho_1 := \min\{\rho_{u^i} : i = 1 \dots, m\}$. By (38), for every $(x, y) \in (B(\hat{x}, \rho_1) \cap \mathcal{K}_L) \times B(\tilde{y}, \rho_1)$ we have

$$\begin{aligned} \delta_1 f(x, y)(u') &\geq \frac{\kappa}{2}, \\ \forall u' &\in B(u^i, \rho_{u^i}) \cap K_L^\perp \cap S, \quad \forall i = 1, \dots, m. \end{aligned}$$

Combining this result with (39) shows that

$$\begin{aligned} (x, y) \in (B(\hat{x}, \rho_1) \cap \mathcal{K}_L) \times B(\tilde{y}, \rho_1) \\ \Rightarrow \delta_1 f(x, y)(u) \geq \frac{\kappa}{2}, \quad \forall u \in K_L^\perp \cap S. \end{aligned} \quad (40)$$

By Assumption 2, there exists $\rho_2 \in (0, \rho_1]$ such that if $(x, y) \in (B(\hat{x}, \rho_2) \cap \mathcal{K}_L) \times B(\tilde{y}, \rho_2)$, then

$$\begin{aligned} \sigma \in (0, \rho_2) \Rightarrow \left| \inf_{t \in (0, \sigma)} \frac{f(x + tu, y) - f(x, y)}{t} \right. \\ \left. - \delta_1 f(x, y)(u) \right| < \frac{\kappa}{4}, \quad \forall u \in K_L^\perp \cap S. \end{aligned}$$

It follows that for every $(x, y) \in (B(\hat{x}, \rho_2) \cap \mathcal{K}_L) \times B(\tilde{y}, \rho_2)$,

$$\begin{aligned} t \in (0, \rho_2) \Rightarrow \frac{f(x + tu, y) - f(x, y)}{t} \\ \geq \delta_1 f(x, y)(u) - \frac{\kappa}{4} \geq \frac{\kappa}{4}, \quad \forall u \in K_L^\perp \cap S, \end{aligned}$$

where the last lower bound is due to (40). Consequently,

$$\begin{aligned} (x, y) \in (B(\hat{x}, \rho_2) \cap \mathcal{K}_L) \times B(\tilde{y}, \rho_2) \Rightarrow f(x + u, y) \\ - f(x, y) \geq \frac{\kappa}{4} \|u\|, \quad \forall u \in K_L^\perp \cap B(0, \rho_2). \end{aligned} \quad (41)$$

Now, condition (b) shows that there exists $\rho_3 \in (0, \rho_2]$ such that

$$y \in B(\tilde{y}, \rho_3) \Rightarrow \mathcal{X}_L(y) \in B\left(\hat{x}, \frac{\rho_2}{2}\right). \quad (42)$$

Put

$$\mathcal{X}(y) := \mathcal{X}_L(y), \quad \forall y \in B(\tilde{y}, \rho_3). \quad (43)$$

For an arbitrary $y \in B(\tilde{y}, \rho_3)$, we consider

$$\begin{aligned} & f(\mathcal{X}(y) + u, y) - f(\mathcal{X}(y), y) \\ &= f(\mathcal{X}(y) + u_0 + u_\perp, y) - f(\mathcal{X}(y) + u_0, y) \\ & \quad + f(\mathcal{X}(y) + u_0, y) - f(\mathcal{X}(y), y), \end{aligned} \quad (44)$$

where $u \in E$ is an arbitrary direction which is decomposed into

$$u = u_0 + u_\perp \quad \text{with } u_0 \in K_L \quad \text{and} \quad u_\perp \in K_L^\perp.$$

Let $u \in B(0, \rho_2/2)$, then $\mathcal{X}(y) + u_0 \in B(\hat{x}, \rho_2) \cap \mathcal{K}_L$ by (42) and (43). The latter, jointly with the facts that $y \in B(\tilde{y}, \rho_2)$ and $u_\perp \in K_L^\perp \cap B(0, \rho_2)$, allows us to apply (41) by identifying $\mathcal{X}(y) + u_0$ with x and u_\perp with u :

$$\begin{aligned} & f(\mathcal{X}(y) + u_0 + u_\perp, y) - f(\mathcal{X}(y) + u_0, y) \\ & \geq \frac{\kappa}{4} \|u_\perp\| > 0, \forall u \in B\left(0, \frac{\rho_2}{2}\right) \quad \text{with } u_\perp \neq 0. \end{aligned} \quad (45)$$

Condition (b) shows also that there exists $\sigma_4 \in (0, \rho_2/2]$, such that

$$\begin{aligned} & f(\mathcal{X}(y) + u_0, y) - f(\mathcal{X}(y), y) \geq 0, \\ & \forall u \in B(0, \sigma_4) \quad \text{with } u_0 \neq 0, \end{aligned} \quad (46)$$

where the inequality is strict if \mathcal{X}_L is a strict local minimizer function. Introducing (45) and (46) into (44) shows that $f(\mathcal{X}(y) + u, y) - f(\mathcal{X}(y), y) > 0$, for every $u \in B(0, \sigma_4) \setminus K_L$, and $f(\mathcal{X}(y) + u, y) - f(\mathcal{X}(y), y) \geq 0$, for every $u \in B(0, \sigma_4) \cap K_L$. Hence, $f(\cdot, y)$ reaches a local minimum at $\mathcal{X}(y)$, and the latter is strict whenever \mathcal{X}_L is a strict local minimizer function. The same conclusion will be obtained for all $\mathcal{X}(y)$ corresponding to $y \in B(\tilde{y}, \rho_3)$, hence \mathcal{X} is a local minimizer function. Since for every $y \in B(\tilde{y}, \rho_3)$ we have $\mathcal{X}(y) \in \mathcal{K}_L$, (18) is satisfied and $V_L \supset B(\tilde{y}, \rho_3)$. \square

