

# Energy minimization methods

Mila NIKOLOVA

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

## Table of content

1. Introduction	1
2. Preliminaries	5
3. Regularity results	9
4. Nonconvex regularization	14
5. Nonsmooth regularization	19
6. Nonsmooth data fidelity	25
7. Nonsmooth data fidelity and regularization	32
8. Conclusion	37
9. Cross-references	38
10. References	38

## 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. The energies under consideration can be differentiable or not, convex or not. Analytical results on the minimizers of different energies are provided that reveal salient features of the images recovered in this way, as a function of the shape of the energy itself. An intrinsic mutual relationship between energy minimization and modeling via the choice of the energy is thus established. Examples and illustrations corroborate the presented results. Applications that take benefit from these results are presented as well.

## 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

$$v = A(u_o) \text{ with noise,} \quad (1)$$

where  $A : \mathbb{R}^p \rightarrow \mathbb{R}^q$  is a (linear or nonlinear) transform. When  $u$  is an  $m \times n$  image, its pixels are arranged columnwise into a  $p$ -length real vector, where  $p = mn$  and the original  $u[i, j]$  is identified with  $u[(i-1)m + j]$ . Some typical applications are, for instance, denoising, deblurring, segmentation, zooming and super-resolution, reconstruction in inverse problems, coding and compression, feature

selection, and compressive sensing. 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

$$\text{Find } \hat{u} \text{ such that } \mathcal{F}(\hat{u}, v) = \min_{u \in U} \mathcal{F}(u, v), \quad (2)$$

$$\mathcal{F}(u, v) = \Psi(u, v) + \beta \Phi(u), \quad (3)$$

where  $\mathcal{F} : \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}$  is called an energy (or an objective),  $U \subset \mathbb{R}^p$  is a set of constraints,  $\Psi$  is a data fidelity term,  $\Phi$  brings prior information on  $u_o$ , and  $\beta > 0$  is a parameter which controls the trade-off between  $\Psi$  and  $\Phi$ .

The term  $\Psi$  ensures that  $\hat{u}$  satisfies (1) quite faithfully according to an appropriate measure. The noise  $n$  is random and a natural way to derive  $\Psi$  from (1) is to use probabilities; see, e.g., [5, 32, 37, 56]. More precisely, if  $\pi(v|u)$  is the likelihood of data  $v$ , the usual choice is

$$\Psi(u, v) = -\log \pi(v|u). \quad (4)$$

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

$$\Psi(u, v) \propto \|Au - v\|_2^2. \quad (5)$$

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

The role of  $\Phi$  in (3) is to push the solution to exhibit some a priori known or desired features. It is called prior or regularization or penalty term. In many image processing applications,  $\Phi$  is of the form

$$\Phi(u) = \sum_{i=1}^r \phi(\|D_i u\|), \quad (6)$$

where for any  $i \in \{1, \dots, r\}$ ,  $D_i : \mathbb{R}^p \rightarrow \mathbb{R}^s$ , for  $s$  an integer  $s \geq 1$ , are linear operators and  $\|\cdot\|$  is usually the  $\ell_1$  or the  $\ell_2$  norm. For instance, the family  $\{D_i\} \equiv \{D_i : i \in \{1, \dots, r\}\}$  can represent the discrete approximation of the gradient or the Laplacian operator on  $u$  or the finite differences of various orders, or the combination of any of these with the synthesis operator of a frame transform, or the vectors of the canonical basis of  $\mathbb{R}^r$ . Note that  $s = 1$  if  $\{D_i\}$  are finite differences or a discrete Laplacian; then

$$s = 1 \quad \Rightarrow \quad \phi(\|D_i u\|) = \phi(|D_i u|).$$

And if  $\{D_i\}$  are the basis vectors of  $\mathbb{R}^r$ , one has  $\phi(|D_i u|) = \phi(|u[i]|)$ . In (6),  $\phi : \mathbb{R}_+ \mapsto \mathbb{R}$  is quite a “general” function, often called a *potential function* (PF). A very standard assumption is that

**H 1**  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$  is proper, lower semicontinuous (l.s.c.) and increasing on  $\mathbb{R}_+$ , with  $\phi(t) > \phi(0)$  for any  $t > 0$ .

Some typical examples for  $\phi$  are given in Table 1 and their plots in Fig. 1.

**Remark 1** If  $\phi'(0^+) > 0$  the function  $t \rightarrow \phi(|t|)$  is nonsmooth at zero in which case  $\Phi$  is nonsmooth on  $\cup_{i=1}^r [w \in \mathbb{R}^p : D_i w = 0]$ . Conversely,  $\phi'(0^+) = 0$  leads to a smooth at zero  $t \rightarrow \phi(|t|)$ . With the PF (f13),  $\Phi$  leads to the counting function, commonly called the  $\ell_0$ -norm.

Convex PFs	
$\phi'(0^+) = 0$	$\phi'(0^+) > 0$
(f1) $\phi(t) = t^\alpha, 1 < \alpha \leq 2$	(f5) $\phi(t) = t$
(f2) $\phi(t) = \sqrt{\alpha + t^2}$	
(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) $\phi(t) = \frac{\alpha t^2}{1 + \alpha t^2}$	(f11) $\phi(t) = \frac{\alpha t}{1 + \alpha t}$
(f8) $\phi(t) = \log(\alpha t^2 + 1)$	(f12) $\phi(t) = \log(\alpha t + 1)$
(f9) $\phi(t) = 1 - \exp(-\alpha t^2)$	(f13) $\phi(0) = 0, \phi(t) = 1 \text{ if } t \neq 0$

Table 1: Commonly used PFs  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$  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. And all nonconvex PFs with  $\phi'(0^+) > 0$  are *concave* on  $\mathbb{R}_+$ . Recall that (f6) is the discrete equivalent of the Mumford-Shah (MS) prior [17, 72].

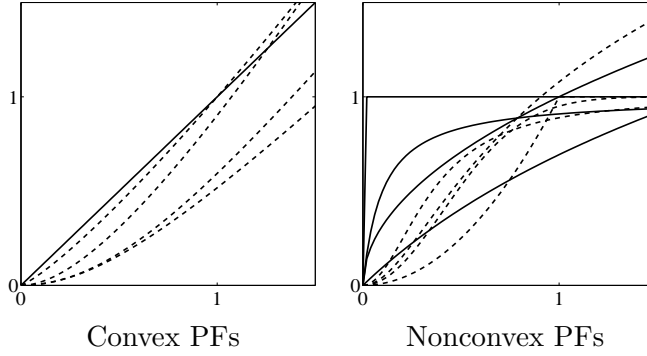


Figure 1: Plots of the PFs given in Table 1. PFs with  $\phi'(0^+) = 0$  (- - -), PFs with  $\phi'(0^+) > 0$  (—).

For the human vision, *an important requirement is that the prior  $\Phi$  promotes smoothing inside homogeneous regions but preserves sharp edges*. According to a fine analysis conducted in the 1990s, and summarized in [5],  $\phi$  preserves edges if *H1 holds as if H2, stated below, holds true* as well:

$$\mathbf{H\,2} \quad \lim_{t \rightarrow \infty} \frac{\phi'(t)}{t} = 0.$$

This assumption is satisfied by all PFs in Table 1 except for (f1) in case if  $\alpha = 2$ . Note that there are numerous other heuristics for edge preservation.

## 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,  $\hat{u}$  minimizes with respect to  $u$  the energy

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

Identifying these first term above with  $\Psi(\cdot, v)$  and the second one with  $\Phi$  shows the basis of the equivalence. Classical papers on MAP energies using MRF priors are [14–16, 20, 51, 56]. Since the pioneering work of Geman and Geman [56], various *nonconvex* PFs  $\phi$  were explored in order to produce images involving neat edges, see, e.g., [54, 55, 65]. MAP energies involving MRF priors are also considered in many books, such as [32, 53, 64]. For a pedagogical account, see [96].

- Regularization for ill-posed inverse problems was initiated in the book of Tikhonov and Arsenin [93] in 1977. The main idea can be stated in terms of the stabilization of this kind of problems. Useful textbooks in this direction are, e.g., [61, 69, 94] and especially the recent [91]. This methodology and its most recent achievements are nicely discussed from quite a general point of view in Chap. “Regularization Methods for Ill-Posed Problems” in this handbook.
- Variational methods are related to PDE restoration methods and are naturally developed for signals and images defined on a continuous subset  $\Omega \subset \mathbb{R}^d$ ,  $d = 1, 2, \dots$ ; for images  $d = 2$ . Originally, the data fidelity term is of the form (5) for  $A = \text{Id}$  and  $\Phi(u) = \int_{\Omega} \phi(\|Du\|_2) dx$ , where  $\phi$  is a *convex function* as those given in Table 1 (top). Since the beginning of the 1990s, a remarkable effort was done to find heuristics on  $\phi$  that enable *to recover edges and breakpoints in restored images and signals while smoothing the regions between them*; see, e.g., [5, 13, 26, 31, 59, 64, 73, 85, 87]. One of the most successful is the *Total Variation* (TV) regularization corresponding to  $\phi(t) = t$ , which was proposed by Rudin, Osher, and Fatemi in [87]. Variational methods were rapidly applied along with data fidelity terms  $\Psi$ . The use of differential operators  $D^k$  of various orders  $k \geq 2$  in the prior  $\Phi$  has been recently investigated, see, e.g., [22, 23]. More details on variational methods for image processing can be found in several textbooks like [3, 5, 91].

For numerical implementation, the variational functional is discretized and  $\Phi$  takes the form of (6). Different discretization approaches are considered; see e.g., [2, 27, 95]

The *equivalence between these approaches* has been considered in several seminal papers, see, e.g., [37, 63]. The state of the art and the relationship among all these methodologies are nicely outlined in the recent book of Scherzer et al. [91]. 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 frequently data fidelity and priors are modeled separately. It is hence necessary to check if the minimizer  $\hat{u}$  of  $\mathcal{F}(\cdot, v)$  obeys all information contained in the data model  $\Psi$  as well as in the prior  $\Phi$ . Hence the question:

how the prior  $\Phi$  and the data fidelity  $\Psi$  are *effectively* involved in  $\hat{u}$  – a minimizer of  $\mathcal{F}(\cdot, v)$ . This leads to formulate the following *inverse 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.

(7)

This problem was posed in a systematic way and studied since [74, 75]. The *point of view* provided by (7) is actually adopted by many authors. Problem (7) is totally general and involves crucial stakes:

- It 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 fulfill the requirements in applications.
- This kind of results can help to derive numerical schemes using knowledge on the solutions.

Problem (7) remains open. 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  $\mathcal{F}$  which can be convex or nonconvex – or smooth or nonsmooth, and results address local and global minimizers*.

### 1.3 Organization of the Chapter

Some preliminary notions and results that help the reading the chapter are sketched in Sect. 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 nonconvex regularization. In Sect. 5 it is shown that nonsmooth regularization leads typically to minimizers that are sparse in the space spanned by  $\{D_i\}$ . Conversely, Sect. 6 exhibits that the minimizers relevant to nonsmooth data fidelity achieve an exact fit for numerous data samples. Section 7 considers results when both  $\Psi$  and  $\Phi$  are nonsmooth. 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 Notation

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.

- $D_j^n$  – The differential operator of order  $n$  with respect to the  $j$ th component of a function.

- $v[i]$  – The  $i$ th entry of vector  $v$ .
- $\#J$  – 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  $K \subset \mathbb{R}^n$ .
- $A^*$  – The transpose 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)
- $\mathbb{1}_n \in \mathbb{R}^n$  – The  $n$ -length vector composed of ones, i.e.,  $\mathbb{1}_n[i] = 1, 1 \leq i \leq n$ .
- $\mathbb{L}^n$  – The Lebesgue measure on  $\mathbb{R}^n$ .
- $\text{Id}$  – The identity operator.
- $\|\cdot\|_\rho$  – A vector or a matrix  $\rho$ -norm.
- $\mathbb{R}_+ \stackrel{\text{def}}{=} \{t \in \mathbb{R} : t \geq 0\}$  and  $\mathbb{R}_+^* \stackrel{\text{def}}{=} \{t \in \mathbb{R} : t > 0\}$ .
- $\text{TV}$  – Total Variation.
- $\{e_1, \dots, e_n\}$  – 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  $\mathcal{F} : \mathbb{R}^p \rightarrow \mathbb{R}$  is coercive if  $\lim_{\|u\| \rightarrow \infty} \mathcal{F}(u) = +\infty$ .

A special attention being dedicated to nonsmooth functions, we recall some basic facts.

**Definition 2** Given  $v \in \mathbb{R}^q$ , the function  $\mathcal{F}(\cdot, v) : \mathbb{R}^p \rightarrow \mathbb{R}$  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  $\delta_1$  means that derivatives with respect to the first variable of  $\mathcal{F}$  are addressed.

Here  $\delta_1 \mathcal{F}(\hat{u}, v)(w)$  is a *right-side* derivative; the *left-side* derivative is  $-\delta_1 \mathcal{F}(\hat{u}, v)(-w)$ . If  $\mathcal{F}(\cdot, v)$  is differentiable at  $\hat{u}$ , then  $\delta_1 \mathcal{F}(\hat{u}, v)(w) = D_1 \mathcal{F}(\hat{u}, v)w$ . For  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$ , we denote by  $\phi'(t^-)$  and  $\phi'(t^+)$  its left-side and right-side derivatives, respectively.

The classical necessary condition for a local minimum of a (nonsmooth) function is recalled [60, 86]:

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

If  $\mathcal{F}(\cdot, v)$  is Fréchet differentiable at  $\hat{u}$ , one finds  $D_1\mathcal{F}(\hat{u}, v) = 0$ .

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 [60, 86]. Hence  $\mathcal{F}(\cdot, v)$  is differentiable at almost every  $u$ . However, when  $\mathcal{F}(\cdot, v)$  is nondifferentiable, its minimizers are typically located at points where  $\mathcal{F}(\cdot, v)$  is nondifferentiable; see, e.g., Example 1 below.

**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 as

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

Clearly,  $\mathcal{F}(\cdot, v)$  is not Fréchet differentiable only at zero. For any  $|v| \leq \beta$ , the minimizer of  $\mathcal{F}(\cdot, v)$  is located precisely at zero.

The next corollary shows what can happen if the necessary condition in Theorem 1 fails.

