(to appear) Handbook of Mathematical Methods in Imaging, Chapter 5, Springer
Editor: Otmar Scherzer

Energy minimization methods

Mila NIKOLOVA

CMLA, ENS Cachan, CNRS, UniverSud, 61 Av. President Wilson, F-94230 Cachan, France
email: Mila.Nikolova@cmla.ens-cachan.fr

Table of content

Abstract

Energy minimization methods are a very popular tool in image and signal processing. This chapter deals with images defined on a discrete finite set. Energy minimization methods are presented from a non classical standpoint: we provide analytical results on their minimizers that reveal salient features of the images recovered in this way, as a function of the shape of the energy itself. The energies under consideration can be differentiable or not, convex or not. Examples and illustrations corroborate the presented results. Applications that take benefit from these results are presented as well.

Index terms: Analysis of minimizers, Denoising, Edge restoration, Inverse problems, Image restoration, Minimizer function, Non-convex analysis, Non-smooth analysis, Optimization, Perturbation analysis, Proximal analysis, Regularization, Signal and image processing, Signal restoration, Stability analysis, Total variation, Variational methods.

1 Introduction

In numerous applications, an unknown image or a signal $u_o \in \mathbb{R}^p$ is represented by data $v \in \mathbb{R}^q$ according to an observation model, called also forward model

$$v = A(u_o) \odot n, \tag{1}$$

where $A : \mathbb{R}^p \to \mathbb{R}^q$ is a (linear or non linear) transform and n represents perturbations acting via an operation $\odot$. When u is an $m \times n$ image, it is supposed that its pixels are arranged into a p-length real vector[1], where $p = mn$. Some typical applications are for instance denoising, deblurring, segmentation, zooming and super-resolution, reconstruction in inverse problems, coding and compression. In all these cases, recovering a good estimate $\hat{u}$ for u_o needs to combine the observation along with a prior and desiderata on the unknown u_o. A common way to define such an estimate is

$$\underline{\text{Find } \hat{u} \text{ such that}} \quad \underline{\mathcal{F}(\hat{u}, v)} = \underline{\min_{u \in U} \mathcal{F}(u, v)}, \tag{3}$$

$$\underline{\mathcal{F}(u, v)} = \underline{\Psi(u, v) + \beta \Phi(u)}, \tag{4}$$

where $\mathcal{F} : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ is called an energy, $U \subset \mathbb{R}^p$ is a set of constraints, Ψ is a data-fidelity term, Φ brings priors on u_o and $\beta > 0$ is a parameter which controls the trade-off between Ψ and Φ.

The term Ψ ensures that $\hat{u}$ satisfies (1) quite faithfully according to an appropriate measure. The noise n is random and a natural way to derive Ψ from (1) is to use probabilities; see e.g. [17,28,32,50]. More precisely, if $\pi(v|u)$ is the likelihood of data v, the usual choice is

$$\Psi(u, v) = -\log \pi(v|u). \tag{5}$$

For instance, if A is a linear operator and $v = Au + n$ where n is additive independent and identically distributed (i. i. d.) zero-mean Gaussian noise one finds that

$$\Psi(u, v) \propto \|Au - v\|_2^2. \tag{6}$$

[1] Consider an $m \times n$ image u. For instance, its columns can be concatenated, which can be seen as

$$\begin{bmatrix} u[1] & u[m+1] & \cdots & & & \cdots & u[(n-1)m+1] \\ u[2] & u[m+2] & \cdots & & & \cdots & u[(n-1)m+2] \\ & & & \cdots & & & \\ \vdots & \vdots & \cdots & & u[i-1] & \cdots & \\ \vdots & \vdots & \cdots & u[i-m] & u[i] & \cdots & \\ & & & \cdots & & & \\ u[m] & u[2m] & \cdots & & & \cdots & u[p] \end{bmatrix} \tag{2}$$

In this case, the original $u[i, j]$ is identified with $u[(i-1)m + j]$ in (2).

Convex PFs			
$\phi'(0^+) = 0$		$\phi'(0^+) > 0$	
(f1)	$\phi(t) = t^\alpha,\ 1 < \alpha \leqslant 2$	(f5)	$\phi(t) = t$
(f2)	$\phi(t) = \sqrt{\alpha + t^2} - \sqrt{\alpha}$		
(f3)	$\phi(t) = \log(\cosh(\alpha t))$		
(f4)	$\phi(t) = t/\alpha - \log(1 + t/\alpha)$		
Nonconvex PFs			
$\phi'(0^+) = 0$		$\phi'(0^+) > 0$	
(f6)	$\phi(t) = \min\{\alpha t^2, 1\}$	(f10)	$\phi(t) = t^\alpha,\ 0 < \alpha < 1$
(f7)	$\phi(t) = \dfrac{\alpha t^2}{1 + \alpha t^2}$	(f11)	$\phi(t) = \dfrac{\alpha t}{1 + \alpha t}$
(f8)	$\phi(t) = \log(\alpha t^2 + 1)$	(f12)	$\phi(t) = \log(\alpha t + 1)$
(f9)	$\phi(t) = 1 - \exp(-\alpha t^2)$	(f13)	$\phi(0) = 0,\ \phi(t) = 1$ if $t \neq 0$

Table 1: Commonly used PFs $\phi : \mathbb{R}_+ \to \mathbb{R}$ where $\alpha > 0$ is a parameter. Note that among the nonconvex PFs, (f8), (f10) and (f12) are coercive while the remaining PFs, namely (f6), (f7), (f9), (f11) and (f13), are bounded.

This remains quite a common choice partly because it simplifies calculations.

The role of Φ in (4) is to push the solution to exhibit some a priori known or desired features. It is called prior term, or regularization, or penalty term, and so on. In many image processing applications, Φ is of the form

$$\Phi(u) = \sum_{i=1}^{r} \phi(\|\mathrm{D}_i u\|_2), \tag{7}$$

where for any $i \in \{1, \cdots, r\}$, $\mathrm{D}_i : \mathbb{R}^p \to \mathbb{R}^s$, for s an integer $s \geqslant 1$, are linear operators. For instance, the family $\{\mathrm{D}_i\} \equiv \big\{\mathrm{D}_i :\in \{1, \cdots, r\}\big\}$ can represent the discrete approximation of the gradient or the Laplacian operator on u, or finite differences of various orders, or the combination of any of these with the synthesis operator of a frame transform. Note that $s = 1$ if $\{\mathrm{D}_i\}$ are finite differences or a discrete Laplacian; then

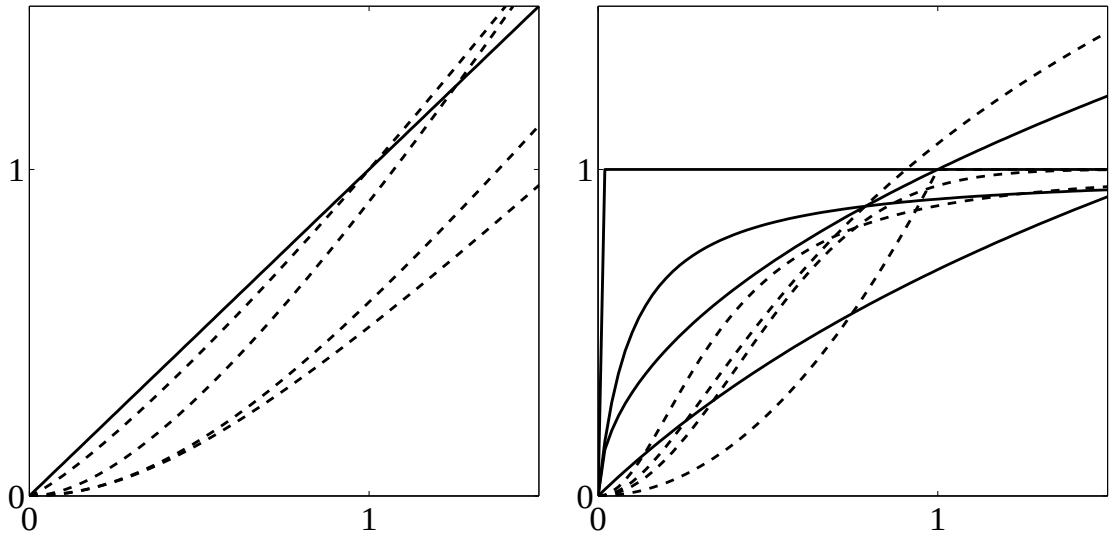
$$s = 1 \quad \Rightarrow \quad \phi(\|\mathrm{D}_i u\|_2) = \phi(|\mathrm{D}_i u|).$$

In (7), $\phi : \mathbb{R}_+ \mapsto \mathbb{R}$ is quite a "general" function, often called a potential function (PF). A very standard assumption is that

H1 $\phi : \mathbb{R}_+ \to \mathbb{R}$ *is increasing and nonconstant on* $\mathbb{R}_+$*, lower semi-continuous and for simplicity,* $\phi(0) = 0$.

Several typical examples for ϕ are given in Table 1 and their plots are seen in Fig. 1.

Remark 1 *Note that if* $\phi'(0^+) > 0$ *the function* $t \to \phi(|t|)$ *is nonsmooth at zero in which case* Φ *is nonsmooth on* $\cup_{i=1}^{r}[w \in \mathbb{R}^p : \mathrm{D}_i w = 0]$. *Conversely,* $\phi'(0^+) = 0$ *leads to a smooth at zero* $t \to \phi(|t|)$.



Convex PFs Nonconvex PFs

Figure 1: Plots of the PFs given in Table 1. PFs with $\phi'(0^+) = 0$ (- - -), PFs with $\phi'(0^+) > 0$ (—).

According to the rules of human vision, an important requirement is that the prior, i.e. Φ should promote smoothing inside homogeneous regions but preserve sharp edges.

1.1 Background

Energy minimization methods, as described here, are at the crossroad of several well established methodologies that are briefly sketched below.

- Bayesian *maximum a posteriori* (MAP) estimation using Markov random field (MRF) priors. Such an estimation is based on the maximization of the posterior distribution

$$\pi(u|v) = \pi(v|u)\pi(u)/Z,$$

where $\pi(u)$ is the prior model for u_o and $Z = \pi(v)$ can be seen as a constant. Equivalently, it minimizes with respect to u the energy

$$\mathcal{F}(u, v) = -\ln \pi(v|u) - \ln \pi(u).$$

Identifying the first term above with $\Psi(\cdot, v)$ and the second one with Φ shows the fundamentals of the equivalence. Key papers on MAP energies involving MRF priors are [11–13, 17, 44, 50].

Since the pioneering work of Geman and Geman [50], various nonconvex PFs ϕ were explored in order to produce images involving neat edges, see e.g. [48, 49, 63]. MAP energies involving MRF priors are also considered in a large amount of books, such as [28, 47, 59, 62]. A recent pedagogical account is found in [96].

- Regularization for ill-posed inverse problems was initiated in the book of Tikhonov and Arsenin [93] in 1977. The background idea can be stated in terms of the stabilization of this kind of problems. Useful textbooks in this direction are e.g. [56, 67, 94] and especially the very recent [88]. This methodology and its most recent achievements are nicely discussed from quite a general point of view in the chapter "Regularization Methods for Ill-Posed problems" inside this Handbook.

- Variational methods are originally related to PDE restoration methods and are naturally developed for signals and images defined on a continuous subset $\Omega \subset \mathbb{R}^d$, $d = 1, 2, \cdots$. For images $d = 2$. Originally, the data-fidelity term is of the form (6) for $A = \mathrm{Id}$ and $\Phi(u) = \int_\Omega \phi(\|Du\|_2)dx$, where ϕ is a convex function as those given in Table 1. Since the beginning of the 90s, a remarkable effort was done to find heuristics on ϕ that enable to recover edges and breakpoints in restored images and signals while smoothing the regions between them (see [5, 10, 22, 27, 62, 86] to name just a few from a huge literature) with a particular emphasis on convex PFs. Up to now, the most successful seems to be the Total Variation (TV) regularization corresponding to $\phi(t) = t$, which was proposed by Rudin, Osher and Fatemi in [86]. Variational methods were rapidly applied along with various linear operators A and more generally, with various data-fidelity terms Ψ. Among all important papers we evoke [22, 27, 53, 71, 82, 92]. The use of differential operators D^k of orders $k \geqslant 2$ in the prior Φ has been rarely investigated; see [19, 24]. Let us remind that whenever D is a differential operator and ϕ is nonconvex, the minimization problem on $u : \Omega \subset \mathbb{R}^d \to \mathbb{R}$ does not admit a solution and no convergence result can be exhibited. More details on variational methods for image processing can be found in several textbooks like [3, 5, 88].

  For numerical implementation, the variational functional is discretized. Using a rearrangement of a discretized finite u into a p-length vector, Φ takes the form[2] of (7) where $r = p$ and $\mathrm{D}_i \in \mathbb{R}^{s\times p}$ for $s = 2$, $1 \leqslant i \leqslant p$.

[2] By a commonly used discretization (see e.g. [23]), using the representation of an image as a vector according to (2), we have

$$\|\mathrm{D}_i u\|_2 = \sqrt{\big(u[i] - u[i-1]\big)^2 + \big(u[i] - u[i-p]\big)^2}, \tag{8}$$

The equivalence between these approaches is considered in several seminal papers, see e.g. [32, 60] as well as the numerous references therein. The state of the art and the relationship among all these methodologies is nicely outlined in the recent book of Scherzer, Grasmair *et al.* [88]. This book gives a brief historical overview of these methodologies and attaches a great importance to the functional analysis of the presented results.

1.2 The main features of the minimizers as a function of the energy

Pushing curiosity ahead leads to various additional questions. One observes that usually data-fidelity and priors are modeled in a separate way. It is hence necessary to check if the minimizer $\hat{u}$ of $\mathcal{F}(\cdot, v)$ meets properly all information contained in the data production model Ψ as well as in the prior Φ. Hence the question: how the prior Φ and the data-fidelity Ψ are they effectively involved in $\hat{u}$—a minimizer of $\mathcal{F}(\cdot, v)$. This leads to formulate the following backward modeling problem:

> **Analyze the mutual relationship between the salient features exhibited by the minimizers $\hat{u}$ of an energy $\mathcal{F}(\cdot, v)$ and the shape of the energy itself.** (9)

This problem was posed in a systematic way and knowingly studied for the first time in [72, 73]. The point of view provided by (9) is actually adopted many authors. Problem (9) is totally general and involves crucial stakes:

- (9) yields rigorous and strong results on the minimizers $\hat{u}$.
- Such a knowledge enables a real control on the solution—the reconstructed image or signal $\hat{u}$.
- Conversely, it opens new perspectives for modeling.
- It enables the conception of specialized energies $\mathcal{F}$ that meet the requirements in various applications.
- This kind of results can help to derive numerical schemes that use what is known about the solution.

Problem (9) remains open and is intrinsically tortuous (which properties to look for, how to conduct the analysis...) The results presented here concern images, signals and data living on finite grids. In this practical framework, the results in this chapter are quite general since they hold for energies

along with appropriate boundary conditions. Other discretization approaches have also been considered, see e.g. [2, 95].

$\mathcal{F}$ which can be convex or non-convex, smooth or non-smooth, and results address local and global minimizers.

1.3 Organization of the chapter

Some preliminary notions and results that help the reading of the chapter are sketched in Section 2. Section 3 is devoted to the regularity of the (local) minimizers of $\mathcal{F}(\cdot, v)$ with a special focus on nonconvex regularization. Section 4 shows how edges are enhanced using non-convex regularization. In section 5 it is shown that non-smooth regularization leads typically to minimizers involving constant patches. Conversely, section 6 exhibits that the minimizers relevant to non-smooth data-fidelity achieve an exact fit for numerous data samples. In all cases, illustrations and applications are presented.

2 Preliminaries

In this section we set the notations and recall some classical definitions and results on minimization problems.

2.1 Notations

We systematically denote by $\hat{u}$ a (local) minimizer of $\mathcal{F}(\cdot, v)$. It is explicitly specified when $\hat{u}$ is a global minimizer. Below n is an integer bigger than one.

- D_j^n—the differential operator of order n with respect to the jth component of a function.
- $v[i]$—the ith entry of vector v.
- $A[i, j]$—the component located at row i and column j of matrix A.
- $\#J$—the cardinality of the set J.
- $J^c = I \setminus J$—the complement of $J \subset I$ in I where I is a set.
- $K^\perp$ – the orthogonal complement of a sub vector space $K \subset \mathbb{R}^n$.
- A^* – transposed of a matrix (or a vector) where A is real-valued.
- $A \succ 0$ ($A \succeq 0$)—the matrix A is positive definite (positive semi-definite)
- $\chi_\Sigma(x) = \begin{cases} 1 & \text{si} \quad x \in \Sigma, \\ 0 & \text{si} \quad x \notin \Sigma. \end{cases}$ – the characteristic function of a set Σ.
- $\mathbb{1}_n \in \mathbb{R}^n$ with $\mathbb{1}_n[i] = 1$, $1 \leqslant i \leqslant n$.

- $\mathbb{L}^n$—the Lebesgue measure on $\mathbb{R}^n$.
- Id—the identity operator.
- $\|.\|_\rho$—a vector or a matrix ρ-norm.
- $\mathbb{R}_+ \stackrel{\text{def}}{=} \{t \in \mathbb{R} \ : \ t \geqslant 0\}$ and $\mathbb{R}_+^* \stackrel{\text{def}}{=} \{t \in \mathbb{R} \ : \ t > 0\}$.
- TV—Total Variation.
- $\{e_1, \ldots, e_n\}$—the canonical basis of $\mathbb{R}^n$, i.e. $e_i[i] = 1$ and $e_i[j] = 0$ if $i \neq j$.

2.2 Reminders and definitions

Definition 1 *A function $\mathcal{F} : \mathbb{R}^p \to \mathbb{R}$ is coercive if $\lim_{\|u\| \to \infty} \mathcal{F}(u) = +\infty$.*

Definition 2 *A function $\mathcal{F}$ on $\mathbb{R}^p$ is proper if $\mathcal{F} : \mathbb{R}^p \to \ ]-\infty, +\infty]$ and if it is not identically equal to $+\infty$.*

A special attention being dedicated to non-smooth functions, we recall some basic facts.

Definition 3 *Given $v \in \mathbb{R}^q$, the function $\mathcal{F}(\cdot, v) : \mathbb{R}^p \to \mathbb{R}$ admits at $\hat{u} \in \mathbb{R}^p$ a one-sided derivative in a direction $w \in \mathbb{R}^p$, denoted $\delta_1\mathcal{F}(\hat{u}, v)(w)$, if the following limit exists:*

$$\delta_1\mathcal{F}(\hat{u}, v)(w) = \lim_{t \searrow 0} \frac{\mathcal{F}(\hat{u} + tw, v) - \mathcal{F}(\hat{u}, v)}{t},$$

where the index 1 in δ_1 specifies that we address derivatives with respect to the first variable of $\mathcal{F}$.

In fact, $\delta_1\mathcal{F}(\hat{u}, v)(w)$ is a right-side derivative; the relevant left-side derivative is $-\delta_1\mathcal{F}(\hat{u}, v)(-w)$. If $\mathcal{F}(\cdot, v)$ is differentiable at $\hat{u}$, then $\delta_1\mathcal{F}(\hat{u}, v)(w) = D_1\mathcal{F}(\hat{u}, v)w$. In particular, for $\phi : \mathbb{R}_+ \to \mathbb{R}$ we have

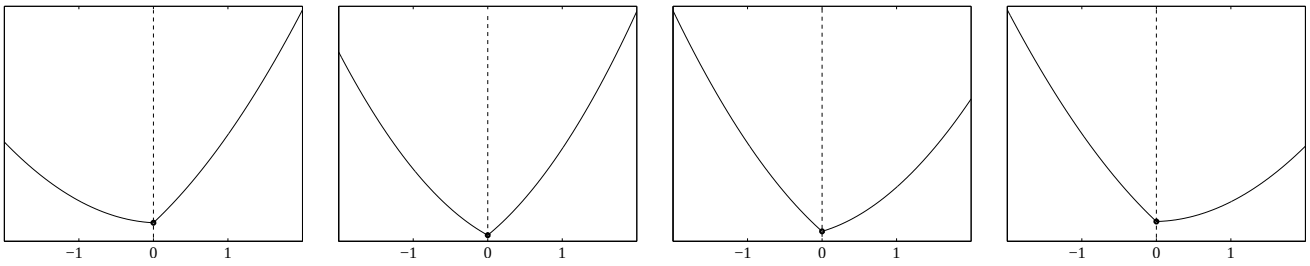
$$\phi'(\theta^+) \stackrel{\text{def}}{=} \delta\phi(\theta)(1) = \lim_{t \searrow 0} \frac{\phi(\theta + t) - \phi(\theta)}{t}, \quad \theta \geqslant 0 \quad \text{and} \quad \phi'(\theta^-) \stackrel{\text{def}}{=} -\delta\phi(\theta)(-1).$$

Next we recall the classical necessary condition for a local minimum of a possibly non-smooth function [55, 85].