Proof of Proposition 1: If L is empty, the domain of $\delta_1 f$ is empty. Consider in the following that L is nonempty. Define the function $(x, y, t, u') \rightarrow \Delta_1 f(x, y)(t, u')$ by

$$\Delta_1 f(x, y)(t, u') = \frac{f(x + tu', y) - f(x, y)}{t}.$$

Consider an arbitrary $u \in K_L^\perp \cap S$. Let us fix an arbitrary $\varepsilon > 0$. By Definition 5, there exists $\rho_1 > 0$, such that $\overline{B(\hat{x}, \rho_1) \cap \mathcal{K}_L} \subset N_{\hat{x}}$ and $\overline{B(\tilde{y}, \rho_1)} \subset N$, and

$$\begin{aligned} t \in (0, \rho_1) \Rightarrow & |\delta_1 f(x, y)(u') - \Delta_1 f(x, y)(t, u')| \\ & \leq \frac{\varepsilon}{3}, \quad \forall x \in N_{\hat{x}}, \quad \forall y \in N, \quad \forall u' \in K_L^\perp \cap S. \end{aligned} \quad (47)$$

Choose $t_0 \in (0, \rho_1/2)$. Being continuous, f is uniformly continuous on the compact $\overline{B(\hat{x}, \rho_1) \cap \mathcal{K}_L} \times \overline{B(\tilde{y}, \rho_1)}$. Then there exists $\rho_2 \in (0, \rho_1/2)$ such that if $(x, y) \in (B(\hat{x}, \rho_2) \cap \mathcal{K}_L) \times B(\tilde{y}, \rho_2)$ and $u' \in K_L^\perp \cap S$, then $x + t_0 u' \in B(\hat{x}, \rho_1) \cap \mathcal{K}_L$ and we have

$$\begin{aligned} |f(\hat{x} + t_0 u, \tilde{y}) - f(x + t_0 u', y)| & \leq \frac{t_0 \varepsilon}{6} \quad \text{and} \\ |f(\hat{x}, \tilde{y}) - f(x, y)| & \leq \frac{t_0 \varepsilon}{6}. \end{aligned}$$

It follows that for every $(x, y) \in (B(\hat{x}, \rho_2) \cap \mathcal{K}_L) \times B(\tilde{y}, \rho_2)$ and $u' \in K_L^\perp \cap S$ we have

$$\begin{aligned} & |\Delta_1 f(\hat{x}, \tilde{y})(t_0, u) - \Delta_1 f(x, y)(t_0, u')| \\ &= \frac{1}{t_0} |f(\hat{x} + t_0 u, \tilde{y}) - f(\hat{x}, \tilde{y}) - f(x + t_0 u', y) \\ & \quad + f(x, y)| \leq \frac{\varepsilon}{3}, \end{aligned}$$

and then, using (47),

$$\begin{aligned} & |\delta_1 f(\hat{x}, \tilde{y})(u) - \delta_1 f(x, y)(u')| \\ & \leq |\delta_1 f(\hat{x}, \tilde{y})(u) - \Delta_1 f(\hat{x}, \tilde{y})(t_0, u)| \\ & \quad + |\Delta_1 f(\hat{x}, \tilde{y})(t_0, u) - \Delta_1 f(x, y)(t_0, u')| \\ & \quad + |\delta_1 f(x, y)(u') - \Delta_1 f(x, y)(t_0, u')| \\ & \leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3}. \end{aligned}$$

Then $(x, y, u') \rightarrow \delta_1 f(x, y)(u')$ is continuous at (\hat{x}, \tilde{y}, u) . The same holds for every $u \in K_L^\perp \cap S$. \square

Proof of Proposition 2: If $L = \emptyset$, then $\mathcal{K}_L = \mathcal{E}$ and $f = f_0$ is \mathcal{C}^m . We will consider that L is nonempty. Let $\hat{x} \in \mathcal{K}_L \cap N_x$ be such that $G_i \hat{x} \neq \theta_i$ for all $i \in L^c$. Let $\tilde{y} \in N_y$. Choose an arbitrary $\varepsilon > 0$. Consider the decomposition of f given in (21). First we focus on f_0 . If L^c is nonempty, put

$$\rho_1 := \left(\min_{i \in L^c} \|G_i \hat{x} - \theta_i\| \right) \left(\max_{i \in L^c} \sup_{u \in S} \|G_i u\| \right)^{-1}. \quad (48)$$

Then $\rho_1 > 0$ and by H1, every φ_i , corresponding to $i \in L^c$, is C^m on $B(\hat{x}, \rho_1) \cap \mathcal{E}$. If $L^c = \emptyset$, put $\rho_1 = 1$. Moreover, there is $\rho_2 \in (0, \rho_1)$ such that $(\overline{B(\hat{x}, \rho_2)} \cap \mathcal{E}) \times \overline{B(\tilde{y}, \rho_2)} \subset N_x \times N_y$. Then f_0 is C^m on $(\overline{B(\hat{x}, \rho_2)} \cap \mathcal{E}) \times \overline{B(\tilde{y}, \rho_2)}$. For arbitrary $y \in B(\tilde{y}, \rho_2)$ and $u \in K_L^\perp \cap S$, consider the second-order Taylor expansion of $f_0(\cdot, y)$ about an arbitrary $x \in B(\hat{x}, \rho_2/2) \cap \mathcal{K}_L$,

$$\begin{aligned} & \frac{f_0(x + tu, y) - f_0(x, y)}{t} - D_1 f_0(x, y)u \\ &= t \int_0^1 (1-s) D_1^2 f_0(x + stu, y)(u, u) ds, \end{aligned}$$

where $x + tu \in (B(\hat{x}, \rho_2) \cap \mathcal{E})$ if $t \in [0, \rho_2/2]$. The function

$$(t, u, x, y) \rightarrow \left| \int_0^1 (1-s) D_1^2 f_0(x + stu, y)(u, u) ds \right|$$

is continuous on $[0, \rho_2/2] \times (K_L^\perp \cap S) \times \overline{B(\hat{x}, \rho_2/2)} \times \overline{B(\tilde{y}, \rho_2)}$. The latter set being compact, the function above is upper bounded on this set by a constant $\kappa_1 > 0$. Put

$$\rho_3 := \min \left\{ \frac{\rho_2}{2}, \frac{\varepsilon}{2\kappa_1} \right\}.$$