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

$$\Theta_0 \stackrel{\text{def}}{=} \{(u, v) \in \mathbb{R}^p \times \mathbb{R}^q : \exists w \in \mathbb{R}^p, -\delta_1\mathcal{F}(u, v)(-w) > \delta_1\mathcal{F}(u, v)(w)\}. \quad (8)$$

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,  $\delta_1\mathcal{F}(\hat{u}, v)(-w) \geq 0$ , hence

$$-\delta_1\mathcal{F}(\hat{u}, v)(-w) \leq 0 \leq \delta_1\mathcal{F}(\hat{u}, v)(w), \quad \forall w \in \mathbb{R}^p. \quad (9)$$

If  $(\hat{u}, v) \in \Theta_0$ , the necessary condition (9) cannot hold.  $\square$

**Example 2** Suppose that  $\Psi$  in (3) is a differentiable function for any  $v \in \mathbb{R}^q$ . For a finite set of positive numbers, say  $\theta_1, \dots, \theta_k$ , suppose that the PF  $\phi$  is differentiable on  $\mathbb{R}_+ \setminus \cup_{j=1}^k \theta_j$  and that

$$\phi'(\theta_j^-) > \phi'(\theta_j^+), \quad 1 \leq j \leq k. \quad (10)$$

Given a (local) minimizer  $\hat{u}$ , denote

$$I = \{1, \dots, r\} \quad \text{and} \quad I_{\hat{u}} = \{i \in I : \|D_i\hat{u}\|_2 = \theta_j, 1 \leq j \leq k\}.$$

Define  $F(\hat{u}, v) = \Psi(\hat{u}, v) + \beta \sum_{i \in I \setminus I_{\hat{u}}} \phi(\|D_i\hat{u}\|_2)$ , which is differentiable at  $\hat{u}$ . Clearly,  $\mathcal{F}(\hat{u}, v) = F(\hat{u}, v) + \beta \sum_{i \in I_{\hat{u}}} \phi(\|D_i\hat{u}\|_2)$ . Applying the necessary condition (9) for  $w = \hat{u}$  yields

$$\beta \sum_{i \in I_{\hat{u}}} \phi'(\|D_i\hat{u}\|_2^-) \leq -D_1F(\hat{u}, v)(\hat{u}) \leq \beta \sum_{i \in I_{\hat{u}}} \phi'(\|D_i\hat{u}\|_2^+).$$

In particular, one has  $\sum_{i \in I_{\hat{u}}} \phi'(\|D_i \hat{u}\|_2^-) \leq \sum_{i \in I_{\hat{u}}} \phi'(\|D_i \hat{u}\|_2^+)$ , which contradicts the assumption on  $\phi'$  in (10). It follows that if  $\hat{u}$  is a (local) minimizer of  $\mathcal{F}(\cdot, v)$ , then  $I_{\hat{u}} = \emptyset$  and

$$\|D_i \hat{u}\|_2 \neq \theta_j, \quad 1 \leq j \leq k, \quad \forall i \in I.$$

A typical case is the PF (f6) in Table 1, namely  $\phi(t) = \min\{\alpha t^2, 1\}$ . Then  $k = 1$  and  $\theta_1 = \frac{1}{\sqrt{\alpha}}$ .

The following existence theorem can be found, e.g., in the textbook [35].

**Theorem 2** For  $v \in \mathbb{R}^q$ , let  $U \subset \mathbb{R}^p$  be a nonempty and closed subset and  $\mathcal{F}(\cdot, v) : U \rightarrow \mathbb{R}$  a lower semicontinuous (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  $\mathcal{F}(\hat{u}, v) = \inf_{u \in U} \mathcal{F}(u, v)$ .

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  $\mathcal{F} : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$  involve (f6) in Table 1 and read

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

For any  $v$ ,  $\mathcal{F}(\cdot, v)$  is not coercive since it is bounded by  $\beta$  in the direction spanned by  $\{(0, u[2])\}$ . However, its global minimum is strict and is reached for  $\hat{u}[1] = \hat{u}[2] = v[1]$  with  $\mathcal{F}(\hat{u}, v) = 0$ .

To prove the existence of optimal solutions for more general energies, we refer to the textbook [8].

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 3** Let  $\mathcal{F} : \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}$  and  $O \subseteq \mathbb{R}^q$ . We say that  $\mathcal{U} : O \rightarrow \mathbb{R}^p$  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 [35, 49].

**Theorem 3** Let  $\mathcal{F}(\cdot, v) : \mathbb{R}^p \rightarrow \mathbb{R}$  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 closed convex set of minimizers  $\{\hat{\mathcal{U}}(v)\} \stackrel{\text{def}}{=} \left\{ \hat{u} \in \mathbb{R}^p : \mathcal{F}(\hat{u}, v) = \inf_{u \in U} \mathcal{F}(u, v) \right\}$ .
- (ii) If in addition  $\mathcal{F}(\cdot, v)$  is strictly convex, then there is a unique minimizer  $\hat{u} = \mathcal{U}(v)$  (which is also global). So  $\mathcal{F}(\mathbb{R}^p, v)$  has a unique minimizer function  $v \mapsto \mathcal{U}(v)$ .

The next lemma, which can be found, e.g., in [52], 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  $\mathcal{C}^m$ ,  $m \geq 2$ , 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  $\mathcal{U} : O \rightarrow \mathbb{R}^p$ , such that  $D_1^2 \mathcal{F}(\mathcal{U}(v), v) \succ 0$  for every  $v \in O$  and  $\mathcal{U}(v) = \hat{u}$ .

This lemma is extended in several directions in this chapter.

**Definition 4** Let  $\phi : [0, +\infty) \rightarrow \mathbb{R}$  and  $m \geq 0$  an integer. We say that  $\phi$  is  $\mathcal{C}^m$  on  $\mathbb{R}_+$ , or equivalently that  $\phi \in \mathcal{C}^m(\mathbb{R}_+)$ , if and only if  $t \mapsto \phi(|t|)$  is  $\mathcal{C}^m$  on  $\mathbb{R}$ .

By this definition,  $\phi'(0) = 0$ . In Table 1 – left,  $\phi \in \mathcal{C}^1(\mathbb{R}_+)$  for (f1) if  $\alpha < 2$ ,  $\phi \in \mathcal{C}^2(\mathbb{R}_+)$  for (f4), while for (f2), (f3) and (f7)-(f9) we find  $\phi \in \mathcal{C}^\infty(\mathbb{R}_+)$ .

### 3 Regularity Results

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

$$\begin{aligned} \mathcal{F}(u, v) &= \|Au - v\|_2^2 + \beta \sum_{i \in I} \phi(\|D_i u\|_2), \\ I &\stackrel{\text{def}}{=} \{1, \dots, r\}, \end{aligned} \tag{11}$$

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

$$D \stackrel{\text{def}}{=} \begin{bmatrix} D_1 \\ \dots \\ D_r \end{bmatrix}.$$

When  $A$  in (11) is not injective, a standard assumption in order to have regularization is

**H 3**  $\ker(A) \cap \ker(D) = \{0\}$ .

H3 is trivial if  $\text{rank } A = p$  or  $\text{rank } D = p$ . Often,  $\ker(D) = \text{span}(\mathbb{1}_p)$  and  $A\mathbb{1}_p \neq 0$ , so H3 holds.

#### 3.1 Some General Results

We first verify the conditions on  $\mathcal{F}(\cdot, v)$  in (11) that enable Theorems 2 and 3 to be applied. Since H1 holds,  $\mathcal{F}(\cdot, v)$  in (11) is l.s.c. and proper.

1.  $\mathcal{F}(\cdot, v)$  in (11) is *coercive* for any  $v \in \mathbb{R}^q$  at least in *one* of the following cases:
  - $\text{Rank}(A) = p$  and  $\phi : \mathbb{R}_+ \mapsto \mathbb{R}_+$  is nondecreasing.
  - H1 and H3 hold and  $\lim_{t \nearrow \infty} \phi(t) = \infty$  (e.g., (f1)-(f5), (f8), (f10), and (f12) in Table 1).

By Theorem 2,  $\mathcal{F}(\cdot, v)$  has minimizers.

2. For any  $v \in \mathbb{R}^q$ , the energy  $\mathcal{F}(\cdot, v)$  in (11) is *convex and coercive* if H1 and H3 hold for a *convex*  $\phi$ . Then the claim in Theorem 3(i) holds true.

3. Further,  $\mathcal{F}(\cdot, v)$  in (11) is *strictly convex and coercive* for any  $v \in \mathbb{R}^q$  if  $\phi$  satisfies H1 and if *one* of the following assumptions holds:

- Rank( $A$ ) =  $p$  and  $\phi$  is convex.
- H3 holds and  $\phi$  is strictly convex.

Then the claim in Theorem 3(ii) holds. Further, if  $\mathcal{F}$  is  $\mathcal{C}^m$  for  $m \geq 2$ , then the minimizer function  $\mathcal{U} : \mathbb{R}^q \rightarrow \mathbb{R}^p$  (see Definition 3) is  $\mathcal{C}^{m-1}$  by Lemma 1.

However, the PFs involved in (11) used for signal and image processing are often nonconvex, bounded or nondifferentiable. One extension of the standard results is given in the next Sect. 3.2.

### 3.2 Stability of the Minimizers of Energies with Possibly Nonconvex Priors

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 [42]. Other results have been established on the stability of the local minimizers of general *smooth* energies [52]. Typically, these results are quite abstract to be applied directly to energies of the form (11).

Here the assumptions stated below are considered.

**H 4** The operator  $A$  in (11) satisfies rank  $A = p$ , i.e.,  $A^*A$  is invertible.

**H 5** The PF  $\phi$  in (11) is  $\mathcal{C}^0(\mathbb{R}_+)$  and  $\mathcal{C}^m$ ,  $m \geq 2$ , on  $\mathbb{R}_+^*$  with  $0 \leq \phi'(0^+) < \infty$ .

Under H1, H2, H4, and H5, the prior  $\Phi$  (and hence  $\mathcal{F}(\cdot, v)$ ) in (11) can be *nonconvex and in addition nonsmooth*. By H1 and H4,  $\mathcal{F}(\cdot, v)$  in (11) admits a global minimum  $v \in \mathbb{R}^q$  – see item 1 in Sec. 3.1. However,  $\mathcal{F}(\cdot, v)$  can present *numerous local minima*.

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

The results summarized in this Sec. 3.2 provide the state of the art for energies of the form (11).

#### 3.2.1 Local Minimizers

The stability of local minimizers is an important matter in its own right for several reasons. Often, a nonconvex energy is minimized only locally, in the vicinity of some initial guess. Second, the 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.

The statements below are a simplified version of the results established in [44].

**Theorem 4** Let  $\mathcal{F}(\cdot, v)$  in (11) 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 3)  $\mathcal{U} : O \rightarrow \mathbb{R}^p$  which is  $\mathcal{C}^{m-1}$  on  $O$  and fulfills  $\hat{u} = \mathcal{U}(v)$ .

Since  $\Theta$  is closed in  $\mathbb{R}^q$  and  $\mathbb{L}^q(\Theta) = 0$ , the stated properties are *generic*.

**Commentary on the assumptions.** All assumptions H1, H2, and H5 bearing on the PF  $\phi$  are nonrestrictive; they address all PFs in Table 1 except for (f13) which is discontinuous at zero. The assumption H4 cannot be avoided, as seen in Example 4.

**Example 4** Consider  $\mathcal{F} : \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}$  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.

$$\begin{aligned} v > \frac{1}{2} & \quad \hat{u} = \left(c, c - v + \frac{1}{2}\right) \quad \text{for any } c \in \left[0, v - \frac{1}{2}\right] \quad (\text{nonstrict minimizer}). \\ |v| \leq \frac{1}{2} & \quad \hat{u} = 0 \quad (\text{unique minimizer}) \\ v < -\frac{1}{2} & \quad \hat{u} = \left(c, c - v - \frac{1}{2}\right) \quad \text{for any } c \in \left[v + \frac{1}{2}, 0\right] \quad (\text{nonstrict minimizer}). \end{aligned}$$

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

**Other results.** The derivations in [44] reveal several other practical results.

1. If  $\phi \in \mathcal{C}^2(\mathbb{R}_+)$ , see Definition 4, then  $\forall v \in \mathbb{R}^q \setminus \Theta$ , every local minimizer  $\hat{u}$  of  $\mathcal{F}(u, v)$  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 is a generic property 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, the statement of Theorem 4 holds true if  $\phi'(0^+) = 0$  and if there is  $\tau > 0$  such that  $\phi'(\tau^-) > \phi'(\tau^+)$ . This is the case of the PF (f6) in Table 1.
3. If  $\phi'(0^+) > 0$ , define

$$\hat{J} \stackrel{\text{def}}{=} \{i \in I : D_i \hat{u} = 0\} \quad \text{and} \quad K_{\hat{J}} \stackrel{\text{def}}{=} \left\{w \in \mathbb{R}^p : D_i w = 0, \forall i \in \hat{J}\right\}. \quad (12)$$

Then  $\forall v \in \mathbb{R}^q \setminus \Theta$ , every local minimizer  $\hat{u}$  of  $\mathcal{F}(u, v)$  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  $K_{\hat{J}}^\perp$ .

Here (a) and (b) provide a sufficient condition for a strict (local) minimum of  $\mathcal{F}(\cdot, v)$  at  $\hat{u}$  (a direct consequence of [80, Theorem 1]). These conditions are satisfied at the (local) minimizers  $\hat{u}$  of  $\mathcal{F}(\cdot, v)$  for every  $v \in \mathbb{R}^q$ , except for a negligible subset of  $\mathbb{R}^q$ , in which case Lemma 1 can be applied.

One can interpret these results as 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 (11) that does not result from a  $\mathcal{C}^{m-1}$  local minimizer function, is null.

### 3.2.2 Global Minimizers of Energies with for Possibly Nonconvex Priors

The results on the global minimizers of (11) presented next are extracted from [45].

**Theorem 5** *Assume that  $\mathcal{F}(\cdot, v)$  in (11) 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  $\hat{\mathcal{U}} : \mathbb{R}^q \setminus \hat{\Theta} \rightarrow \mathbb{R}^p$  is  $\mathcal{C}^{m-1}$  on an open subset of  $\mathbb{R}^q \setminus \hat{\Theta}$  which is dense in  $\mathbb{R}^q$ .*

Otherwise said, in a real-world problem there is no chance of getting data  $v$  such that the energy  $\mathcal{F}(\cdot, v)$  (11) has more than one global minimizer.

Nonetheless,  $\hat{\Theta}$  plays a crucial role for the recovery of edges; this issue is developed in Sect. 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 (11) that hold for all PFs  $\phi$  in Table 1. Related questions have mainly been considered in particular cases or asymptotically; see, e.g., [4, 71, 92]. In [51] the mean and the variance of the minimizers  $\hat{u}$  for strictly convex and differentiable functions  $\phi$  have been explored.

The bounds provided below are of practical interest for the initialization and the convergence analysis of numerical schemes. The statements given below are extracted from [82].

**Bounds on the restored data.** One compares the “restored” data  $A\hat{u}$  with the given data  $v$ .

**H 6** *Consider the alternative assumptions:*

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

The set  $\Theta_0$  allows us to address the PF given in (f6). Let us emphasize that under H1 and H6 the PF  $\phi$  can be convex or nonconvex.