Theorem 1 *If $\mathcal{F}(\cdot, v)$ has a local minimum at $\hat{u} \in \mathbb{R}^p$, then $\delta_1\mathcal{F}(\hat{u}, v)(w) \geqslant 0$, for every $w \in \mathbb{R}^p$.*

If $\mathcal{F}(\cdot, v)$ is Fréchet-differentiable at $\hat{u}$, one easily deduce that $D_1\mathcal{F}(\hat{u}, v) = 0$ at a local minimizer $\hat{u}$.

Rademacher's theorem states that if $\mathcal{F}$ is proper and Lipschitz continuous on $\mathbb{R}^p$, then the set of points in $\mathbb{R}^p$ at which $\mathcal{F}$ is not Fréchet-differentiable forms a set of Lebesgue measure zero [55, 85]. Hence $\mathcal{F}$ is differentiable at almost every u. However, when $\mathcal{F}(\cdot, v)$ is non differentiable, its minimizers are typically located at points where $\mathcal{F}(\cdot, v)$ is non differentiable. See e.g. Example 1 and Fig. 2 below.



$\beta = 1,\ v = -0.9$ $\beta = 1,\ v = -0.2$ $\beta = 1,\ v = 0.5$ $\beta = 1,\ v = 0.95$

Figure 2: For the set of values for v given above, $\mathcal{F}(\cdot, v)$ is plotted with "—" while its minimizer $\hat{u}$ is marked with "o". In all these cases $\hat{u}$ lies at a point where $\mathcal{F}(\cdot, v)$ is non differentiable.

Example 1 *Consider* $\mathcal{F}(u, v) = \frac{1}{2}\|u - v\|^2 + \beta|u|$ *for* $\beta > 0$ *and* $u, v \in \mathbb{R}$. *The minimizer* $\hat{u}$ *of* $\mathcal{F}(\cdot, v)$ *reads*

$$\hat{u} = \begin{cases} v + \beta & \text{if} \quad v < -\beta \\ 0 & \text{if} \quad |v| \leqslant \beta \\ v - \beta & \text{if} \quad v > \beta \end{cases} \qquad (\hat{u} \text{ is shrunk w.r.t. } v.)$$

$\mathcal{F}(\cdot, v)$ *and* $\hat{u}$ *are plotted in Fig. 2 for several values of* v.

The next corollary tells us what can happen if the necessary condition in Theorem 1 does not hold.

Corollary 1 *Let* $\mathcal{F}$ *be differentiable on* $(\mathbb{R}^p \times \mathbb{R}^q) \setminus \Theta_0$ *where*

$$\Theta_0 \stackrel{\text{def}}{=} \{(u, v) \in \mathbb{R}^p \times \mathbb{R}^q : \exists w \in \mathbb{R}^p,\ -\delta_1\mathcal{F}(u, v)(-w) > \delta_1\mathcal{F}(u, v)(w)\}. \tag{10}$$

Given $v \in \mathbb{R}^q$, *if* $\hat{u}$ *is a (local) minimizer of* $\mathcal{F}(\cdot, v)$ *then*

$$(\hat{u}, v) \notin \Theta_0.$$

Proof. If $\hat{u}$ is a local minimizer, then by Theorem 1, $\delta_1\mathcal{F}(\hat{u}, v)(-w) \geqslant 0$, hence

$$-\delta_1\mathcal{F}(\hat{u}, v)(-w) \leqslant 0 \leqslant \delta_1\mathcal{F}(\hat{u}, v)(w), \quad \forall w \in \mathbb{R}^p. \tag{11}$$

If $(\hat{u}, v) \in \Theta_0$, the necessary condition (11) cannot hold. □

Example 2 *Suppose that* Ψ *in* (4) *is a differentiable function for any* $v \in \mathbb{R}^q$. *Let the PF* ϕ *be such that for some positive numbers, say* $\theta_1, \cdots, \theta_k$, *its left-hand-side derivative is strictly higher than its right-hand side derivative i.e.* $\phi'(\theta_j^-) > \phi'(\theta_j^+)$, $1 \leqslant j \leqslant k$ *for* k *finite, and that* ϕ *is differentiable beyond this set of numbers. Given a (local) minimizer* $\hat{u}$, *denote*

$$I = \{1, \cdots, r\} \quad \text{and} \quad I_{\hat{u}} = \{i \in I : \|\mathrm{D}_i\hat{u}\|_2 = \theta_j, 1 \leqslant j \leqslant k\}.$$

Define $F(\hat{u}, v) = \Psi(\hat{u}, v) + \beta \sum_{i \in I \setminus I_{\hat{u}}} \phi(\|\mathrm{D}_i \hat{u}\|_2)$. *Note that* $F(\cdot, v)$ *is differentiable at* $\hat{u}$. *The energy* $\mathcal{F}(\cdot, v)$ *at* $\hat{u}$ *reads* $\mathcal{F}(\hat{u}, v) = F(\hat{u}, v) + \beta \sum_{i \in I_{\hat{u}}} \phi(\|\mathrm{D}_i \hat{u}\|_2)$. *Applying the necessary condition* (11) *for* $w = \hat{u}$ *yields*

$$\beta \sum_{i \in I_{\hat{u}}} \phi'(\|\mathrm{D}_i \hat{u}\|_2^-) \leqslant -D_1 F(\hat{u}, v)(\hat{u}) \leqslant \beta \sum_{i \in I_{\hat{u}}} \phi'(\|\mathrm{D}_i \hat{u}\|_2^+).$$

In particular, we must have $\sum_{i \in I_{\hat{u}}} \phi'(\|\mathrm{D}_i \hat{u}\|_2^-) \leqslant \sum_{i \in I_{\hat{u}}} \phi'(\|\mathrm{D}_i \hat{u}\|_2^+)$, *which contradicts the assumption on* ϕ. *It follows that if* $\hat{u}$ *is a (local) minimizer of* $\mathcal{F}(\cdot, v)$, *then* $I_{\hat{u}} = \varnothing$ *and hence*

$$\|\mathrm{D}_j \hat{u}\|_2 \neq \theta_j, \quad 1 \leqslant j \leqslant k.$$

A typical case is the PF (f6) in Table 1, namely $\phi(t) = \min\{\alpha t^2, 1\}$. *Then* $k = 1$ *and* $\theta = \theta_1 = \frac{1}{\sqrt{\alpha}}$. *Remind that (f6) is the discrete equivalent of the Mumford-Shah prior [14].*

The following existence theorem can be find e.g. in [30].

Theorem 2 *For* $v \in \mathbb{R}^q$, *let* $U \subset \mathbb{R}^p$ *be a non-empty and closed subset and* $\mathcal{F}(\cdot, v) : U \to \mathbb{R}$ *a lower semi-continuous (l.s.c.) proper function. If* U *is unbounded (with possibly* $U = \mathbb{R}^p$*), suppose that* $\mathcal{F}(\cdot, v)$ *is coercive. Then there exists* $\hat{u} \in U$ *such that* $\mathcal{F}(\hat{u}, v) = \inf_{u \in U} \mathcal{F}(u, v)$.

We should emphasize that this theorem gives only sufficient conditions for the existence of a minimizer. They are not necessary, as seen in the example below.

Example 3 *Let* $\mathcal{F} : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ *involve (f6) in Table 1 and read*

$$\mathcal{F}(u, v) = (u[1] - v[1])^2 + \beta \phi(|\, u[1] - u[2]\,|) \quad \textit{for} \quad \phi(t) = \max\{\alpha t^2, 1\}, \quad 0 < \beta < \infty.$$

For any v, *it is obvious that* $\mathcal{F}(\cdot, v)$ *is not coercive since it is bounded by* β *in the direction* $\mathrm{span}\{(0, u[2])\}$. *Nevertheless, its global minimum is strict and is reached for* $\hat{u}[1] = \hat{u}[2] = v[1]$. *At the global minimum,* $\mathcal{F}(\cdot, v)$ *gets its minimal value, namely* $\mathcal{F}(\hat{u}, v) = 0$.

Most of the results summarized in this chapter exhibit the behavior of the minimizer points $\hat{u}$ of $\mathcal{F}(\cdot, v)$ under variations of v. In words, they deal with local minimizer functions.

Definition 4 *Let* $\mathcal{F} : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ *and* $O \subseteq \mathbb{R}^q$. *We say that* $\mathcal{U} : O \to \mathbb{R}^p$ *is a local minimizer function for the family of functions* $\mathcal{F}(\cdot, O) = \{\mathcal{F}(\cdot, v) : v \in O\}$ *if for any* $v \in O$, *the function* $\mathcal{F}(\cdot, v)$ *reaches a strict local minimum at* $\mathcal{U}(v)$.

When $\mathcal{F}(\cdot, v)$ is proper, l.s.c. and convex, the standard results below can be evoked, see [30, 43].

Theorem 3 *Let $\mathcal{F}(\cdot, v) : \mathbb{R}^p \to \mathbb{R}$ be proper, convex, l.s.c. and coercive for every $v \in \mathbb{R}^q$.*

(i) Then $\mathcal{F}(\cdot, v)$ has a unique (global) minimum which is reached for a convex and closed set of minimizers $\{\widehat{\mathcal{U}}(v)\} = \left\{ \hat{u} \in \mathbb{R}^p : \mathcal{F}(\hat{u}, v) = \inf_{u \in U} \mathcal{F}(u, v) \right\}$;

(ii) If in addition $\mathcal{F}(\cdot, v)$ is strictly convex, then the minimizer $\hat{u} = \mathcal{U}(v)$ is unique.

Moreover, the minimizer function $v \mapsto \mathcal{U}(v)$ is unique (hence it is global) and it is continuous if $\mathcal{F}$ is continuous [17, Lemmas 1-2, p. 307].

The next lemma, which can be found e.g. in [45], addresses the regularity of the local minimizer functions when $\mathcal{F}$ is smooth. It can be seen as a variant of the Implicit functions theorem.

Lemma 1 *Let $\mathcal{F}$ be $\mathcal{C}^m$, $m \geqslant 2$, on a neighborhood of $(\hat{u}, v) \in \mathbb{R}^p \times \mathbb{R}^q$. Suppose that $\mathcal{F}(\cdot, v)$ reaches at $\hat{u}$ a local minimum such that $D_1^2\mathcal{F}(\hat{u}, v) \succ 0$. Then there are a neighborhood $O \subset \mathbb{R}^q$ containing v and a unique $\mathcal{C}^{m-1}$ local minimizer function $\mathcal{U} : O \to \mathbb{R}^p$, such that $D_1^2\mathcal{F}(\mathcal{U}(\nu), \nu) \succ 0$ for every $\nu \in O$ and $\mathcal{U}(v) = \hat{u}$.*

This lemma is extended in several directions in this chapter.

According to a fine analysis conducted in the 90s and nicely summarized in [5], ϕ preserves edges if H1 holds as if H2, stated below, holds true as well:

H2 $\lim_{t \to \infty} \dfrac{\phi'(t)}{t} = 0.$

This assumption is satisfied by all PFs in Table 1 except for (f1) in case if $\alpha = 2$. We do not evoke the numerous other heuristics for edge preservation as far as they will not be used explicitly in this chapter.

Definition 5 *Let $\phi : [0, +\infty) \to \mathbb{R}$ and $m \geqslant 0$ an integer. We say that ϕ is $\mathcal{C}^m$ on $\mathbb{R}_+$, or equivalently that $\phi \in \mathcal{C}^m(\mathbb{R}_+)$ if and only if the following conditions hold:*

(i) ϕ is $\mathcal{C}^m$ on $(0, +\infty)$;

(ii) the function $t \mapsto \phi\big(|t|\big)$ is $\mathcal{C}^m$ at zero.

Using this definition, for the PF (f1) in Table 1 we see that ϕ is $\mathcal{C}^1$ on $[0, +\infty)$, that $\phi \in \mathcal{C}^2(\mathbb{R}_+)$ for (f4), while for the other differentiable functions satisfying $\phi'(0^+) = 0$ we find $\phi \in \mathcal{C}^\infty(\mathbb{R}_+)$.

3 Regularity results

Here we focus on the regularity of the minimizers of $\mathcal{F} : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ of the form

$$\begin{aligned} \mathcal{F}(u, v) &= \|Au - v\|_2^2 + \beta \sum_{i \in I} \phi(\|\mathrm{D}_i u\|_2), \\ I &\stackrel{\text{def}}{=} \{1, \cdots, r\}, \end{aligned} \tag{12}$$

where $A \in \mathbb{R}^{q \times p}$, and for any $i \in I$ we have $\mathrm{D}_i \in \mathbb{R}^{s \times p}$ for $s \geqslant 1$. Let us denote by D the following $rs \times p$ matrix:

$$\mathrm{D} \stackrel{\text{def}}{=} \begin{bmatrix} \mathrm{D}_1 \\ \cdots \\ \mathrm{D}_r \end{bmatrix}.$$

When A in (12) is not injective, a standard assumption in order to have regularization is

H3 $\ker(A) \cap \ker(\mathrm{D}) = \{0\}$.

Notice that H3 trivially holds when $\operatorname{rank} A = p$. In typical cases $\ker(\mathrm{D}) = \operatorname{span}(\mathbb{1}_p)$, whereas usually $A\mathbb{1}_p \neq 0$, so H3 holds again. Examples for ϕ are seen in Table 1.

3.1 Some general results

We first check the conditions on $\mathcal{F}(\cdot, v)$ in (12) that enable Theorems 2 and 3 to be applied. It is useful to remind that since H1 holds, $\mathcal{F}(\cdot, v)$ in (12) is l.s.c. and proper.

1. Note that $\mathcal{F}(\cdot, v)$ in (12) is coercive for any $v \in \mathbb{R}^q$ at least one of the following cases:

   - $\operatorname{rank}(A) = p$ and $\phi : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is non-decreasing;
   - H1 and H3 hold and ϕ is coercive in addition (e.g. as (f1)-(f5), (f8), (f10) and (f12) in Table 1).

   In these cases, Theorem 2 can be applied and shows that $\mathcal{F}(\cdot, v)$ does admit minimizers.

2. For any $v \in \mathbb{R}^q$, the energy $\mathcal{F}(\cdot, v)$ in (12) is convex and coercive if H1 and H3 hold for a convex ϕ. Then statement (i) of Theorem 3 holds true.

3. Furthermore, $\mathcal{F}(\cdot, v)$ in (12) is strictly convex and coercive for any $v \in \mathbb{R}^q$ if ϕ satisfies H1 and if one of the following assumptions holds true

   - $\operatorname{rank}(A) = p$ and ϕ is convex ;
   - H3 holds and ϕ is strictly convex.

Then statement (*ii*) of Theorem 3 can be applied. In particular, for any $v \in \mathbb{R}^q$, $\mathcal{F}(\cdot, v)$ has a unique strict minimizer and there is a unique local minimizer function $\mathcal{U} : \mathbb{R}^q \to \mathbb{R}^p$ which is continuous (remind Definition 4 on p. 10).

However, the PFs involved in (12) used for signal and image processing are often nonconvex or nondifferentiable. An extension of the standard results given above is hence necessary. This is the goal of the subsequent § 3.2.

3.2 Stability of the minimizers of energies with possibly nonconvex priors

In this subsection the assumptions stated below are considered.

H4 *The operator A in* (12) *satisfies* $\operatorname{rank} A = p$, *i.e. A^*A is invertible.*

H5 *The PF ϕ in* (12) *is $\mathcal{C}^0(\mathbb{R}_+)$ and $\mathcal{C}^m$, $m \geqslant 2$, on $\mathbb{R}_+^*$ with* $0 \leqslant \phi'(0^+) < \infty$;
if $\phi'(0^+) = 0$ it is required also that ϕ is $\mathcal{C}^m$ on $\mathbb{R}_+$ (see Definition 5).

Under H1 (p. 3), H2, H4 and H5, the prior (and hence $\mathcal{F}(\cdot, v)$) in (12) can be nonconvex and in addition nonsmooth. Thanks to H1 and H3, Theorem 2 ensures that for any $v \in \mathbb{R}^q$, $\mathcal{F}(\cdot, v)$ admits a global minimum. However, it can present numerous local minima.

Energies $\mathcal{F}$ with nonconvex and possibly non differentiable PFs ϕ are frequently used in engineering problems since they were observed to give rise to high quality solutions $\hat{u}$. It is hence critically important to have good knowledge on the stability of the obtained solutions.

Even though established under restrictions on A, the results summarized in this subsection provide the state of the art on this subject. Further research is highly desirable to assess the stability of broader classes of energies.

3.2.1 Local minimizers

The stability of local minimizers is a matter of critical importance in its own right for several reasons. In many applications, the estimation of the original signal or image u_o is performed by only locally minimizing a nonconvex energy in the vicinity of some initial guess. Second, it is worth recalling that minimization schemes that guarantee the finding of the global minimum of a nonconvex objective function are exceptional. The practically obtained solutions are usually only local minimizers, hence the importance of knowing their behavior.

The theorem below is a simplified version of the results established in [39].

Theorem 4 *Let $\mathcal{F}(\cdot, v)$ in (12) satisfy H1, H2, H4 and H5. Then there exists a closed subset $\Theta \subset \mathbb{R}^q$ whose Lebesgue measure is $\mathbb{L}^q(\Theta) = 0$ such that for any $v \in \mathbb{R}^q \setminus \Theta$, there exists an open subset $O \subset \mathbb{R}^q$ with $v \in O$ and a local minimizer function (see Definition 4, p. 10) $\mathcal{U} : O \to \mathbb{R}^p$ which is $\mathcal{C}^{m-1}$ on O and meets $\hat{u} = \mathcal{U}(v)$.*

Related questions have been considered in critical point theory, sometimes in semi-definite programming; the well-posedness of some classes of smooth optimization problems was addressed in [37]. A lot of results have been established on the stability of the local minimizers of general smooth energies [45]. It worths noting that these results are quite abstract to be applied directly to our energy in (12).

Commentary on the assumptions. All assumptions H1, H2 and H5 bearing on the PF ϕ are non restrictive at all since they address all nonconvex PFs in Table 1 except for (f13) which is discontinuous at zero. The assumption H4 may, or may not, be satisfied—it depends on the application in mind. This assumption is difficult to avoid, as seen in Example 4.

Example 4 *Consider $\mathcal{F} : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$ given by*

$$\mathcal{F}(u, v) = (u[1] - u[2] - v)^2 + |u[1]| + |u[2]|.$$

where $v \equiv v[1]$. The minimum is obtained after a simple computation.

$$\begin{aligned}
v > \frac{1}{2} \qquad & \{\widehat{\mathcal{U}}(v)\} = \left(c, c - v + \frac{1}{2}\right) \ \text{ for any } \ c \in \left[0, v - \frac{1}{2}\right] \quad \textit{(non-strict minimizer)}; \\
|v| \leqslant \frac{1}{2} \qquad & \hat{u} = 0 \quad \textit{(unique minimizer)} \\
v < -\frac{1}{2} \qquad & \{\widehat{\mathcal{U}}(v)\} = \left(c, c - v - \frac{1}{2}\right) \ \text{ for any } \ c \in \left[v + \frac{1}{2}, 0\right] \quad \textit{(non-strict minimizer)}.
\end{aligned}$$

In this case assumption H4 is violated and there is a local minimizer function only for $v \in \left[-\frac{1}{2}, \frac{1}{2}\right]$.

Intermediate results. The derivations in [39] reveal a series of important intermediate results.

1. If $\phi'(0^+) = 0$ and ϕ is $\mathcal{C}^2(\mathbb{R}_+)$—remind Definition 5—then $\forall v \in \mathbb{R}^q \setminus \Theta$, every local minimizer $\hat{u}$ of $\mathcal{F}(\cdot, v)$ is strict and $D_1^2\mathcal{F}(\hat{u}, v) \succ 0$. Consequently, Lemma 1 is extended since the statement holds true $\forall v \in \mathbb{R}^q \setminus \Theta$.