Then for every $(x, y) \in (B(\hat{x}, \rho_3) \cap \mathcal{K}_L) \times B(\tilde{y}, \rho_3)$ and for every $u \in K_L^\perp \cap S$,

$$\begin{aligned} t \in (0, \rho_3) \Rightarrow & \left| \frac{f_0(x + tu, y) - f_0(x, y)}{t} \right. \\ & \left. - D_1 f_0(x, y) \cdot u \right| \leq \kappa_1 t \leq \frac{\varepsilon}{2}. \end{aligned} \quad (49)$$

Next we focus on f_L . By (20) and (21), for every $u \in E$ and $x \in \mathcal{K}_L$ we have

$$\begin{aligned} f_L(x) &= \sum_{i \in L} \varphi_i(0) \quad \text{and} \\ f_L(x + tu) &= \beta \sum_{i \in L} \varphi_i(tG_i u), \quad \forall x \in \mathcal{K}_L. \end{aligned}$$

Using H2 and Definition 3, for every $x \in \mathcal{K}_L$, the application δf_L is well defined and reads

$$\delta f_L(x)(u) = \beta \sum_{i \in L} \delta \varphi_i(0)(G_i u), \quad \forall u \in E, \quad (50)$$

where $\delta \varphi_i(0)(G_i u) = 0$ if $G_i u = 0$ for some $i \in L$. By H2, for every $i \in L$ there is $\rho_i > 0$, such that

$$\begin{aligned} t \in (0, \rho_i) \Rightarrow & \left| \frac{\varphi_i(tv) - \varphi_i(0)}{t} - \delta \varphi_i(0)(v) \right| \\ & < \frac{\varepsilon}{2r\beta}, \quad \forall v \in \mathbb{R}^s \cap S. \end{aligned}$$

Put

$$\rho_4 := \min \left\{ \rho_3, \left(\min_{i \in L} \rho_i \right) \left(\max_{i \in L} \sup_{u \in S} \|G_i u\| \right)^{-1} \right\}.$$

If $t \in (0, \rho_4)$, then for all $i \in L$ we have $t\|G_i u\| < \rho_i$, for all $u \in S$. Hence

$$\begin{aligned} t \in (0, \rho_4) \Rightarrow & \left| \frac{\varphi_i(tG_i u) - \varphi_i(0)}{t} - \delta \varphi_i(0)(G_i u) \right| \\ & < \frac{\varepsilon}{2r\beta}, \quad \forall u \in K_L^\perp \cap S, \quad \forall i \in L. \end{aligned}$$

This, combined with (50), shows that for all $x \in \mathcal{K}_L$ and for all $u \in K_L^\perp \cap S$,

$$\begin{aligned} t \in (0, \rho_4) \Rightarrow & \left| \frac{f_L(x + tu) - f_L(x)}{t} - \delta f_L(x)(u) \right| \\ &= \beta \left| \sum_{i \in L} \left(\frac{\varphi_i(tG_i u) - \varphi_i(0)}{t} - \delta \varphi_i(0)(G_i u) \right) \right| \\ &\leq \beta \sum_{i \in L} \left| \frac{\varphi_i(tG_i u) - \varphi_i(0)}{t} - \delta \varphi_i(0)(G_i u) \right| \\ &< \beta \#L \frac{\varepsilon}{2r\beta} < \frac{\varepsilon}{2}. \end{aligned} \quad (51)$$

Combining (49) and (51) entails that for every $(x, y) \in (B(\hat{x}, \rho_4) \cap \mathcal{K}_L) \times B(\tilde{y}, \rho_4)$,

$$\begin{aligned} t \in (0, \rho_4) \Rightarrow & \left| \frac{f(x + tu, y) - f(x, y)}{t} - \delta_1 f(x, y)(u) \right| \\ &\leq \left| \frac{f_0(x + tu, y) - f_0(x, y)}{t} - D_1 f(x, y) \cdot u \right| \\ &\quad + \left| \frac{f_L(x + tu) - f_L(x)}{t} - \delta f_L(x)(u) \right| \\ &\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2}, \quad \forall u \in K_L^\perp \cap S. \end{aligned}$$

We can take $N_{\hat{x}} = \mathcal{K}_L \cap B(\hat{x}, \rho_4)$ and $N = B(\tilde{y}, \rho_4)$. \square

Proof of Theorem 2: The proof consists in showing how to apply Theorem 1. Let us fix an arbitrary $x_0 \in E$. For every $i = 1, \dots, r$, define $g_i : E \rightarrow \mathbb{R}^s$ by

$$g_i(x) = G_i x_0 + G_i \Pi_E(x - x_0), \quad \forall x \in E,$$

where Π_E is the orthogonal projection on E . Then for every $x \in \mathcal{E}$, and for all $i = 1, \dots, r$, we have the equivalence: $g_i(x) = \theta_i$ if, and only if, $G_i x = \theta_i$. The applications $\{g_i\}$ are of the form of those considered in Theorem 1. Furthermore, (3) and (6) yield the same set $L = \ell(x)$, for all $x \in \mathcal{E}$, and (4) and (20) define the same affine subspace \mathcal{K}_L .