**Theorem 6** *Consider  $\mathcal{F}(\cdot, v)$  of the form (11) where H1, H3, and H6 hold. For every  $v \in \mathbb{R}^q$ , if  $\mathcal{F}(\cdot, v)$  has a (local) minimum at  $\hat{u}$ , then*

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

**Comments on the results.** This bound holds for every (local) minimizer of  $\mathcal{F}(\cdot, v)$ . If  $A$  is a uniform tight frame (i.e.,  $A^*A = \text{Id}$ ), one has

$$\|\hat{u}\|_2 \leq \|v\|_2.$$

**The mean of restored data.** In many applications, the noise corrupting the data can be supposed to have a mean equal to zero. When  $A = \text{Id}$ , it is well known that  $\text{mean}(\hat{u}) = \text{mean}(v)$ , see, e.g., [5]. However, for a general  $A$  one has

$$A\mathbb{1}_p \propto \mathbb{1}_q \quad \Rightarrow \quad \text{mean}(\hat{u}) = \text{mean}(v). \quad (13)$$

The requirement  $A\mathbb{1}_p \propto \mathbb{1}_q$  is quite restrictive. In the simple case when  $\phi(t) = t^2$ ,  $\ker(D) = \mathbb{1}_{rs}$  and  $A$  is square and invertible, it is easy to see that this is also a *sufficient* condition. Finally, if  $A \neq \text{Id}$ , then generally  $\text{mean}(\hat{u}) \neq \text{mean}(v)$ .

**The residuals for edge preserving regularization.** A bound on the data fidelity term at a (local) minimizer  $\hat{u}$  of  $\mathcal{F}(\cdot, v)$  shall be given. The edge preserving H2 (see Sect. 1) is replaced by a stronger edge preserving assumption:

$$\mathbf{H7} \quad \|\phi'\|_\infty \stackrel{\text{def}}{=} \max \left\{ \sup_{t \geq 0} |\phi'(t^+)|, \sup_{t > 0} |\phi'(t^-)| \right\} < \infty.$$

Except for (f1) and (f13), all other PFs in Table 1 satisfy H7. Note that when  $\phi'(0^+) > 0$  and H7 hold, one usually has  $\|\phi'\|_\infty = \phi'(0^+)$ .

**Theorem 7** *Let  $\mathcal{F}(\cdot, v)$  be of the form (11) where  $\text{rank}(A) = q \leq p$ , and H1, H3, H6, and H7 hold. For every  $v \in \mathbb{R}^q$ , if  $\mathcal{F}(\cdot, v)$  has a (local) minimum at  $\hat{u}$ , then*

$$\|A\hat{u} - v\|_\infty \leq \frac{\beta}{2} \|\phi'\|_\infty \|(AA^*)^{-1}A\|_\infty \|D\|_1. \quad (14)$$

Let us emphasize that the bound in (14) is *independent of data  $v$*  and that it is satisfied for *any local or global minimizer  $\hat{u}$*  of  $\mathcal{F}(\cdot, v)$ . (Recall that for a real matrix  $C$  with entries  $C[i, j]$ , one has  $\|C\|_1 = \max_j \sum_i |C[i, j]|$  and  $\|C\|_\infty = \max_i \sum_j |C[i, j]|$ ; see, e.g., [35].)

If  $D$  corresponds to a discrete gradient operator for a two-dimensional image,  $\|D\|_1 = 4$ . If in addition  $A = \text{Id}$ , (14) yields

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

The result of this theorem may seem surprising. In a statistical setting, the quadratic data fidelity term  $\|Au - v\|_2^2$  in (11) corresponds to white Gaussian noise on the data, *which is unbounded*. However, if  $\phi$  is edge preserving according to H7, any (local) minimizer  $\hat{u}$  of  $\mathcal{F}(\cdot, v)$  gives rise to a *noise estimate*  $(v - A\hat{u})[i]$ ,  $1 \leq i \leq q$  that is *tightly bounded* as stated in (14).

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

When  $\mathcal{F}(\cdot, v)$  is convex and coercive, (14) shows that a good initialization for a minimization algorithm should be a point  $u_0$  such that  $Au_0 = v$ , e.g., the minimum norm solution of  $\|v - \hat{u}\|_2$  given by  $u_0 = A^*(AA^*)^{-1}v$ .

## 4 Nonconvex 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 and Geman [56], various *nonconvex*  $\Phi$  in (3) have been proposed [15, 54, 55, 64, 68, 72, 85]. Indeed, the relevant minimizers exhibit neat edges between homogeneous regions. However, these nonconvex energies are tiresome to control and to minimize (only few algorithms are proved to find the global minimizer in particular cases). In order to avoid these numerical intricacies, since the 1990s, an important effort was done to derive *convex* edge preserving PFs, see, e.g., [20, 31, 57, 64, 87] and [5] for an overview. The most popular convex edge preserving PF was derived by Rudin, Osher, and Fatemi [87]: it amounts to  $\phi = t$ , for  $\{D_i\}$  yielding the discrete gradient operator, the  $\ell_2$ -norm in (6) (see Sec. 1) and the relevant  $\Phi$  is called *the Total Variation (TV) regularization*.

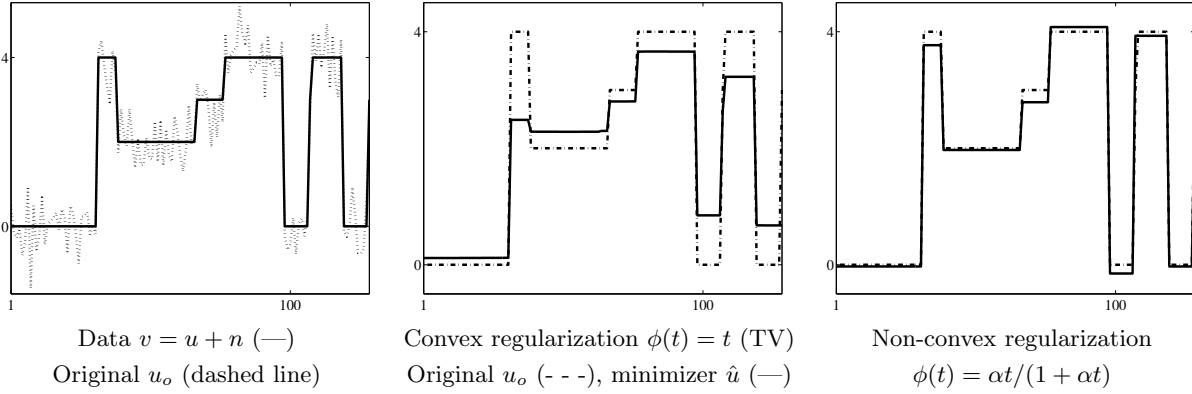


Figure 2: Minimizers of  $\mathcal{F}(u, v) = \|u - v\|_2^2 + \beta \sum_{i=1}^{p-1} \phi(|u[i] - u[i+1]|)$ .

In Fig. 2 one sees that the height of the edges is better recovered when  $\phi$  is nonconvex, compared to the convex TV regularization. The same effect can also be observed, e.g., in Figs. 7, 8 and 10.

A considerable progress in nonconvex minimization has been realized. For energies of the form (2)-(3) we refer to [9, 19, 88, 89].

This section is devoted to explain why edges are nicely recovered using a nonconvex  $\phi$ .

### 4.2 Assumptions on Potential Functions $\phi$

Consider  $\mathcal{F}(\cdot, v)$  of the form (11) where  $D_i : \mathbb{R}^p \rightarrow \mathbb{R}^1$ ,  $i \in I = \{1, \dots, r\}$ , i.e.,

$$\mathcal{F}(u, v) = \|Au - v\|_2^2 + \beta \sum_{i \in I} \phi(|D_i u|), \quad (15)$$

and  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  satisfies H1 (see Sect. 1), H6 Sect. 3.3, and H8 given below

**H 8**  $\phi$  is  $\mathcal{C}^2$  and  $\phi'(t) \geq 0$  on  $\mathbb{R}_+^*$ , and  $\inf_{t \in \mathbb{R}_+^*} \phi''(t) < 0$  with  $\lim_{t \rightarrow \infty} \phi''(t) = 0$ ;

as well as *one* of the following assumptions:

**H 9**  $\phi'(0^+) = 0$ , and there are two numbers  $\tau > 0$  and  $\mathcal{T} \in (\tau, \infty)$  such that  $\phi''(t) \geq 0$  on  $[0, \tau]$ ,  $\phi''(t) < 0$  on  $(\tau, \infty)$ ,  $\phi''(t)$  decreases on  $(\tau, \mathcal{T})$  and increases on  $(\mathcal{T}, \infty)$ .

**H 10**  $\phi'(0^+) > 0$ , and  $\lim_{t \rightarrow 0} \phi''(t) < 0$  is well defined and  $\phi''(t) < 0$  strictly increases on  $(0, \infty)$ .

These assumptions are illustrated in Fig. 3. They hold for all nonconvex PFs in Table 1, except for (f6) and (f13) which are presented separately. Further, these assumptions are easy to relax.

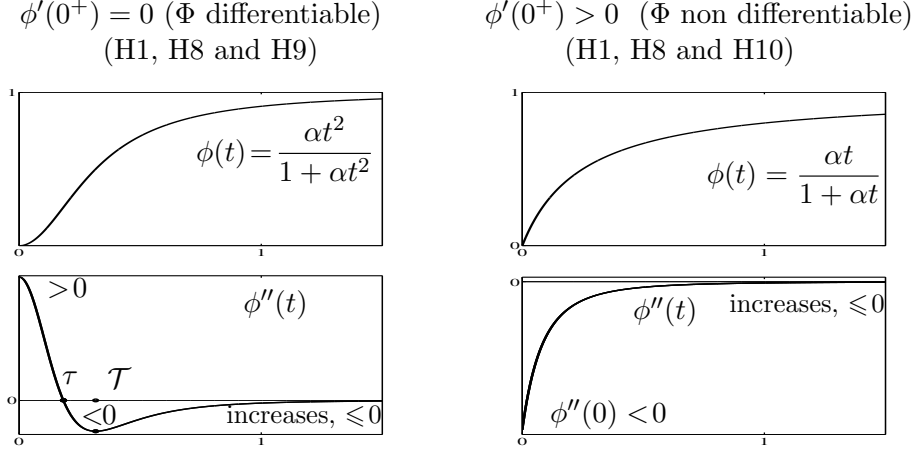


Figure 3: Illustration of the assumptions in two typical cases – (f7) and (f11) in Table 1.

The results presented below come from [81].

### 4.3 How it Works on $\mathbb{R}$

This example illustrates the main facts that explain why edges are enhanced when  $\phi$  is non-convex, satisfying H1, and H8 along with either H9 or H10.

Let  $\mathcal{F} : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  be given by

$$\mathcal{F}(u, v) = \frac{1}{2}(u - v)^2 + \beta\phi(u) \text{ for } \begin{cases} \beta > \frac{1}{|\phi''(\mathcal{T})|} & \text{if } \phi'(0^+) = 0 \text{ (H1, H8 and H9)} \\ \beta > \frac{1}{|\lim_{t \searrow 0} \phi''(t)|} & \text{if } \phi'(0^+) > 0 \text{ (H1, H8 and H10)} \end{cases}$$

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

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

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

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

One has  $\theta_0 = 0$  if  $\phi'(0^+) > 0$  and  $0 < \theta_0 < \mathcal{T} < \theta_1$  if  $\phi'(0^+) = 0$ . A few calculations yield

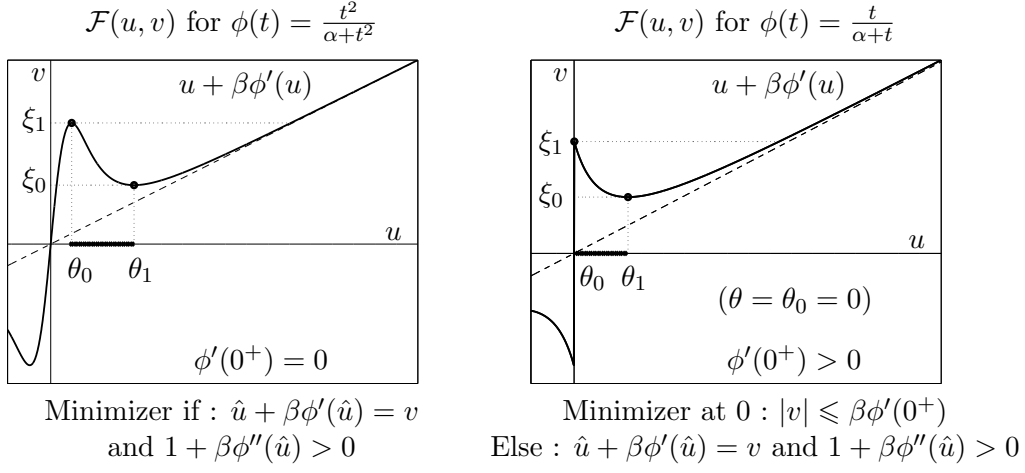


Figure 4: The curve of  $u \mapsto (D_1\mathcal{F}(u, v) - v)$  on  $\mathbb{R} \setminus \{0\}$ . All assumptions mentioned before hold.

1. For every  $v \in \mathbb{R}_+$  no minimizer lives in  $(\theta_0, \theta_1)$  (cf. Fig. 4).
2. One computes  $0 < \xi_0 < \xi_1$  such that (cf. Fig. 4)
  - (a) If  $0 \leq v \leq \xi_1$ ,  $\mathcal{F}(\cdot, v)$  has a (local) minimizer  $\hat{u}_0 \in [0, \theta_0]$ , hence  $\hat{u}_0$  is subject to a strong smoothing.
  - (b) If  $v \geq \xi_0$ ,  $\mathcal{F}(\cdot, v)$  has a (local) minimizer  $\hat{u}_1 \geq \theta_1$ , 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. 5;
  - (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  $\xi$  and  $\mathcal{C}^1$ -smooth on  $\mathbb{R}_+ \setminus \{\xi\}$ .

Item 1 is the key for the recovery of either homogeneous regions or 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 Sect. 3.2.2.

#### 4.4 Either Smoothing or Edge Enhancement

**(A) Case  $\phi'(0^+) = 0$ .** Below the case depicted in Figs. 4–left and 5–left is extended to  $\mathbb{R}^p$ .

**Theorem 8** *Let  $\mathcal{F}(\cdot, v)$  be of the form (15) where H1, H3, H8 and H9 hold, and  $\{D_i : i \in I\}$  are linearly independent. Set  $\mu \stackrel{\text{def}}{=} \max_{1 \leq i \leq r} \|D^*(DD^*)^{-1}e_i\|_2$ . For  $\beta > \frac{2\mu^2 \|A^*A\|_2}{|\phi''(\mathcal{T})|}$ , 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*

$$\text{either } |D_i \hat{u}| \leq \theta_0, \quad \text{or } |D_i \hat{u}| \geq \theta_1, \quad \forall i \in I. \quad (16)$$



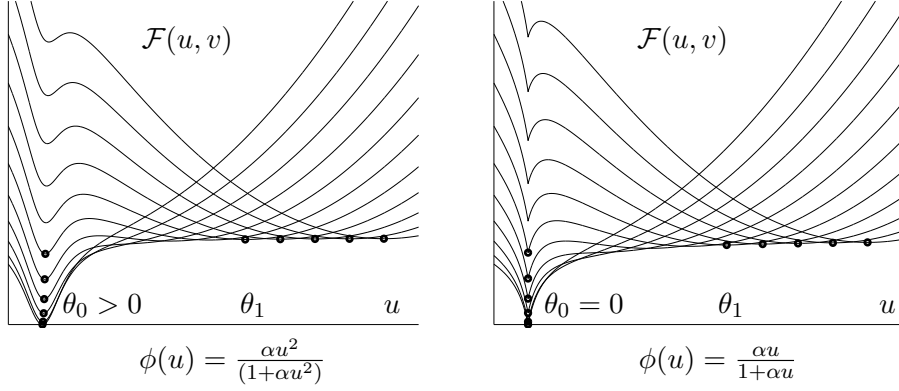


Figure 5: Each curve represents  $\mathcal{F}(u, v) = \frac{1}{2}(u - v)^2 + \beta\phi(u)$  for an increasing sequence  $v \in [0, \xi_1]$ . The global minimizer of each  $\mathcal{F}(\cdot, v)$  is marked with “•”. No (local) minimizer lives in  $(\theta_0, \theta_1)$ .