   For real data v—a random sample of $\mathbb{R}^q$—whenever $\mathcal{F}(\cdot, v)$ is differentiable and satisfies the assumptions of Theorem 4, it as almost sure that local minimizers $\hat{u}$ are strict and their Hessians $D_1^2\mathcal{F}(\hat{u}, v)$ are positive definite.

2. Using Corollary 1, p. 9, the statement of Theorem 4 holds true if $\phi'(0^+) = 0$ and there is $\tau > 0$ such that $\phi'(\tau^-) > \phi'(\tau^+)$. This is the case of the PF (f6) in Table 1 (p. 3) which is the discrete version of the Mumford-Shah regularization.

3. If $\phi'(0^+) > 0$, define

$$\hat{J} \overset{\text{def}}{=} \big\{ i \in I \ : \ \mathrm{D}_i \hat{u} = 0 \big\} \ \text{ and } \ K_{\hat{J}} \overset{\text{def}}{=} \{ w \in \mathbb{R}^p \ : \ \mathrm{D}_i w = 0, \ \forall i \in \hat{J} \}. \tag{13}$$

Then $\forall v \in \mathbb{R}^q \setminus \Theta$, every local minimizer $\hat{u}$ of $\mathcal{F}(\cdot, v)$ is strict and

(a) $D_1\mathcal{F}|_{K_{\hat{J}}}(\hat{u}, v) = 0$ and $D_1^2\mathcal{F}|_{K_{\hat{J}}}(\hat{u}, v) \succ 0$—a sufficient condition for a strict minimum on $K_{\hat{J}}$;

(b) $\delta_1\mathcal{F}(\hat{u}, v)(w) > 0, \ \ \forall w \in K_{\hat{J}}^{\perp} \setminus \{0\}$—a sufficient condition for a strict minimum on $K_{\hat{J}}^{\perp}$.

Let us emphasize that (a) and (b) provide a sufficient condition for a strict (local) minimum of $\mathcal{F}(\cdot, v)$ at $\hat{u}$ (a straightforward consequence of [78, Theorem 1] and Lemma 1). Hence these conditions are satisfied at the (local) minimizers $\hat{u}$ of $\mathcal{F}(\cdot, v)$ for almost every $v \in \mathbb{R}^q$.

We can interpret all these results as it follows:

Under the assumptions H1, H2, H4 and H5, given real data $v \in \mathbb{R}^q$, the chance to get a nonstrict (local) minimizer or a (local) minimizer of the energy in (12) that does not result from a $\mathcal{C}^{m-1}$ local minimizer function, is null.

3.2.2 Global minimizers of energies with possibly nonconvex priors

An overview of the results on global minimizers for several classes of functions can be found in [37]. The setting being quite abstract, the results presented there are difficult to apply to the energy in (12). The results on the global minimizers of (12) presented next are a extracted from [40].

Theorem 5 *Assume that $\mathcal{F}(\cdot, v)$ in (12) satisfy H1, H2, H4 and H5. Then there exists a subset $\hat{\Theta} \subset \mathbb{R}^q$ such that $\mathbb{L}^q(\hat{\Theta}) = 0$ and the interior of $\mathbb{R}^q \setminus \hat{\Theta}$ is dense in $\mathbb{R}^q$, and for any $v \in \mathbb{R}^q \setminus \hat{\Theta}$ the energy $\mathcal{F}(\cdot, v)$ has a unique global minimizer. Furthermore, the global minimizer function $\hat{\mathcal{U}} : \mathbb{R}^q \setminus \hat{\Theta} \to \mathbb{R}^p$ is $\mathcal{C}^{m-1}$ on an open subset of $\mathbb{R}^q \setminus \hat{\Theta}$ which is dense in $\mathbb{R}^q$.*

This means that in a real-world problem there is no chance of getting data v such that the energy $\mathcal{F}(\cdot, v)$ has more than one global minimizer.

We anticipate mentioning that even though $\hat{\Theta}$ is negligible, it plays a crucial role for the recovery of edges; this issue is developed in section 4.

3.3 Nonasymptotic bounds on minimizers

The aim here is to give nonasymptotic analytical bounds on the local and the global minimizers $\hat{u}$ of $\mathcal{F}(\cdot, v)$ in (12) that hold for all PFs ϕ in Table 1. Related questions have mainly been considered in particular situations, such as $A = \mathrm{Id}$, for some particular ϕ, or when v is a special noise-free function, or in the context of the separable regularization of wavelet coefficients, or in asymptotic conditions when one of the terms in (12) vanishes—let us cite among others [4,69,91]. An outstanding paper [44] explores the mean and the variance of the minimizers $\hat{u}$ for strictly convex and differentiable functions ϕ. The bounds provided below are of practical interest for the initialization and the convergence analysis of numerical schemes.

H6 *ϕ is $\mathcal{C}^0$ on $\mathbb{R}_+$ (cf. Definition 5) with $\phi(0) = 0$ and $\phi(t) > 0$ for any $t > 0$.*

H7 *There are two alternative assumptions:*

- *$\phi'(0^+) = 0$ and ϕ is $\mathcal{C}^1$ on $\mathbb{R}_+^* \setminus \Theta_0$ where the set $\Theta_0 = \{t > 0 : \phi'(t^-) > \phi'(t^+)\}$ is at most finite.*
- *$\phi'(0^+) > 0$ and ϕ is $\mathcal{C}^1$ on $\mathbb{R}_+^*$.*

The conditions on Θ_0 in this assumption allows us to address the PF given in (f6). Let us emphasize that under H1 (p. 3), H6 and H7 the PF ϕ can be convex or nonconvex.

The statements given below were derived in [80] where one can find additional bounds and details.

Theorem 6 *Consider $\mathcal{F}$ of the form (12), and let H1, H6 and H7 hold.*

(i) Let one of the following assumptions hold:

(a) $\mathrm{rank}\,(A) = p$

(b) ϕ is strictly increasing on $\mathbb{R}_+$ and H3 holds.

For every $v \in \mathbb{R}^q$, if $\mathcal{F}(\cdot, v)$ reaches a (local) minimum at $\hat{u}$, then

$$\|A\hat{u}\|_2 \leqslant \|v\|_2.$$

(ii) Assume that $\mathrm{rank}\,(A) = p \geqslant 2$, $\ker(\mathrm{D}) = \mathrm{span}(\mathbb{1}_p)$ *and ϕ is strictly increasing on $\mathbb{R}_+$. There is a closed set $N \subset \mathbb{R}^q$ with $\mathbb{L}^q(N) = 0$ such that $\forall v \in \mathbb{R}^q \setminus N$, if $\mathcal{F}(\cdot, v)$ reaches a (local) minimum at $\hat{u}$, then*

$$\|A\hat{u}\|_2 < \|v\|_2.$$

A full description of the set N can be found in [80].

Comments on the results. If A is orthonormal (e.g. $A = \mathrm{Id}$), the obtained results yield

$$\begin{aligned} (i) \quad &\Rightarrow \qquad \|\hat{u}\|_2 \leqslant \|v\|_2; \\ (ii) \quad &\Rightarrow \qquad \|\hat{u}\|_2 < \|v\|_2. \end{aligned}$$

These provide sharper bounds than the one available in [5].

When the least eigenvalue $\lambda^2_{\min}$ of A^*A is positive, it is obvious that

$$\begin{aligned} (i) \quad &\Rightarrow \qquad \|\hat{u}\|_2 \leqslant \left|\lambda^{-1}_{\min}\right| \|v\|_2; \\ (ii) \quad &\Rightarrow \qquad \|\hat{u}\|_2 < \left|\lambda^{-1}_{\min}\right| \|v\|_2. \end{aligned}$$

In the case of noise-free data and $\mathrm{rank}\,(A) = p$, one naturally wishes to recover the original (unknown) u_o. It is hence necessary that $\|A\hat{u}\|_2 = \|v\|_2$. Comparing the results obtained in (i) and (ii) show that such a goal is unreachable if ϕ is strictly increasing on $\mathbb{R}_+$.

It follows that exact recovery needs that ϕ is constant for $t \geqslant \tau$, for a constant $\tau \geqslant 0$.

The mean of restored data. In many applications, the noise corrupting the data can be supposed to have a mean equal to zero. When $A = \mathrm{Id}$, it is well known that $\mathrm{mean}(\hat{u}) = \mathrm{mean}(v)$, see e.g. [5]. It is shown in [80, Proposition 2] that for a general A

$$A\mathbb{1}_p \propto \mathbb{1}_q \tag{14}$$

$$\Rightarrow \qquad \mathrm{mean}(\hat{u}) = \mathrm{mean}(v). \tag{15}$$

However, (14) is quite a restrictive requirement. In the simple case when $\phi(t) = t^2$, $\ker(\mathrm{D}) = \mathbb{1}_{rs}$ and A is square and invertible, it is easy to see that the restrictive requirement (14) is also sufficient [80, Remark 2]. It turns out that if A is no longer equal to Id, the natural requirement (15) is generally false. A way to remedy for this situation is to minimize $\mathcal{F}(\cdot, v)$ under the explicit constraint derived from (15).

The residuals for edge-preserving regularization. The goal here is to provide bounds that characterize the data-fidelity term at a (local) minimizer $\hat{u}$ of $\mathcal{F}(\cdot, v)$. More precisely, the focus is on edge-preserving PFs satisfying

H8 $\quad \|\phi'\|_\infty \stackrel{\text{def}}{=} \max\left\{\sup_{t \geqslant 0} |\phi'(t^+)|, \sup_{t>0} |\phi'(t^-)|\right\} < \infty.$

Comparing with H2 shows that ϕ under H8 is edge-preserving. Observe that except for (f1) and (f13), all PFs given in Table 1 satisfy H8. Even though (f13) is edge-preserving, ϕ' is not well defined at zero. Note that when $\phi'(0^+) > 0$ is finite we usually have $\|\phi'\|_\infty = \phi'(0^+)$.

The statement given below is established in [80, Theorem 3.1.].

Theorem 7 *Consider $\mathcal{F}$ of the form* (12) *where* $\operatorname{rank}(A) = q \leqslant p$ *and let H1, H6, H7 and H8 hold. Suppose that $\|\phi'\|_\infty = 1$. Then for every $v \in \mathbb{R}^q$, if $\mathcal{F}(\cdot, v)$ has a (local) minimum at $\hat{u}$, we have* [3]

$$\|A\hat{u} - v\|_\infty \leqslant \frac{\beta}{2}\|\phi'\|_\infty \|(AA^*)^{-1}A\|_\infty \|\mathrm{D}\|_1. \tag{16}$$

In particular, if $A = \mathrm{Id}$ and D corresponds to the discrete gradient operator on a two-dimensional image, $\|\mathrm{D}\|_1 = 4$ and we find

$$\|v - \hat{u}\|_\infty \leqslant 2\beta\|\phi'\|_\infty.$$

The result of this theorem may seem surprising. In a statistical setting, the quadratic data-fidelity term $\|Au - v\|_2^2$ in (12) corresponds to white Gaussian noise on the data, which noise is unbounded. However, if ϕ is edge-preserving with $\|\phi'\|_\infty$ bounded, the (local) minimizers $\hat{u}$ of $\mathcal{F}(\cdot, v)$ give rise to noise estimates $(v - A\hat{u})[i]$, $1 \leqslant i \leqslant q$ that are tightly bounded as stated in (16).

Hence the assumption for Gaussian noise on the data v is distorted by the solution $\hat{u}$.

The proof of the theorem reveals that this behavior is due to the boundedness of the gradient of the regularization term. Let us emphasize that the bound in (16) is independent of data v and that it is satisfied for any local or global minimizer $\hat{u}$ of $\mathcal{F}(\cdot, v)$.

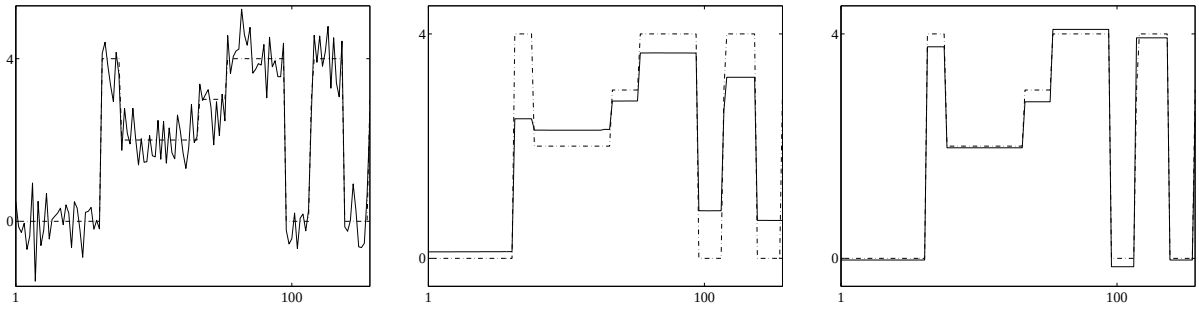
4 Non-convex regularization

4.1 Motivation

A permanent requirement is that the energy $\mathcal{F}$ favors the recovery of neat edges. Since the pioneering work of Geman & Geman [50], various nonconvex Φ in (4) have been proposed [12,48,49,62,66,70,82]. Indeed, the relevant minimizers exhibit neat edges between homogeneous regions. However, they are tiresome to control and to reach (only a few algorithms are proved to find the global minimizer of particular energies within an acceptable time). In order to avoid the numerical intricacies arising with nonconvex regularization, since [52, 61, 90] in the 1990s, an important effort was done to derive

[3] Let us remind that for any $m \times n$ real matrix C with components $C[i,j]$, $1 \leqslant m$, $1 \leqslant j \leqslant n$, we have $\|C\|_1 = \max_j \sum_{i=1}^{m} |C[i,j]|$ and $\|C\|_\infty = \max_i \sum_{j=1}^{n} |C[i,j]|$, see e.g. [51].

convex edge-preserving PFs, see e.g. [17,27,62,86] and [5] for an excellent account. The most popular convex edge-preserving PF was derived by Rudin, Osher and Fatemy [86]: it amounts to $\phi = t$, for $\{\mathrm{D}_i\}$ yielding the discrete gradient operator (see (2) and (8)) and the relevant Φ is called the Total Variation (TV) regularization.



Data $v = u + n$ (—)
Original u_o (dashed line)

Convex regularization $\phi(t) = t$ (TV)
Original u_o (- - -), minimizer $\hat{u}$ (—)

Non-convex regularization
$\phi(t) = \alpha t/(1 + \alpha t)$

Figure 3: Minimizers of $\mathcal{F}(u, v) = \|u - v\|_2^2 + \beta \sum_{i=1}^{p-1} \phi(|u[i] - u[i + 1]|)$.

Fig. 3 nicely shows that the hight of the edges is much more faithful when ϕ is nonconvex, compared to the convex TV regularization. The same effect can also be observed e.g. in Figs. 8, 9 and 11.

This section is devoted to explain why edges are nicely recovered using a nonconvex ϕ.

4.2 Assumptions on potential functions ϕ

Consider $\mathcal{F}(\cdot, v)$ of the form (12) where $\mathrm{D}_i : \mathbb{R}^p \to \mathbb{R}^1$, $i \in I = \{1, \cdots, r\}$, i.e.

$$\mathcal{F}(u, v) = \|Au - v\|_2^2 + \beta \sum_{i \in I} \phi(|\mathrm{D}_i u|), \tag{17}$$

where $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ satisfies H1 (p. 3), H6 (p. 16) and H9 given below

H9 *ϕ is $\mathcal{C}^2$ on $\mathbb{R}_+^*$, $\phi'(0^+) \geqslant 0$ $\forall t \geqslant 0$, $\inf_{t \in \mathbb{R}_+^*} \phi''(t) < 0$ and $\lim_{t \to \infty} \phi''(t) = 0$;*

as well as one of the following assumptions:

H10 *If $\phi'(0^+) = 0$, then $\exists \tau > 0$ and $\exists \mathcal{T} \in (\tau, \infty)$ such that $\phi''(t^+) \geqslant 0$, $\forall t \in [0, \tau]$, while $\phi''(t) \leqslant 0$, $\forall t > \tau$, ϕ'' strictly decreases on $(\tau, \mathcal{T}]$ and increases on $[\mathcal{T}, \infty)$;*

H11 *If $\phi'(0^+) > 0$ then $\lim_{t \searrow 0} \phi''(t) < 0$ is well defined and $\phi''(t) \leqslant 0$ is increasing on $(0, \infty)$.*

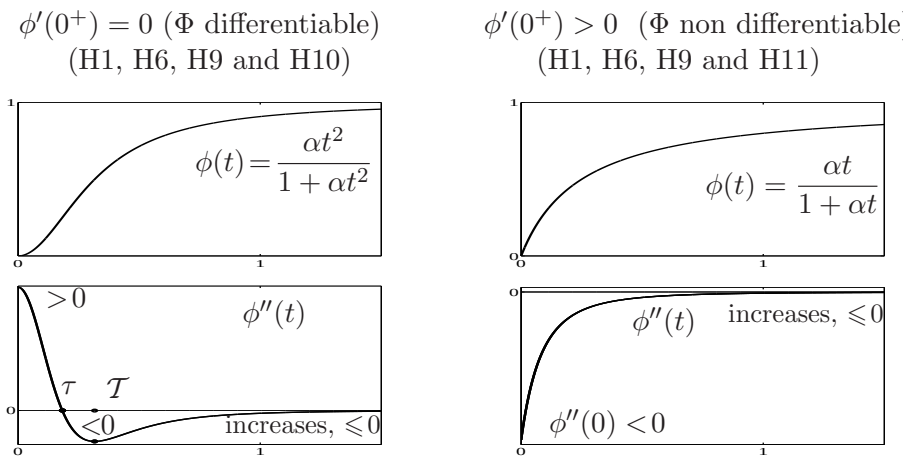


Figure 4: Illustration of the assumptions in two typical cases.

These assumptions are illustrated in Fig. 4. Even though they might seem tricky, they hold true for all nonconvex PFs in Table 1, except for (f6) and (f13). The "irregular cases" (f6) and f(13) are considered separately.

The results presented below are published in [79].

4.3 How it works on $\mathbb{R}$

This example shows the main phenomena underlying the theory on edge-enhancement using nonconvex ϕ satisfying H1, H6 and H9 along with either H10 or H11.

Let $\mathcal{F} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ read (see [Sec. 2, p. 963] [79]).

$$\mathcal{F}(u,v) = \frac{1}{2}(u-v)^2 + \beta\phi(u) \text{ for } \begin{cases} \beta > -1/\phi''(\mathcal{T}) & \text{if } \phi'(0^+) = 0 \text{ (H1, H6, H9 and H10)} \\ \beta > -1/\lim_{t\searrow 0}\phi''(t) & \text{if } \phi'(0^+) > 0 \text{ (H1, H6, H9 and H11)} \end{cases}$$

The (local) minimality conditions for $\hat{u}$ of $\mathcal{F}(\cdot, v)$ read

- If $\phi'(0^+) = 0$ or $\left[\phi'(0^+) > 0 \text{ and } \hat{u} \neq 0\right]$: $\hat{u} + \beta\phi'(\hat{u}) = v$ and $1 + \beta\phi''(\hat{u}) \geqslant 0$;
- If $\phi'(0^+) > 0$ and $\hat{u} = 0$: $|v| \leqslant \beta\phi'(0^+)$.

To simplify, we assume that $v \geqslant 0$. Define

$$\theta_0 = \inf C_\beta \text{ and } \theta_1 = \sup C_\beta,$$
$$\text{for} \quad C_\beta = \{u \in \mathbb{R}_+^* : D_1^2\mathcal{F}(u,v) < 0\} = \left\{u \in \mathbb{R}_+^* : \phi''(u) < -1/\beta\right\}.$$

We have $\theta_0 = 0$ if $\phi'(0^+) > 0$ and $0 < \theta_0 < \mathcal{T} < \theta_1$ if $\phi'(0^+) = 0$. After some calculations one finds:

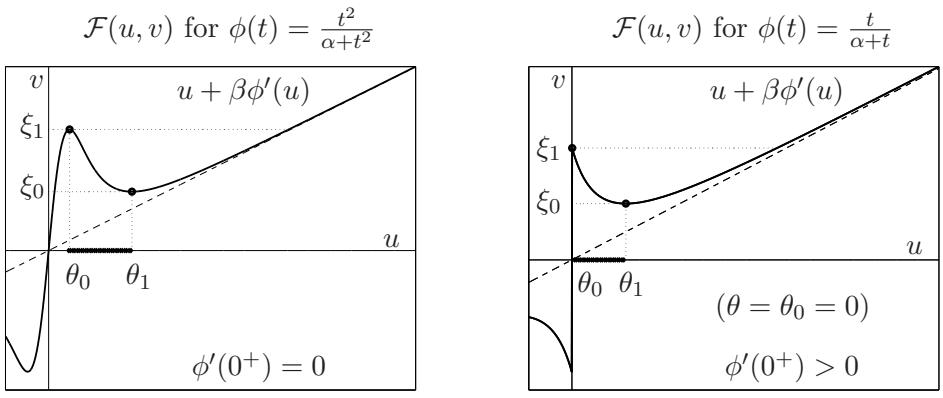


Figure 5: The curve of $u \mapsto \big(D_1\mathcal{F}(u,v) - v\big)$ on $\mathbb{R} \setminus \{0\}$. Note that all assumptions mentioned before do hold.