By Proposition 2, (\hat{x}, \tilde{y}) is contained in a neighborhood $(N_{\hat{x}} \times N') \subset (N_x \cap \mathcal{K}_L) \times N$ such that f admits a one-sided derivative application $\delta_1 f$ which is uniformly defined on $N_{\hat{x}} \times N' \times (K_L^\perp \cap S)$. Combining this with Proposition 1 shows that Assumptions 1 and 2 of Theorem 1 are satisfied. Furthermore, using (50), for every $u \in K_L^\perp \cap S$ we have

$$\begin{aligned} \delta_1 f(\hat{x}, \tilde{y})(u) &= D_1 f_0(\hat{x}, \tilde{y})u + \delta f_L(\hat{x})(u) \\ &= D_1 f_0(\hat{x}, \tilde{y})u + \beta \sum_{i \in L} \delta \varphi_i(0)(G_i u). \end{aligned}$$

Condition (a) is equivalent to $\delta_1 f(\hat{x}, \tilde{y})(u) > 0$, for all $u \in K_L^\perp \cap S$. The result follows from Theorem 1. \square

Proof of Proposition 3: Clearly, φ satisfies H1. Furthermore, for any $u \in \mathbb{R}^s$ we have

$$\delta \varphi(0)(u) = \lim_{t \searrow 0} \frac{\varphi(t\|u\|) - \varphi(0)}{t} = \varphi'(0^+) \|u\| > 0,$$

which shows that $\delta \varphi(0)$ is well defined. By the definition of $\varphi'(0^+)$, for $\varepsilon > 0$, there is $\rho > 0$ such that

$$\left| \frac{\varphi(t) - \varphi(0)}{t} - \varphi'(0^+) \right| < \varepsilon, \quad \forall t \in (0, \rho).$$

The above inequality holds if we write $t\|u\|$ for $u \in \mathbb{R}^s \cap S$ in place of t :

$$\begin{aligned} &\left| \frac{\varphi(t\|u\|) - \varphi(0)}{t} - \varphi'(0^+) \|u\| \right| \\ &= \left| \frac{\varphi(tu) - \varphi(0)}{t} - \delta \varphi(0)(u) \right| < \varepsilon. \end{aligned}$$

Hence $(\varphi(tu) - \varphi(0))/t$ converges towards $\delta \varphi(0)(u)$ when $t \searrow 0$ uniformly on S . \square

Proof of Lemma 3: Since \mathcal{X} is a local minimizer function, $D_1 f(\mathcal{X}(y), y) = 0$, for all $y \in N$. The set N being open, we differentiate with respect to y both sides of the last equation:

$$D_1^2 f(\mathcal{X}(y), y) D\mathcal{X}(y) = -D_{12} f(\mathcal{X}(y), y), \quad \forall y \in N.$$

Identifying the above differentials with their respective Jacobian matrices, we have

$$\begin{aligned} &\text{rank } D_{12} f(\mathcal{X}(y), y) \\ &\leq \min\{\text{rank } D_1^2 f(\mathcal{X}(y), y), \text{rank } D\mathcal{X}(y)\}, \quad \forall y \in N. \end{aligned}$$

Consequently, $\text{rank } D\mathcal{X}(y) \geq \text{rank } D_{12} f(\mathcal{X}(y), y)$, for all $y \in N$.

Proof of Theorem 3: Let $G \in \mathcal{L}(E, \mathbb{R}^s)$ be the linear operator corresponding with g . Then $Dg(\mathcal{X}(y)) = GD\mathcal{X}(y) \in \mathcal{L}(F, \mathbb{R}^s)$, for all $y \in N$. Using a rank inequality [25] and Lemma 3, we find that for all $y \in N$,

$$\begin{aligned} &\text{rank } (GD\mathcal{X}(y)) \geq \text{rank } G + \text{rank } D\mathcal{X}(y) - p \\ &\geq \text{rank } G + \text{rank } D_{12} f(\mathcal{X}(y), y) - p \\ &= \text{rank } Dg(\mathcal{X}(y)) + \text{rank } D_{12} f(\mathcal{X}(y), y) - p, \end{aligned}$$

since $\text{rank } G = \text{rank } Dg(x)$, for all $x \in \mathcal{E}$. Then from condition (29),

$$\text{rank } (GD\mathcal{X}(y)) \geq 1, \quad \forall y \in N. \quad (52)$$

Consider that $\overset{\circ}{V}_N$ is nonempty. From (30), $g(\mathcal{X}(y)) = \theta$, for all $y \in \overset{\circ}{V}_N$. Differentiating with respect to y both sides of this identity shows that for every $y \in \overset{\circ}{V}_N$, we have $GD\mathcal{X}(y) = 0$ and hence $\text{rank } (GD\mathcal{X}(y)) = 0$. But this contradicts (52) since $\overset{\circ}{V}_N \subset N$. It follows that $\overset{\circ}{V}_N = \emptyset$. This proves (i).