In many imaging problems,  $\{D_i\}$  are not linearly independent. If  $\{D_i\}$  are linearly dependent, the result (16) holds true for all (local) minimizers  $\hat{u}$  that are locally homogeneous on regions that are connected with respect to  $\{D_i\}$ . Otherwise, one recovers both high edges and smooth transitions, as seen in Fig. 8a. When  $\phi$  is convex, all edges are smoothed, as one can observe in Fig. 7a.

The PF  $\phi(t) = \min\{\alpha t^2, 1\}$  (f6) in Table 1, does not satisfy assumptions H8 and H9. From Corollary 1 and Example 2 (Sec. 2.2), any (local) minimizer  $\hat{u}$  of  $\mathcal{F}(\cdot, v)$  obeys

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

Propositions 1 below addresses only the *global minimizers* of  $\mathcal{F}(\cdot, v)$ .

**Proposition 1** *Let  $\mathcal{F}(\cdot, v)$  be given by (15) where  $\phi(t) = \min\{\alpha t^2, 1\}$ ,  $\{D : i \in I\}$  are linearly independent and  $\text{rank}(A) \geq p - r \geq 1$ . If  $\mathcal{F}(\cdot, v)$  has a global minimizer at  $\hat{u}$ , then*

$$\text{either } |D_i \hat{u}| \leq \frac{1}{\sqrt{\alpha}} \Gamma_i, \quad \text{or } |D_i \hat{u}| \geq \frac{1}{\sqrt{\alpha}} \Gamma_i, \quad \text{for } \Gamma_i = \sqrt{\frac{\|Be_i\|_2^2}{\|Be_i\|_2^2 + \alpha\beta}} < 1, \quad \forall i \in I, \quad (17)$$

where  $B$  is a matrix depending only on  $A$  and  $D$ .

If  $D = \text{Id}$ , then  $B = A$ . If  $u$  one-dimensional signal and  $D_i u = u[i] - u[i + 1]$ ,  $1 \leq i \leq p - 1$ , one has  $B = (\text{Id} - \frac{1}{p} \mathbb{1}\mathbb{1}^T)AH$  where  $H \in \mathbb{R}^{p \times p}$  is upper triangular composed of ones.

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

Let us define the following subsets:

$$\hat{J}_0 \stackrel{\text{def}}{=} \{i \in I : |D_i \hat{u}| \leq \theta_0\} \quad \text{and} \quad \hat{J}_1 \stackrel{\text{def}}{=} I \setminus \hat{J}_0 = \{i \in I : |D_i \hat{u}| \geq \theta_1\}. \quad (18)$$

One can interpret the results of Theorem 8 and Propositions 1 as follows:

The pixels in  $\hat{J}_0$  form homogeneous regions with respect to  $\{D_i\}$ , whereas the pixels in  $\hat{J}_1$  are break points.

In particular, if  $\{D_i\}$  correspond to first-order differences,  $\hat{J}_0$  addresses smoothly varying regions where  $|D_i \hat{u}| \leq \theta_0$ , while  $\hat{J}_1$  corresponds to edges higher than  $\theta_1 - \theta_0$ .

**(B) Case  $\phi'(0^+) > 0$ .** Here the results are stronger without assumptions on  $\{D_i\}$ . This case corresponds to Figs. 4–right and 5–right.

**Theorem 9** Consider  $\mathcal{F}(\cdot, v)$  of the form (15) where H3 holds and  $\phi$  satisfies H1, H8 and H10. Let  $\beta > \frac{2\mu^2 \|A^*A\|_2}{|\lim_{t \searrow 0} \phi''(t)|}$ , where  $\mu > 0$  is a constant depending only on  $\{D_i\}$ . Then  $\exists \theta_1 > 0$  such that  $\forall v \in \mathbb{R}^q$ , every (local) minimizer  $\hat{u}$  of  $\mathcal{F}(\cdot, v)$  satisfies

$$\text{either } |D_i \hat{u}| = 0, \quad \text{or } |D_i \hat{u}| \geq \theta_1, \quad \forall i \in I. \quad (19)$$

The results of Theorem 9 were extended to energies involving box constraints in [33].

The “0-1” PF (f13) in Table 1 does not satisfy H8 and H10 since it is discontinuous at 0.

**Proposition 2** Let  $\mathcal{F}(\cdot, v)$  in (15) be defined for  $\phi(0) = 0$ ,  $\phi(t) = 1$  if  $t > 0$ , i.e., (f13),  $\{D_i\}$  be linearly independent and  $\text{rank } A \geq p - r \geq 1$ . If  $\mathcal{F}(\cdot, v)$  has a global minimum at  $\hat{u}$ , then

$$\text{either } |D_i \hat{u}| = 0 \quad \text{or} \quad |D_i \hat{u}| \geq \frac{\sqrt{\beta}}{\|Be_i\|_2}, \quad \forall i \in I, \quad (20)$$

where the matrix  $B$  depends only on  $D$  and on  $A$ .

In (20),  $B$  is the same as in Proposition 1. For  $\theta_1 \stackrel{\text{def}}{=} \min_{i \in I} \frac{\sqrt{\beta}}{\|Be_i\|}$ , it is clear that (20) holds. Let

$$\hat{J}_0 \stackrel{\text{def}}{=} \{i : |D_i \hat{u}| = 0\} \text{ and } \hat{J}_1 \stackrel{\text{def}}{=} I \setminus \hat{J}_0 = \{i : |D_i \hat{u}| \geq \theta_1\}.$$

Using this notation, the results of Theorem 9 and Proposition 2 show that:

The indexes in  $\hat{J}_0$  address regions in  $\hat{u}$  that can be called *strongly homogeneous* (since  $|D_i \hat{u}| = 0$ ) while  $\hat{J}_1$  addresses breakpoints where  $|D_i \hat{u}| \geq \theta_1$ .

If  $\{D_i\}$  are first-order differences,  $\hat{u}$  is *neatly segmented*:  $\hat{J}_0$  corresponds to *constant* regions while  $\hat{J}_1$  describes all edges and they are higher than  $\theta_1$ .

Direct segmentation of an image from data transformed via a general (nondiagonal) operator  $A$  remains a difficult task using standard methods. The result in (19), Theorem 9, tells us that such a segmentation is naturally involved in the minimizers  $\hat{u}$  of  $\mathcal{F}(\cdot, v)$ , for *any* operator  $A$ . This effect can be observed, e.g., on Figs. 8b,d and 11d.

**(C) Illustration: Deblurring of an image from noisy data.** The original image  $u_o$  in Fig. 6(a) presents smoothly varying regions, constant regions and sharp edges. Data in Fig. 6(b) correspond to  $v = a * u_o + n$ , where  $a$  is a blur with entries  $a_{i,j} = \exp(-(i^2 + j^2)/12.5)$  for  $-4 \leq i, j \leq 4$ , and  $n$  is white Gaussian noise yielding 20 dB of SNR. The amplitudes of the original image are in the range of  $[0, 1.32]$  and those of the data in  $[-5, 50]$ . In all restored images,  $\{D_i\}$  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.

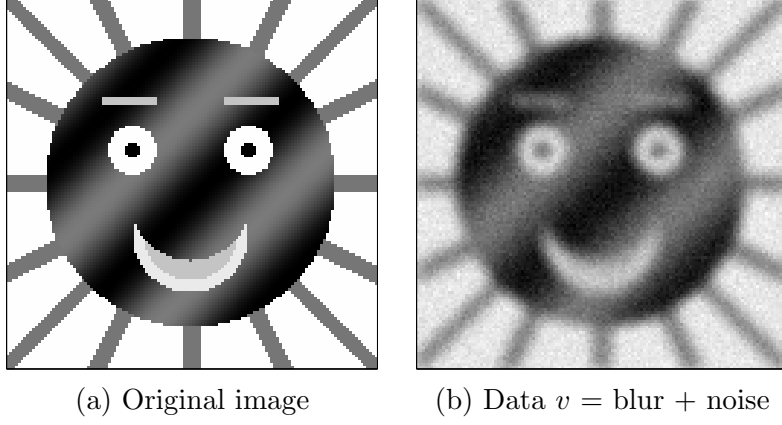


Figure 6: Data  $v = a \star u_o + n$ , where  $a$  is a blur and  $n$  is white Gaussian noise, 20 dB of SNR.

The restorations in Fig. 7 are obtained using *convex* PFs  $\phi$  while those in Fig. 8 using *nonconvex* PFs  $\phi$ . Edges are sharp and high in Fig. 8 where  $\phi$  is nonconvex, which corroborates the results in paragraphs (A) and (B). In Fig. 8(b) and (d)  $\phi$  is nonconvex and  $\phi'(0^+) > 0$  in addition. As stated in Theorem 9, in spite of the fact that  $A$  is nondiagonal (and ill-conditioned), the restored images are fully segmented and the edges between constant pieces are high.

## 5 Nonsmooth Regularization

Observe that the minimizers corresponding to  $\phi'(0^+) > 0$  (nonsmooth regularization) in Figs. 2 (b)-(c), 7(b), and 8(b) and (d), 10(a), (b) and (c), 11(d), are constant on numerous regions. This section is aimed to explain and to generalize this observation.

Consider

$$\mathcal{F}(u, v) = \Psi(u, v) + \beta\Phi(u) \quad (21)$$

$$\Phi(u) = \sum_{i=1}^r \phi(\|D_i u\|_2), \quad (22)$$

where  $\Psi : \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}$  is any explicit or implicit  $\mathcal{C}^m$ -smooth function for  $m \geq 2$  and  $D_i : \mathbb{R}^p \mapsto \mathbb{R}^s$ ,  $\forall i \in I = \{1, \dots, r\}$ , are general linear operators for any integer  $s \geq 1$ . It is assumed that  $\phi$  satisfies H1 along with

**H 11**  $\phi$  is  $\mathcal{C}^2$ -smooth on  $\mathbb{R}_+^*$  and  $\phi'(0^+) > 0$ .

Note that  $\Psi$  and  $\phi$  can be convex or nonconvex. Let us define the set-valued function  $\mathcal{J}$  on  $\mathbb{R}^p$  by

$$\mathcal{J}(u) = \left\{ i \in I : \|D_i u\|_2 = 0 \right\}. \quad (23)$$

Given  $u \in \mathbb{R}^p$ ,  $\mathcal{J}(u)$  indicates all regions where  $D_i u = 0$ . Such regions are called *strongly homogeneous* with respect to  $\{D_i\}$ . (The adverb “strongly” is used to emphasize the difference with just “homogeneous regions” where  $\|D_i u\|_2 \approx 0$ .) In particular, if  $\{D_i\}$  correspond to first-order differences between neighboring samples of  $u$  or to discrete gradients,  $\mathcal{J}(u)$  indicates all constant regions in  $u$ .

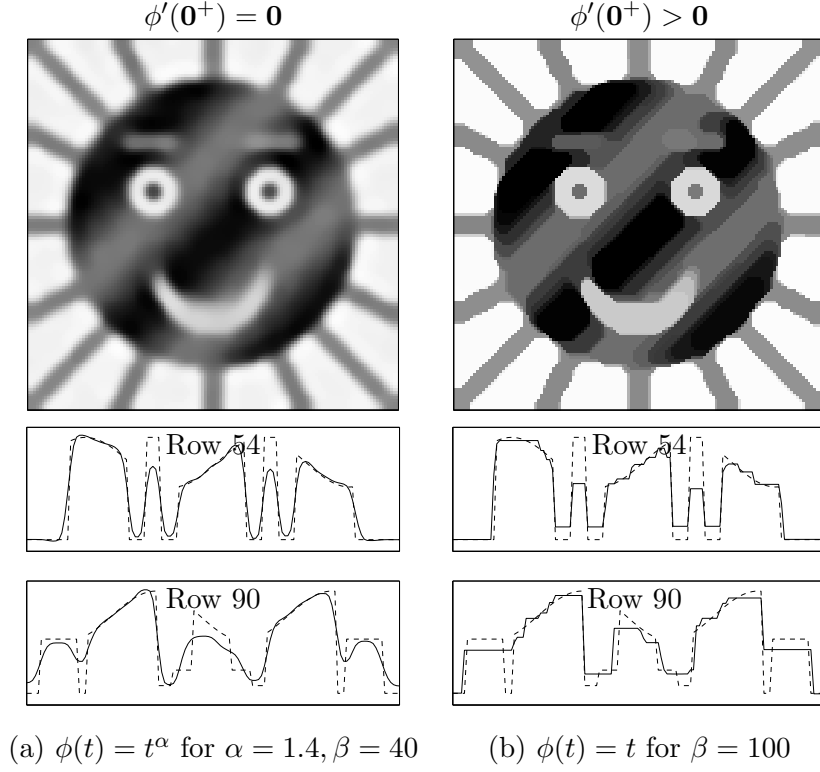


Figure 7: Restoration using convex PFs.

## 5.1 Main Theoretical Result

The results presented below are extracted from [80].

**Theorem 10** *Given  $v \in \mathbb{R}^q$ , assume that  $\mathcal{F}(\cdot, v)$  in (21)–(22) is such that  $\Psi$  is  $\mathcal{C}^m$ ,  $m \geq 2$  on  $\mathbb{R}^p \times \mathbb{R}^q$ , and that  $\phi$  satisfies H1, and H11. 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*

$$K_{\hat{J}} = \left\{ u \in \mathbb{R}^p : D_i u = 0, \forall i \in \hat{J} \right\}. \quad (24)$$

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  $\mathcal{F}|_{K_{\hat{J}}}(\cdot, O'_{\hat{J}})$  has a local minimizer function  $\mathcal{U}_{\hat{J}} : O'_{\hat{J}} \rightarrow K_{\hat{J}}$  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  $\mathcal{U} : O_{\hat{J}} \rightarrow \mathbb{R}^p$  which satisfies  $\mathcal{U}(v) = \hat{u}$ ,  $\mathcal{U}|_{K_{\hat{J}}} = \mathcal{U}_{\hat{J}}$  and

$$\nu \in O_{\hat{J}} \Rightarrow D_i \mathcal{U}(\nu) = 0, \text{ for all } i \in \hat{J}. \quad (25)$$

Note that  $\hat{J}$  and  $K_{\hat{J}}$  are the same as those introduced in (12) Sec 3.2.1.

