Either Fit to Data Entries or Locally to Prior: The Minimizers of Objectives with Nonsmooth Nonconvex Data Fidelity and Regularization

Mila Nikolova

CMLA CNRS ENS Cachan UniverSud
61 av. du Prsident Wilson, 94235 Cachan Cedex, France
nikolova@cmla.ens-cachan.fr
http://www.cmla.ens-cachan.fr/~nikolova/

Abstract. We investigate coercive objective functions composed of a data-fidelity term and a regularization term. Both of these terms are non differentiable and non convex, at least one of them being strictly non convex. The regularization term is defined on a class of linear operators including finite differences. Their minimizers exhibit amazing properties. Each minimizer is the exact solution of an (overdetermined) linear system composed partly of linear operators from the data term, partly of linear operators involved in the regularization term. This is a strong property that is useful when we know that some of the data entries are faithful and the linear operators in the regularization term provide a correct modeling of the sought-after image or signal. It can be used to tune numerical schemes as well. Beacon applications include super resolution, restoration using frame representations, inpainting, morphologic component analysis, and so on. Various examples illustrate the theory and show the interest of this new class of objectives.

Keywords: Image processing, Inverse problems, Non-smooth analysis, Non-convex analysis, Regularization, Signal processing, Variational methods.

1 Introduction

We consider general linear problems where observed data $v[i],\ 1 \leqslant i \leqslant q$, are related to an object of interest $u \in \mathbb{R}^p$ according to

$$v[i] = \langle a_i, u\rangle \text{ with perturbations}, \quad 1 \leqslant i \leqslant q\ .$$

The object u can be a signal or an $n \times m$ image rearranged into a p-length vector. The family of linear operators $\{a_i \in \mathbb{R}^p,\ 1 \leqslant i \leqslant q\}$ can be any. For instance, it can describe direct observation, optical blurring, sub-sampling, missing data problems, a Radon or a Fourier transform (e.g. in computational tomography), and so on [6], [4], [1]. Following a regularization approach, see e.g. [10], [3], [5],

A.M. Bruckstein et al. (Eds.): SSVM 2011, LNCS 6667, pp. 110–121, 2012.

[2], [9], given data $v \in \mathbb{R}^q$, the sought-after solution $\hat{u}$ is defined as a minimizer of an objective $\mathfrak{F}(\cdot, v) : \mathbb{R}^p \mapsto \mathbb{R}$ of the form

$$\mathfrak{F}(u,v) = \sum_{I} \psi\big(\langle a_i, u\rangle - v[i]\big) + \beta \sum_{j\in J} \varphi(\langle g_j, u\rangle), \quad \beta > 0 \tag{1}$$

$$I = \{1, \cdots, q\} \text{ and } J = \{1, \cdots, r\}. \tag{2}$$

The linear operators $\{g_j \in \mathbb{R}^p,\ j \in J\}$ can be any. In practice they produce finite differences of various orders, or discrete Laplacian operators. Let $\{d_j \in \mathbb{R}^p,\ j \in J\}$ denote one of these difference operators; another case of interest is when $g_j = (W^*)^\top d_j$ where W^* is the synthesis operator of a tight frame transform W. To avoid trivialities, it is assumed that

$$a_i \neq 0,\ \forall i \in I \quad \text{and} \quad g_j \neq 0,\ \forall j \in J\ .$$

Let us denote by $A \in \mathbb{R}^{q\times p}$ and $G \in \mathbb{R}^{r\times p}$ the matrices whose rows are all $a_i^\top$ and all $g_i^\top$, respectively:

$$A = [a_1, \cdots, a_q]^\top \quad \text{and} \quad G = [g_1, \cdots, g_r]^\top\ ,$$

where the superscript $^\top$ stands for transposed. We assume that

H1 $\ker A \cap \ker G = \{0\}$.

We adopt the classical notation

$$\mathbb{R}_+ = \{t \in \mathbb{R} : t \geqslant 0\} \quad \text{and} \quad \mathbb{R}_+^* = \{t \in \mathbb{R} : t > 0\}\ .$$

We investigate the case when *both* $\psi : \mathbb{R} \to \mathbb{R}_+$ *and* $\varphi : \mathbb{R} \to \mathbb{R}_+$ *are even nondifferentiable at zero and concave on* $\mathbb{R}_+$*, where at least one of them is strictly concave on* $\mathbb{R}_+$. Thus ψ and φ share some features. The precise assumptions on these functions are presented jointly.

H2 *For* $f = \psi$ *and* $f = \varphi$*, we have*

1. $f : \mathbb{R} \to \mathbb{R}_+$ *is* even, $\mathcal{C}^2$ *on* $\mathbb{R} \setminus \{0\}$ *and* $f(t) > f(0) = 0$ *if* $|t| \neq 0$*;*
2. $f'(0^+) > 0$ *and* $f'(t) > 0$ *on* $\mathbb{R}_+^*$*;*
3. f'' *is increasing on* $\mathbb{R}_+^*$, $f''(t) \leqslant 0, \forall t > 0$ *and* $\lim_{t \searrow 0} f''(t)$ *is well defined.*

H3 *At least one of the functions* $f = \psi$ *or* $f = \varphi$ *satisfy*
f *is strictly concave on* $\mathbb{R}_+$*:* $f''(t) < 0, \forall t > 0$ *and* $\lim_{t\searrow 0} f''(t) < 0$.

Several examples of functions f are shown in Table 1 and plotted in Fig. 1.