Consider an arbitrary compact $N' \subset N$. Let the components of g and G be denoted $g_i : \mathcal{E} \rightarrow \mathbb{R}$ and $G_i \in \mathcal{L}(E, \mathbb{R})$, respectively, for $i = 1, \dots, s$. Focus on some $\tilde{y} \in N'$. From (52), there is an index $j_{\tilde{y}} \in \{1, \dots, s\}$ such that $G_{j_{\tilde{y}}} D\mathcal{X}(\tilde{y}) \neq 0$. Since \mathcal{X} is \mathcal{C}^1 , there is $\rho_{\tilde{y}} > 0$, with $B(\tilde{y}, \rho_{\tilde{y}}) \subset N$, such that

$$y \in B(\tilde{y}, \rho_{\tilde{y}}) \Rightarrow G_{j_{\tilde{y}}} D\mathcal{X}(y) \neq 0,$$

and hence $\text{rank}(G_{j_{\tilde{y}}} D\mathcal{X}(y)) = 1$. It follows that the set

$$\tilde{V}_{\tilde{y}} := \{y \in B(\tilde{y}, \rho_{\tilde{y}}) : g_{j_{\tilde{y}}}(\mathcal{X}(y)) = \theta_{j_{\tilde{y}}}\}$$

is a \mathcal{C}^1 -manifold of dimension $p - 1$ [23, 30, 41]. By the continuity of \mathcal{X} on N , the set $\tilde{V}_{\tilde{y}} \cap N'$ is closed in \mathcal{F} . Such a \tilde{V}_y can be found for every $y \in N'$. The resulting set $\{B(y, \rho_y) : y \in N'\}$ is an open covering of N' . Since N' is compact, there is a finite set of points, say $y^i \in N'$, for $i = 1, \dots, m$, such that

$$N' \subset \bigcup_{i=1}^m B(y^i, \rho_i), \quad \text{where } \rho_i := \rho_{y^i}.$$

Put also $j_i := j_{y^i}$ and $\tilde{V}_i := \tilde{V}_{y^i}$, for every $i = 1, \dots, m$. Using (30), for every $i = 1, \dots, m$,

$$\begin{aligned} V_{N'} \cap B(y^i, \rho_i) &= \{y \in N' \cap B(y^i, \rho_i) : g(\mathcal{X}(y)) = \theta\} \\ &\subseteq \{y \in N' \cap B(y^i, \rho_i) : g_{j_i}(\mathcal{X}(y)) = \theta_{j_i}\} \\ &= \tilde{V}_i \cap N' \subset \tilde{V}_i. \end{aligned}$$

Noticing that $V_{N'} = \bigcup_{i=1}^m (V_{N'} \cap B(y^i, \rho_i))$, we find that

$$V_{N'} \subseteq \bigcup_{i=1}^m \tilde{V}_i \cap N' \subset \bigcup_{i=1}^m \tilde{V}_i.$$

The term in the middle is closed in \mathcal{F} and the term on the right is a finite union of manifolds of dimension $p - 1$. Hence the result. \square

Proof of Proposition 4: Suppose that there is $(\hat{x}, \tilde{y}) \in M$ such that $f(\cdot, \tilde{y})$ has a local minimum at \hat{x} . From (31), there is a direction $u \in E$ for which $-\delta_1 f(\hat{x}, \tilde{y})(-u) > \delta_1 f(\hat{x}, \tilde{y})(u)$. But Lemma 1 implies that $-\delta_1 f(\hat{x}, \tilde{y})(-u) \leq 0 \leq \delta_1 f(\hat{x}, \tilde{y})(u)$, which contradicts the former inequality. It follows that $(\hat{x}, \tilde{y}) \notin M$ and hence $(\mathcal{X}(N) \times N) \subset M^c$. By assumption, M^c is an open subset of $\mathcal{E} \times \mathcal{F}$, hence there is an open subset $N_x \subset \mathcal{E}$ such that $\mathcal{X}(N) \subset N_x$ and $(N_x \times N) \subset M^c$. Then f is twice differentiable on $N_x \times N$. Theorem 3 can be applied and yields the result. \square

Proof of Theorem 4: For some fixed $x_0 \in \mathcal{E}$, consider the affine operator $g : \mathcal{E} \rightarrow \mathbb{R}^s$ given by

$$g(x) = Gx_0 + G\Pi_E(x - x_0), \quad \forall x \in \mathcal{E}.$$

Then $\text{rank} Dg(x) = \text{rank}(G\Pi_E)$, for all $x \in \mathcal{E}$. If we identify f with $f|_{\mathcal{E}}$ and p with p_E , we see that the condition in (29) is satisfied, since $D_{12}f|_{\mathcal{E}} = D_{12}\psi|_{\mathcal{E}}$. Using that $\mathcal{X}(y) \in \mathcal{E}$, for all $y \in N'$, (34) reads

$$\begin{aligned} V_{L, N'} &= \{y \in N' : G\Pi_E(\mathcal{X}(y) - x_0) = \theta - Gx_0\} \\ &= \{y \in N' : g(\mathcal{X}(y)) = \theta\}. \end{aligned}$$

If $T_i = \emptyset$, for all $i = 1, \dots, r$, the result follows from Theorem 3.

Otherwise, the set M introduced in (31) is nonempty and reads

$$M = \bigcup_{i=1}^r M_i \times \mathbb{R}^q, \quad (53)$$

$$\text{where } M_i = \{x \in \mathcal{E} : G_i x - \theta_i \in T_i\}.$$

Let us focus on some $i \in \{1, \dots, r\}$ for which $T_i \neq \emptyset$. For an arbitrary $\tilde{x} \in M_i$, we have $\tilde{z} := G_i \tilde{x} - \theta_i \in T_i$. Since $T_i \subset \mathbb{R}^s$ is a \mathcal{C}^1 -manifold of dimension $s - 1$ (see H4), with this \tilde{z} there are associated a neighborhood $\tilde{N} \subset \mathbb{R}^s$, with $\tilde{z} \in \tilde{N}$, and a \mathcal{C}^1 -function $k : \tilde{N} \rightarrow \mathbb{R}$, so that [23, 30, 41]

$$z \in \tilde{N} \cap T_i \Leftrightarrow \begin{cases} k(z) = 0, \\ \text{rank } Dk(z) = 1. \end{cases} \quad (54)$$

Furthermore, there is $\rho > 0$ such that every $x \in B(\tilde{x}, \rho) \cap \mathcal{E}$ yields $G_i x - \theta_i \in \tilde{N}$. Define now the function

$$\tilde{k}(x) = k(G_i x - \theta_i), \quad \text{for every } x \in B(\tilde{x}, \rho) \cap \mathcal{E}.$$