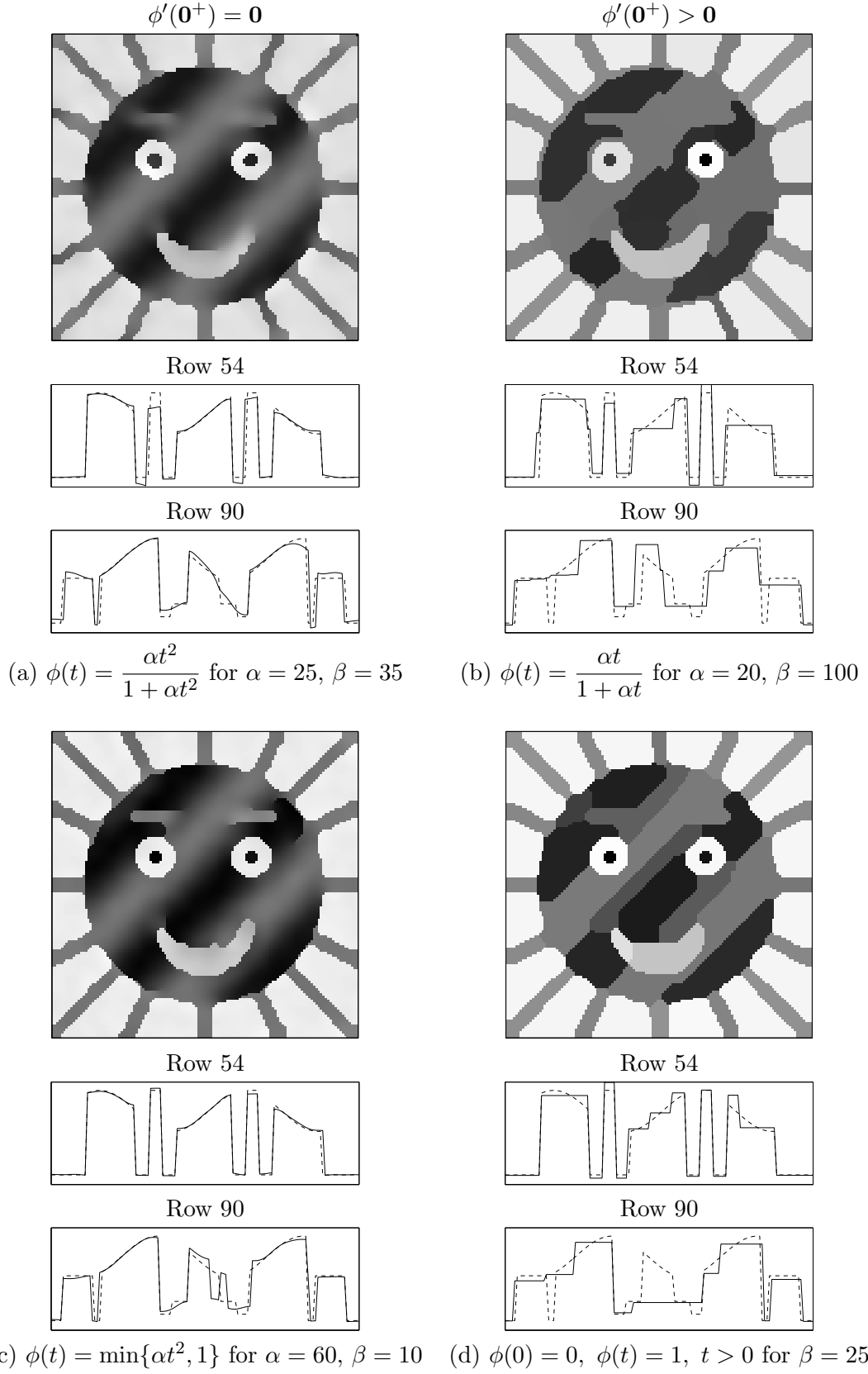


Figure 8: Restoration using non-convex PFs.

**Commentary on the assumptions.** Since  $\mathcal{F}(\cdot, v)$  has a local minimum at  $\hat{u}$ , by Theorem 1 one has  $\delta_1 \mathcal{F}(\hat{u}, v)(w) \geq 0$ , for all  $w \in K_{\hat{J}}^\perp \setminus \{0\}$  and if for some  $w$  the inequality becomes inequality, then the inequality is strict for  $-w$ . So (a) is not a strong requirement. Condition (b) amounts to Lemma 1 (Sec. 2.2) applied to  $\mathcal{F}|_{K_{\hat{J}}}$  which is  $\mathcal{C}^m$  on a neighborhood of  $(\hat{u}, v)$  belonging to  $K_{\hat{J}} \times \mathbb{R}^q$ .

If  $\mathcal{F}(\cdot, v)$  (possibly nonconvex) is of the form (11) and assumption H4 (Sect. 3.2) holds, Theorem 4 and the other results given next show that (a) and (b) are satisfied for any  $v \in \mathbb{R}^q \setminus \Theta$  where  $\Theta$  is closed and  $\mathbb{L}^q(\Theta) = 0$ .

**Significance of the results.** Using the definition of  $\mathcal{J}$  in (23), the conclusion of the theorem can be reformulated as

$$v \in O_{\hat{J}} \Rightarrow \mathcal{J}(\mathcal{U}(v)) \supseteq \hat{J} \Leftrightarrow \mathcal{U}(v) \in K_{\hat{J}}. \quad (26)$$

Minimizers involving large subsets  $\hat{J}$  are observed in Figs. 2 (b)-(c), 7(b), and 8(b) and (d), 10(a), (b) and (c), 11(d). It was seen in Examples 1 and 4, as well as in Sect. 4.3 (case  $\phi'(0^+) > 0$ ), that  $\hat{J}$  is nonempty for data  $v$  living in an open  $O_{\hat{J}}$ . Note also that there is an open subset  $\tilde{O}_{\hat{J}} \subset O_{\hat{J}}$  such that  $\mathcal{J}(\mathcal{U}(v)) = \hat{J}$  for all  $v \in \tilde{O}_{\hat{J}}$ . These sets  $\tilde{O}_{\hat{J}}$  are described in Example 5.

Observe that (26) 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  $\mathbb{R}^q$ .

Focus on a (local) minimizer function  $\mathcal{U} : O \rightarrow \mathbb{R}^p$  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  $\mathbb{R}^q$ . When data  $\nu$  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 form a covering of  $O$ . When  $\nu \in \tilde{O}_{J_j}$ , we find a minimizer  $\hat{u} = \mathcal{U}(\nu)$  satisfying  $\mathcal{J}(\hat{u}) = J_j$ .

Energies with nonsmooth regularization terms as those considered here, exhibit local minimizers which *generically* satisfy constraints of the form  $\mathcal{J}(\hat{u}) \neq \emptyset$ .

In particular, if  $\{D_i\}$  are discrete gradients or first-order difference operators, minimizers  $\hat{u}$  are typically constant on many regions. For example., if  $\phi(t) = t$ , we have  $\Phi(u) = \text{TV}(u)$ , and this explains the stair-casing effect observed in TV methods on discrete images and signals [30, 39].

## 5.2 Examples and Discussion

The subsection begins with an illustration of Theorem 10 and its meaning.

**Restoration of a noisy signal.** Fig. 9 shows a piecewise constant signal  $u_o$  corrupted with two different noises. Fig. 10 depicts the restoration from these two noisy data samples by minimizing an energy of the form  $\mathcal{F}(u, v) = \|u - v\|^2 + \beta \sum_{i=1}^{p-1} \phi(|u[i] - u[i+1]|)$ . The minimizers shown in Figs. 10a-c correspond to functions  $\phi$  such that  $\phi'(0^+) > 0$  and they are constant on large segments. The reader is invited to compare the subsets where these minimizers are constant. The function  $\phi$  in Fig. 10d satisfies  $\phi'(0^+) = 0$  and the resultant minimizers are nowhere constant.

Example 5 below gives a rich geometric interpretation of Theorem 25.

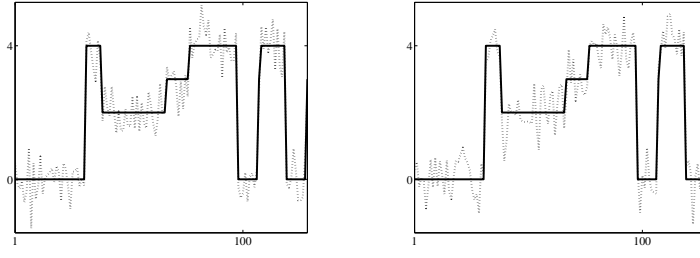


Figure 9: Data  $v = u_o + n$  (—) corresponding to the original  $u_o$  (-.-.) contaminated with two different noise samples  $n$  on the left and on the right.

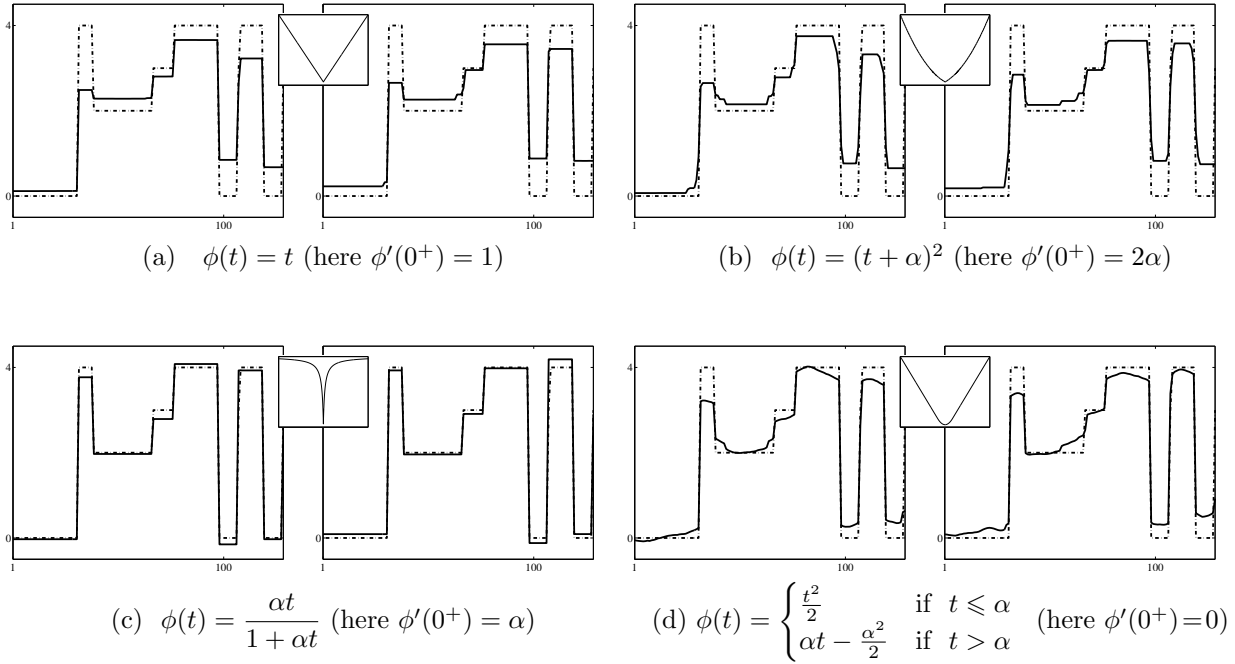


Figure 10: Restoration using different functions  $\phi$ . Original  $u_o$  (-.-.), minimizer  $\hat{u}$  (—). Each figure from (a) to (d) shows the two minimizers  $\hat{u}$  corresponding to the two data sets in Fig. 9 (left and right), while the shape of  $\phi$  is plotted in the middle.

**Example 5 (1D TV Regularization)** Let  $\mathcal{F} : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$  be given by

$$\mathcal{F}(u, v) = \|Au - v\|_2^2 + \beta \sum_{i=1}^{p-1} |u[i] - u[i+1]|, \quad \beta > 0, \quad (27)$$

where  $A \in \mathbb{R}^{p \times p}$  is invertible. Clearly, there is a unique minimizer function  $\mathcal{U}$  for  $\mathcal{F}(\cdot, \mathbb{R}^p)$ . Two striking phenomena concerning the sets  $\tilde{O}_J$  are described next:

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  $J \subset \{1, \dots, p-1\}$ , 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  $i \in J^c$ . Their closure forms a covering of  $\mathbb{R}^p$ .

The next Remark 2 deserves to be combined with the conclusions of Sec. 3.3.

**Remark 2** The energy in (27) has a straightforward Bayesian interpretation in terms of maximum a posteriori (MAP) estimation (see Sec. 1.1, first item). The quadratic data fidelity term corresponds to an observation 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  $\sigma^2$ . 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  $u[i] - u[i+1]$ ,  $1 \leq i \leq p-1$ , that is  $\exp(-\lambda|t|)$  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  $\hat{u}$ .

### 5.3 Applications

The use of nondifferentiable (and also nonconvex) regularization in compressive sensing is actually extremely abundant; readers can check e.g. the textbook [50].

**Image reconstruction is Computed Tomography.** The concentration of an isotope in a part of the body provides an image characterizing metabolic functions and local blood flow [21,62]. 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  $v[i] \geq 0$  reaching each detector,  $i = 1, \dots, q$ . The observed photon counts  $v$  have a Poissonian distribution [21,90]. Their mean is determined using projection operators  $\{a_i, i = 1, 2, \dots, q\}$  and a constant  $\rho > 0$ . The data fidelity  $\Psi$  derived from the log-likelihood function is nonstrictly convex and reads:

$$\Psi(u, v) = \rho \left\langle \sum_{i=1}^q a_i, u \right\rangle - \sum_{i=1}^q v[i] \ln(\langle a_i, u \rangle). \quad (28)$$



Figure 11 presents image reconstruction from simulated ECT data by minimizing and energy of the form (21) and (22) where  $\Psi$  is given by (28) and  $\{D_i\}$  yield the first-order differences between each pixel and its eight nearest neighbors. One observes, yet again, that a PF  $\phi$  which is nonconvex with  $\phi'(0^+) > 0$  leads to a nicely segmented piecewise constant reconstruction.

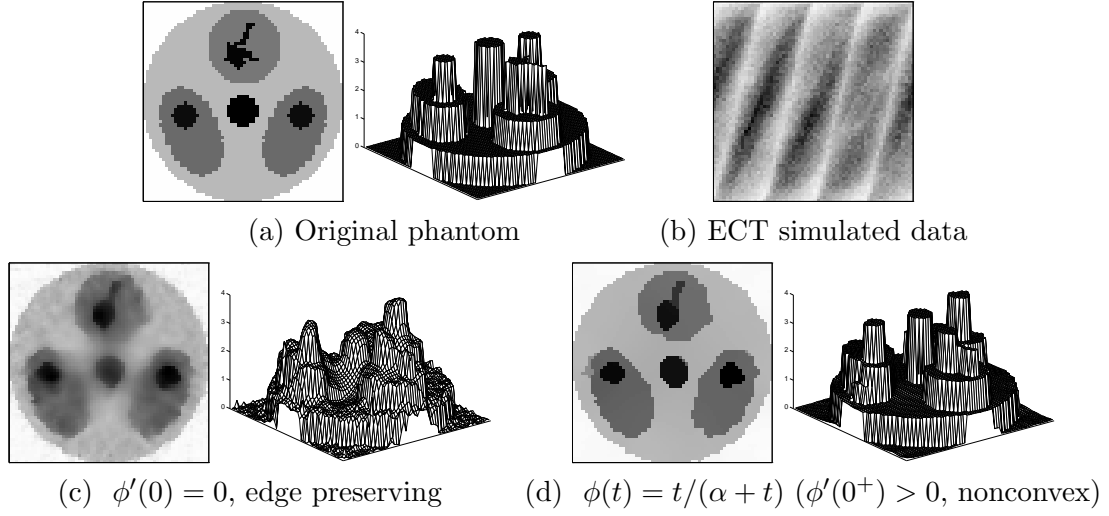


Figure 11: ECT.  $\mathcal{F}(u, v) = \Psi(u, v) + \beta \sum_{i \in I} \phi(|D_i u|)$ .