1.1 Motivation

An illustration of a minimizer of $\mathfrak{F}(\cdot, v)$ in (1) for $A = \mathrm{I}$, and (ψ, φ) satisfying H2 and H3, is given in Fig. 2. One observes that restored samples either fit data samples exactly or form constant patches.

Table 1. Functions $f\big|_{\mathbb{R}_+} : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying H2. All functions except ($\mathfrak{f}$6) satisfy H3 as well. The functions ($\mathfrak{f}$3), ($\mathfrak{f}$4), ($\mathfrak{f}$5) and ($\mathfrak{f}$6) are coercive.

	($\mathfrak{f}$1)	($\mathfrak{f}$2)	($\mathfrak{f}$3)	($\mathfrak{f}$4)	($\mathfrak{f}$5)	($\mathfrak{f}$6)
$f\big\|_{\mathbb{R}_+}$	$\frac{\alpha\, t}{\alpha\, t+1}$ $\alpha > 0$	$1-\alpha^t$ $0<\alpha<1$	$\ln(\alpha t+1)$ $\alpha > 0$	$(t+\varepsilon)^\alpha - \varepsilon^\alpha$ $0<\alpha<1, \varepsilon>0$	t^α $0<\alpha<1$	t $\cdot$
$f'\big\|_{\mathbb{R}_+^*}$	$\frac{\alpha}{(\alpha t+1)^2}$	$-\alpha^t \ln\alpha$	$\frac{\alpha}{\alpha t+1}$	$\alpha(t+\varepsilon)^{\alpha-1}$	$\alpha t^{\alpha-1}$	1
$f'(0^+)$	α	$-\ln\alpha$	α	$\alpha\varepsilon^{\alpha-1}$	$+\infty$	1
$f''\big\|_{\mathbb{R}_+^*}$	$\frac{-2\alpha^2}{(\alpha t+1)^3}$	$-\alpha^t(\ln\alpha)^2$	$\frac{-\alpha^2}{(\alpha t+1)^2}$	$\alpha(\alpha-1)(t+\varepsilon)^{\alpha-2}$	$\alpha(\alpha-1)t^{\alpha-2}$	0
$\lim_{t\searrow 0} f''(t)$	$-2\alpha^2$	$-(\ln\alpha)^2$	$-\alpha^2$	$\alpha(\alpha-1)\varepsilon^{\alpha-2}$	$-\infty$	0

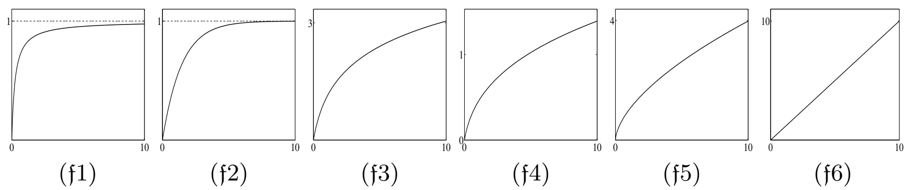


Fig. 1. Plots of the PFs $f\big|_{\mathbb{R}_+}$ given in Table 1

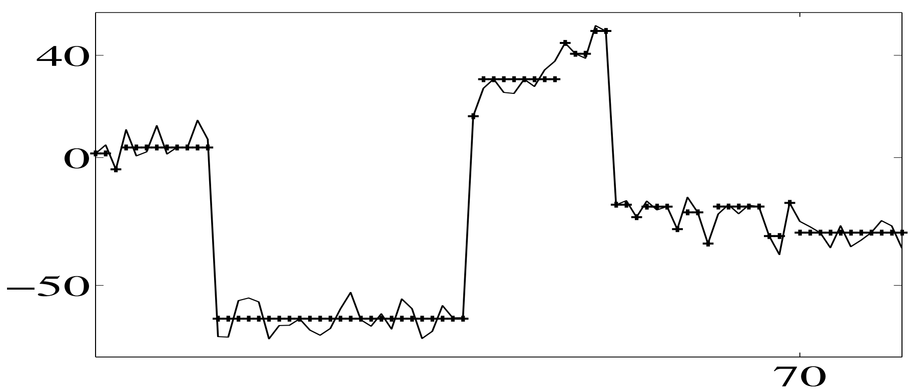


Fig. 2. $\mathfrak{F}(u,v) = \sum_{i=1}^{p} \psi(u[i]-v[i]) + \beta \sum_{i=1}^{p-1} \varphi(u[i+1]-u[i])$ for $\psi(t) = |t|^{0.7}$ and $\varphi(t) = \frac{\alpha\ |t|}{\alpha\ |t|+1}$. Note that H1 is satisfied and that (ψ,φ) satisfy H2 and H3. Data v are plotted with "—", each sample of the minimizer $\hat{u}$ is marked with "+".

Example 1. This example is quite illuminating. Given $v \in \mathbb{R} \setminus \{0\}$, consider $\mathfrak{F}(\cdot, v) : \mathbb{R} \mapsto \mathbb{R}$ for $A = \mathrm{I}$, and a pair of functions (ψ, φ) satisfying H2 and H3:

$$\mathfrak{F}(u, v) = \psi(u - v) + \beta\varphi(u) \ , \forall u \in \mathbb{R}, \tag{3}$$

$$F(u, v) = \mathfrak{F}(u, v), \qquad \forall u \in \mathbb{R} \setminus \{0, v\}. \tag{4}$$

Note that F is the restriction of $\mathfrak{F}$ on $\mathbb{R} \setminus \{0, v\}$.