Since by (53), every $x \in B(\tilde{x}, \rho) \cap M_i$ yields $G_i x - \theta_i \in T_i \cap \tilde{N}$, then (54) shows that

$$\tilde{k}(x) = 0, \quad \text{for every } x \in M_i \cap B(\tilde{x}, \rho). \quad (55)$$

We have $D\tilde{k} \in \mathcal{L}(E, \mathbb{R})$ and

$$\begin{aligned} D\tilde{k}(x) &= Dk(G_i x - \theta_i) (G_i \Pi_E), \\ &\text{for every } x \in B(\tilde{x}, \rho) \cap \mathcal{E}. \end{aligned}$$

Suppose that for some $x \in B(\tilde{x}, \rho) \cap \mathcal{E}$ we have $D\tilde{k}(x) = 0$. Since $Dk(z) \in \mathcal{L}(\mathbb{R}^s, \mathbb{R})$, for $z \in \tilde{N}$, and $\text{rank}(G_i \Pi_E) = s$, it follows that $Dk(G_i x - \theta_i) = 0$. However, if $x \in M_i \cap B(\tilde{x}, \rho)$, then $G_i x - \theta_i \in T_i \cap \tilde{N}$

and (54) ensures that $Dk(G_i x - \theta_i) \neq 0$. We deduce that

$$\text{rank } D\tilde{k}(x) = 1, \quad \forall x \in M_i \cap B(\tilde{x}, \rho). \quad (56)$$

Combining (55) and (56) shows that $M_i \cap B(\tilde{x}, \rho)$ is a manifold of dimension $p_E - 1$, included in \mathcal{E} . Since the same considerations can be developed for every $\tilde{x} \in M_i$, it follows that M_i is a manifold of dimension $p_E - 1$, included in \mathcal{E} . Hence, M is the union of r manifolds of dimension $p_E - 1 + q$, included in $\mathcal{E} \times \mathbb{R}^q$, hence it is a closed, negligible subset of $\mathcal{E} \times \mathbb{R}^q$. The conclusion follows from Proposition 4. \square

Example 3 (Remainder). We first simplify the expression of f using the following change of variables:

$$z = Gx \quad \text{such that} \quad \begin{cases} z_i = x_i - x_{i+1} & \text{for} \\ i = 1, \dots, p-1, \\ z_p = x_p. \end{cases}$$

Consider then the function $\tilde{f}(z, y) := f(G^{-1}z, y)$,

$$\tilde{f}(z, y) = \|Bz - y\|^2 + \beta \sum_{i=1}^{p-1} |z_i| \quad (57)$$

$$\text{where } B = A G^{-1}.$$

Although we analyze \tilde{f} , the results about f follow from the fact that for every $y \in \mathbb{R}^q$,

$$\hat{x} = \arg \min_{x \in \mathbb{R}^p} f(x, y) \Leftrightarrow \hat{z} = G\hat{x} = \arg \min_{z \in \mathbb{R}^p} \tilde{f}(z, y).$$

Correspondingly, the minimizer function \mathcal{Z} for $\tilde{f}(\cdot, \mathbb{R}^q)$ satisfies $\mathcal{Z} = G\mathcal{X}$. For \tilde{f} as given in (58), $\ell(z) = \{i \in \{1, \dots, p-1\} : z_i = 0\}$. Choose some $L \subset \{1, \dots, p-1\}$ such that $0 < \#L < p-1$. Choose then $\hat{z} \in \mathbb{R}^p$ such that $\ell(\hat{z}) = L$, that is $\hat{z}_i = 0$ for $i \in L$ and $\hat{z}_i \neq 0$ for $i \in L^c$. Put b_i to denote the i th column of B , for $i = 1, \dots, p$. Let the matrix B_L be obtained from B by equating to zero all columns of B whose indexes belong to L . Then for any $z \in \mathbb{R}^p$,

$$B_L z = \sum_{i \in L^c} b_i z_i + b_p z_p.$$

With these notations, define the sets

$$I_{\hat{x}} := \left\{ y \in \mathbb{R}^q : \begin{array}{l} 2b_i^T y = 2b_i^T B_L \hat{z} \\ + \beta \text{sign}(\hat{z}_i), \quad \forall i \in L^c \\ 2b_p^T y = 2b_p^T B_L \hat{z} \end{array} \right\}, \quad (58)$$

$$J_{\hat{x}} := \{y \in \mathbb{R}^q : -\beta + 2b_i^T B_L \hat{z} < 2b_i^T y < \beta + 2b_i^T B_L \hat{z}, \quad \forall i \in L\}. \quad (59)$$

Since B is invertible, $I_{\hat{x}}$ is an affine subspace of dimension $p - \#L^c - 1 = \#L$, $J_{\hat{x}}$ is a polyhedron of \mathbb{R}^q , and their intersection,

$$W_{\hat{x}} = I_{\hat{x}} \cap J_{\hat{x}},$$

is a polyhedron of dimension $\#L$. Next we verify that our initial point \hat{z} is the minimizer of $\tilde{f}(\cdot, y)$ for every $y \in W_{\hat{x}}$. With the present notations, $\mathcal{K}_L = K_L = \{u \in \mathbb{R}^p : u_i = 0, \forall i \in L\}$ and $K_L^\perp = \{u \in \mathbb{R}^p : u_i = 0, \forall i \in L^c\}$. The restriction of $\tilde{f}(\cdot, y)$ to \mathcal{K}_L reads

$$\tilde{f}|_{\mathcal{K}_L}(z, y) = \|B_L z - y\|^2 + \sum_{i \in L^c} |z_i|.$$

Since $\hat{z}_i \neq 0, \forall i \in L^c$, there is $\rho > 0$ such that $z_i \neq 0, \forall i \in L^c$, for all $z \in B(\hat{z}, \rho)$. Then $\tilde{f}|_{\mathcal{K}_L}(\cdot, y)$ is C^∞ on $B(\hat{z}, \rho)$ and reaches its minimum at the point where its differential is null. By (58),

$$y \in I_{\hat{x}} \Leftrightarrow D_1(\tilde{f}|_{\mathcal{K}_L})(\hat{z}, y) = 0. \quad (60)$$