1. For every $v \in \mathbb{R}_+$ no minimizer lives in (θ_0, θ_1) (*cf.* Fig. 5);

2. One computes $0 < \xi_0 < \xi_1$ such that (*cf.* Fig. 5)

   (a) if $0 \leqslant v \leqslant \xi_1$, $\mathcal{F}(\cdot, v)$ has a (local) minimizer $\hat{u}_0 \in [0, \theta_0]$, hence $\hat{u}_0$ is subject to a strong smoothing;

   (b) if $v \geqslant \xi_0$, $\mathcal{F}(\cdot, v)$ has a (local) minimizer $\hat{u}_1 \geqslant \theta_1$, hence $\hat{u}_1$ is subject to a weak smoothing;

   (c) if $v \in [\xi_0, \xi_1]$ then $\mathcal{F}(\cdot, v)$ has two local minimizers, $\hat{u}_0$ and $\hat{u}_1$;

3. There is $\xi \in (\xi_0, \xi_1)$ such that $\mathcal{F}(\cdot, \xi)$ has two global minimizers, $\mathcal{F}(\hat{u}_0, \xi) = \mathcal{F}(\hat{u}_1, \xi)$, as seen in Fig. 6;

   (a) if $0 < v < \xi$ the unique global minimizer is $\hat{u} = \hat{u}_0$;

   (b) if $v > \xi$ the unique global minimizer is $\hat{u} = \hat{u}_1$;

4. The global minimizer function $v \mapsto \mathcal{U}(v)$ is discontinuous at ξ and $\mathcal{C}^1$-smooth on $\mathbb{R}_+ \setminus \{\xi\}$.

Item 1 is the key for the recovery either of homogeneous regions or of high edges. The minimizer $\hat{u}_0$ (see Items 2a, 3a) corresponds to the restoration of homogeneous regions, while $\hat{u}_1$ (see Items 2b, 3b) corresponds to edges. Item 3 corresponds to a decision for the presence of an edge at the global minimizer. Since $\{\xi\}$ is closed and $\mathbb{L}^1\{\xi\} = 0$, Item 4 confirms the results of § 3.2.2. The detailed calculations are outlined in [79, sec. 2.].

The theory presented next is a generalization of these facts.

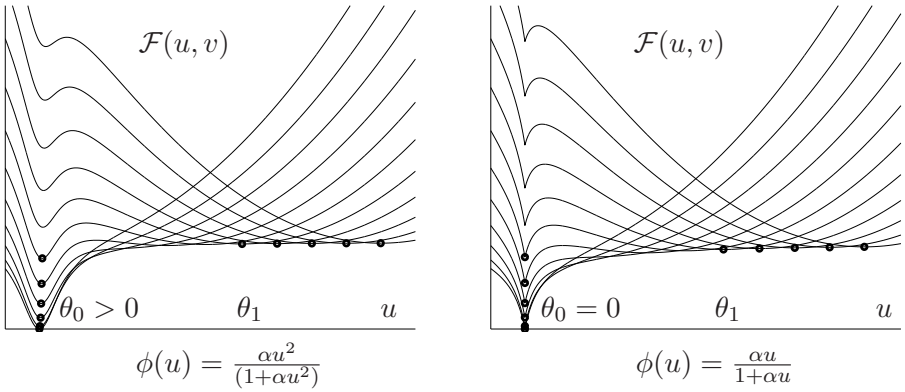


Figure 6: Each curve represents $\mathcal{F}(u,v) = \frac{1}{2}(u-v)^2 + \beta\phi(u)$ for an increasing sequence $v \in [0, \xi_1)$. The global minimizer of each $\mathcal{F}(\cdot, v)$ is emphasized with "•". Observe that no (local) minimizer lives in (θ_0, θ_1).

4.4 Either smoothing or edge enhancement

We adopt the hypotheses formulated in § 4.2. Given v, let $\hat{u}$ be a local (or global) minimizer of $\mathcal{F}(\cdot, v)$. The results presented here are extracted essentially from [79, section 3].

(A) Case $\phi'(0^+) = 0$. The theorem below as well as Proposition 1 are established in [79, § 3.1].

Theorem 8 *Let H1, H6, H9 and H10 hold. Let $\{\mathrm{D}_i : i \in I\}$ be linearly independent and*

$$\mu \stackrel{\text{def}}{=} \max_{1 \leqslant i \leqslant r} \|\mathrm{D}^*(\mathrm{D}\mathrm{D}^*)^{-1} e_i\|_2.$$

If $\beta > \dfrac{2\mu^2 \, \|A^*A\|_2}{|\phi''(\mathcal{T})|}$, there are $\theta_0 \in (\tau, \mathcal{T})$ and $\theta_1 \in (\mathcal{T}, \infty)$ such that $\forall v \in \mathbb{R}^q$, if $\hat{u}$ is a (local) minimizer of $\mathcal{F}(\cdot, v)$, then

$$\textit{either} \quad |\mathrm{D}_i \hat{u}| \leqslant \theta_0, \quad \textit{or} \quad |\mathrm{D}_i \hat{u}| \geqslant \theta_1, \quad \forall i \in I. \tag{18}$$

In imaging problems, $\{\mathrm{D}_i\}$ are generally not linearly independents. Note that if $\{\mathrm{D}_i\}$ are linearly dependent, the result (18) holds true for all (local) minimizers $\hat{u}$ that are locally homogeneous on connected regions[4]. However, if this is not the case, one recovers both high edges and smooth transitions, as seen in Fig. 9(a). When ϕ is convex, all edges are smoothed, as one can observe in Fig. 8(a).

The PF $\phi(t) = \min\{\alpha t^2, 1\}$ (the discrete version of Mumford-Shah functional), (f6) in Table 1, does not satisfy assumptions H9 and H10. In particular,

$$2\sqrt{\alpha} = \phi'\left(\frac{1}{\sqrt{\alpha}}^-\right) > \phi'\left(\frac{1}{\sqrt{\alpha}}^+\right) = 0.$$

[4]More precisely, connected with respect to $\{\mathrm{D}_i\}$.

A straightforward consequence of Corollary 1 (p. 9) is that for any (local) minimizer $\hat{u}$ of $\mathcal{F}(\cdot, v)$ we have

$$|\mathrm{D}_i \hat{u}| \neq \frac{1}{\sqrt{\alpha}}, \quad \forall i \in I.$$

Propositions 1 and 2 below address only global minimizers under specific conditions on $\{\mathrm{D}_i\}$.

Proposition 1 *Let $\phi(t) = \min\{\alpha t^2, 1\}$, the set $\{\mathrm{D}_i : i \in I\}$ be linearly independent and $\mathrm{rank}\,(A) \geqslant p - r \geqslant 1$. Assume that $\mathcal{F}(\cdot, v)$ has a global minimizer at $\hat{u}$. Then*

$$\textit{either} \quad |\mathrm{D}_i \hat{u}| \leqslant \frac{1}{\sqrt{\alpha}}\,\Gamma_i, \quad \textit{or} \quad |\mathrm{D}_i \hat{u}| \geqslant \frac{1}{\sqrt{\alpha}\,\Gamma_i}, \quad \textit{for} \quad \Gamma_i = \sqrt{\frac{\|Be_i\|_2^2}{\|Be_i\|_2^2 + \alpha\beta}} < 1, \quad \forall i \in I, \tag{19}$$

where B is a matrix[5] depending only on A and D. Moreover, the inequalities in (19) are strict if the global minimizer $\hat{u}$ of $\mathcal{F}(\cdot, v)$ is unique.

In the case when u is an one-dimensional signal, the following result is exhibited in [74].

Proposition 2 *Let $\phi(t) = \min\{\alpha t^2, 1\}$, $\mathrm{D}_i u = u[i] - u[i+1]$, $1 \leqslant i \leqslant p-1$ and $A\mathbb{1}_p \neq 0$ with $\mathrm{rank}\,(A) \geqslant 1$. Then for any global minimizer $\hat{u}$ of $\mathcal{F}(\cdot, v)$ we have*

$$\textit{either} \quad |\mathrm{D}_i \hat{u}| \leqslant \frac{1}{\sqrt{\alpha}}\,\Gamma_i, \quad \textit{or} \quad |\mathrm{D}_i \hat{u}| \geqslant \frac{1}{\sqrt{\alpha}\,\Gamma_i}, \quad \forall i \in I, \tag{20}$$

$$\textit{where} \qquad \Gamma_i = \sqrt{\frac{\|B \sum_{j=i+1}^{p} e_j\|_2^2}{\|B \sum_{j=i+1}^{p} e_j\|_2^2 + \alpha\beta}} < 1,$$

and B is a matrix[6] depending only on A. Moreover, the inequalities in (20) are strict if the global minimizer $\hat{u}$ of $\mathcal{F}(\cdot, v)$ is unique.

In both Propositions 1 and 2, set $\theta_0 = \dfrac{\gamma}{\sqrt{\alpha}}$ and $\theta_1 = \dfrac{1}{\sqrt{\alpha}\gamma}$ for $\gamma \stackrel{\text{def}}{=} \max_{i \in I} \Gamma_i < 1$.

Let us define the following subsets:

$$\hat{J}_0 \stackrel{\text{def}}{=} \{i \in I \ : \ |\mathrm{D}_i \hat{u}| \leqslant \theta_0\} \ \text{ and } \ \hat{J}_1 \stackrel{\text{def}}{=} I \setminus \hat{J}_0 = \{i \in I \ : \ |\mathrm{D}_i \hat{u}| \geqslant \theta_1\}. \tag{21}$$

Using these notations, one can unify the interpretation of Theorem 8 and Propositions 1 and 2.

[5] From the assumptions, $r \leqslant p$ in all cases. If $r = p$, we have $B = \mathrm{Id}$. If $r < p$, we choose matrices $H \in \mathbb{R}^{r \times p}$, $H_a \in \mathbb{R}^{p \times p - r}$ and $\mathrm{D}_a \in \mathbb{R}^{p-r \times p}$ such that for any $u \in \mathbb{R}^p$ we have $u = H\mathrm{D}u + H_a \mathrm{D}_a u$ and $\mathrm{rank}\,(AH_a) = p - r$. Denote $M_a = AH_a \in \mathbb{R}^{q \times p - r}$. Then $B = \mathrm{Id} - M_a (M_a^* M_a)^{-1} M_a^*$.

[6] In this case, B reads

$$A^* \left(\mathrm{Id} - \frac{A\mathbb{1}_p \mathbb{1}_p^* A^*}{\|A\mathbb{1}_p\|_2^2} \right) A.$$

Since $\theta_0 < \theta_1$, a natural interpretation of the results of Theorem 8, and Propositions 1 and 2, is that $\big[|\mathrm{D}_i\hat{u}| \ : \ i \in \hat{J}_0\big]$ are homogeneous regions with respect to $\{\mathrm{D}_i\}$ while $\{|\mathrm{D}_i\hat{u}| \ : \ i \in \hat{J}_1\}$ are break points in $\mathrm{D}_i\hat{u}$.

In particular, if $\{\mathrm{D}_i\}$ correspond to first-order differences, $\hat{J}_0$ addresses smoothly varying regions while $\hat{J}_1$ corresponds to edges higher than $\theta_1 - \theta_0$.

(B) Case $\phi'(0^+) > 0$. Here the results are stronger without any assumption on $\{\mathrm{D}_i\}$. The next Theorem 9 and Proposition 3 are proven in [79, §3.2.].

Theorem 9 *Let H1, H6, H9 and H11 hold. Let* $\beta > \dfrac{2\mu^2 \, \|A^*A\|_2}{|\lim_{t\searrow 0} \phi''(t)|}$, *where* $\mu > 0$ *is a constant depending only on* $\{\mathrm{D}_i\}$. *Then* $\exists \theta_1 > 0$ *such that* $\forall v \in \mathbb{R}^q$, *every (local) minimizer* $\hat{u}$ *of* $\mathcal{F}(\cdot, v)$ *satisfies*

$$\textit{either} \quad |\mathrm{D}_i\hat{u}| = 0, \quad \textit{or} \quad |\mathrm{D}_i\hat{u}| \geqslant \theta_1, \quad \forall i \in I. \tag{22}$$

The "0-1" PF (f13) in Table 1, $\phi(0) = 0$, $\phi(t) = 1$ if $t > 0$ does not satisfy assumptions H6, H9 and H11 since it is discontinuous at 0.

Proposition 3 *Let* ϕ *be the "0-1" PF, i.e. (f13) in Table 1, the set* $\{\mathrm{D}_i : i \in I\}$ *be linearly independent and* $\operatorname{rank} A \geqslant p - r \geqslant 1$. *If* $\mathcal{F}(\cdot, v)$ *has a global minimum at* $\hat{u}$, *then*

$$\textit{either} \quad |\mathrm{D}_i\hat{u}| = 0 \quad \textit{or} \quad |\mathrm{D}_i\hat{u}| \geqslant \frac{\sqrt{\beta}}{\|Be_i\|_2}, \quad \forall i \in I, \tag{23}$$

where B *is the same as in Proposition 1. The inequality in* (23) *is strict if* $\mathcal{F}(\cdot, v)$ *has a unique global minimizer.*

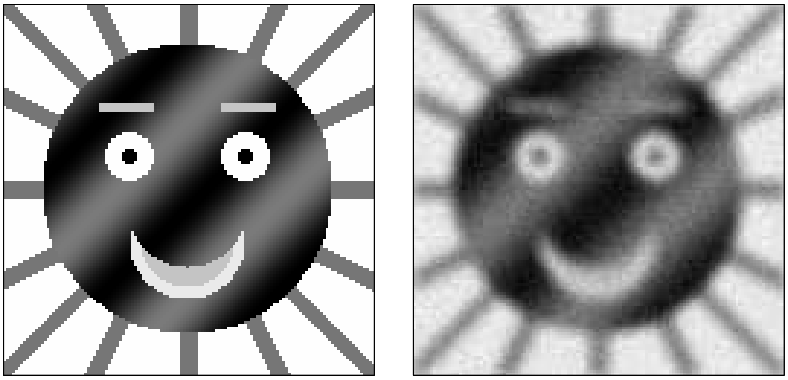
Note that (23) holds true if we set $\theta_1 = \min\limits_{i \in I} \dfrac{\sqrt{\beta}}{\|Be_i\|}$. Let

$$\hat{J}_0 \stackrel{\text{def}}{=} \{i \ : \ |\mathrm{D}_i\hat{u}| = 0\} \text{ and } \hat{J}_1 \stackrel{\text{def}}{=} I \setminus \hat{J}_0 = \{i \ : \ |\mathrm{D}_i\hat{u}| \geqslant \theta_1\} \ .$$

With the help of these notations, the results established in Theorem 9 and Proposition 3 allow the relevant solutions $\hat{u}$ to be characterized as stated below.

By Theorem 9 and Proposition 3, the set $\hat{J}_0$ addresses regions in $\hat{u}$ that can be called strongly homogeneous as far as $\big[\ |\mathrm{D}_i\hat{u}| = 0 \ \Leftrightarrow \ i \in \hat{J}_0\big]$ while $\hat{J}_1$ addresses break-points in $|\mathrm{D}_i\hat{u}|$ larger than θ_1 since $\big[\ |\mathrm{D}_i\hat{u}| > \theta_1 \ \Leftrightarrow \ i \in \hat{J}_1\big]$.

If D corresponds to first order differences or discrete gradients, $\hat{u}$ is neatly segmented with respect to $\{\mathrm{D}_i\}$: $\hat{J}_0$ corresponds to constant regions while $\hat{J}_1$ describes all edges and the latter are higher than θ_1.



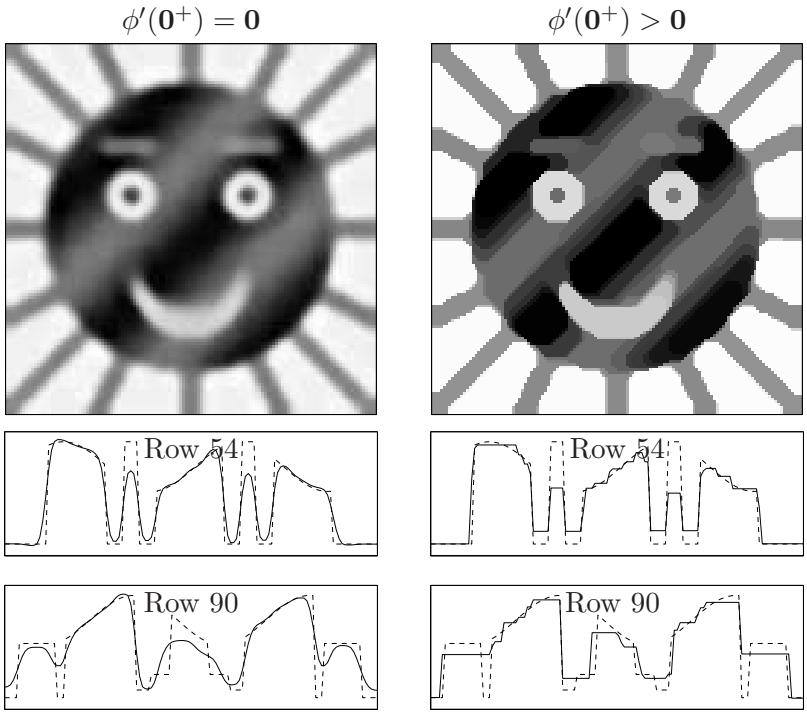
(a) Original image (b) Data v = blur + noise

Figure 7: Data $v = a \star u_o + n$, where a is a blur and n is white Gaussian noise, 20 dB of SNR.

Let us remind that direct segmentation of an image from data transformed via a general (non diagonal) operator A remains a tortuous task using standard methods. The result in (22), Theorem 9, tells us that such a segmentation is naturally involved in the minimizer $\hat{u}$ of $\mathcal{F}(\cdot, v)$ in the context of this theorem. Let us emphasize that this segmentation effect holds for any operator A. This can be observed, e.g., on Figs. 9b,d and 12d.

(C) Illustration: Deblurring of an image from noisy data. The original image u_o in Fig. 7(a) presents smoothly varying regions, constant regions and sharp edges. Data in Fig. 7(b) correspond to $v = a * u_o + n$, where a is a blur with entries $a_{i,j} = \exp\left(-(i^2 + j^2)/12.5\right)$ for $-4 \leqslant i, j \leqslant 4$, 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 $[-5, 50]$. In all restored images, $\{\mathrm{D}_i\}$ correspond to the first-order differences of each pixel with its 8 nearest neighbors. In all figures, the obtained minimizers are displayed on the top. Just below, the sections corresponding to rows 54 and 90 of the restored images are compared with the same rows of the original image. Note that these rows cross the delicate locations of the eyes and the mouth in the image.

The restorations in Fig. 8 are obtained using convex PFs ϕ while those in Fig. 9 using nonconvex PFs ϕ. According to the theory presented in paragraphs (A) and (B) here above, edges are sharp and high in Fig. 9 where ϕ is nonconvex while they are underestimated in Fig. 8 where ϕ is convex. In Fig. 9(b) and (d) ϕ is nonconvex and $\phi'(0^+) > 0$ in addition. As stated in Theorem 9, in spite of the fact that A is nondiagonal (and ill-conditionned), the restored images are fully segmented and the edges between constant pieces are high. Even though Proposition 3 assumes that $\{\mathrm{D}_i\}$ are linearly independent, the segmentation effect using linearly dependent $\{\mathrm{D}_i\}$ as described above is often neat.