The differential of order j of a function f with respect to its k-th argument is denoted by $D_k^j f$. Since $\mathfrak{F}$ is coercive, it does admit minimizers. Let $\hat{u}$ be a minimizer of $\mathfrak{F}(\cdot, v)$. The necessary conditions for $\mathfrak{F}$ to have a (local) minimum at $\hat{u} \neq 0$ and $\hat{u} \neq v$, or equivalently, for F to have a (local) minimum at $\hat{u}$, namely $D_1 F(\hat{u}, v) = 0$ and $D_1^2 F(\hat{u}, v) \geqslant 0$, do not hold. Indeed, by H3, the second derivatives on $\mathbb{R} \setminus \{0, v\}$ of ψ and φ are non positive and at least one of them is negative. So

$$D_1^2 F(u, v) = \psi''(u - v) + \beta\varphi''(u) < 0 \quad \forall u \in \mathbb{R} \setminus \{0, v\} \ .$$

Hence there is no minimizer such that $\hat{u} \neq 0$ *and* $\hat{u} \neq v$. In this way, $F(\cdot, v)$ in (4) does not have minimizers. It follows that any minimizer of $\mathfrak{F}(\cdot, v)$ in (3) satisfies

$$\hat{u} \in \{0, v\}.$$

Example 2. Given $v \in \mathbb{R}$, consider $\mathfrak{F}(\cdot, v) : \mathbb{R}^2 \mapsto \mathbb{R}$ as given below:

$$\mathfrak{F}(u, v) = \psi\big(u[1] + u[2] - v\big) + \beta\big(\varphi(u[1]) + \varphi(u[2])\big), \quad 0 < \beta < 1 \ .$$

Let $\psi = \varphi$ satisfy H2 and H3. Then $\mathfrak{F}(\cdot, v)$ has two strict global minimizers

$$\hat{u}_1 = [v, \ 0]^\top \quad \text{and} \quad \hat{u}_2 = [0, \ v]^\top$$

yielding $\mathfrak{F}(\hat{u}_1, v) = \mathfrak{F}(\hat{u}_2, v) = \beta\varphi(v) < \varphi(v) = \psi(v) = \mathfrak{F}(0, v)$. When ψ and φ are nonsmooth and strictly nonconvex on $\mathbb{R}_+$, we have two strict global (sparse) minimizers.

If $\psi(t) = \varphi(t) = |t|$, then $\mathfrak{F}(\cdot, v)$ is convex and reaches its minimum for

$$\hat{u}_t = (1 - t)\,[v, \ 0]^\top + t\,[0, \ v]^\top, \quad 0 \leqslant t \leqslant 1 \ .$$

This yields $\mathfrak{F}(\hat{u}_t, v) = \beta\big|v\big|, \quad 0 \leqslant t \leqslant 1$. The minimum is hence *nonstrict.*

1.2 Notations

Given a $K \times p$ matrix B with rows $b_i^\top$, $1 \leqslant i \leqslant K$, a K-length vector w and a strictly increasing subsequence $\varpi \subset \{1, \cdots, K\}$, say $\varpi = (\varpi[1], \cdots, \varpi[n])$ with $\varpi[1] < \cdots < \varpi[n]$, where $n = \sharp\varpi$, we *systematically* denote

$$B_\varpi = [b_{\varpi[1]}, \cdots, b_{\varpi[n]}]^\top \quad \text{and} \quad w_\varpi[i] = w\big[\varpi[i]\big], \ 1 \leqslant i \leqslant n \ . \tag{5}$$

We write $B_\varpi^\top$ for the transposed of B_ϖ. The range of B_ϖ reads $\mathcal{R}(B_\varpi)$. We denote by $\mathbb{1}$ a column vector of whatever length appropriate to the context composed of ones. If necessary, $\mathbb{1}_K$ specifies that the vector is of length K.

The canonical basis of $\mathbb{R}^K$ is denoted $\big\{e_i, \, i \in \{1, \cdots, K\}\big\}$.

1.3 Outline of the Paper

Existence and strictness of local minimizers are shown in section 2. Section 3 reveals that a strict (local) minimizer is the unique solution of a linear system. Stability of minimizers is studied in section 4. Section 5 focuses on the case when ψ and φ are coercive and strictly nonconvex on $\mathbb{R}_+$. The numerical examples in section 6 confirm the theoretical results. All proofs can be found in [7].

2 Preliminaries

2.1 The Objective $\mathfrak{F}$ Is Not Too Bad

Even though nonconvex and nonsmooth, $\mathfrak{F}(\cdot, v)$ does have minimizers. A general strong sufficient condition is evoked below.

Lemma 1. *Let (ψ, φ) satisfy H2 and H3. Let* one *of the following assumptions hold true:*

(a) $\operatorname{rank}(A) = p$ *and ψ is coercive, i.e.* $\lim_{t\to\infty} \psi(t) = \infty$*;*
(b) H1 holds, and ψ and φ are coercive.

Then $\forall\, v \in \mathbb{R}^q$ and $\forall\, \beta > 0$, the function $\mathfrak{F}(., v)$ in (1) *does admit a minimum.*

We should emphasize that Lemma 1 gives only strong sufficient conditions for the existence of a minimizer. They are not necessary, as illustrated by the example given below.