## 6 Nonsmooth Data-fidelity

Figure 12 shows that there is a striking distinction in the behavior of the minimizers relevant to nonsmooth data fidelity terms (b) with respect to nonsmooth 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.

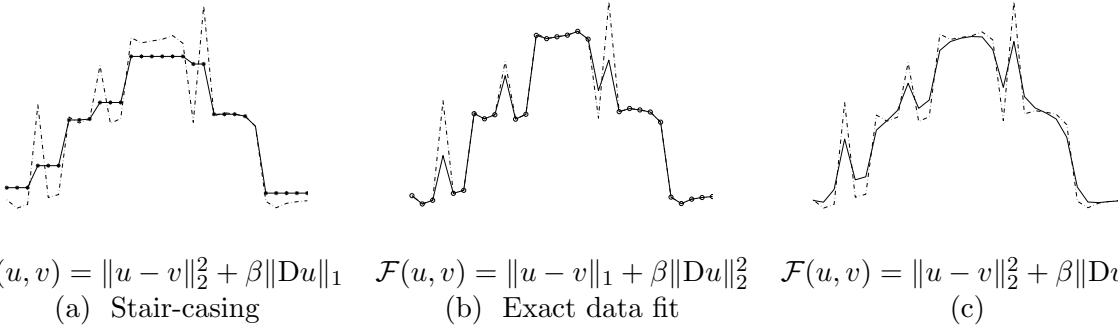


Figure 12:  $D$  is a first-order difference operator, i.e.  $D_i u = u[i] - u[i + 1]$ ,  $1 \leq i \leq p - 1$ . Data (---), Restored signal (—). Constant pieces in (a) are emphasized using “\*” while data samples that are equal to the relevant samples of the minimizer in (b) are emphasized using “o”.

Consider

$$\mathcal{F}(u, v) = \Psi(u, v) + \beta\Phi(u), \quad (29)$$

$$\Psi(u, v) = \sum_{i=1}^q \psi(|\langle a_i, u \rangle - v[i]|), \quad (30)$$

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

**H 12**  $\psi$  is  $\mathcal{C}^m$ ,  $m \geq 2$  on  $\mathbb{R}_+^*$  and  $\psi'(0^+) > 0$  is finite.

By this condition,  $t \mapsto \psi(|t|)$  is continuous on  $\mathbb{R}$ . Let  $A \in \mathbb{R}^{q \times p}$  denote the matrix such that for any  $i = 1, \dots, q$ , its  $i$ th row reads  $a_i^*$ .

Nonsmooth data fidelity terms  $\Psi$  in energies of the form (29) and (30) were introduced in *image processing* in 2001 [77].

## 6.1 General Results

Here we present some results on the minimizers  $\hat{u}$  of  $\mathcal{F}$  as given in (29) and (30), where  $\Psi$  is *nondifferentiable*, obtained in [78, 79]. An additional assumption is that

**H 13** The regularization term  $\Phi : \mathbb{R}^p \rightarrow \mathbb{R}$  in (29) is  $\mathcal{C}^m$ ,  $m \geq 2$ .

Note that  $\Phi$  in (29) can be convex or nonconvex. To analyze the observation in Fig. 12b, the following set-valued function  $\mathcal{J}$  will be useful:

$$(u, v) \in (\mathbb{R}^p \times \mathbb{R}^q) \mapsto \mathcal{J}(u, v) = \left\{ i \in \{1, \dots, q\} : \langle a_i, u \rangle = v[i] \right\}. \quad (31)$$

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  $\hat{J}^c = \{1, \dots, q\} \setminus \hat{J}$ .

**Theorem 11** Let  $\mathcal{F}$  be of the form (29)–(30) where assumptions H12, and H13 hold. Given  $v \in \mathbb{R}^q$  let  $\hat{u} \in \mathbb{R}^p$  be a (local) minimizer of  $\mathcal{F}(\cdot, v)$ . For  $\hat{J} = \mathcal{J}(\hat{u}, v)$ , where  $\mathcal{J}$  is defined according to (31), 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  $\mathcal{U} : O_{\hat{J}} \rightarrow \mathbb{R}^p$  relevant to  $\mathcal{F}(\cdot, O_{\hat{J}})$ , yielding in particular  $\hat{u} = \mathcal{U}(v)$ , and

$$\nu \in O_{\hat{J}} \Rightarrow \begin{cases} \langle a_i, \mathcal{U}(\nu) \rangle = \nu[i] & \text{if } i \in \hat{J}, \\ \langle a_i, \mathcal{U}(\nu) \rangle \neq \nu[i] & \text{if } i \in \hat{J}^c. \end{cases} \quad (32)$$

The result in (32) means that  $\mathcal{J}(\mathcal{U}(\nu), \nu) = \hat{J}$  is constant on  $O_{\hat{J}}$ .

Note that for every  $v$  and  $J \neq \emptyset$ , 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 (24) may be instructive. Compare also (b) and (c) in Theorem 11 with (a) and (b) in Theorem 10. By the way, conditions (b) and (c) in Theorem 11 ensure that  $\mathcal{F}(\cdot, v)$  reaches a strict minimum at  $\hat{u}$  [78, 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  $\hat{u}$ ,

$$\begin{aligned} \delta_1 \mathcal{F}(\hat{u}, v)(w) &= \phi'(0^+) \sum_{i \in \hat{J}} |\langle a_i, w \rangle| + \sum_{i \in \hat{J}^c} \psi'(\langle a_i, \hat{u} \rangle - v[i]) \langle a_i, w \rangle + \beta D\Phi(\hat{u})w \geq 0, \\ &\text{for any } w \in \mathbb{R}^p \end{aligned} \tag{33}$$

**Commentary on the assumptions.** Assumption (a) does not require the independence of the whole set  $\{a_i : i \in \{1, \dots, q\}\}$ . It is easy to check 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; see Lemma 1 (Sec. 2.2). If an arbitrary function  $\mathcal{F}(\cdot, v) : \mathbb{R}^p \rightarrow \mathbb{R}$  has a minimum at  $\hat{u}$ , then necessarily  $\delta_1 \mathcal{F}(\hat{u}, v)(w) \geq 0$  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 Sect. 3.2.1 can provide a starting point.

**Corollary 2** *Let  $\mathcal{F}$  be of the form (29)–(30) where  $p = q$ , and H12, and H13 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  $\{a_i : 1 \leq i \leq q\}$  is linearly independent.*

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

*Then*

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

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

$$\nu \in O_{\hat{J}} \Rightarrow \langle a_i, \mathcal{U}(\nu) \rangle = \nu[i] \quad \forall i \in \hat{J} = \{1, \dots, q\}. \tag{34}$$

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

In the context of Corollary 2,  $A$  is invertible. Combining this with (33) 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

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

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  $\#J \geq 1$ . The result in (32) means that the set-valued function  $v \rightarrow \mathcal{J}(\mathcal{U}(v), v)$  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  $O_{\hat{J}}$ .

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 *nonsmooth data fidelity terms generically yield minimizers fitting exactly a certain number of the data entries*. The resultant numerical effect is observed in Fig. 12b as well as in Figs. 14 and 15.

**Remark 3 (stability of minimizers)** *The fact that there is a  $\mathcal{C}^{m-1}$  local minimizer function shows that, in spite of the nonsmoothness of  $\mathcal{F}$ , for any  $v$ , all 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 6** *Let  $\mathcal{F}$  read*

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

*It is easy to see that there is a unique local minimizer function  $\mathcal{U}$  which is given by*

$$\begin{aligned}\mathcal{U}(v)[i] &= \frac{1}{\beta} \text{sign}(v[i]) & \text{if} & \quad |v[i]| > \frac{1}{\beta}, \\ \mathcal{U}(v)[i] &= v[i] & \text{if} & \quad |v[i]| \leq \frac{1}{\beta}.\end{aligned}$$

*Condition (c) in Theorem 11 fails to hold only for  $\left\{ v \in \mathbb{R}^q : |v[i]| = \frac{1}{\beta}, \forall i \in \hat{J} \right\}$ . This set is of Lebesgue measure zero in  $\mathbb{R}^q$ . For any  $J \in \{1, \dots, q\}$  put*

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

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

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

*Each set  $O_J$  has a positive Lebesgue measure in  $\mathbb{R}^q$ . Moreover, the union of all  $O_J$  when  $J$  ranges on all subsets  $J \subset \{1, \dots, q\}$  (including the empty set) forms a partition of  $\mathbb{R}^q$ .*

**Numerical experiment.** The original image  $u_o$  is shown in Fig. 13(a). Data  $v$  in Fig. 13(b) are obtained by replacing some pixels of  $u_o$  by aberrant impulses, called *outliers*. In all Figs. 14, 15, 16

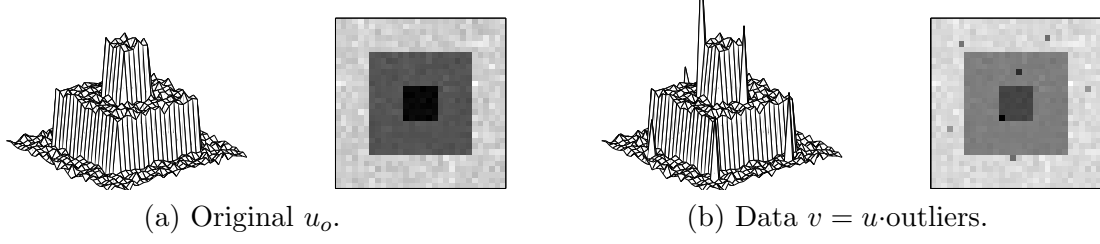


Figure 13: Original  $u_o$  and data  $v$  degraded by outliers.

and 17,  $\{D_i\}$  correspond to the first-order differences between each pixel and its four nearest neighbors. Fig. 14(a) corresponds to an  $\ell_1$  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. 14(b), is null everywhere except at the positions of the outliers in  $v$ . The pixels corresponding to nonzero residuals (i.e. the elements of  $\hat{J}^c$ ) provide a good estimate of the locations of the outliers in  $v$ . Fig. 15(a) shows 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. 15(b) is null only on restricted areas – but has a very small magnitude everywhere beyond the outliers.

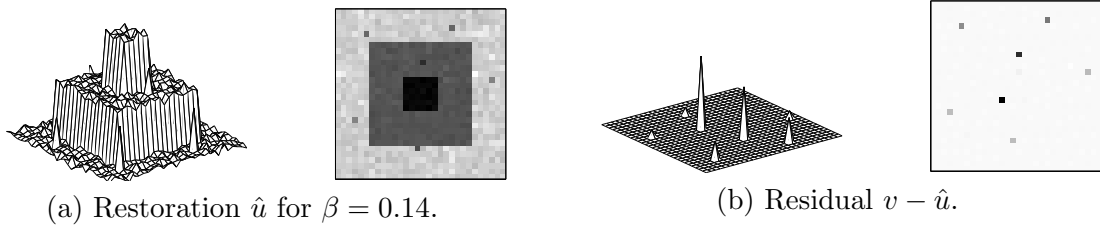


Figure 14: Restoration using  $\mathcal{F}(u, v) = \sum_i |u[i] - v[i]| + \beta \sum_{i \in I} |D_i u|^\alpha$   $\alpha = 1.1$  and  $\beta = 0.14$ .

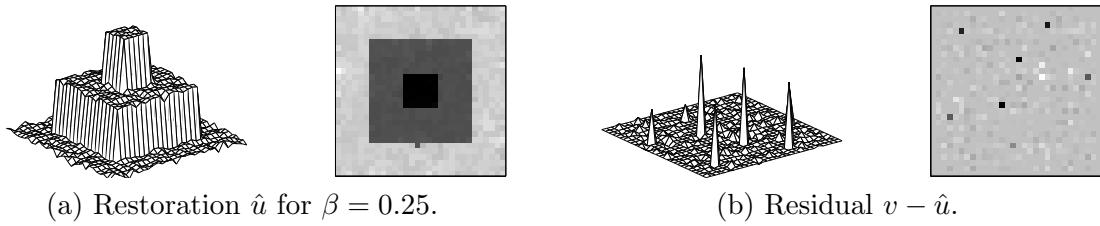


Figure 15: Restoration using  $\mathcal{F}(u, v) = \sum_i |u[i] - v[i]| + \beta \sum_{i \in I} |D_i u|^\alpha$  for  $\alpha = 1.1$  and  $\beta = 0.25$ .

The minimizers of two different cost-functions  $\mathcal{F}$  involving a *smooth* data fidelity term  $\Psi$ , shown in Figs. 16 and 17, do not fit any data entry. In particular, the restoration in Fig. 17 corresponds to a nonsmooth regularization and it is constant over large regions; this effect was explained in section 5.

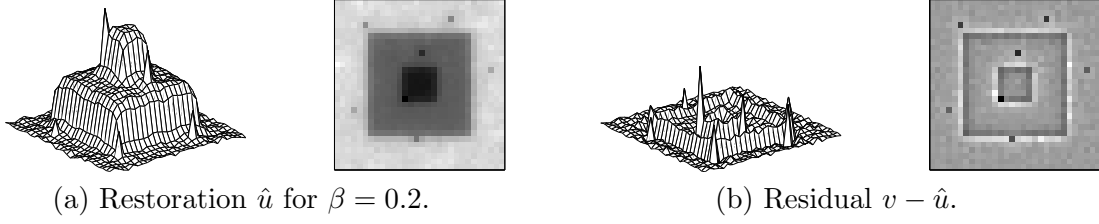


Figure 16: Restoration using a smooth energy,  $\mathcal{F}(u, v) = \sum_i (u[i] - v[i])^2 + \beta \sum_i (|D_i u|)^2$ ,  $\beta = 0.2$ .

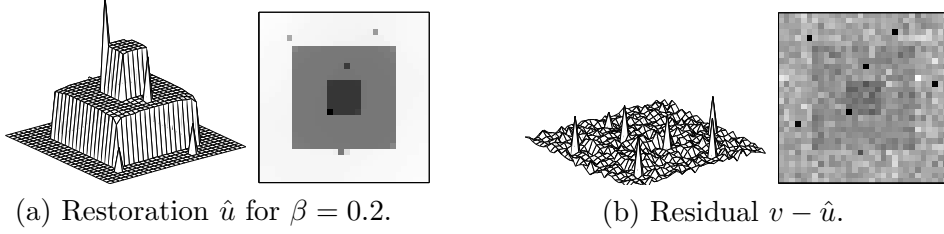


Figure 17: Restoration using nonsmooth regularization  $\mathcal{F}(u, v) = \sum_i (u[i] - v[i])^2 + \beta \sum_i |D_i u|$ ,  $\beta = 0.2$ .