(a) $\phi(t) = t^\alpha$ for $\alpha = 1.4, \beta = 40$ (b) $\phi(t) = t$ for $\beta = 100$

Figure 8: Restoration using convex PFs.

4.5 Selection for the global minimum

Let the original image u_o be of the form

$$u_o = \eta \chi_\Sigma, \quad \eta > 0, \tag{24}$$

where the sets $\Sigma \subset \{1, \ldots, p\}$ and Σ^c are nonempty and $\chi_\Sigma \in \mathbb{R}^p$ is the characteristic function of Σ, i.e.

$$\chi_\Sigma[i] = \begin{cases} 1 & \text{if } \ i \in \Sigma, \\ 0 & \text{if } \ i \in \Sigma^c. \end{cases}$$

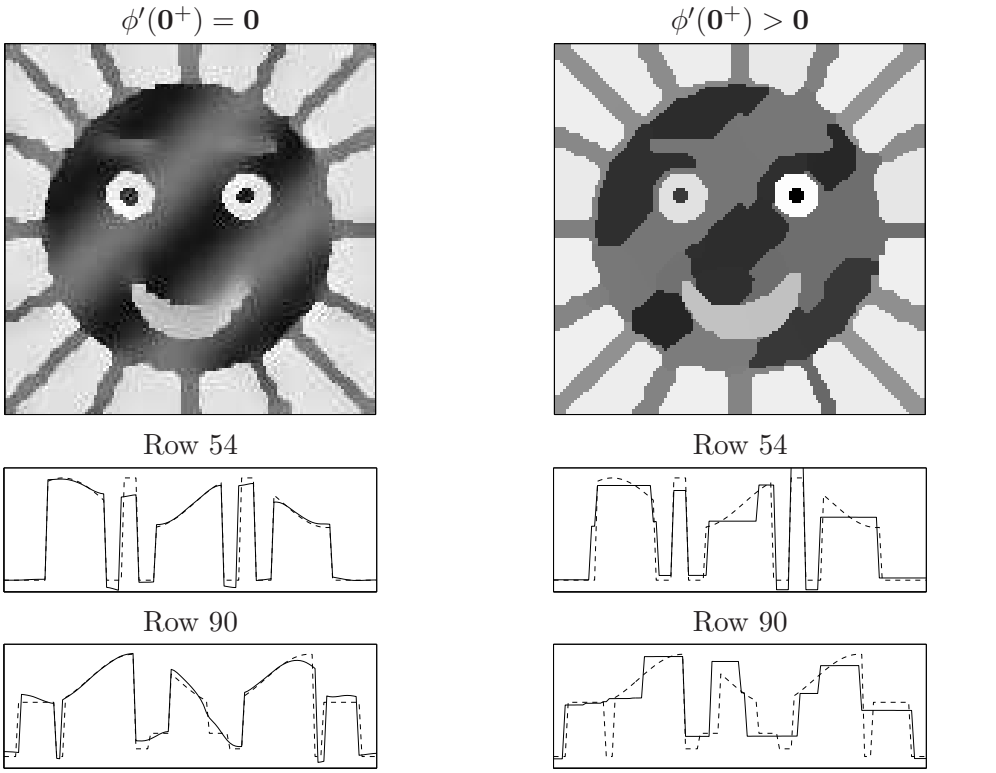
Let data read

$$v = Au_o = A\ \eta \chi_\Sigma.$$

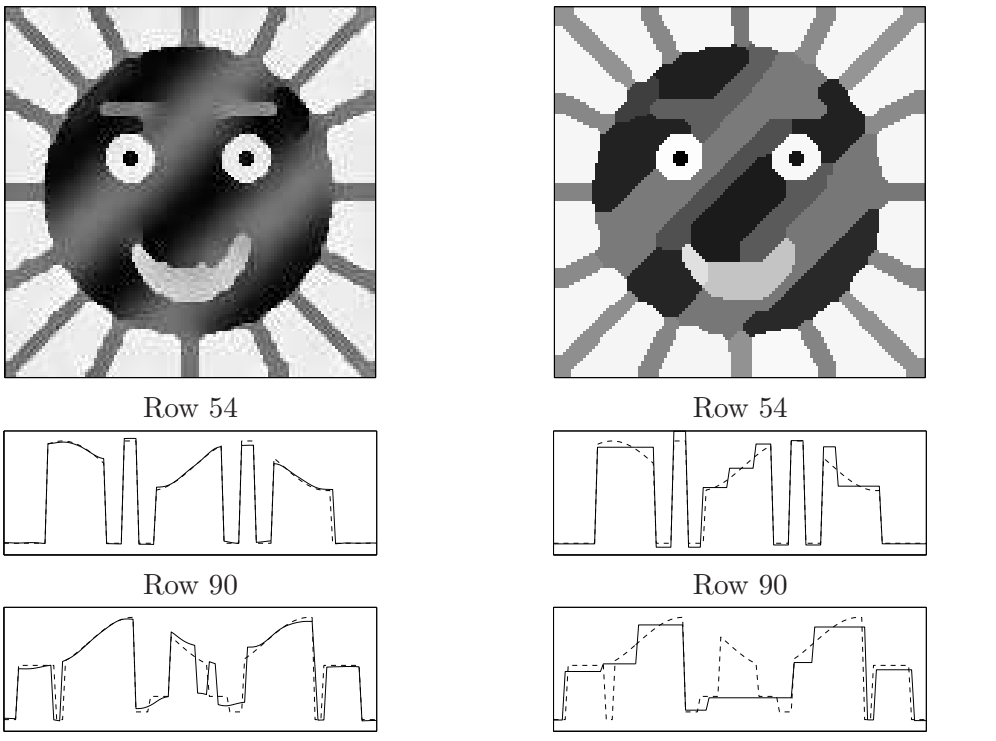
Consider that $\mathcal{F}(\cdot, A\ \eta\chi_\Sigma)$ is of the form (17) and focus on its global minimizer $\hat{u}_\eta$. The question discussed here is:

How to characterize the global minimizer $\hat{u}_\eta$ of $\mathcal{F}(\cdot, A\ \eta\chi_\Sigma)$ according to the value of $\eta > 0$?

The results sketched below were established in [79, section 4]. In order to answer the question formulated above, two additional assumption may be taken into account.



(a) $\phi(t) = \dfrac{\alpha t^2}{1+\alpha t^2}$ for $\alpha = 25$, $\beta = 35$ (b) $\phi(t) = \dfrac{\alpha t}{1+\alpha t}$ for $\alpha = 20$, $\beta = 100$



(c) $\phi(t) = \min\{\alpha t^2, 1\}$ for $\alpha = 60$, $\beta = 10$ (d) $\phi(0) = 0,\ \phi(t) = 1,\ t > 0$ for $\beta = 25$

Figure 9: Restoration using non-convex PFs.

H12 *For any $i \in I$, D_i yields (possibly weighted) pairwise differences and* $\ker(\mathrm{D}) = \mathrm{span}(\mathbb{1}_p)$*;*

H13 *For any $t \in \mathbb{R}_+$, there is a constant*[7] $0 < c < +\infty$ *such that $\underline{\phi(t) \leqslant c}$; for simplicity, fix $\underline{c = 1}$.*

Moreover, it is assumed that A satisfies H4 (p. 13). Remind that by H4 and H1 (p. 3), $\mathcal{F}(\cdot, v)$ admits a global minimum and that the latter is reached.

From now on, we denote

$$J_1 = \Big\{ i \in \{1, \ldots, r\} : |\mathrm{D}_i u_o| = |\mathrm{D}_i \eta \chi_\Sigma| > 0 \Big\} \quad \text{and} \quad J_1^c = I \setminus J_1. \tag{25}$$

Note that $\underline{J_1 \text{ addresses the edges in } u_o = \eta\chi_\Sigma}$.

Proposition 4 (**$\phi'(0^+) = 0$, $\mathcal{F}(\cdot, v)$ is $\mathcal{C}^2(\mathbb{R}_+)$**) *. Assume H1 (p. 3), H4, H6 (p. 16), H9, H10, H12 and H13. Let every (local) minimizer of $\mathcal{F}(\cdot, A\eta\chi_\Sigma)$ satisfies the property stated in* (18) *(p. 22). Then there are two constants $\eta_0 > 0$ and $\eta_1 > \eta_0$ such that*

$$\eta \in [0, \eta_0) \quad \Rightarrow \quad |\mathrm{D}_i \hat{u}_\eta| \leqslant \theta_0, \quad \forall i \in I \qquad \underline{\textit{($\hat{u}_\eta$ is fully smooth)}} \tag{26}$$

whereas

$$\eta \geqslant \eta_1 \quad \Rightarrow \quad \begin{array}{ll} |\mathrm{D}_i \hat{u}_\eta| \leqslant \theta_0, & \forall i \in J_1^c, \\ |\mathrm{D}_i \hat{u}_\eta| \geqslant \theta_1, & \forall i \in J_1, \end{array} \qquad \underline{\textit{(the edges in $\hat{u}_\eta$ are correct)}}.$$

This result corroborates the interpretation of θ_0 and θ_1 as thresholds for the detection of smooth differences and edges, respectively—see (21) and the comments following this equation.

Proposition 5 (Truncated quadratic PF) *Let $\phi(t) = \min\{\alpha t^2, 1\}$—see (f6) in Table 1. Assume that H4 and H13 are satisfied. Define $\omega_\Sigma \in \mathbb{R}^p$ by*

$$\omega_\Sigma = \left(A^*A + \beta\alpha \mathrm{D}^*\mathrm{D}\right)^{-1} A^* A \chi_\Sigma. \tag{27}$$

Then there are $\eta_0 > 0$ and $\eta_1 > \eta_0$ such that

$$\eta \in [0, \eta_0) \quad \Rightarrow \quad \hat{u}_\eta = \eta\, \omega_\Sigma, \qquad \underline{\textit{($\hat{u}_\eta$ is fully smooth)}} \tag{28}$$

$$\eta \geqslant \eta_1 \quad \Rightarrow \quad \hat{u}_\eta = \eta\, \chi_\Sigma, \qquad \underline{\textit{(exact recovery, $\hat{u}_\eta = u_o$)}} \tag{29}$$

Moreover, $\hat{u}_\eta$ in (28) and (29) is the unique global minimizer of the relevant $\mathcal{F}(\cdot, \eta A\chi_\Sigma)$.

Observe that $\eta\omega_\Sigma$ in (28) is the regularized least-squares solution, hence it does not involve edges. For $\eta \geqslant \eta_1$ the global minimizer $\hat{u}_\eta$ is equal to the original u_o.

[7] Note that H1 and H13 entail $\lim_{t \to \infty} \phi'(t) = 0$. Then the edge-preservation necessary condition H2 (p. 11) is trivially satisfied.

Proposition 6 ($0 < \phi'(0^+) < +\infty$) *Assume H1, H4, H6 (p. 16), H9, H11, H12 and H13. Let every (local) minimizer of $\mathcal{F}(\cdot, A\eta\chi_\Sigma)$ satisfies the property stated in* (22) *(p. 24). Then there exist $\eta_0 > 0$ and $\eta_1 > \eta_0$ such that*

$$\eta \in [0, \eta_0) \quad \Rightarrow \quad \hat{u}_\eta = \eta\zeta \, \mathbb{1}_p, \quad \textit{where} \quad \zeta = \frac{(A\mathbb{1}_p)^* A\chi_\Sigma}{\|A\mathbb{1}_p\|_2^2} \qquad \underline{\textit{($\hat{u}_\eta$ is constant)}} \tag{30}$$

whereas

$$\eta > \eta_1 \quad \Rightarrow \quad \begin{array}{ll} \mathrm{D}_i\hat{u}_\eta = 0, & \forall i \in J_1^c, \\ |\mathrm{D}_i\hat{u}_\eta| \geqslant \theta_1, & \forall i \in J_1. \end{array} \qquad \underline{\textit{($\hat{u}_\eta$ is piecewise constant with correct edges)}}$$

If η is small, the global solution $\hat{u}_\eta$ is constant, while for η large enough, $\hat{u}_\eta$ has the same edges and the same constant regions as the original $u_o = \eta\chi_\Sigma$. Moreover, if Σ and Σ^c are connected with respect to $\{\mathrm{D}_i : i \in I\}$, there are $\hat{s}_\eta \in (0, \eta]$ and $\hat{c}_\eta \in \mathbb{R}$ such that

$$\hat{u}_\eta = \hat{s}_\eta\chi_\Sigma + \hat{c}_\eta\mathbb{1}_p, \qquad \underline{\text{(asymptotically exact recovery, } \hat{u}_\eta \overset{\eta\to\infty}{\to} u_o)} \tag{31}$$

and $\hat{s}_\eta \to \eta$ and $\hat{c}_\eta \to 0$ as $\eta \to \infty$. Hence $\hat{u}_\eta$ provides a faithful restoration of the original $u_o = \eta\chi_\Sigma$.

Proposition 7 ("0-1" PF) *Let ϕ be given by (f13) in Table 1. Assume that H4 and H12 are satisfied. Then there are $\eta_0 > 0$ and $\eta_1 > \eta_0$ such that*

$$\eta \in [0, \eta_0) \quad \Rightarrow \quad \hat{u}_\eta = \eta\zeta \, \mathbb{1}_p \qquad \underline{\textit{($\hat{u}_\eta$ is constant)}}, \tag{32}$$

$$\eta > \eta_1 \quad \Rightarrow \quad \hat{u}_\eta = \eta \, \chi_\Sigma, \qquad \underline{\textit{(exact recovery, $\hat{u}_\eta = u_o$)}} \tag{33}$$

where ζ is given in (30). Moreover, $\hat{u}_\eta$ in (32) and (33) is the unique global minimizer of $\mathcal{F}(\cdot, A\eta\chi_\Sigma)$.

By way of conclusion, non convexity and boundedness of ϕ can ensure correct edge recovery as well as (possibly asymptotically) correct recovery of u_o.

The results presented here can be extended to other forms of finite differences. At this stage, the assumption H4 (i.e. that A^*A is invertible) seems difficult to avoid. A further development is necessary to characterize the global minimizer $\hat{u}_\eta$ when data are corrupted with some perturbations.

5 Minimizers under non-smooth regularization

Observe that the minimizers corresponding to $\underline{\phi'(0^+) > 0}$ (non smooth regularization) in Figs. 3 (b)-(c), 8(b), and 9(b) and (d), 11(a), (b) and (c), 12(d), are constant on numerous regions. This section is aimed to explain and to generalize this phenomenon.

Consider

$$\mathcal{F}(u,v) = \Psi(u,v) + \beta\Phi(u) \tag{34}$$

$$\Phi(u) = \sum_{i=1}^{r} \phi\left(\left\|\mathrm{D}_i u\right\|_2\right), \tag{35}$$

where $\Psi : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ is any explicit or implicit $\mathcal{C}^m$-smooth function for $m \geqslant 2$ and $\mathrm{D}_i : \mathbb{R}^p \mapsto \mathbb{R}^s$, $\forall i \in I \stackrel{\text{def}}{=} \{1, \cdots, r\}$ are general linear operators for any integer $s \geqslant 1$. It is assumed that ϕ satisfies H1 (p. 3) and H6 (p. 16) along with

H14 *ϕ is $\mathcal{C}^2$-smooth on $\mathbb{R}_+^*$ and $\phi'(0^+) > 0$.*

It worths emphasizing that Ψ and ϕ can be convex or nonconvex. Let us define the set-valued function $\mathcal{J}$ on $\mathbb{R}^p$ by

$$\mathcal{J}(u) = \Big\{ i \in I \,:\, \|\mathrm{D}_i u\|_2 = 0 \Big\}. \tag{36}$$

Given $u \in \mathbb{R}^p$, $\mathcal{J}(u)$ indicates all regions where $\mathrm{D}_i u = 0$. Such regions are called[8] strongly homogeneous with respect to $\{\mathrm{D}_i\}$. In particular, if $\{\mathrm{D}_i\}$ correspond to first-order differences between neighboring samples of u or to discrete gradients, $\mathcal{J}(u)$ indicates all constant regions in u.

5.1 Main theoretical result

The results presented below are extracted from [78].

Theorem 10 *Given $v \in \mathbb{R}^q$, assume that $\mathcal{F}(\cdot, v)$ in (34)-(35) is such that Ψ is $\mathcal{C}^m$, $m \geqslant 2$ on $\mathbb{R}^p \times \mathbb{R}^q$, and that ϕ satisfies H1, H6 and H14. Let $\hat{u} \in \mathbb{R}^p$ be a (local) minimizer of $\mathcal{F}(\cdot, v)$. For $\hat{J} \stackrel{\text{def}}{=} \mathcal{J}(\hat{u})$, let $K_{\hat{J}}$ be the vector subspace*

$$K_{\hat{J}} = \Big\{ u \in \mathbb{R}^p : \mathrm{D}_i u = 0, \forall i \in \hat{J} \Big\}. \tag{37}$$

Suppose also that

(a) $\delta_1 \mathcal{F}(\hat{u}, v)(w) > 0$, for every $w \in K_{\hat{J}}^{\perp} \setminus \{0\}$;

(b) there is an open subset $O'_{\hat{J}} \subset \mathbb{R}^q$ such that $\mathcal{F}|_{K_{\hat{J}}}(., O'_{\hat{J}})$ has a local minimizer function $\mathcal{U}_{\hat{J}} : O'_{\hat{J}} \to K_{\hat{J}}$ which is $\mathcal{C}^{m-1}$ continuous at v and $\hat{u} = \mathcal{U}_{\hat{J}}(v)$.

Then there is an open neighborhood $O_{\hat{J}} \subset O'_{\hat{J}}$ of v such that $\mathcal{F}(\cdot, O_{\hat{J}})$ admits a $\mathcal{C}^{m-1}$ local minimizer function $\mathcal{U} : O_{\hat{J}} \to \mathbb{R}^p$ which satisfies $\mathcal{U}(v) = \hat{u}$, $\mathcal{U}|_{K_{\hat{J}}} = \mathcal{U}_{\hat{J}}$ and

$$\nu \in O_{\hat{J}} \quad \Rightarrow \quad \mathrm{D}_i \mathcal{U}(\nu) = 0, \ \textit{for all } i \in \hat{J}. \tag{38}$$

[8] The adverb "strongly" is used in order to emphasize the difference with just "homogeneous regions" that are characterized by $\|\mathrm{D}_i u\|_2 \approx 0$.

It can be shown that the results of Theorem 10 hold true also for irregular functions ϕ of the form (f13) in Table 1. Remind that $\hat{J}$ and $K_{\hat{J}}$ are the same as those introduced in (13).

Commentary on the assumptions. Since $\mathcal{F}(\cdot, v)$ has a local minimum at $\hat{u}$, Theorem 1 (p. 8) tells us that $\delta_1\mathcal{F}(\hat{u}, v)(w) \geqslant 0$, for all $w \in K_{\hat{J}}^{\perp}$ and this inequality cannot be strict unless $\mathcal{F}$ is non-smooth. When Φ is non-smooth as specified above, it is easy to see that (a) is not a strong requirement. By Lemma 1, condition (b) holds if $\mathcal{F}|_{K_{\hat{J}}}$ is $\mathcal{C}^m$ on a neighborhood of $(\hat{u}, v)$ belonging to $K_{\hat{J}} \times \mathbb{R}^q$, and if $D_1(\mathcal{F}|_{K_{\hat{J}}})(\hat{u}, v) = 0$ and $D_1^2(\mathcal{F}|_{K_{\hat{J}}})(\hat{u}, v) \succ 0$, which is the classical sufficient condition for a strict (local) minimizer.

If $\mathcal{F}$ is (possibly nonconvex) of the form (12) and assumption H4 (p. 13) holds as well, the intermediate results given in item 3 next to Theorem 4 (p. 14) show that (a) and (b) are satisfied for any $v \in \mathbb{R}^q \setminus \Theta$ where Θ is closed and $\mathbb{L}^q(\Theta) = 0$. In these conditions, real-world data have no real chance[9] to belong to Θ so they lead to (local) minimizers that satisfy conditions (a) and (b).