Noticing that $\hat{z} \in \mathcal{K}_L$ entails that $B\hat{z} = B_L \hat{z}$, the side-derivative of $\tilde{f}(\cdot, y)$ at \hat{z} for any $u \in K_L^\perp$ reads

$$\delta_1 \tilde{f}(\hat{z}, y)(u) = 2(B_L \hat{z} - y)^T B u + \beta \sum_{i \in L} |u_i|.$$

Then saying that $\delta_1 \tilde{f}(\hat{z}, y)(u) > 0$, for every $u \in K_L^\perp$, is equivalent to

$$|2(B_L \hat{z} - y)^T B_L u| < \beta \sum_{i \in L} |u_i|, \quad \forall u \in K_L^\perp,$$

and hence to

$$\begin{aligned} -\beta \sum_{i \in L} |u_i| + 2(B_L \hat{z})^T B u &< 2y^T B u < \\ \beta \sum_{i \in L} |u_i| + 2(B_L \hat{z})^T B u, &\quad \forall u \in K_L^\perp. \end{aligned}$$

Let $\{e_i : i = 1, \dots, p\}$ denote the canonical basis of \mathbb{R}^p . Noticing that $K_L^\perp = \text{span}\{e_i, i \in L\}$, we consider the above inequality for $u = e_i$, for all $i \in L$. Recalling that $B e_i = b_i$, for every $i \in L$, (59) shows that

$$y \in J_{\hat{x}} \Leftrightarrow \delta_1 \tilde{f}(\hat{z}, y)(u) > 0, \quad \forall u \in K_L^\perp. \quad (61)$$

By Theorem 2, the right sides of (60) and (61) show that \hat{z} is the minimizer of $\tilde{f}(\cdot, y)$ for every $y \in W_{\hat{x}}$. Equivalently, $\hat{x} = G^{-1}\hat{z}$ is the minimizer of $f(\cdot, y)$ for every $y \in W_{\hat{x}}$. Hence Assertion 1.

For $\sigma \in \{-1, +1\}^p$, put

$$\begin{aligned} Q_{L,\sigma} &:= \{z \in \mathcal{K}_L : z_i \neq 0, \forall i \in L^c\} \\ &= \left\{ z \in \mathbb{R}^p : \begin{array}{l} z_i = 0, \forall i \in L \\ \sigma_i z_i > 0, \forall i \in L^c \end{array} \right\}. \end{aligned}$$

Clearly, each $Q_{L,\sigma}$ is a quadrant of \mathcal{K}_L^\perp , so its dimension is $p - \#L$. Then define the set

$$V_{L,\sigma} := \left\{ y \in \mathbb{R}^q : \begin{array}{l} 2b_i^T y = 2b_i^T B_L Q_{L,\sigma} + \sigma_i \beta, \forall i \in L^c, \\ 2b_i^T y = 2b_i^T B_L Q_{L,\sigma}, \\ -\beta + 2b_i^T B_L Q_{L,\sigma} < 2b_i^T y < 2b_i^T B_L Q_{L,\sigma} + \beta, \forall i \in L. \end{array} \right\}$$

Every such $V_{L,\sigma}$ is an unbounded polyhedron of \mathbb{R}^q . Based on (58) and (59), the set \tilde{V}_L mentioned in Assertion 2 can be taken of the form

$$\tilde{V}_L = \bigcup_{i \in L^c} \bigcup_{\sigma_i \in \{-1, 1\}} V_{L,\sigma}.$$

Each V_L is hence the union of $2^{p-\#L-1}$ unbounded polyhedra of \mathbb{R}^q . Notice also that

$$\begin{aligned} \tilde{V}_{\{1, \dots, p-1\}} &= \{y \in \mathbb{R}^q : |2b_i^T y| < \beta, \forall i = 1, \dots, p-1\} \\ \text{and } \tilde{V}_\emptyset &= \{y \in \mathbb{R}^q : |2b_i^T y| > \beta, \forall i = 1, \dots, p-1\}. \end{aligned}$$

For the particular case when B is the identity, we find

$$\tilde{V}_L = \left\{ y \in \mathbb{R}^q : \begin{array}{l} |y_i| < \beta/2 \quad \text{if } i \in L, \\ |y_i| > \beta/2 \quad \text{if } i \in L^c. \end{array} \right\}$$

Notice that the set mentioned in (26) reads $V_L = \bigcup_{L' \supset L} \tilde{V}_{L'}$.