## 6.2 Applications

*The possibility to keep some data samples unchanged by using nonsmooth data fidelity is a precious property in various application fields.* Nonsmooth data fidelities are good to detect and smooth outliers. This property was exploited for deblurring under impulse noise contamination, see, e.g., [10–12].

**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 Sect. 4, a systematic default of the images and signals restored using *convex* edge preserving PFs  $\phi$  is that the amplitude of edges is underestimated.

Shrinkage estimators operate on a decomposition of data  $v$  into a frame of  $\ell^2$ , 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  $\{\tilde{w}_i : i \in J\}$ . The frame coefficients of  $v$  read  $y = Wv$  and are contaminated with noise  $Wn$ . The inaugural work of Donoho and Johnstone [40] considers two different shrinkage estimators: given  $T > 0$ , *hard thresholding* corresponds to

$$y_T[i] = \begin{cases} y[i] & \text{if } i \in J_1, \\ 0 & \text{if } i \in J_0, \end{cases} \quad \text{where} \quad \begin{cases} J_0 = \{i \in J : |y[i]| \leq T\}; \\ J_1 = J \setminus J_0, \end{cases} \quad (35)$$

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  $\sigma$  and

$$T = \sigma \sqrt{2 \log_e p}. \quad (36)$$

This threshold is difficult to use in practice because it increases with the size of  $u$ . Numerous improvements were realized, see, e.g., [4, 13, 24, 34, 38, 66, 70]. In all cases, the main problem is that smoothing large coefficients oversmooths edges while thresholding small coefficients can generate Gibbs-like oscillations near edges, see Figs. 18c, d. If shrinkage is weak, noisy coefficients (outliers) remain almost unchanged and produce artifacts having the shape of  $\{\tilde{w}_i\}$ , see Figs. 18c–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 sought-after signal or image [18, 25, 36, 43, 67]. A critical analysis was presented in [46].

A specialized hybrid method involving  $\ell_1$  data fidelity on frame coefficients is proposed in [46]. Data are initially hard thresholded – see (35) – using a *suboptimal* threshold  $T$  in order to keep as much as possible information. (The use of another shrinkage estimator would alter all coefficients, which is not desired.) 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.

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:

$$F(x, y) = \lambda_1 \sum_{i \in J_1} |x[i] - y[i]| + \lambda_0 \sum_{i \in J_0} |x[i]| + \sum_{i \in I} \phi \left( \|D_i \widetilde{W} x\|_2 \right), \quad \lambda_{0,1} > 0, \quad (37)$$

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

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

Several properties relevant to the minimizers of  $F$  in (37), the parameters  $\lambda_i$ ,  $i \in \{0, 1\}$  and the solution  $\hat{u}$  are outlined in [46].

Noisy data  $v$  are shown along with the original  $u_o$  in Fig. 18a. The restoration in Fig. 18b minimizes  $\mathcal{F}(u) = \|Au - v\|_2^2 + \beta \text{TV}$  – homogeneous regions remain noisy, edges are smoothed and spikes are eroded. Fig. 18c is obtained using the *sure-shrink* method [41] from the toolbox WaveLab. The other restorations use thresholded Daubechies wavelet coefficients with eight vanishing moments. The optimal value for the hard thresholding obtained using (36) is  $T = 35$ . The relevant restoration – Fig. 18d – exhibits important Gibbs-like oscillations as well as wavelet-shaped artifacts. For  $T = 23$  the coefficients have a richer information content but  $\widetilde{W} y_T$ , shown in Fig. 18e manifests Gibbs artifacts and many wavelet-shaped artifacts. Introducing the thresholded coefficients of Fig. 18e in the specialized energy  $F$  in (37) leads to Fig. 18f: edges are clean and piecewise polynomial parts are well recovered.

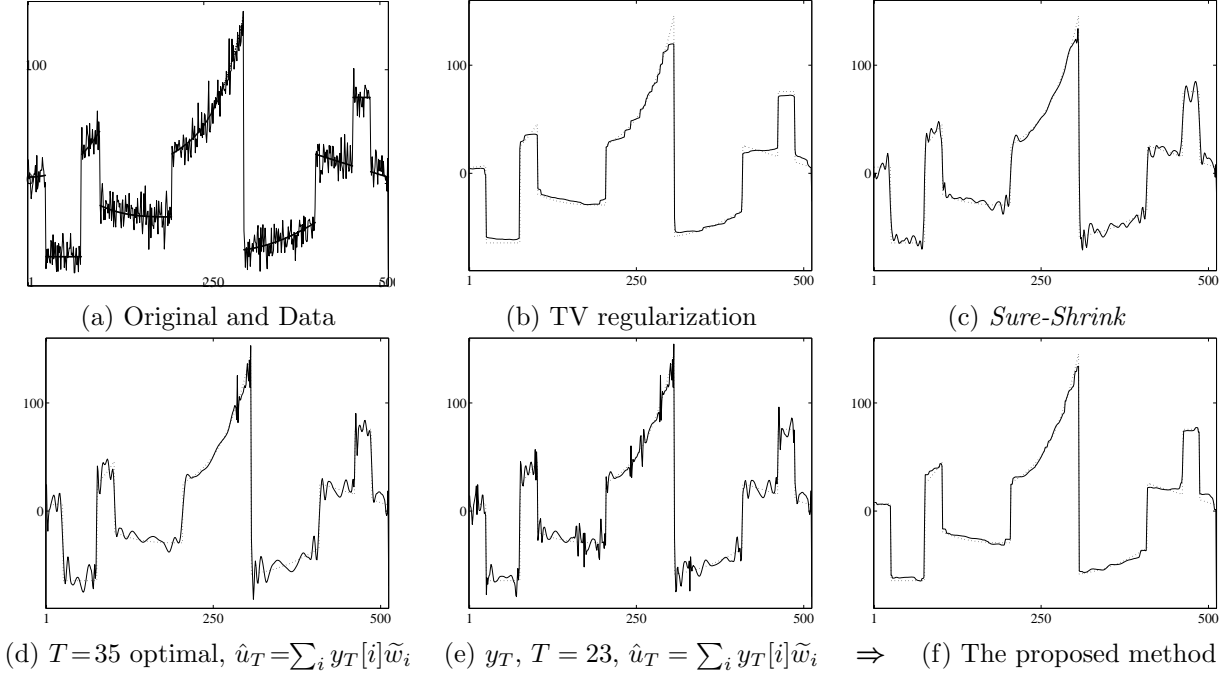


Figure 18: Methods to restore the noisy signal in (a). Restored signal (—), Original signal (---).

## 7 Nonsmooth data fidelity and regularization

### 7.1 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].

Following [1, 78, 79], S. Esedoglu and T. Chan explored in [29] the minimizers of the  $L_1$ -TV functional given below

$$\mathcal{F}(u, v) = \int_{\mathbb{R}^d} |u(x) - v(x)| dx + \beta \int_{\mathbb{R}^d} |\nabla u(x)| dx, \quad (38)$$

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 [29] is based on a representation of  $\mathcal{F}$  in (38) in terms of the level sets of  $u$  and  $v$ . Most of the results are established for data  $v$  given by the characteristic function  $\chi_\Sigma$  of a bounded domain  $\Sigma \subset \mathbb{R}^d$ . Theorem 5.2 in [29] says that if  $v = \chi_\Sigma$ , 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 [29] states that if in addition  $\Sigma$  is *convex*, then for almost every  $\beta \geq 0$ ,  $\mathcal{F}(\cdot, v)$  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  $\beta$  while for a larger  $\beta$  they suddenly disappear.

#### 7.1.1 Denoising of binary images and convex relaxation

Many problems such as text denoising and document processing, two-phase image segmentation, shape restoration, fairing of surfaces in computer graphics are naturally stated as the minimization of an



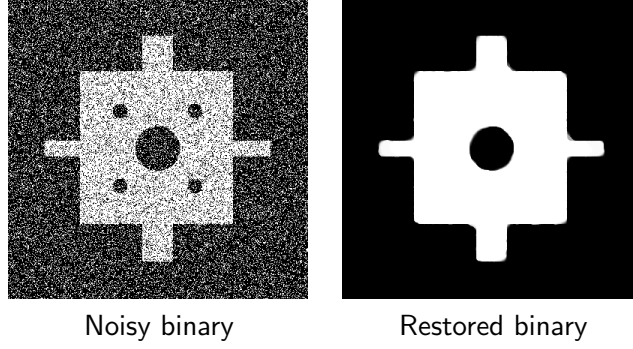


Figure 19: Restoration of a binary noisy image by minimizing  $L_1$ -TV.

energy over the set of the binary images. These energies are obviously nonconvex since the constraint set is finite. Their global minimizer was shown in [28] to be also the minimizer of the convex  $L_1$ -TV functional which is convex. This result yielded much simpler algorithms for binary image restoration. An illustration is given on Fig. 19.

Since then,  $L_1$ -TV relaxations have become a common tool for convex relaxations, see e.g. among many others [84] and the references therein.

Also,  $L_1$ -TV energies were revealed very successful in image decomposition, see e.g., [7, 48].

### 7.1.2 Multiplicative Noise Removal

In various active imaging systems, such as synthetic aperture radar (SAR), laser or ultrasound imaging, the data representing the underlying (unknown image)  $S_0$  are corrupted with multiplicative noise. Such a noise is a severe degradation, see Fig. 20. When possible, a few independent measurements for the same scene are realized,  $\Sigma_k = S_0 \eta_k$  for  $k \in \{1, \dots, K\}$ , where the noise  $\eta_k$  is typically modeled by the one-sided exponential distribution. The data  $\Sigma$  used for denoising is the average of the set of all  $K$  measurements:

$$\Sigma = \frac{1}{K} \sum_{k=1}^K \Sigma_k = S_0 \eta. \quad (39)$$

The combined multiplicative noise follows a Gamma distribution. Adaptive filtering works only if the noise is weak. For strong noise, variational methods often use TV regularization. In [6] the log-likelihood of the raw data (39) is regularized using TV. Instead, the properties of  $L_1$ -TV are used to design an energy in [47]. First, the log-data  $v = \log \Sigma$  is decomposed into a curvelet transform yielding noisy coefficients  $y = Wv$ . A suboptimal hard thresholding is applied for  $T$  adapted to the expectation of the log-noise. Let  $I_0 = \{i : |y[i]| \leq T\}$  and  $I_1 = \{i : |y[i]| > T\}$ . Since the threshold is low,  $I_1$  contains outliers. Coefficients  $\hat{x}$  are restored by minimizing

$$F(x) = \lambda_1 \sum_{i \in I_1} |(x - y)[i]| + \lambda_0 \sum_{i \in I_0} |x[i]| + \text{TV}(x).$$

The restored image  $\hat{S}$ , shown in Fig. 20 is obtained as  $\hat{S} = \exp(\widetilde{W}(\hat{x})) B$  where  $\widetilde{W}$  is a left inverse of  $W$  and  $B$  is a bias correction.



Noisy:  $K = 4$  in (39)



[6]: PSNR=24.55, MAE=10.06



[47]: PSNR=25.84, MAE=9.09



Original (512x512)

Figure 20: Aerial image of the town of Nîmes ( $512 \times 512$ ) for  $K = 4$  in (39). Restorations using different methods. Parameters: [6] for  $\rho = 120$ ; [47]  $T = 2\sqrt{\psi_1(K)}$ ,  $\lambda_0 = 1.5$ ,  $\lambda_1 = 10$ .

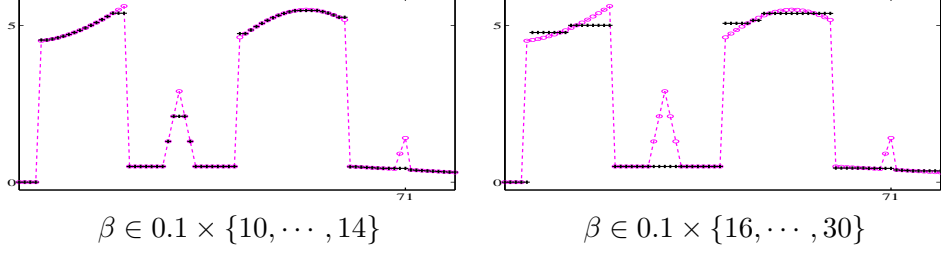


Figure 21: Minimizers of  $\mathcal{F}(\cdot, v)$  as given in (40) for  $\phi(t) = \ln(\alpha t + 1)$ ,  $\alpha = 2$  and different values of  $\beta$ . Data samples ( $\circ\circ\circ$ ), Minimizer samples  $\hat{u}[i]$  ( $+++$ ).

## 7.2 $\ell_1$ Data Fidelity with Regularization Concave on $\mathbb{R}_+$

One could expect that  $\ell_1$  data fidelity regularized with a PF concave on  $\mathbb{R}_+$  should somehow reinforce the properties of  $\ell_1$  - TV. The question was recently examined in [83]. Consider the energy

$$\begin{aligned} \mathcal{F}(u, v) &= \sum_{i \in I} |\langle a_i, u \rangle - v[i]| + \beta \sum_{j \in J} \phi(|D_j u|) \\ \text{for } I &\stackrel{\text{def}}{=} \{1, \dots, q\} \quad \text{and} \quad J \stackrel{\text{def}}{=} \{1, \dots, r\} \end{aligned} \quad (40)$$

where  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is continuous and concave on  $\mathbb{R}_+$  (e.g., (f10), (f11) and (f12) in Table 1).

### 7.2.1 Motivation

Figs. 21 and 22 depict (the global) minimizers of  $\mathcal{F}(u, v)$  in (40) for a one-dimensional signal where  $A = \text{Id}$ ,  $\{D_i\}$  are first-order differences and  $\phi$  is smooth and concave on  $\mathbb{R}_+$ .

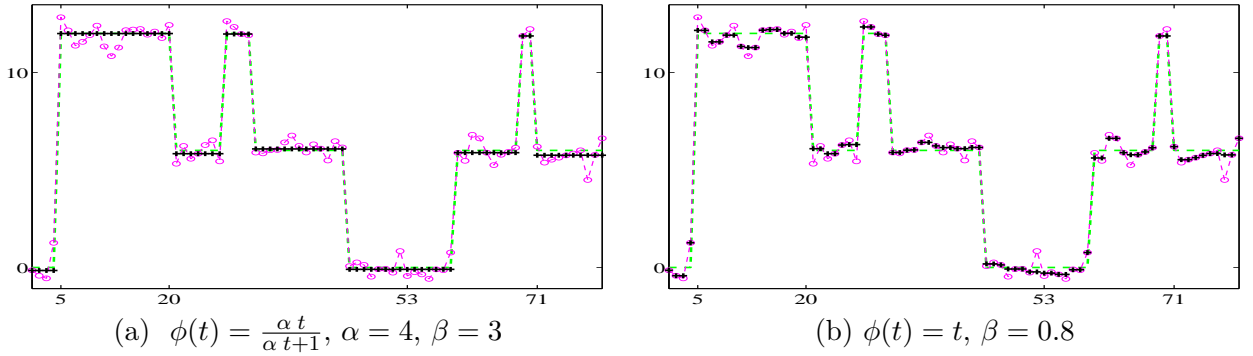


Figure 22: Minimizers of  $\mathcal{F}(\cdot, v)$  as given in (40) for different PFs  $\phi$ . Data are corrupted with Gaussian noise. Data samples  $v[i]$  are marked with ( $\circ\circ\circ$ ), samples  $\hat{u}[i]$  of the minimizer—with ( $+++$ ). The original signal is reminded in ( $---$ ).