Significance of the results. Using the definition of $\mathcal{J}$ in (36), the conclusion of the theorem can be reformulated as

$$v \in O_{\hat{J}} \quad \Rightarrow \quad \mathcal{J}\left(\mathcal{U}(v)\right) \supseteq \hat{J} \quad \Leftrightarrow \quad \mathcal{U}(v) \in K_{\hat{J}}. \tag{39}$$

Minimizers involving large subsets $\hat{J}$ are observed in Figs. 3 (b)-(c), 8(b), and 9(b) and (d), 11(a), (b) and (c), 12(d). It was seen in Examples 1 and 4, as well as in § 4.3 (case $\phi'(0^+) > 0$), that $\hat{J}$ is nonempty for data v living in an open $O_{\hat{J}}$. An analytical example will be presented in § 5.2. So consider that $\underline{\#\hat{J} \geqslant 1}$. Then (39) is a severe restriction since $K_{\hat{J}}$ is a closed and negligible subset of $\mathbb{R}^p$ whereas data v vary on open subsets $O_{\hat{J}}$ of $\mathbb{R}^q$. (The converse situation where a local minimizer $\hat{u}$ of $\mathcal{F}(\cdot, v)$ satisfies $\mathrm{D}_i\hat{u} \neq 0$, for all i seems quite natural, especially for noisy data.) Note also that there is an open subset $\tilde{O}_{\hat{J}} \subset O_{\hat{J}}$ such that $\mathcal{J}\left(\mathcal{U}(v)\right) = \hat{J}$ for all $v \in \tilde{O}_{\hat{J}}$.

Focus on a (local) minimizer function $\mathcal{U} : O \to \mathbb{R}^p$ for $\mathcal{F}(\cdot, O)$ and put $\hat{J} = \mathcal{J}(\mathcal{U}(v))$ for some $v \in O$. By Theorem 10, the sets $O_{\hat{J}}$ and $\tilde{O}_{\hat{J}}$ are of positive measure in $\mathbb{R}^q$. The chance that random points ν (namely noisy data) come across $O_{\hat{J}}$, or $\tilde{O}_{\hat{J}}$, is real[10]. When data ν range over O, the set-valued function $(\mathcal{J} \circ \mathcal{U})$ generally takes several distinct values, say $\{J_j\}$. Thus, with a (local) minimizer function $\mathcal{U}$, defined on an open set O, there is associated a family of subsets $\{\tilde{O}_{J_j}\}$ which

[9] More precisely, the probability that real data (a random sample of $\mathbb{R}^q$) do belong to Θ is null.

[10] The reason for this claim is that probability that $\nu \in O_{\hat{J}}$, or that $\nu \in \tilde{O}_{\hat{J}}$, is strictly positive.

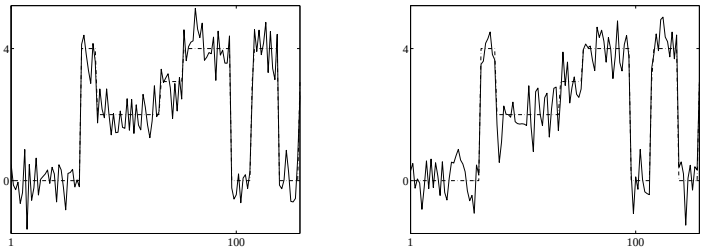


Figure 10: Data $v = u_o + n$ (—) corresponding to the original u_o (-.-.) contaminated with two different noise samples n on the left and on the right.

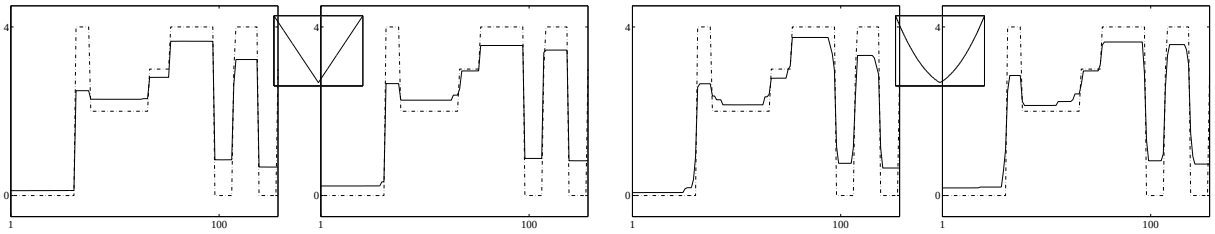
form a covering of O. When $\nu \in \tilde{O}_{J_j}$, we find a minimizer $\hat{u} = \mathcal{U}(\nu)$ satisfying $\mathcal{J}(\hat{u}) = J_j$. This underlies the conclusion stated next:

> Energies with non-smooth regularization terms as those considered here, exhibit local minimizers which generically satisfy constraints of the form $\mathcal{J}(\hat{u}) \neq \varnothing$.
> In particular, if $\{\mathrm{D}_i\}$ are discrete gradients or first-order difference operators, minimizers $\hat{u}$ are typically constant on many regions. E.g., if $\phi(t) = t$, we have $\Phi(u) = \mathrm{TV}(u)$ and this explains the stair-casing effect observed in TV methods on discrete images and signals [26, 34].

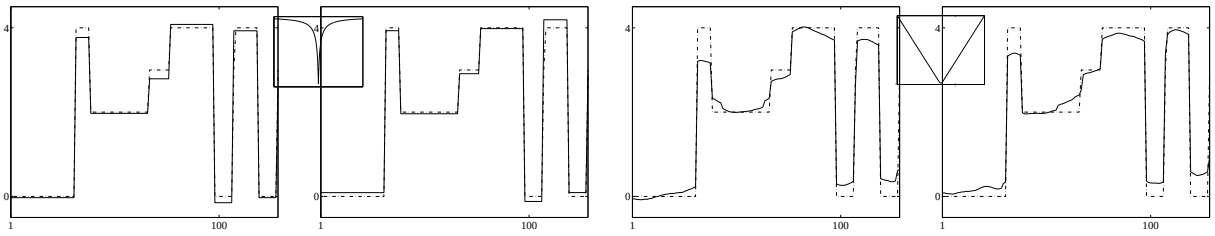
Restoration of a noisy signal. In Figs. 10 and 11 we consider the restoration of a piecewise constant signal u_o from noisy data $v = u_o + n$ by minimizing $\mathcal{F}(u, v) = \|u - v\|^2 + \beta \sum_{i=1}^{p-1} \phi(|u[i] - u[i+1]|)$. In order to evaluate the ability of different functions ϕ to recover, and to conserve, the locally constant zones yielded by minimizing the relevant $\mathcal{F}(\cdot, v)$, we process in the same numerical conditions two data sets, contaminated by two very different noise realizations plotted in Fig. 10. The minimizers shown in Figs. 11(a), (b) and (c) correspond to functions ϕ such that $\phi'(0^+) > 0$. In accordance with the above theoretical results, they are constant on large segments. In each one of these figures, the reader is invited to compare the subsets where the minimizers corresponding to the two data sets (Fig 10) are constant. In contrast, the function ϕ in Fig. 11(d) satisfies $\phi'(0^+) = 0$ and the resultant minimizers are nowhere constant.

5.2 The 1D TV regularization

The example below describes the sets $\tilde{O}_J$, for *every* $J \subset \{1, \ldots, r\}$, in the context of the one-dimensional discrete TV regularization. It provides a rich geometric interpretation of Theorem 10.



(a) $\phi(t) = t$ (here $\phi'(0^+) = 1$) (b) $\phi(t) = (t+\alpha)^2$ (here $\phi'(0^+) = 2\alpha$)



(c) $\phi(t) = \dfrac{\alpha t}{1+\alpha t}$ (here $\phi'(0^+) = \alpha$) (d) $\phi(t) = \begin{cases} \frac{t^2}{2} & \text{if } t \leqslant \alpha \\ \alpha t - \frac{\alpha^2}{2} & \text{if } t > \alpha \end{cases}$ (here $\phi'(0^+) = 0$)

Figure 11: Restoration using different functions ϕ. Original u_o (-.-.), minimizer $\hat{u}$ (—). Each figure from (a) to (d) shows the two minimizers $\hat{u}$ corresponding to the two data sets in Fig. 10 (left and right), while the shape of ϕ is plotted in the middle.

Let $\mathcal{F} : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ be given by

$$\mathcal{F}(u, v) = \|Au - v\|_2^2 + \beta \sum_{i=1}^{p-1} \big|u[i] - u[i+1]\big|, \tag{40}$$

where $A \in \mathbb{R}^{p \times p}$ is invertible and $\beta > 0$. It is easy to see that there is a unique minimizer function $\mathcal{U}$ for $\mathcal{F}(\cdot, \mathbb{R}^p)$. We have two striking phenomena (see [78] for details):

1. For every point $\hat{u} \in \mathbb{R}^p$, there is a polyhedron $Q_{\hat{u}} \subset \mathbb{R}^p$ of dimension $\#\mathcal{J}(\hat{u})$, such that for every $v \in Q_{\hat{u}}$, the same point $\mathcal{U}(v) = \hat{u}$ is the unique minimizer of $\mathcal{F}(\cdot, v)$;

2. For every $J \subset \{1, \ldots, p-1\}$, there is a subset $\tilde{O}_J \subset \mathbb{R}^p$, composed of $2^{p-\#J-1}$ unbounded polyhedra (of dimension p) of $\mathbb{R}^p$, such that for every $v \in \tilde{O}_J$, the minimizer $\hat{u}$ of $\mathcal{F}(\cdot, v)$ satisfies $\hat{u}_i = \hat{u}_{i+1}$ for all $i \in J$ and $\hat{u}_i \neq \hat{u}_{i+1}$ for all $i \in J^c$. A description of these polyhedra is given in the appendix of [78]. Moreover, their closure forms a covering of $\mathbb{R}^p$.

Remark 2 *The energy in* (40) *has a straightforward Bayesian interpretation in terms of maximum a posteriori (MAP) estimation (see p. 4). The quadratic data-fidelity term corresponds to a forward model of the form $v = Au_o + n$ where n is independent identically distributed (i.i.d.) Gaussian noise*

with mean zero and variance denoted by σ^2. *The likelihood reads* $\pi(v|u) = \exp\left(-\frac{1}{2\sigma^2}\|Au - v\|_2^2\right)$. *The regularization term corresponds to an i.i.d. Laplacian prior on each difference* $u[i] - u[i+1]$, $1 \leqslant i \leqslant p-1$. *More precisely, each difference has a distribution of the form* $\exp(-\lambda|t|)$ *for* $\lambda = \frac{\beta}{2\sigma^2}$. *Since this density is continuous on* $\mathbb{R}$, *the probability to get a null sample* $t = u[i] - u[i+1] = 0$, *is equal to zero. However, the results presented above show that for the minimizer* $\hat{u}$ *of* $\mathcal{F}(\cdot, v)$, *the probability to have* $\hat{u}[i] - \hat{u}[i+1] = 0$ *for a certain amount of indexes* i *is strictly positive. This means that the Laplacian prior on the differences* $u[i] - u[i+1]$ *is far from being incorporated in the MAP solution* $\hat{u}$. *On the other hand, given that* $\|\phi'\|_\infty = 1$ *and that* $\|\mathrm{D}\|_1 = 2$, *Theorem 7 tells us that* $\|A\hat{u} - v\|_\infty \leqslant \beta\|(AA^*)^{-1}A\|_\infty$, *hence the recovered noise* $(A\hat{u} - v)[i]$, $1 \leqslant i \leqslant q$ *is bounded. However, the noise* n *in the forward model is unbounded. The distribution of the original noise* n *is not incorporated in the MAP estimate* $\hat{u}$ *neither.*

5.3 An application to Computed Tomography

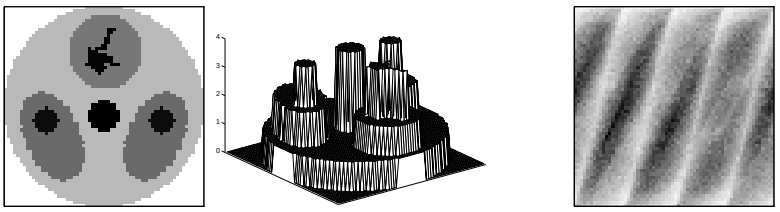
The concentration of an isotope in a part of the body provides an image characterizing metabolic functions and local blood flow [18, 54, 58]. In Emission Computed Tomography (ECT), a radioactive drug is introduced in a region of the body and the emitted photons are recorded around it. Data are formed by the number of photons $v[i] \geqslant 0$ reaching each detector, $i = 1, \cdots, q$. The observed photon counts v have a Poissonian distribution [18, 87]. Their mean is determined using projection operators $\{a_i,\, i = 1, 2, \ldots, q\}$ and a constant $\rho > 0$. The data-fidelity Ψ derived from the log-likelihood function is non strictly convex and reads:

$$\Psi(u, v) = \rho \left\langle \sum_{i=1}^{q} a_i,\ u \right\rangle - \sum_{i=1}^{q} v[i] \ln(\langle a_i, u\rangle). \tag{41}$$

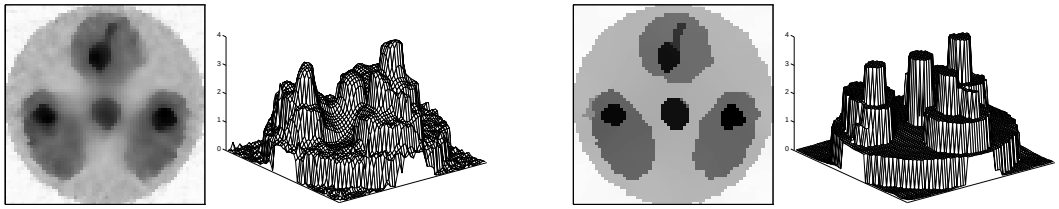
Fig. 12 presents image reconstruction from simulated ECT data by minimizing and energy of the form (34)-(35) where Ψ is given by (41) and $\{\mathrm{D}_i\}$ yield the first-order differences between each pixel and its 8 nearest neighbors. One observes, yet again, that a PF ϕ which is nonconvex with $\phi'(0^+) > 0$ leads to a nicely segmented piecewise constant reconstruction.

6 Minimizers relevant to non-smooth data-fidelity

Fig. 13 shows that there is a striking distinction in the behavior of the minimizers relevant to nonsmooth data-fidelity terms (b) with respect to non-smooth regularization (a). More precisely, many data samples are fitted exactly when the data-fidelity term is nonsmooth. This particular behavior is explained and generalized in the present section.

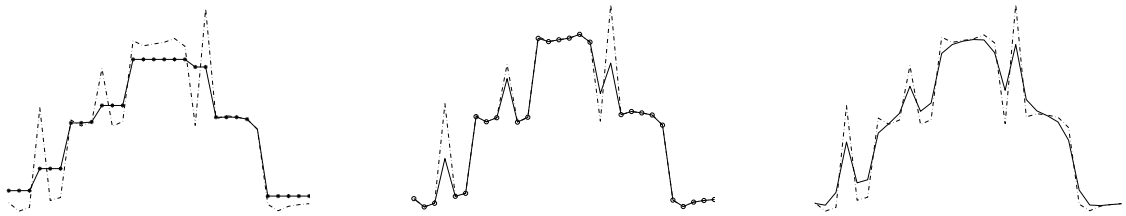


(a) Original phantom (b) ECT simulated data



(c) $\phi'(0) = 0$, edge-preserving (d) $\phi(t) = t/(\alpha + t)$ ($\phi'(0^+) > 0$, non convex)

Figure 12: ECT. $\mathcal{F}(u, v) = \Psi(u, v) + \beta \sum_{i \in I} \phi(|\mathrm{D}_i u|)$.



$\mathcal{F}(u, v) = \|u - v\|_2^2 + \beta \|\mathrm{D}u\|_1$ $\mathcal{F}(u, v) = \|u - v\|_1 + \beta \|\mathrm{D}u\|_2^2$ $\mathcal{F}(u, v) = \|u - v\|_2^2 + \beta \|\mathrm{D}u\|_2^2$

(a) Stair-casing (b) Exact data-fit (c)

Figure 13: D is a first-order difference operator, i.e. $\mathrm{D}_i u = u[i] - u[i+1]$, $1 \leqslant i \leqslant p-1$. Data (- - -), Restored signal (—). Constant pieces in (a) are emphasized using "$*$" while data samples that are equal to the relevant samples of the minimizer in (b) are emphasized using "$\circ$".

Consider

$$\mathcal{F}(u, v) = \Psi(u, v) + \beta \Phi(u), \tag{42}$$

$$\Psi(u, v) = \sum_{i=1}^{q} \psi\left(|\langle a_i, u \rangle - v[i]|\right), \tag{43}$$

where $a_i \in \mathbb{R}^p$ for all $i \in \{1, \cdots, q\}$ and $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ is a function satisfying

H15 *$\psi(0) = 0$, ψ is increasing and not identically null on $\mathbb{R}_+$, and $\psi \in \mathcal{C}^0(\mathbb{R}_+)$.*

Remind that the later condition means that $t \mapsto \psi(|t|)$ is continuous on $\mathbb{R}$ (*cf.* Definition 5, p. 11). Let $A \in \mathbb{R}^{q \times p}$ denote the matrix such that for any $i = 1, \cdots, q$, its ith row reads a_i^*.

Recall that many papers are dedicated to the minimization of $\Psi(u, v) = \|Au - v\|_\rho^\rho$ alone i.e., $\mathcal{F} = \Psi$, mainly for $\rho = 2$ (least-squares problems) [57], often for $\rho = 1$ (least absolute deviations) [15],

but also for $\rho \in (0, \infty]$ [83, 84]. Nonsmooth data-fidelity terms Ψ in energies of the form (42)-(43) were introduced in image processing in 2003 [75].

6.1 General theory

Here we present some results on the minimizers $\hat{u}$ of $\mathcal{F}$ as given in (42)-(43), where Ψ is non differentiable, obtained in [76, 77]. Additional assumptions are that

H16 *ψ is $\mathcal{C}^m$, $m \geqslant 2$ on $\mathbb{R}_+^*$ and $\psi'(0^+) > 0$ is finite.*

H17 *The regularization term $\Phi : \mathbb{R}^p \to \mathbb{R}$ in (42) is $\mathcal{C}^m$, $m \geqslant 2$.*

Note that Φ in (42) can be convex or nonconvex.

To analyze the phenomenon observed in Fig. 13(b), the following set-valued function $\mathcal{J}$ will be useful:

$$(u, v) \in (\mathbb{R}^p \times \mathbb{R}^q) \quad \mapsto \quad \mathcal{J}(u, v) = \Big\{ i \in \{1, \ldots, q\} : \langle a_i, u \rangle = v[i] \Big\}. \tag{44}$$

Given v and a (local) minimizer $\hat{u}$ of $\mathcal{F}(\cdot, v)$, the set of all data entries $v[i]$ that are fitted exactly by $(A\hat{u})[i]$ reads $\hat{J} = \mathcal{J}(\hat{u}, v)$. Its complement is $\hat{J}^c = \{1, \cdots, q\} \setminus \hat{J}$.

Theorem 11 *Let $\mathcal{F}$ be of the form (42)-(43) where assumptions H15, H16 and H17 hold true. Given $v \in \mathbb{R}^q$ let $\hat{u} \in \mathbb{R}^p$ be a (local) minimizer of $\mathcal{F}(\cdot, v)$. For $\hat{J} = \mathcal{J}(\hat{u}, y)$, where $\mathcal{J}$ is defined according to (44), let*

$$\mathcal{K}_{\hat{J}}(v) = \{u \in \mathbb{R}^p : \ \langle a_i, u \rangle = v[i] \ \forall i \in \hat{J} \text{ and } \langle a_i, u \rangle \neq v[i] \ \forall i \in \hat{J}^c\},$$

and let $K_{\hat{J}}$ be its tangent. Suppose the following:

(a) The set $\{a_i : i \in \hat{J}\}$ is linearly independent;

(b) $\forall w \in K_{\hat{J}} \setminus \{0\}$ we have $D_1(\mathcal{F}|_{\overline{\mathcal{K}_{\hat{J}}(v)}})(\hat{u}, v)w = 0$ and $D_1^2(\mathcal{F}|_{\overline{\mathcal{K}_{\hat{J}}(v)}})(\hat{u}, v)(w, w) > 0$;

(c) $\forall w \in K_{\hat{J}}^{\perp} \setminus \{0\}$ we have $\delta_1 \mathcal{F}(\hat{u}, v)(w) > 0$.