References

1. S. Alliney and S.A. Ruzinsky, "An algorithm for the minimization of mixed l_1 and l_2 norms with application to Bayesian estimation," *IEEE Transactions on Signal Processing*, Vol. SP-42, pp. 618–627, 1994.
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*, Vol. 29, pp. 845–866, 1992.
3. J.E. Besag, "Statistical analysis of non-lattice data," *The Statistician*, Vol. 24, pp. 179–195, 1975.
4. J.E. Besag, "Digital image processing: Towards Bayesian image analysis," *Journal of Applied Statistics*, Vol. 16, pp. 395–407, 1989.
5. M. Black and A. Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications to early vision," *International Journal of Computer Vision*, Vol. 19, pp. 57–91, 1996.
6. A. Blake and A. Zisserman, *Visual Reconstruction*, The MIT Press: Cambridge, 1987.
7. J. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer-Verlag: New York, 2000.
8. C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Transactions on Image Processing*, Vol. 2, pp. 296–310, 1993.
9. C. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Transactions on Image Processing*, Vol. 5, pp. 480–492, 1996.
10. V. Caselles, J.-M. Morel, and C. Sbert, "An axiomatic approach to image interpolation," *IEEE Transactions on Image Processing*, Vol. 7, pp. 59–83, 1998.
11. 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*, Vol. 29, pp. 182–193, 1992.
12. A. Chambolle and P.-L. Lions, "Image recovery via total variation minimization and related problems," *Numerische Mathematik*, Vol. 76, pp. 167–188, 1997.
13. T.F. Chan and C.K. Wong, "Total variation blind deconvolution," *IEEE Transactions on Image Processing*, Vol. 7, pp. 370–375, 1998.
14. F. Clarke, *Optimization and Nonsmooth Analysis*, SIAM: Philadelphia, 1990.
15. G. Demoment, "Image reconstruction and restoration: Overview of common estimation structure and problems," Vol. *IEEE Transactions on Acoustics Speech and Signal Processing*, ASSP-37, pp. 2024–2036, 1989.
16. D. Dobson and F. Santosa, "Recovery of blocky images from noisy and blurred data," *SIAM Journal on Applied Mathematics*, Vol. 56, pp. 1181–1199, 1996.
17. D. Donoho, I. Johnstone, J. Hoch, and A. Stern, "Maximum entropy and the nearly black object," *Journal of the Royal Statistical Society B*, Vol. 54, pp. 41–81, 1992.
18. S. Durand and M. Nikolova, "Stability of minimizers of regularized least squares objective functions I: Study of the local behaviour," Report CMLA No. 2002-08, Cachan, France, 2002.
19. A. Fiacco and G. McCormic, *Nonlinear Programming: Classics in Applied Mathematics*, SIAM: Philadelphia, 1968.
20. D. Geman and G. Reynolds, "Constrained restoration and recovery of discontinuities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-14, pp. 367–383, 1992.
21. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-6, pp. 721–741, 1984.
22. 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.

23. E. Hebey, *Introduction à l'analyse non linéaire des variétés*, Fondations, Cambridge University Press: Paris, Diderot ed., 1997.
24. J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and Minimization Algorithms*, Vol. I and II," Springer-Verlag: Berlin, 1996.
25. R. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
26. T. Kailath, "A view of three decades of linear filtering theory," *IEEE Transactions on Information Theory*, Vol. IT-20, pp. 146–181, 1974.
27. A.C. Kak, "Computerized tomography with X-ray, emission, and ultrasound sources," *Proceedings of the IEEE*, Vol. 67, pp. 1245–1272, 1979.
28. C. Lemaréchal and C. Sagastizábal, "Practical aspects of the Moreau-Yosida regularization I: Theoretical properties," *SIAM Journal on Optimization*, Vol. 7, pp. 367–385, 1997.
29. K.S. Miller and M.M. Rochwarger, "Least-squares methods for ill-posed problems with a prescribed bound," *SIAM Journal on Mathematical Analysis*, Vol. 1, pp. 52–74 1970.
30. J. Milnor, *Topology from the Differential Point of View*, The University Press of Virginia: Charlottesville, 1965.
31. D. Mumford and J. Shah, "Boundary detection by minimizing functionals," in *Proceedings of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 1985, pp. 22–26.
32. M. Nikolova, "Estimées localement fortement homogènes," *Comptes-Rendus de l'Académie des Sciences*, t. 325, Vol. série 1, pp. 665–670, 1997.
33. M. Nikolova, "Markovian reconstruction using a GNC approach," *IEEE Transactions on Image Processing*, Vol. 8 pp. 1204–1220, 1999.
34. M. Nikolova, "Local strong homogeneity of a regularized estimator," *SIAM Journal on Applied Mathematics*, Vol. 61, pp. 633–658, 2000.
35. P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-12, pp. 629–639, 1990.
36. D. Preiss, "Differentiability of Lipschitz functions on Banach spaces," *Journal of Functional Analysis*.
37. W. Ring, "Structural properties of solutions of total variation regularization problems," *ESSAIM: Mathematical Modelling and Numerical Analysis*, Vol. 34, pp. 799–840, 2000.
38. R.T. Rockafellar, *The Theory of Subgradients and its Application to Problems of Optimization: Convex and Nonconvex Functions*, Helderman Verlag: Berlin, 1981.
39. R.T. Rockafellar and J.B. Wets, *Variational Analysis*, Springer-Verlag: New York, 1997.
40. L. Rudin, S. Osher, and C. Fatemi, "Nonlinear total-variation based noise removal algorithm," *Physica*, Vol. 60 D, pp. 259–268, 1992.
41. L. Schwartz, *Analyse II. Calcul différentiel et équations différentielles*, Vol. 2 of Enseignement des sciences, Hermann, Paris, 1997.
42. N.Z. Shor, *Minimization Methods for Non-Differentiable Functions*, Vol. 3, Springer-Verlag, 1985.
43. A. Tarantola, *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier Science Publishers: Amsterdam, 1987.
44. A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems*, Winston, Washington, DC, 1977.
45. C.R. Vogel and M.E. Oman, "Iterative method for total variation denoising," *SIAM Journal on Scientific Computing*, Vol. 17, pp. 227–238, 1996.
46. J. Weickert, *Anisotropic Diffusion in Image Processing*, B.G. Teubner, Stuttgart, 1998.



Mila Nikolova is researcher with the National Center for Scientific Research (CNRS) in France and currently works in the Center for Mathematics and their Applications (CMLA), ENS de Cachan. Her research interests include inverse problems, mathematical image and signal processing, variational problems and their analysis.