Example 3. Consider $\mathfrak{F}$ of the form (1) for $p = 3$ and $q = 2$ where

$$A = \begin{bmatrix} 1\,0\,0 \\ 0\,0\,1 \end{bmatrix}, \quad v = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad \begin{array}{l} g_1 = [1\ \ -1\ \ \ 0]^\top, \\ g_2 = [0\ \ \ 1\ \ -1]^\top, \end{array} \quad \psi(t) = |t|\ , \quad \varphi(t) = \frac{\alpha|t|}{\alpha|t|+1}. \tag{6}$$

Assumptions H1, H2 and H3 are satisfied. The objective $\mathfrak{F}$ reads

$$\mathfrak{F}(u, v) = \big|u[1] - v[1]\big| + \big|u[3] - v[2]\big| + \beta\Big(\varphi(u[1] - u[2]) + \varphi(u[2] - u[3])\Big)\ .$$

Clearly, $\mathfrak{F}(., v)$ does not meet the conditions of Lemma 1 since $\operatorname{rank}(A) = 2 < p = 3$ and φ is not coercive. Nevertheless, one computes that for $\alpha = 1$ and $\beta = 2$ the global minimizer of $\mathfrak{F}(., v)$ reads

$$\hat{u} = [1\ \ 1\ \ 3]^\top\ . \tag{7}$$

Below we show that if a (local) minimizer $\hat{u}$ fits exactly some data entries, the relevant rows of A are almost surely linearly independent.

Lemma 2. *For $\nu \subseteq \{1, \cdots, \operatorname{rank} A\}$ such that* $\operatorname{rank} A_\nu < \sharp\nu$*, consider the subset $V_\nu \overset{\text{def}}{=} \{w \in \mathbb{R}^{\sharp\nu}\ :\ w \in \mathcal{R}(A_\nu)\ \}$. We have*

(i) $V_\nu \subsetneq \mathbb{R}^{\sharp\nu}$ is closed and $\mathbb{L}^{\sharp\nu}(V_\nu) = 0$.

(ii) Given $v \in \mathbb{R}^q$ such that $v_\nu \in \mathbb{R}^{\sharp\nu} \setminus V_\nu$, let $\hat{u}$ be a (local) minimizer of $u \mapsto \mathfrak{F}(u,v)$ satisfying $\big\{i \in I : \langle a_i, \hat{u}\rangle = v[i]\big\} = \nu$. Then $\operatorname{rank} A_\nu = \sharp\nu$.

Given $v \in \mathbb{R}^q$, let $\hat{u}$ be a (local) minimizer of $u \mapsto \mathfrak{F}(u,v)$. With each such $\hat{u}$ we systematically associate the following subsets:

$$\nu = \{i \in I : \langle a_i, \hat{u}\rangle = v[i]\} \quad \text{and} \quad \nu^c = I \setminus \nu = \{i \in I : \langle a_i, \hat{u}\rangle \neq v[i]\}\ , \tag{8}$$

$$\sigma = \{i \in J : \langle g_i, \hat{u}\rangle = 0\} \quad \text{and} \quad \sigma^c = J \setminus \sigma = \{i \in J : \langle g_i, \hat{u}\rangle \neq 0\}\ . \tag{9}$$

In the case of *Example 3*, we have $\nu = \{1,2\} = I$ and $\sigma = \{1\}$, so $\nu^c = \varnothing$ and $\sigma^c = \{2\}$.

For $(u,v) \in \mathbb{R}^p \times \mathbb{R}^q$, denote

$$\psi_i(u) = \psi\big(\langle a_i, u\rangle - v[i]\big), \quad \forall\ i \in I, \tag{10}$$
$$\varphi_i(u) = \varphi(\langle g_i, u\rangle), \qquad \forall\ i \in J. \tag{11}$$

Since (ψ, φ) are $\mathcal{C}^2$ on $\mathbb{R} \setminus \{0\}$, one can expect that ψ_i and φ_j in (10)-(11) are locally $\mathcal{C}^2$ provided that $i \notin \nu$ and $j \notin \sigma$.

Lemma 3. *Given $v \in \mathbb{R}^q$, let $\mathfrak{F}(\cdot, v)$ reach a (local) minimum at $\hat{u}$. Let H2 and H3 hold. Put*

$$\rho = \min\left\{\min_{i\in\nu^c} \frac{|\langle a_i, \hat{u}\rangle - v[i]|}{\|a_i\|_2},\ \min_{j\in\sigma^c} \frac{|\langle g_j, \hat{u}\rangle|}{\|g_j\|_2}\right\}.$$

We have $\rho > 0$. Let $u \in B(\hat{u}, \rho) \stackrel{\text{def}}{=} \{w \in \mathbb{R}^p : \|w - \hat{u}\|_2 < \rho\}$ then

$$i \in \nu^c \Rightarrow \psi_i(u) \in \mathcal{C}^2\big(B(\hat{u}, \rho)\big)\ , \tag{12}$$
$$j \in \sigma^c \Rightarrow \varphi_i(u) \in \mathcal{C}^2\big(B(\hat{u}, \rho)\big)\ . \tag{13}$$

2.2 (Local) Minimizers Are Strict

A local minimizer $\hat{u}$ is *strict* if there is a neighborhood $\mathcal{O} \subset \mathbb{R}^N$, containing $\hat{u}$, such that $\mathfrak{F}(\hat{u}, v) < \mathfrak{F}(w, v)$ for any $w \in \mathcal{O}$. Such a minimizer is isolated.