Then there are a neighborhood $O_{\hat{J}} \subset \mathbb{R}^q$ containing v and a $\mathcal{C}^{m-1}$ local minimizer function $\mathcal{U} : O_{\hat{J}} \to \mathbb{R}^p$ relevant to $\mathcal{F}(\cdot, O_{\hat{J}})$, yielding in particular $\hat{u} = \mathcal{U}(v)$, and

$$\nu \in O_{\hat{J}} \quad \Rightarrow \quad \begin{cases} \langle a_i, \mathcal{U}(\nu) \rangle = \nu[i] & \text{if} \quad i \in \hat{J}, \\ \langle a_i, \mathcal{U}(\nu) \rangle \neq \nu[i] & \text{if} \quad i \in \hat{J}^c. \end{cases} \tag{45}$$

The latter means that $\mathcal{J}(\mathcal{U}(\nu), \nu) = \hat{J}$ is constant on $O_{\hat{J}}$.

Note that for every v and $J \neq \varnothing$, the set $\mathcal{K}_J(v)$ is a finite union of connected components, whereas its closure $\overline{\mathcal{K}_J(v)}$ is an affine subspace. Its tangent $K_{\hat{J}}$ reads

$$K_{\hat{J}} = \{u \in \mathbb{R}^p : \ \langle a_i, u\rangle = 0 \ \forall i \in \hat{J}\}.$$

A comparison with $K_{\hat{J}}$ in (37) may be instructive. Compare also (b)-(c) in Theorem 11 with (a)-(b) in Theorem 10, p. 30. By the way, conditions (b)-(c) in Theorem 11 ensure that $\mathcal{F}(\cdot, v)$ reaches a strict minimum at $\hat{u}$ [76, Proposition 1]. Observe that this sufficient condition for strict minimum involves the behavior of $\mathcal{F}(\cdot, v)$ on two orthogonal subspaces separately. This occurs because of the nonsmoothness of $t \mapsto \psi(|t|)$ at zero. It can be useful to note that at a minimizer $\hat{u}$,

$$\delta_1\mathcal{F}(\hat{u}, v)(w) = \phi'(0^+)\sum_{i\in\hat{J}} |\langle a_i, w\rangle| + \sum_{i\in\hat{J}^c} \psi'(\langle a_i, \hat{u}\rangle - v[i])\langle a_i, w\rangle + \beta D\Phi(\hat{u})w \geqslant 0, \tag{46}$$

for any $w \in \mathbb{R}^p$.

Commentary on the assumptions. Assumption (a) does not require the independence of the whole set $\{a_i : i \in \{1, \ldots, q\}\}$. It is shown (Remark 6 in [76]) that this assumption fails to hold only for some v is included in a subspace of dimension strictly smaller than q. Hence, assumption (a) is satisfied for almost all $v \in \mathbb{R}^q$ and the theorem addresses any matrix A, whether it be singular or invertible.

Assumption (b) is the classical sufficient condition for a strict local minimum of a smooth function over an affine subspace. If an arbitrary function $\mathcal{F}(\cdot, v) : \mathbb{R}^p \to \mathbb{R}$ has a minimum at $\hat{u}$, then necessarily $\delta_1\mathcal{F}(\hat{u}, v)(w) \geqslant 0$ for all $w \in K_{\hat{J}}^{\perp}$, see Theorem 1. In comparison, (c) requires only that the latter inequality be strict.

It will be interesting to characterize the sets of data v for which (b) and (c) may fail at some (local) minimizers. Some ideas from § 3.2.1 might provide a starting point.

Corollary 2 *Let $\mathcal{F}$ be of the form* (42)-(43) *where $p = q$, and H15, H16 and H17 hold true. Given $v \in \mathbb{R}^q$ let $\hat{u} \in \mathbb{R}^p$ be a (local) minimizer of $\mathcal{F}(\cdot, v)$. Suppose the following:*

(a) The set $\{a_i : 1 \leqslant i \leqslant q\}$ is linearly independent;

(b) $\forall w \in \mathbb{R}^q$ satisfying $\|w\|_2 = 1$ we have $\beta\big|D\Phi(\hat{u})w\big| < \psi'(0^+)\sum_{i=1}^{q} |\langle a_i, w\rangle|$.

Then

$$\hat{J} = \{1, \cdots, q\}$$

and there are a neighborhood $O_{\hat{J}} \subset \mathbb{R}^q$ *containing* v *and a* $\mathcal{C}^{m-1}$ *local minimizer function* $\mathcal{U} : O_{\hat{J}} \to \mathbb{R}^p$ *relevant to* $\mathcal{F}(\cdot, O_{\hat{J}})$*, yielding in particular* $\hat{u} = \mathcal{U}(v)$*, and*

$$\nu \in O_{\hat{J}} \quad \Rightarrow \quad \langle a_i, \mathcal{U}(\nu) \rangle = \nu[i] \quad \forall i \in \hat{J} = \{1, \cdots, q\}. \tag{47}$$

More precisely, $\mathcal{U}(\nu) = A^{-1}\nu$ *for any* $\nu \in O_{\hat{J}}$*.*

Note that in the context of Corollary 2, A is invertible. Combining this with (46) and (b) shows that

$$\begin{aligned} \mathcal{K}_{\hat{J}}(v) &= \{u \in \mathbb{R}^p : \ Au = v\} = A^{-1}v, \\ K_{\hat{J}} &= \ker(A) = \{0\}. \end{aligned}$$

Then

$$\left\{ v \in \mathbb{R}^q : \beta \left| D\Phi(A^{-1}v)w \right| < \psi'(0^+) \sum_{i=1}^{q} |\langle a_i, w \rangle|, \ \forall w \in \mathbb{R}^q \setminus \{0\}, \|w\|_2 = 1 \right\} \subset O_{\hat{J}} \equiv O_{\{1,\cdots,q\}}.$$

The subset on the left contains an open subset of $\mathbb{R}^q$ by the continuity of $v \mapsto D\Phi(A^{-1}v)$ combined with (b).

Significance of the results. Consider that $\#J \geqslant 1$. The result in (45) means that the set-valued function $v \to \mathcal{J}(\mathcal{U}(v), v)$ is constant on $O_{\hat{J}}$, i.e., that $\mathcal{J}$ is constant under small perturbations of v. Equivalently, all residuals $\langle a_i, \mathcal{U}(v) \rangle - v[i]$ for $i \in \hat{J}$ are null on $O_{\hat{J}}$. Intuitively, this may seem unlikely, especially for noisy data.

Theorem 11 shows that $\mathbb{R}^q$ contains *volumes of positive measure* composed of data that lead to local minimizers which fit exactly the data entries belonging to the same set. In general, there are volumes corresponding to various $\hat{J}$ so that noisy data come across them. That is why non-smooth data-fidelity terms generically yield minimizers fitting exactly a certain number of the data entries. The resultant numerical effect is observed in Fig. 13(b) as well as in Figs. 15 and 16.

Remark 3 (stability of minimizers) The fact that there is a $\mathcal{C}^{m-1}$ local minimizer function shows that, in spite of the non-smoothness of Ψ (and hence of $\mathcal{F}$), for any v, all the local minimizers of $\mathcal{F}(\cdot, v)$ which satisfy the conditions of the theorem are *stable under weak perturbations of data* v. This result extends Lemma 1.

Example. Let $\mathcal{F}$ read

$$\mathcal{F}(u,v) = \sum_{i=1}^{q} \big|u[i] - v[i]\big| + \frac{\beta}{2}\sum_{i=1}^{q}(u[i])^2, \quad \beta > 0.$$

It is easy to compute (see [76, p. 978.]) that there is a local minimizer function $\mathcal{U}$ whose entries read

$$\mathcal{U}(v)[i] = \frac{1}{\beta}\text{sign}(v[i]) \quad \text{if} \quad |v[i]| > \frac{1}{\beta},$$
$$\mathcal{U}(v)[i] = v[i] \quad \text{if} \quad \big|v[i]\big| \leqslant \frac{1}{\beta}.$$

Condition (c) in Theorem 11 fails to hold only for $\Big\{v \in \mathbb{R}^q : \big|v[i]\big| = \frac{1}{\beta},\ \forall i \in \hat{J}\Big\}$. The latter set is of Lebesgue measure zero in $\mathbb{R}^q$.

For any $J \in \{1, \cdots, q\}$ put

$$O_J = \left\{v \in \mathbb{R}^q : \big|v[i]\big| \leqslant \frac{1}{\beta},\ \forall i \in J \ \text{ and } \ \big|v[i]\big| > \frac{1}{\beta},\ \forall i \in J^c\right\}.$$

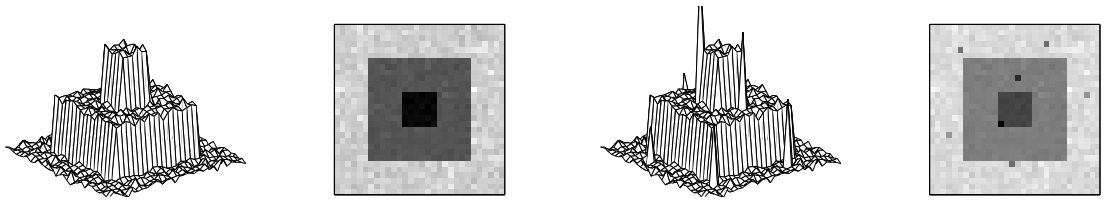
Obviously, every $v \in O_J$ gives rise to a minimizer $\hat{u}$ satisfying

$$\hat{u}[i] = v[i], \quad \forall i \in J \ \text{ and } \ \hat{u}[i] \neq v[i], \quad \forall i \in J^c.$$

Note that for every $J \subset \{1, \cdots, q\}$, the set O_J has a positive Lebesque measure in $\mathbb{R}^q$. Moreover, the union of all O_J when J ranges on all subsets $J \subset \{1, \cdots, q\}$ (including the empty set) forms a partition of $\mathbb{R}^q$.

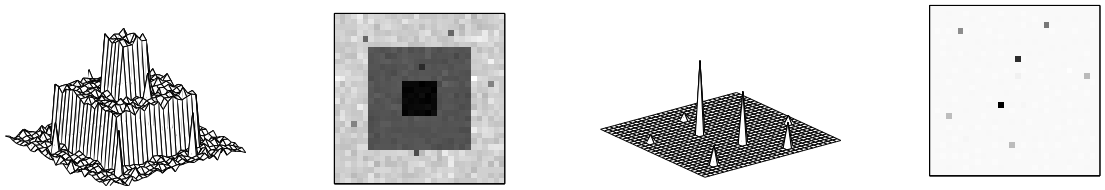
Numerical experiment. The original image u_o in Fig. 14(a) can be supposed to be a noisy version of an ideal piecewise constant image. Data v in Fig. 14(b) are obtained by replacing some pixels of u_o, whose locations are seen in Fig. 17-left, by aberrant impulsions, called outliers. In all Figs. 15, 16, 18 and 19, $\{\mathrm{D}_i\}$ correspond to the first-order differences between each pixel and its four nearest neighbors. The image in Fig. 15(a) corresponds to an ℓ_1 data-fidelity term for $\beta = 0.14$. The outliers are well visible although their amplitudes are clearly reduced. The image of the residuals $v - \hat{u}$, shown in Fig. 15(b), is null everywhere except at the positions of the outliers in v. The pixels corresponding to non-zero residuals (i.e. the elements of $\hat{J}^c$) provide a faithful estimate of the locations of the outliers in v, as seen in Fig. 17-middle. Next, in Fig. 16(a) we show a minimizer $\hat{u}$ of the same $\mathcal{F}(\cdot, v)$ obtained for $\beta = 0.25$. This minimizer does not contain visible outliers and is very close to the original image u_o. The image of the residuals $v - \hat{u}$ in Fig. 16(b) is null only on restricted areas, but has a very small magnitude everywhere beyond the outliers. However, applying the above detection rule now leads to numerous false detections, as seen in Fig. 17-right.

The minimizers of two different cost-function $\mathcal{F}$ involving a *smooth* data-fidelity term Ψ, shown in Figs. 18 and 19, do not fit any data entry. In particular, the restoration in Fig. 19 corresponds to a nonsmooth regularization and it is constant over large regions; this effect was explained in section 5.



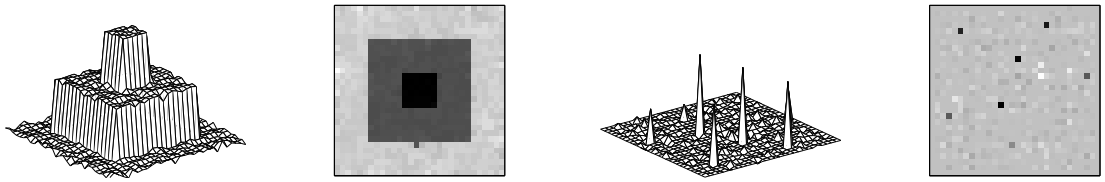
(a) Original u_o. (b) Data $v = u \cdot$outliers.

Figure 14: Original u_o and data v degraded by outliers.



(a) Restoration $\hat{u}$ for $\beta = 0.14$. (b) Residual $v - \hat{u}$.

Figure 15: Restoration using $\mathcal{F}(u, v) = \sum_i |u[i] - v[i]| + \beta \sum_{i \in I} |\mathrm{D}_i u|^\alpha$ $\alpha = 1.1$ and $\beta = 0.14$.



(a) Restoration $\hat{u}$ for $\beta = 0.25$. (b) Residual $v - \hat{u}$.

Figure 16: Restoration using $\mathcal{F}(u, v) = \sum_i |u[i] - v[i]| + \beta \sum_{i \in I} |\mathrm{D}_i u|^\alpha$ for $\alpha = 1.1$ and $\beta = 0.25$.

More details and information can be found in [76].

6.2 Applications

The possibility to keep some data samples unchanged by using nonsmooth data fidelity is a precious property in various application fields. Non-smooth data-fidelities are good to detect and smooth outliers. This was applied to impulse noise removal in [77] and to separate impulse from Gaussian noise in [81]. This property was extensively exploited for deblurring under impulse noise contamination, see e.g. [7–9].

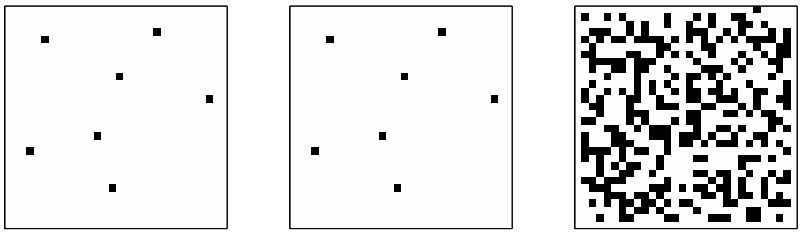
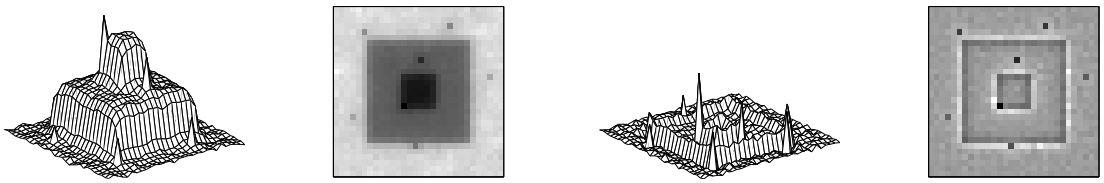


Figure 17: Left: the locations of the outliers in v. Next—the locations of the pixels of a minimizer $\hat{u}$ at which $\hat{u}[i] \neq v[i]$. Middle: these locations for the minimizer obtained for $\beta = 0.14$, Fig. 15. Right: the same locations for the minimizer relevant to $\beta = 0.25$, see Fig. 16.



(a) Restoration $\hat{u}$ for $\beta = 0.2$. (b) Residual $v - \hat{u}$.

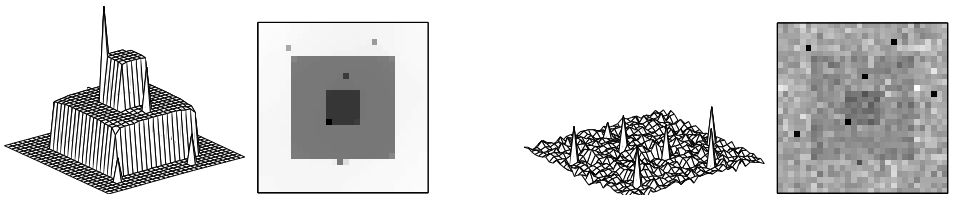
Figure 18: Restoration using a smooth cost-function, $\mathcal{F}(u, v) = \sum_i (u[i]-v[i])^2 + \beta \sum_i (|\mathrm{D}_i u|)^2$, $\beta = 0.2$.

Denoising of frame coefficients. Consider the recovery of an original (unknown) $u_o \in \mathbb{R}^p$— a signal or an image containing smooth zones and edges—from noisy data

$$v = u_o + n,$$

where n represents a perturbation. As discussed in section 4, a systematic default of the images and signals restored using convex edge-preserving PFs ϕ is that the amplitude of edges is underestimated.

Shrinkage estimators operate on a decomposition of data v into a frame of ℓ^2, say $\{w_i : i \in J\}$ where J is a set of indexes. Let W be the corresponding frame operator, i.e. $(Wv)[i] = \langle v, w_i \rangle$, $\forall i \in J$, and $\widetilde{W}$ be a left inverse of W, giving rise to the dual frame $\{\widetilde{w}_i : i \in J\}$. The frame coefficients of v read $y = Wv$ and are contaminated with noise Wn. The inaugural work of Donoho and Johnstone [35]



(a) Restoration $\hat{u}$ for $\beta = 0.2$. (b) Residual $v - \hat{u}$.

Figure 19: Restoration using non-smooth regularization $\mathcal{F}(u, v) = \sum_i (u[i] - v[i])^2 + \beta \sum_i |\mathrm{D}_i u|$, $\beta = 0.2$.

considers two different shrinkage estimators: given $T > 0$, hard thresholding corresponds to

$$y_T[i] = \begin{cases} y[i] & \text{if} \quad i \in J_1, \\ 0 & \text{if} \quad i \in J_0, \end{cases} \qquad \text{where} \qquad \begin{cases} J_0 &= \{i \in J : |y[i]| \leqslant T\}; \\ J_1 &= J \setminus J_0, \end{cases} \tag{48}$$

while in soft-thresholding one takes $y_T[i] = y[i] - T\text{sign}(y[i])$ if $i \in J_1$ and $y_T[i] = 0$ if $i \in J_0$. Both soft and hard thresholding are asymptotically optimal in the minimax sense if n is white Gaussian noise of standard deviation σ and

$$T = \sigma\sqrt{2\log_e p}. \tag{49}$$

This threshold is difficult to use in practice because it increases with the size of u. Numerous other drawbacks were found and important improvements were realized, see e.g. [4, 10, 20, 29, 33, 64, 68, 89]. In all these cases, the main problem is that smoothing large coefficients oversmooths edges while thresholding small coefficients can generate Gibs-like oscillations near edges, see Fig. 20(c) and (d). If shrinkage is weak, noisy coefficients (outliers) remain almost unchanged and produce artifacts having the shape of $\{\widetilde{w}_i\}$, see Fig. 20(c), (d), (e).

In order to alleviate these difficulties, several authors proposed hybrid methods where the information contained in important coefficients $y[i]$ is combined with priors in the domain of the sough-after signal or image [16, 21, 31, 38, 46, 65]. A critical analysis of these preexisting methods is presented in [41].

The key idea in [41] is to construct a specialized hybrid method involving ℓ_1 data fidelity on frame coefficients. More precisely, data are initially hard-thresholded—see (48)—using a suboptimal threshold T in order to keep as much as possible information. (The use of another more sophisticated shrinkage estimator would alter all coefficients, that is why a hard-thresholding is preferred.) Then

1. J_1 is composed of
   - large coefficients bearing the main features of u_o that one wishes to preserve intact;
   - aberrant coefficients (outliers) that must be restored using the regularization term.
2. J_0 is composed of
   - noise coefficients that must be kept null;
   - coefficients $y[i]$ corresponding to edges and other details in u_o—these need to be restored in accordance with the prior incorporated in the regularization term.