With a (local) minimizer $\hat{u}$ of $\mathfrak{F}(\cdot, v)$ we associate the manifolds given below:

$$\mathcal{K}_{\hat{u}} = \{w \in \mathbb{R}^p : A_\nu w = v_\nu \ \text{ and } \ G_\sigma w = 0\}\ , \tag{14}$$
$$\mathrm{K}_{\hat{u}} = \{w \in \mathbb{R}^p : A_\nu w = 0 \ \text{ and } \ G_\sigma w = 0\}\ , \tag{15}$$

where ν and σ are defined in (8)-(9). Since

$$\hat{u} \in \mathcal{K}_{\hat{u}},$$

we are guaranteed that $\mathcal{K}_{\hat{u}}$ is nonempty. Note that $\mathrm{K}_{\hat{u}}$ is the vector subspace tangent to $\mathcal{K}_{\hat{u}}$. Equivalently, for any $w \in \mathrm{K}_{\hat{u}}$ we have $\hat{u} + w \in \mathcal{K}_{\hat{u}}$: thus $\mathrm{K}_{\hat{u}}$ contains directions in which the (local) minimizer $\hat{u}$ might be nonstrict.

Lemma 4. *Consider* $\mathfrak{F}$ *of the form* (1)*. Let* (ψ,φ) *satisfy H2. For* $v \in \mathbb{R}^q$*, let* $\hat{u}$ *be a (local) minimizer of* $u \mapsto \mathfrak{F}(u,v)$*. The subsets* ν *and* σ *read according to* (8) *and* (9)*, respectively. The vector subspace* $\mathrm{K}_{\hat{u}}$ *is defined in* (15) *and we suppose that*

$$\dim(\mathrm{K}_{\hat{u}}) \geqslant 1 \ .$$

(i) If ψ *satisfies H3 and* $\operatorname{rank} A_\nu < \operatorname{rank} A$*, then* $\exists\, w \in \mathrm{K}_{\hat{u}}$ *such that* $Aw \neq 0$.
(ii) If φ *satisfies H3 and* $\operatorname{rank} G_\sigma < \operatorname{rank} G$*, then* $\exists\, w \in \mathrm{K}_{\hat{u}}$ *such that* $Gw \neq 0$.
(iii) If ψ *and* ϕ *satisfy H3 and we have* $\operatorname{rank} A_\nu < \operatorname{rank} A$ *or* $\operatorname{rank} G_\sigma < \operatorname{rank} G$*, then*

$$\exists\, w \in \mathrm{K}_{\hat{u}} \quad \textit{such that} \quad \Big[Aw \neq 0 \quad \textit{or} \quad Gw \neq 0 \Big] \ .$$

Given $v \in \mathbb{R}^q$, we consider the function given below

$$\begin{aligned} F(\cdot, v) &: \mathcal{K}_{\hat{u}} \mapsto \mathbb{R} \\ F(u,v) &= \sum_{i \in \nu^c} \psi(\langle a_i, u\rangle - v[i]) + \beta \sum_{j \in \sigma^c} \varphi(\langle g_j, u\rangle) \end{aligned} \qquad (16)$$

where $\mathcal{K}_{\hat{u}}$ is defined in (14). Obviously, $F(\cdot,v)$ is the restriction of $\mathfrak{F}(\cdot,v)$ on $\mathcal{K}_{\hat{u}}$. One can remind the function F in Example 1. For any $w \in \mathrm{K}_{\hat{u}}$ we have

$$\langle D_1^2 F(\hat{u},v)w, w\rangle = \sum_{i\in\nu^c} \psi''\big(\langle a_i, \hat{u}\rangle - v[i]\big)\langle a_i, w\rangle^2 + \beta \sum_{i \in \sigma^c} \varphi''(\langle g_j, \hat{u}\rangle)\, \langle g_j, w\rangle^2 \ .$$

Lemma 5. *Let* $\mathfrak{F}$ *be such that* (ψ,φ) *satisfy H2. For* $v \in \mathbb{R}^q$*, let* $\hat{u}$ *be a (local) minimizer of* $\mathfrak{F}(\cdot,v)$*. Suppose that the vector subspace* $\mathrm{K}_{\hat{u}}$ *in* (15) *satisfies*

$$\dim(\mathrm{K}_{\hat{u}}) \geqslant 1 \ .$$

Assume also that one *of the following conditions is met:*

1. ψ *satisfies assumption H3 and* $\operatorname{rank} A_\nu < \operatorname{rank} A$ *;*
2. φ *satisfies assumption H3 and* $\operatorname{rank} G_\sigma < \operatorname{rank} G$ *;*
3. ψ *and* ϕ *satisfy H3, and we have* $\operatorname{rank} A_\nu < \operatorname{rank} A$ *or* $\operatorname{rank} G_\sigma < \operatorname{rank} G$ *.*

Then there exists $w \in \mathrm{K}_{\hat{u}}$ *such that* $\langle D_1^2 F(\hat{u},v)w, w\rangle < 0$ *.*

In general, this lema states quite an unusual result: the restriction of $\mathfrak{F}(\cdot,v)$ on $\mathcal{K}_{\hat{u}}$, namely $F(\cdot,v)$, does not have minimizers. The reader is invited to remind the restricted function F in Example 1 since it does not have minimizers neither.

Next we show that the (local) minimizers of $\mathfrak{F}(\cdot,v)$ are strict in general.

Theorem 1. *Consider* $\mathfrak{F}$ *of the form* (1)*. Let* (ψ,φ) *satisfy H2. For* $v \in \mathbb{R}^q$*, let* $\hat{u}$ *be a (local) minimizer of* $u \mapsto \mathfrak{F}(u,v)$*. The subsets* ν *and* σ *are defined according to* (8) *and* (9)*, respectively, and the vector subspace* $\mathrm{K}_{\hat{u}}$ *is defined in* (15)*. Assume also that* one *of the conditions 1, 2 or 3 in Lemma 5 is met. Then*

$$\mathcal{K}_{\hat{u}} = \{\hat{u}\} \quad \textit{and} \quad \mathrm{K}_{\hat{u}} = \{0\} \ , \qquad (17)$$

so $\mathfrak{F}(\cdot,v)$ *reaches a* strict *(local) minimum at* $\hat{u}$*.*

Example 4. Let us consider again Example 3, p. 114. From the ingredients of $\mathfrak{F}$ given in (6), the minimizer in (7) and the definition of $\mathcal{K}_{\hat{u}}$ in (14), on finds

$$\begin{aligned}\mathcal{K}_{\hat{u}} &= \{w \in \mathbb{R}^3 : \langle a_1, w\rangle = v[1],\ \langle a_2, w\rangle = v[2],\ \langle g_1, w\rangle = 0\}\\ &= \{w \in \mathbb{R}^3 : w[1] = v[1],\ w[3] = v[2],\ w[1] - w[2] = 0\}\\ &= \{w \in \mathbb{R}^3 : w[1] = v[1],\ w[3] = v[2],\ w[2] = w[1]\}\\ &= \{w \in \mathbb{R}^3 : w[1] = w[2] = v[1],\ w[3] = v[2]\}\\ &= \{w \in \mathbb{R}^3 : w[1] = 1,\ w[2] = 1,\ w[3] = 3\}\ =\ \{\hat{u}\}.\end{aligned}$$

Then $\mathrm{K}_{\hat{u}} = \{0\}$.

Let us list the cases when we cannot guarantee that the minimum is strict.

1. $\operatorname{rank} A_\nu = \operatorname{rank} A$ and ψ meets H3 but φ does not.
   Given the fact that A_ν is defined according to (5), the condition given above means that all a_i, $i \in \nu^c$ are linear combinations of $\{a_i, \forall i \in \nu\}$. Then we have the equivalence $\big[A_\nu w = 0 \quad \Leftrightarrow \quad Aw = 0\big]$. In the first instance, this situation occurs when
   $$A\hat{u} = v\ .$$
   In case we wish to change some data equations (e.g. if there is some noise), such a minimizer does not do the job. Otherwise, $\operatorname{rank} A_\nu = \operatorname{rank} A < q$ means that we have reached the maximum among all data entries that can be fitted exactly as far as in general $v \notin \mathcal{R}(A_\nu) = \mathcal{R}(A)$ whose dimension is strictly smaller than the dimension of the data space.
2. $\operatorname{rank} G_\sigma = \operatorname{rank} G$ and φ meets H3 but ψ does not.
   A similar reasoning than above shows that
   $$G_\sigma w = 0 \quad \Leftrightarrow \quad Gw = 0\ .$$
   For instance, if $\{g_j, j \in J\}$ are first-order differences, $G\hat{u} = 0$ means that $\hat{u}$ is constant, i.e. $\hat{u} = c\mathbb{1}$ for any $c \in \mathbb{R} \setminus \{0\}$. Such an $\hat{u}$ is certainly not a meaningful solution.

We conclude that all these cases, excluded from Theorem 1, are quite pathological.

3 Either Fidelity or Prior

3.1 Strict Minimizers Solve Exactly Linear Systems

In spite of the high nonlinearity of the minimization problem, it is shown below that every strict (local) minimizer of $\mathfrak{F}(\cdot, v)$ is the unique solution of a linear system composed out of some elements of $\{a_i, i \in I\}$ and of $\{g_j, j \in J\}$.

Theorem 2. *Let (ψ, φ) satisfy H2 and H3. For $\hat{u}$ a (local) minimizer of $\mathfrak{F}(\cdot, v)$, we posit the definitions of ν and σ in* (8)-(9) *and the one of* $\mathrm{K}_{\hat{u}}$ *in* (15). *Assume* *also that* one *of the conditions 1, 2 or 3 in Lemma 5 is met. Then* $\hat{u}$ *is the* unique *solution of the* linear system of equations *given below:*

$$\begin{aligned} \langle a_i, \hat{u} \rangle &= v[i] \quad \forall i \in \nu \ , \\ \langle g_j, \hat{u} \rangle &= 0 \qquad \forall j \in \sigma \ . \end{aligned} \tag{18}$$

Let $H_{\nu,\sigma} \in \mathbb{R}^{p \times (\sharp\nu + \sharp\sigma)}$ *read*

$$H_{\nu,\sigma} = \begin{bmatrix} A_\nu^\top & G_\sigma^\top \end{bmatrix}^\top \ . \tag{19}$$

We have $\operatorname{rank} H_{\nu,\sigma} = p$ *. Let* $v_{\nu,\sigma} \in \mathbb{R}^{\sharp\nu+\sharp\sigma}$ *have its first subvector equal to* v_ν *and its second* $\sharp\sigma$*-length subvector composed of zeros:* $v_{\nu,\sigma} = \left[v_\nu^\top, (0\, \mathbb{1}_{\sharp\sigma})^\top\right]^\top$*. Then*

$$\hat{u} = (H_{\nu,\sigma}^\top H_{\nu,\sigma})^{-1} H_{\nu,\sigma}^\top v_{\nu,\sigma} \ . \tag{20}$$

Example 5. Let $r = p$ and $a_i = g_i = e_i$ for $i = 1, \cdots, p$. Then $\mathfrak{F}$ reads

$$\mathfrak{F}(u,v) = \sum_{i=1}^{p} \Big(\psi\big(u[i] - v[i]\big) + \beta\varphi\big(u[i]\big)\Big) \ .$$

According to Theorem 2, we have

$$\text{either} \quad \hat{u}[i] = v[i] \quad \text{or} \quad \hat{u}[i] = 0, \quad \forall i \in \{1, \cdots, p\} \ .$$