The theory in [41] is developed for signals and images defined on a subset of $\mathbb{R}^d$ where $d = 1$ or $d = 2$, respectively, and for frames of L^2. To ensure coherence of the chapter, the approach is presented in

the discrete setting. In order to reach the goals formulated in 1 and 2 above, denoised coefficients $\hat{x}$ are defined as a minimizer of the hybrid energy $F(.,y)$ given below:

$$F(x,y) = \lambda_1 \sum_{i\in J_1} \big|x[i] - y[i]\big| + \lambda_0 \sum_{i\in J_0} \big|x[i]\big| + \sum_{i\in I} \phi(\|\mathrm{D}_i \widetilde{W} x\|_2), \quad \lambda_{0,1} > 0, \tag{50}$$

where ϕ is convex and edge-preserving. Then the sought after denoised image or signal is

$$\hat{u} = \widetilde{W}\hat{x} = \sum_{i\in J} \widetilde{w}_i \hat{x}[i].$$

Several important properties relevant to the minimizers of F in (50), the parameters λ_i, $i \in \{0,1\}$ and the solution $\hat{u}$ are outlined in [41].

Noisy data v are shown along with the original u_o in Fig. 20(a). The restoration in Fig. 20 (b) minimizes $\mathcal{F}(u) = \|Au - v\|_2^2 + \beta \sum_i \phi(\|\mathrm{D}_i u\|_2)$ where $\phi(t) = \sqrt{\alpha + t^2}$ for $\alpha = 0.1$, $\beta = 100$—homogeneous regions remain noisy, edges are smoothed and spikes are eroded. Fig. 20(c) is obtained using the sure-shrink method [36] form the toolbox WaveLab. The other restorations use thresholded Daubechies wavelet coefficients with 8 vanishing moments. The optimal value for the hard thresholding obtained using (49) is $T = 35$. The relevant restoration—Fig. 20 (d)—exhibits important Gibbs-like oscillations as well as wavelet-shaped artifacts.

The threshold chosen in [41] is $T = 23$. The corresponding coefficients have a richer information content but $\widetilde{W} y_T$, shown in Fig. 20 (e) manifests Gibbs artifacts and many wavelet-shaped artifacts. Introducing the thresholded coefficients of Fig. 20 (e) in (50) leads to Fig. 20 (f) : edges are clean and piecewise polynomial parts are well recovered.

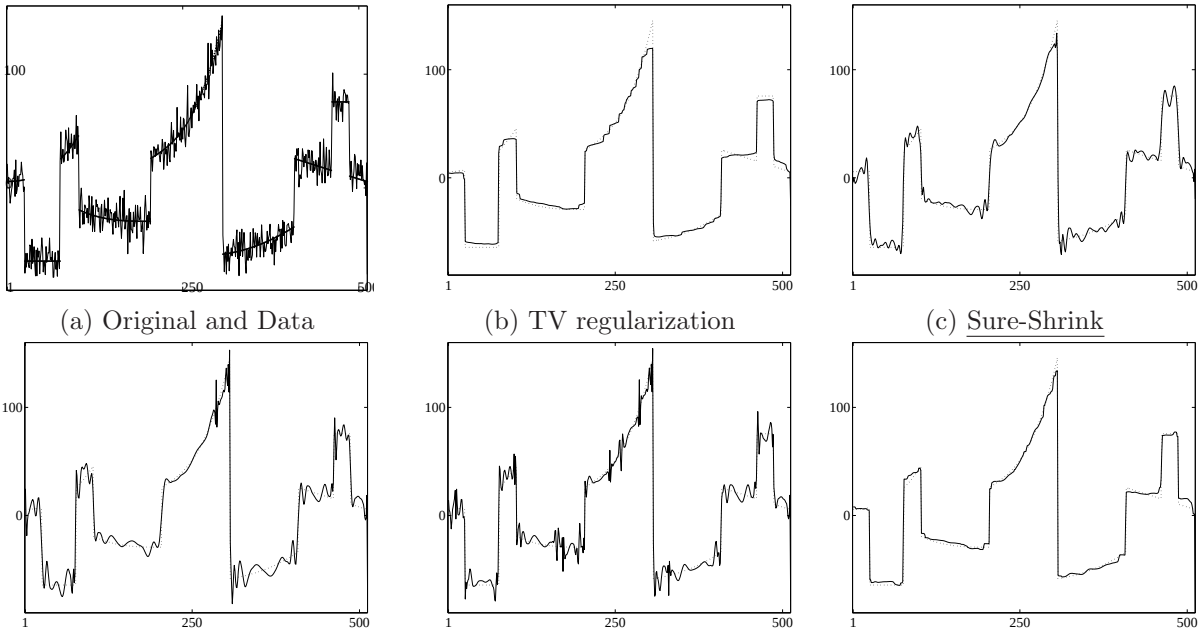
6.3 The L_1-TV case

For discrete signals of finite length, energies of the form $\mathcal{F}(u,v) = \|u - v\|_1 + \beta \sum_{i=1}^{p-1} |u[i+1] - u[i]|$ were considered by Alliney in 1992 [1]. These were exhibited to provide a variational formulation to digital filtering problems.

Following [1, 76, 77], S. Esedoglu and T. Chan explored in [25] the minimizers of the L_1-TV functional given below

$$\mathcal{F}(u,v) = \int_{\mathbb{R}^d} |u(x) - v(x)| dx + \beta \int_{\mathbb{R}^d} |\nabla u(x)| dx, \tag{51}$$

where the sought-after minimizer $\hat{u}$ belongs to the space of bounded variation functions on $\mathbb{R}^d$. The main focus is on images, i.e. $d = 2$. The analysis in [25] is based on a representation of $\mathcal{F}$ in (51) in terms of the level sets of u and v. Most of the results are established for data v given by the characteristic function χ_Σ of a bounded domain $\Sigma \subset \mathbb{R}^d$. Theorem 5.2. in [25] says that if $v = \chi_\Sigma$,



(a) Original and Data (b) TV regularization (c) Sure-Shrink

(d) $T=35$ optimal, $\hat{u}_T = \sum_i y_T[i]\widetilde{w}_i$ (e) y_T, $T = 23$, $\hat{u}_T = \sum_i y_T[i]\widetilde{w}_i$ $\Rightarrow$ (f) The proposed method

Figure 20: Methods to restorer the noisy signal in (a). Restored signal (—), Original signal (- -).

where $\Sigma \subset \mathbb{R}^d$ is bounded, then $\mathcal{F}(\cdot, v)$ admits a minimizer of the form $\hat{u} = \chi_{\hat{\Sigma}}$ (with possibly $\hat{\Sigma} \neq \Sigma$). Furthermore, Corollary 5.3. in [25] states that if in addition Σ is convex, then for almost every $\beta \geqslant 0$, $\mathcal{F}(\cdot, v)$ admits a unique minimizer and $\hat{u} = \chi_{\hat{\Sigma}}$ with $\hat{\Sigma} \subseteq \Sigma$. Moreover, it is shown that small features in the image maintain their contrast intact up to some value of β while for a larger β they suddenly disappear.

Recently, L_1-TV energies were revealed very successful in image decomposition, see e.g. [6, 42].

7 Conclusion

In this chapter we provided some theoretical results relating the shape of the energy $\mathcal{F}$ to minimize and the salient features of its minimizers $\hat{u}$, see (9). These results can serve as a kind of backward modeling: given an inverse problem along with our requirements (priors) on its solution, they guide us how to construct an energy functional whose minimizers properly incorporate all this information. The theoretical results are illustrated using numerical examples. Various application fields can take a benefit from these results. The problematic of such a backward modeling remains open because of the infinite diversity of the inverse problems to solve and the possible energy functionals.

8 Cross-references

- Inverse Scattering
- Large Scale Inverse Problems
- Learning, Classification, Data Mining
- Linear Inverse Problems
- Numerical Methods for Variational Approach in Image Analysis
- Regularization Methods for Ill-Posed Problems
- Segmentation with Priors
- Tomography
- Total Variation in Imaging

References

[1] S. Alliney, Digital filters as absolute norm regularizers, IEEE Transactions on Signal Processing, SP-40 (1992), pp. 1548–1562.

[2] F. Alter, S. Durand, and J. Forment, Adapted total variation for artifact free decompression of JPEG images, Journal of Mathematical Imaging and Vision, 23 (2005), pp. 199–211.

[3] L. Ambrosio, N. Fusco, and D. Pallara, Functions of Bounded Variation and Free Discontinuity Problems, Oxford Mathematical Monographs, Oxford University Press, 2000.

[4] A. Antoniadis and J. Fan, Regularization of wavelet approximations, Journal of Acoustical Society America, 96 (2001), pp. 939–967.

[5] G. Aubert and P. Kornprobst, Mathematical problems in image processing, Springer-Verlag, Berlin, 2 ed., 2006.

[6] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher, Structure-texture image decomposition - modeling, algorithms, and parameter selection, International Journal of Computer Vision, 67 (2006), pp. 111–136.

[7] L. Bar, A. Brook, N. Sochen, and N. Kiryati, Deblurring of color images corrupted by impulsive noise, IEEE Transactions on Image Processing, 16 (2007), pp. 1101–1111.

[8] L. Bar, N. Kiryati, and N. Sochen, Image deblurring in the presence of impulsive noise, International Journal of Computer Vision, 70 (2006), pp. 279–298.

[9] L. Bar, N. Sochen, and N. Kiryati, Image deblurring in the presence of salt-and-pepper noise, in Proceeding of 5th International Conference on Scale Space and PDE methods in Computer Vision, ser. LNCS, vol. 3439, 2005, pp. 107–118.

[10] M. Belge, M. Kilmer, and E. Miller, Wavelet domain image restoration with adaptive edge-preserving regularization, IEEE Transactions on Image Processing, 9 (2000), pp. 597–608.

[11] J. E. Besag, Spatial interaction and the statistical analysis of lattice systems (with discussion), Journal of the Royal Statistical Society B, 36 (1974), pp. 192–236.

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

[13] 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, 19 (1996), pp. 57–91.

[14] A. Blake and A. Zisserman, Visual reconstruction, The MIT Press, Cambridge, 1987.

[15] B. Bloomfield and W. L. Steiger, Least Absolute Deviations: Theory, applications and algorithms, Birkhäuser, Boston, 1983.

[16] Y. Bobichon and A. Bijaoui, Regularized multiresolution methods for astronomical image enhancement, Exper. Astron., (1997), pp. 239–255.

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

[18] ——, A unified approach to statistical tomography using coordinate descent optimization, IEEE Transactions on Image Processing, 5 (1996), pp. 480–492.

[19] K. Bredies, K. Kunich, and T. Pock, Total generalized variation, SIAM Journal on Imaging Sciences (to appear), (2010).

[20] E. J. Candès, D. Donoho, and L. Ying, Fast discrete curvelet transforms, SIAM Multiscale Model. Simul., 5 (2005), pp. 861–899.

[21] E. J. Candès and F. Guo, New multiscale transforms, minimum total variation synthesis. Applications to edge-preserving image reconstruction, Signal Processing, 82 (2002).

[22] F. Catte, T. Coll, P. L. Lions, and J. M. Morel, Image selective smoothing and edge detection by nonlinear diffusion (I), SIAM Journal on Numerical Analysis, 29 (1992), pp. 182–193.

[23] A. Chambolle, An algorithm for total variation minimization and application, Journal of Mathematical Imaging and Vision, 20 (2004).

[24] A. Chambolle and P.-L. Lions, Image recovery via total variation minimization and related problems, Numerische Mathematik, 76 (1997), pp. 167–188.

[25] T. Chan and S. Esedoglu, Aspects of total variation regularized l^1 function approximation, SIAM Journal on Applied Mathematics, 65 (2005), pp. 1817–1837.

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

[27] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, Deterministic edge-preserving regularization in computed imaging, IEEE Transactions on Image Processing, 6 (1997), pp. 298–311.

[28] R. Chellapa and A. Jain, Markov Random Fields: theory and application, Academic Press, Boston, 1993.

[29] C. Chesneau, J. Fadili, and J.-L. Starck, Stein block thresholding for image denoising, tech. report.

[30] P. G. Ciarlet, Introduction to Numerical Linear Algebra and Optimization, Cambridge University Press, 1989.

[31] R. R. Coifman and A. Sowa, Combining the calculus of variations and wavelets for image enhancement, Applied and Computational Harmonic Analysis, 9 (2000).

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

[33] M. N. Do and M. Vetterli, The contourlet transform: an efficient directional multiresolution image representation, IEEE Transactions on Image Processing, 15 (2005), pp. 1916–1933.

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

[35] D. L. Donoho and I. M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Biometrika, 81 (1994), pp. 425–455.

[36] ——, Adapting to unknown smoothness via wavelet shrinkage, Journal of Acoustical Society America, 90 (1995).

[37] A. L. Dontchev and T. Zollezi, Well-posed optimization problems, Springer-Verlag, New York, 1993.

[38] S. Durand and J. Froment, Reconstruction of wavelet coefficients using total variation minimization, SIAM Journal on Scientific Computing, 24 (2003), pp. 1754–1767.

[39] S. Durand and M. Nikolova, Stability of minimizers of regularized least squares objective functions I: study of the local behaviour, Applied Mathematics and Optimization (Springer-Verlag New York), 53 (2006), pp. 185–208.

[40] ——, Stability of minimizers of regularized least squares objective functions II: study of the global behaviour, Applied Mathematics and Optimization (Springer-Verlag New York), 53 (2006), pp. 259–277.

[41] S. Durand and M. Nikolova, Denoising of frame coefficients using ℓ_1 data-fidelity term and edge-preserving regularization, SIAM Journal on Multiscale Modeling and Simulation, 6 (2007), pp. 547–576.

[42] V. Duval, J.-F. Aujol, and Y. Gousseau, The TVL1 model: a geometric point of view, SIAM Journal on Multiscale Modeling and Simulation, 8 (2009), pp. 154–189.

[43] I. Ekeland and R. Temam, Convex Analysis and Variational Problems, SIAM, Amsterdam: North Holland, 1976.

[44] F. Fessler, Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography, IEEE Transactions on Image Processing, 5 (1996), pp. 493–506.

[45] A. Fiacco and G. McCormic, Nonlinear programming, Classics in Applied Mathematics, SIAM, Philadelphia, 1990.

[46] J. Froment and S. Durand, Artifact free signal denoising with wavelets, in Proceedings of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing, vol. 6, 2001.

[47] D. Geman, Random fields and inverse problems in imaging, vol. 1427, École d'Été de Probabilités de Saint-Flour XVIII - 1988, Springer-Verlag, lecture notes in mathematics ed., 1990, pp. 117–193.

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

[49] D. Geman and C. Yang, Nonlinear image recovery with half-quadratic regularization, IEEE Transactions on Image Processing, IP-4 (1995), pp. 932–946.

[50] S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6 (1984), pp. 721–741.

[51] G. Golub and C. Van Loan, Matrix Computations, The Johns Hopkins University Press, Baltimore, 3 ed., 1996.

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

[53] A. Haddad and Y. Meyer, Variational methods in image processing, in "Perspective in Nonlinear Partial Differential equations in Honor of Haïm Brezis," Contemporany Mathematics 446, AMS, 2007, pp. 273–295.

[54] G. Herman, Image rteconstruction from projections. The fundamentals of computerized tomography, Academic Press, 1980.

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

[56] B. Hofmann, Regularization for Applied Inverse and Ill-posed Problems, Teubner, Leipzig, 1986.

[57] T. Kailath, A view of three decades of linear filtering theory, IEEE Transactions on Information Theory, IT-20 (1974), pp. 146–181.

[58] A. Kak and M. Slaney, Principles of Computerized Tomographic Imaging, IEEE Press, New York, NY, 1987.

[59] A. K. e. Katsaggelos, Digital image restoration, Springer-Verlag, New York, 1991.

[60] D. Keren and M. Werman, Probabilistic analysis of regularization, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-15 (1993), pp. 982–995.

[61] K. Lange, Convergence of EM image reconstruction algorithms with Gibbs priors, IEEE Transactions on Medical Imaging, 9 (1990), pp. 439–446.

[62] S. Li, Markov Random Field Modeling in Computer Vision, Springer-Verlag, New York, 1 ed., 1995.

[63] S. Z. Li, On discontinuity-adaptive smoothness priors in computer vision, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-17 (1995), pp. 576–586.

[64] F. Luisier and T. Blu, SURE-LET multichannel image denoising: Interscale orthonormal wavelet thresholding, IEEE Transactions on Image Processing, 17 (2008), pp. 482–492.

[65] F. Malgouyres, Minimizing the total variation under a general convex constraint for image restoration, IEEE Transactions on Image Processing, 11 (2002), pp. 1450–1456.

[66] J.-M. Morel and S. Solimini, Variational Methods in Image Segmentation, Birkhäuser, Basel, 1995.

[67] V. A. Morozov, Regularization Methods for Ill Posed Problems, CRC Press, Boca Raton, 1993.

[68] P. Moulin and J. Liu, Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors, IEEE Transactions on Image Processing, 45 (1999), pp. 909–919.

[69] ——, Statistical imaging and complexity regularization, IEEE Transactions on Information Theory, 46 (2000), pp. 1762–1777.

[70] D. Mumford and J. Shah, Optimal approximations by piecewise smooth functions and associated variational problems, Communications on Pure and Applied Mathematics, (1989), pp. 577–684.

[71] M. Nashed and O. Scherzer, Least squares and bounded variation regularization with nondifferentiable functionals, Numer. Funct. Anal. and Optimiz., 19 (1998), pp. 873–901.

[72] M. Nikolova, Regularisation functions and estimators, in Proceedings of the IEEE International Conference on Image Processing, vol. 2, 1996, pp. 457–460.

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

[74] M. Nikolova, Thresholding implied by truncated quadratic regularization, IEEE Transactions on Image Processing, 48 (2000), pp. 3437–3450.

[75] M. Nikolova, Image restoration by minimizing objective functions with non-smooth data-fidelity terms, in IEEE Int. Conf. on Computer Vision / Workshop on Variational and Level-Set Methods, 2001, pp. 11–18.

[76] ——, Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers, SIAM Journal on Numerical Analysis, 40 (2002), pp. 965–994.

[77] ——, A variational approach to remove outliers and impulse noise, Journal of Mathematical Imaging and Vision, 20 (2004).

[78] ——, Weakly constrained minimization. Application to the estimation of images and signals involving constant regions, Journal of Mathematical Imaging and Vision, 21 (2004), pp. 155–175.

[79] ——, Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares, SIAM Journal on Multiscale Modeling and Simulation, 4 (2005), pp. 960–991.

[80] ——, Analytical bounds on the minimizers of (nonconvex) regularized least-squares, AIMS Journal on Inverse Problems and Imaging, 1 (2007), pp. 661–677.

[81] ——, Semi-explicit solution and fast minimization scheme for an energy with ℓ_1-fitting and tikhonov-like regularization, Journal of Mathematical Imaging and Vision, 34 (2009), pp. 32–47.

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

[83] T. T. Pham and R. J. P. De Figueiredo, Maximum likelihood estimation of a class of non-Gaussian densities with application to l_p deconvolution, IEEE Transactions on Signal Processing, 37 (1989), pp. 73–82.

[84] J. R. Rice and J. S. White, Norms for smoothing and estimation, SIAM Rev., 6 (1964), pp. 243–256.

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

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

[87] K. Sauer and C. Bouman, A local update strategy for iterative reconstruction from projections, IEEE Transactions on Signal Processing, SP-41 (1993), pp. 534–548.

[88] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen, Variational problems in imaging, Springer, New York, 2009.

[89] E. P. Simoncelli and E. H. Adelson, Noise removal via Bayesian wavelet coding, in Proceedings of the IEEE International Conference on Image Processing, Lausanne, Switzerland, Sep. 1996, pp. 379–382.

[90] R. Stevenson and E. Delp, Fitting curves with discontinuities, in Proc. of the 1st Int. Workshop on Robust Comput. Vision, Seattle, WA, 1990, pp. 127–136.

[91] U. Tautenhahn, Error estimates for regularized solutions of non-linear ill-posed problems, Inverse Problems, (1994), pp. 485–500.

[92] S. Teboul, L. Blanc-Féraud, G. Aubert, and M. Barlaud, Variational approach for edge-preserving regularization using coupled PDE's, IEEE Transactions on Image Processing, 7 (1998), pp. 387–397.

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

[94] C. Vogel, Computational Methods for Inverse Problems, Frontiers in Applied Mathematics Series, Number 23, SIAM, 2002.

[95] M. Welk, G. Steidl, and J. Weickert, Locally analytic schemes: a link between diffusion filtering and wavelet shrinkage, Applied and Computational Harmonic Analysis, 24 (2008).

[96] G. Winkler, Image analysis, random fields and Markov chain Monte Carlo methods. A mathematical introduction, Applications of mathematics 27. Stochastic models and applied probability, Springer Verlag, 2 ed., 2006.