Next consider that g_i are as in Fig. 2, i.e.

$$g_i[j] = \begin{cases} -1 & \text{if } \ j = i \\ 1 & \text{if } \ j = i+1 \\ 0 & \text{if } \ j \notin \{i, i+1\} \end{cases} \qquad \text{for} \quad i \in \{1, \cdots, p-1\} \ .$$

Now

$$\mathfrak{F}(u,v) = \sum_{i=1}^{p} \psi\big(u[i] - v[i]\big) + \beta \sum_{i=1}^{p-1} \varphi\big(u[i+1] - u[i]\big) \ .$$

By Theorem 2 we find that

$$\hat{u}[i] = v[i] \quad \text{or} \quad \hat{u}[i] = \hat{u}[i+1], \quad \forall \{i, i+1\} \in I \times I \ .$$

In words, the (local) minimizer is composed partly of constant patches, partly of pixels that fit data samples exactly, as seen in Fig. 2.

On the role of the regularization parameter $\beta > 0$. Theorem 2 and in particular the expression for a (local) minimizer $\hat{u}$ given in (20) does not make an explicit reference to the regularization parameter β. Usually $\mathfrak{F}(\cdot, v)$ has numerous (local) minimizers. According to the same theorem, each one of them is strict and is the unique solution of a linear system of the form (18). Any other such (local) minimizer $\hat{u}'$ corresponds to different subsets $\nu' \subset I$ and $\sigma' \subset J$ and in general, $\mathfrak{F}(\hat{u}, v) \neq \mathfrak{F}(\hat{u}', v)$. As far as a minimizer is determined by the subsets $\nu' \subset I$ and $\sigma' \subset J$, the selection of different local minimizers, including the global minimizer, is controlled by β.

4 Local Stability of Strict Minimizers

Here we study how local minimizers do behave under variations of the data.

Definition 1. *Let* $\mathfrak{F} : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ *and* $\mathcal{O} \subseteq \mathbb{R}^q$ *be open. We say that* $\mathcal{U} : \mathcal{O} \to \mathbb{R}^p$ *is a* (local) minimizer function *for the family of functions* $\mathfrak{F}(\cdot, \mathcal{O}) = \{\mathfrak{F}(\cdot, v) : v \in \mathcal{O}\}$ *if for any* $v \in \mathcal{O}$*, the function* $\mathfrak{F}(\cdot, v)$ *reaches a strict (local) minimum at* $\mathcal{U}(v)$*.*

Theorem 3. *Let* (ψ, φ) *satisfy H2 and H3. For* $v \in \mathbb{R}^q \setminus \{0\}$*, let* $\hat{u}$ *be a (local) minimizer of* $u \mapsto \mathfrak{F}(u, v)$*. We posit the definitions of* ν *and* σ *as given in* (8)-(9), *and of* $\mathrm{K}_{\hat{u}}$ *in* (15)*. Assume also that* one *of the conditions 1, 2 or 3 in Lemma 5 is met. Then there exists* $\varrho > 0$ *and a (local) minimizer function* $\mathcal{U}$

$$\|v' - v\|_2 < \varrho \quad \Rightarrow \quad \hat{u}' = \mathcal{U}(v') \tag{21}$$

$$\mathcal{U}(v') = (H_{\nu,\sigma}^\top H_{\nu,\sigma})^{-1} H_{\nu,\sigma}^\top v'_{\nu,\sigma} \tag{22}$$

where $H_{\nu,\sigma}$ *is defined according to* (19)*.*

Note that the (local) minimizer function $\mathcal{U}$ is linear with respect to data v. The *global* minimizer function is *piecewise linear* with respect to data v.

5 A Special Case

Here we address a particular class of functions (ψ, φ), as given in H4 below.

H4 *Assume the following:*

- ψ *and* φ *satisfy H2 and H3 ;*
- ψ *and* φ *are coercive ;*
- $\psi'(0^+) = +\infty$ *and* $\varphi'(0^+) = +\infty$ *.*

Popular examples are ℓ_p "norms" for $0 < p < 1$, see (f4) in Table 1.

Corollary 1. *Theorems 2 and 3 holds true only under assumption 3.*

It appears that each collection of a_i's and g_j's of rank p corresponds to a (local) minimizer of $\mathfrak{F}(\cdot, v)$. The result can be seen as the inverse of Theorem 2.

Theorem 4. *Given* $v \in \mathbb{R}^q$*, let* $\nu \subset I$ *and* $\sigma \subset J$ *be such that the* system of linear equations *given below does admit a* unique *solution* $\hat{u}$*:*

$$\begin{aligned} \langle a_i, \hat{u} \rangle &= v[i] \quad && \forall i \in \nu \ , \\ \langle g_j, \hat{u} \rangle &= 0 && \forall j \in \sigma \ . \end{aligned} \tag{23}$$

Then for any $\beta > 0$*,* $\hat{u}$ *is a strict (local) minimizer of an objective* $\mathfrak{F}(\cdot, v)$ *of the form* (1) *where* (ψ, φ) *satisfy H4.*

6 Numerical Examples

Here we consider a toy missing data recovery problem using $\mathfrak{F} : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$

$$\mathfrak{F}(u,v) = \sum_{i \in I} \psi(\langle a_i, u\rangle - v[i]) + \beta \sum_{i=1}^{p-2} \varphi\big(u[i+2] - 2u[i+1] + u[i]\big) \tag{24}$$

where $p = 80$ for $\psi(t) = |\,t\,|^{\,0.7}$ and $\varphi(t) = \frac{\alpha\,|t|}{\alpha\,|t|+1}$. Here $\langle g_i, \hat{u}\rangle = 0$ means that

$$\hat{u}[i+2] - 2\hat{u}[i+1] + \hat{u}[i] = 0\ , \tag{25}$$

i.e. that three consecutive pixels form a piece of line. The original is shown in Figs. 3 and 4(c) with a dashed line. It contains large polynomial, nearly affine parts.

In the first experiment in Fig 3 we have $a_i = e_i,\ i \in I$ for $q = \sharp I = 25$. Thus $\langle a_i, u\rangle - v[i] = u[i] - v[i]$ in (24). Data samples are plotted with diamonds. These few data samples are largely enough to interpolate all missing parts by affine pieces. The minimizer is strict because φ meets H3.

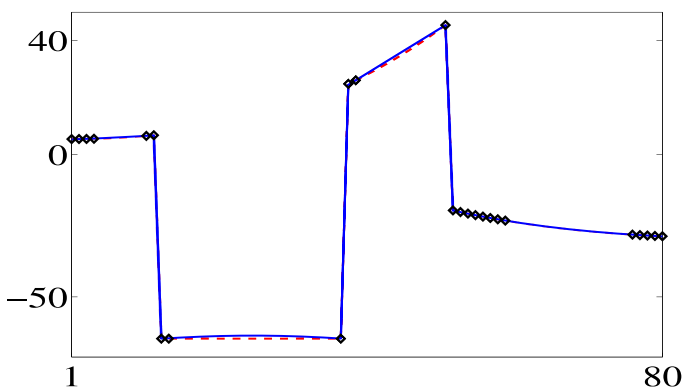
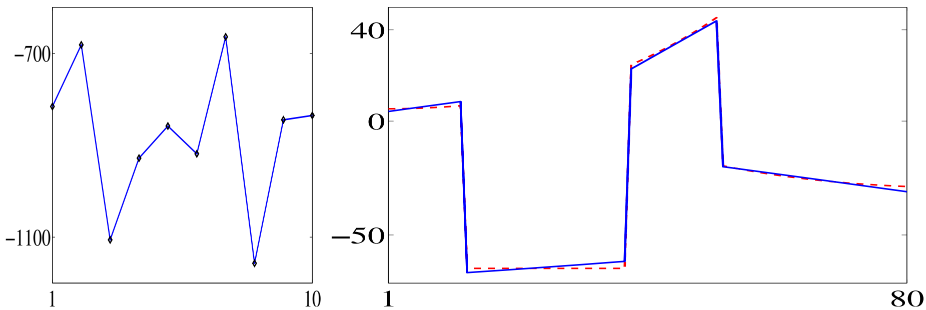


Fig. 3. Data v in $\diamond$, minimizer $\hat{u}$ in thick line, original in dashed line. Results correspond to $\alpha = 4$ and $\beta = 15$.

In the second experiment in Fig. 4 the same original is considered. Ten data samples ($q = \sharp I = 10$) are produced using randomly generated $\{a_i, i \in I\}$. The 10-length data vector v is shown in Fig. 4(a). Yet again, all polynomial parts are interpolated via affine pieces satisfying (25). It is likely that the obtained minimizers $\hat{u}$ yield just a local minimum of $\mathfrak{F}(\cdot, v)$. All data equations are satisfied exactly. Missing parts are fitted using the 2^{nd} order differences in (24). The minimizer is strict because φ meets H3.

The numerical experiments corroborate the theoretical results presented above.



(a) 10 observed data samples v in ($\diamond$), a_i are random

(b) Minimizer $\hat{u}$ in thick line, original in dashed line. Result for $\alpha = 4$ and $\beta = 350$.

Fig. 4. Restoration from 10 random observations

7 Concluding Notes

We show that if ψ and φ are nonconvex and nonsmooth at zero, and at least one of them is strictly nonconvex on $\mathbb{R}_+$, (local) minimizers are generally strict and are given as the unique solution of a linear system composed of linear operators coming from the data term and from the regularization term. This result provides a flexible tool to check if an algorithm minimizing $\mathfrak{F}(\cdot, v)$ has found a strict local minimum.

References

1. Scherzer, O. (ed.): Handbook of Mathematical Methods in Imaging, 1st edn. Springer, Heidelberg (2011)
2. Aubert, G., Kornprobst, P.: Mathematical problems in image processing, 2nd edn. Springer, Berlin (2006)
3. Besag, J.E.: Digital image processing: Towards Bayesian image analysis. Journal of Applied Statistics 16(3), 395–407 (1989)
4. Bovik, A.C.: Handbook of image and video processing. Academic Press, New York (2000)
5. Demoment, G.: Image reconstruction and restoration: Overview of common estimation structure and problems. IEEE Transactions on Acoustics Speech and Signal Processing ASSP-37(12), 2024–2036 (1989)
6. Gonzalez, R., Woods, R.: Digital Image Processing. Addison-Wesley, Reading (1993)
7. Nikolova, M.: Properties of the minimizers of energies with nonsmooth nonconvex data fidelity and regularization. Report (2011)
8. Rockafellar, R.T., Wets, J.B.: Variational analysis. Springer, New York (1997)
9. Scherzer, O., Grasmair, M., Grossauer, H., Haltmeier, M., Lenzen, F.: Variational problems in imaging. Springer, New York (2009)
10. Tikhonov, A., Arsenin, V.: Solutions of Ill-Posed Problems, Winston, Washington DC (1977)