The tests in Figs. 21 and 22 show that a PF concave on  $\mathbb{R}_+$  considerably reinforces the properties of  $\ell_1$  - TV. One observes that the minimizer satisfies *exactly* part of the data term and part of the prior term (corresponding to constant pieces). In Fig. 22(b), the previous  $\ell_1$  - TV model is considered. Fig. 21 shows also that the minimizer remains unchanged for some range of values of  $\beta$  and that after a threshold value, it is simplified.

Example 7 below furnishes a first intuition on the reasons underlying the phenomena observed in Figs. 21 and 22.

**Example 7** Given  $v \in \mathbb{R}$ , consider the function  $\mathcal{F}(\cdot, v) : \mathbb{R} \rightarrow \mathbb{R}$  given below

$$\mathcal{F}(u, v) = |u - v| + \beta\phi(|u|) \quad \text{for } \phi \text{ obeying H} . \quad (41)$$

The necessary conditions for  $\mathcal{F}(\cdot, v)$  to have a (local) minimum at  $\hat{u} \neq 0$  and  $\hat{u} \neq v$  —that its first differential meets  $D_1\mathcal{F}(\hat{u}, v) = 0$  and that its second differential obeys  $D_1^2\mathcal{F}(\hat{u}, v) \geq 0$ —do not hold:

$$\hat{u} \notin \{0, v\} \quad \Rightarrow \quad \begin{cases} D_1\mathcal{F}(\hat{u}, v) = \text{sign}(\hat{u} - v) + \beta\phi'(|\hat{u}|)\text{sign}(\hat{u}) = 0 , \\ D_1^2\mathcal{F}(\hat{u}, v) = \beta\phi''(|\hat{u}|) < 0 , \end{cases}$$

where the last inequality comes from the strict concavity of  $\phi$  on  $\mathbb{R}_+^*$ . Hence,  $\mathcal{F}(\cdot, v)$  cannot have a minimizer such that  $\hat{u} \neq 0$  and  $\hat{u} \neq v$ , for any  $v \in \mathbb{R}$ . Being coercive,  $\mathcal{F}(\cdot, v)$  does have minimizers. Consequently, any (local) minimizer of  $\mathcal{F}$  in (41) satisfies

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

### 7.2.2 Main Theoretical Results

The PFs considered here are concave on  $\mathbb{R}_+$  and smooth on  $\mathbb{R}_+^*$ . More precisely, they satisfy H1 (Sec. 1), H8 and H10 (Sec. 4.2). One can see Fig. 3—right for an illustration of the assumptions.

**Proposition 3** Let  $\mathcal{F}(\cdot, v)$  read as in (40). Assume that H3 (Sec. 3) holds and that  $\phi$  satisfies H1 (Sec. 1), H8 and H10. Then for any  $v$ ,  $\mathcal{F}(\cdot, v)$  has global minimizers.

Given  $v \in \mathbb{R}^q$ , with each  $\hat{u} \in \mathbb{R}^p$  the following subsets are associated:

$$\begin{aligned} \hat{I}_0 &\stackrel{\text{def}}{=} \{i \in I \mid \langle a_i \hat{u} \rangle = v[i]\} & \text{and} & & \hat{I}_0^c &\stackrel{\text{def}}{=} I \setminus \hat{I}_0, \\ \hat{J}_0 &\stackrel{\text{def}}{=} \{i \in J \mid D_i \hat{u} = 0\} & \text{and} & & \hat{J}_0^c &\stackrel{\text{def}}{=} J \setminus \hat{J}_0 . \end{aligned} \quad (42)$$

**Proposition 4** For  $\mathcal{F}(\cdot, v)$  as in (40) satisfying H1, H8 and H10, let  $\hat{u}$  be a (local) minimizer of  $\mathcal{F}(\cdot, v)$ . Then

$$(\hat{I}_0 \cup \hat{J}_0) \neq \emptyset .$$

**H 14** The point  $\hat{u} \in \mathbb{R}^p$  is such that  $\hat{I}_0 \neq \emptyset$  and that

$$w \in \ker D \setminus \{0\} \quad \Rightarrow \quad \exists i \in \hat{I}_0 \quad \text{such that} \quad \langle a_i, w \rangle \neq 0 . \quad (43)$$

If  $\text{rank } D = p$ , then (43) is trivial. Anyway, (43) is not a strong requirement.

**Theorem 12** Consider  $\mathcal{F}(\cdot, v)$ , as given in (40), satisfying H3, as well as H1, H8 and H10. Let  $\hat{u}$  be a (local) minimizer of  $\mathcal{F}(\cdot, v)$  meeting  $\hat{J}_0^c \neq \emptyset$  and H14. Then  $\hat{u}$  is the unique solution of the full column rank linear system given below

$$\begin{cases} \langle a_i, w \rangle = v[i] & \forall i \in \hat{I}_0 , \\ D_j w = 0 & \forall j \in \hat{J}_0 . \end{cases} \quad (44)$$

**Significance of the results.** An immediate consequence of Theorem 12 is the following:

Each (local) minimizer of  $\mathcal{F}(\cdot, v)$  is strict.

Another consequence is that the matrix  $H$  with rows  $(a_i^*, \forall i \in \hat{I}_0$  and  $D_j, \forall j \in \hat{J}_0)$  has *full column rank*. This provides a strong *necessary condition* for a (local) minimizer of  $\mathcal{F}(\cdot, v)$ . And since  $\hat{u}$  in (44) solves a linear system, it involves the same kind of “contrast invariance” as the  $L_1 - \text{TV}$  model. A detailed inspection of the minimizers in Figs. 21 and 22 corroborate Theorem 12. A more practical interpretation of this result reads as follows:

Each pixel of a (local) minimizer  $\hat{u}$  of  $\mathcal{F}(\cdot, v)$  is involved in (at least) one data equation that is fitted exactly  $\langle a_i, \hat{u} \rangle = v[i]$  – or in (at least) one vanishing operator  $D_j \hat{u} = 0$  – or in both types of equations.

This remarkable property can be used in different ways.

### 7.2.3 Applications

An energy  $\mathcal{F}(\cdot, v)$  of the form in (40) with a PF  $\phi$  strictly concave on  $\mathbb{R}_+$  is a good choice when

- There are some nearly faithful data points  $v[i]$ ;
- The matrix  $D$  provides a very reliable prior on the sought-after solution.

A natural way for such a prior is to construct for  $D$  an application-dependent dictionary.

**MR image reconstruction from highly undersampled data.** In the experiment in Fig. 23, only 5% *randomly* chosen noisy data samples in the  $k$ -space (i.e. individual noisy Fourier coefficients) are available, see (a). Data are contaminated with SNR=37 dB white centered *Gaussian noise*. This is a highly underdetermined, ill-posed inverse problem. It can be related to compressed sensing in MRI, see e.g. [58]. The Shepp-Logan phantom being locally constant with oval shapes, the linear operators  $\{D_i\}$  in (40) yield the usual discrete gradient of the image, so that the regularization term provides a correct prior. Indeed,  $Du_o$  is the sparsest linear transform for this image. Clearly,  $A$  is the undersampled Fourier transform corresponding to the 5% randomly chosen  $k$ -samples. For Gaussian noise, an  $\ell_2$  quadratic data fitting term is a classical choice. The  $\ell_2 - \text{TV}$  cost-function  $\|Au - v\|_2^2 + \beta \text{TV}(u)$  is the standard tool to solve this kind of problems. The result is shown in Fig. 23(b).

## 8 Conclusion

This chapter provided some theoretical results relating the shape of the energy  $\mathcal{F}$  to minimize and the salient features of its minimizers  $\hat{u}$  (see (7), Sec. 1.2). These results can serve as a kind of *inverse 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 problem of such a inverse modeling remains open because of the diversity of the inverse problems to solve and the possible energy functionals.

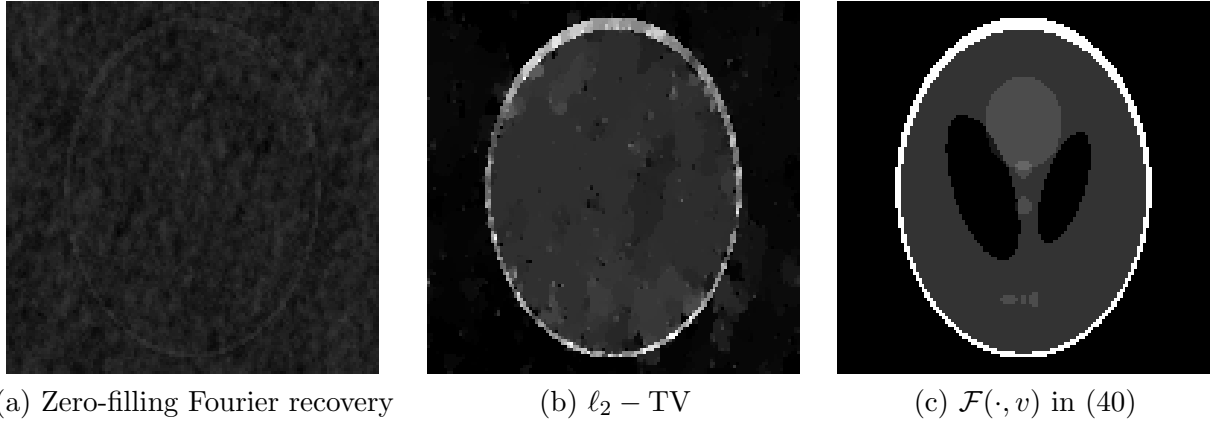


Figure 23: Reconstructed images from **5%** noisy randomly selected samples in the  $k$ -space using different methods.

## 9 Cross-References

- Compressive Sensing;
- Inverse Scattering;
- Iterative Solution Methods;
- Linear Inverse Problems;
- Mathematical Methods in PET and Spect Imaging;
- Mumford and Shah Model and its Applications to Image Segmentation and Image Restoration;
- Regularization Methods for Ill-Posed Problems;
- Statistical Methods in Imaging;
- Total Variation in Imaging.

## 10 References

### References

- [1] Alliney S (1992) Digital filters as absolute norm regularizers. IEEE Trans Signal Process SP-40:1548–1562
- [2] Alter F, Durand S, Forment J (2005) Adapted total variation for artifact free decompression of JPEG images. J Math Imaging Vis 23:199–211
- [3] Ambrosio L, Fusco N, Pallara D (2000) Functions of bounded variation and free discontinuity Problems. Oxford Mathematical Monographs, Oxford University Press

- [4] Antoniadis A, Fan J (2001) Regularization of wavelet approximations. *J Acoust Soc Am* 96: 939–967
- [5] Aubert G, Kornprobst P (2006) *Mathematical problems in image processing*, 2nd edn. Springer, Berlin
- [6] Aubert G, Aujol J.-F. (2008) A variational approach to remove multiplicative noise, *SIAM J. on Appl. Maths.*, 68:925–946
- [7] Aujol J-F, Gilboa G, Chan T, Osher S (2006) Structure-texture image decomposition - modeling, algorithms, and parameter selection. *Int J Comput Vis* 67:111–136
- [8] Auslender A and Teboulle M. (2003) *Asymptotic Cones and Functions in Optimization and Variational Inequalities*. Springer, New York
- [9] Attouch, H., Bolte, J. and Svaiter, B. F. (2013) Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forwardbackward splitting, and regularized GaussSeidel methods. *Math. Program.* 137:911–929
- [10] Bar L, Brook A, Sochen N, Kiryati N (2007) Deblurring of color images corrupted by impulsive noise. *IEEE Trans Image Process* 16:1101–1111
- [11] Bar L, Kiryati N, Sochen N (2006) Image deblurring in the presence of impulsive noise, *International. J Comput Vision* 70:279–298
- [12] Bar L, Sochen N, Kiryati N (2005) 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*, Hofgeismar. LNCS, vol 3459, pp 107–118
- [13] Belge M, Kilmer M, Miller E (2000) Wavelet domain image restoration with adaptive edge-preserving regularization. *IEEE Trans Image Process* 9:597–608
- [14] Besag JE (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *J Roy Stat Soc B* 36:192–236
- [15] Besag JE (1989) Digital image processing : towards Bayesian image analysis. *J Appl Stat* 16:395–407
- [16] Black M, Rangarajan A (1996) On the unification of line processes, outlier rejection, and robust statistics with applications to early vision. *Int J Comput Vis* 19:57–91
- [17] Blake A, Zisserman A (1987) *Visual reconstruction*. MIT Press, Cambridge
- [18] Bobichon Y, Bijaoui A (1997) Regularized multiresolution methods for astronomical image enhancement. *Exp Astron* 7:239–255

- [19] Bolte, J., Sabach, S. and Teboulle, M. (2014) Proximal alternating linearized minimization for nonconvex and nonsmooth problems, *Math. Program. Ser. A*, 146:459–494
- [20] Bouman C, Sauer K (1993) A generalized Gaussian image model for edge-preserving map estimation. *IEEE Trans Image Process* 2:296–310
- [21] Bouman C, Sauer K (1996) A unified approach to statistical tomography using coordinate descent optimization. *IEEE Trans Image Process* 5:480–492
- [22] Bredies K, Kunich K, Pock T (2010) Total generalized variation. *SIAM J Imaging Sci* 3(3): 480–491.
- [23] Bredies K, and Holler, M. (2014) Regularization of linear inverse problems with total generalized variation. *J Inverse and Ill-posed Problems* DOI: 10.1515/jip-2013-0068.
- [24] Candès EJ, Donoho D, Ying L (2005) Fast discrete curvelet transforms. *SIAM Multiscale Model Simul* 5:861–899
- [25] Candès EJ, Guo F (2002) New multiscale transforms, minimum total variation synthesis. Applications to edge-preserving image reconstruction. *Signal Process* 82:1519–1543
- [26] Catte F, Coll T, Lions PL, Morel JM (1992) Image selective smoothing and edge detection by nonlinear diffusion (I). *SIAM J Num Anal* 29:182–193
- [27] Chambolle A (2004) An Algorithm for Total Variation Minimization and Application. *J. Math Imaging Vis* 20:89–98
- [28] Chan T., Esedoglu S. and Nikolova M. (2006) Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J. Appl. Math.* 66:1632–1648
- [29] Chan T, Esedoglu S (2005) Aspects of total variation regularized  $L^1$  function approximation. *SIAM J Appl Math* 65:1817–1837
- [30] Chan TF, Wong CK (1998) Total variation blind deconvolution. *IEEE Trans Image Process* 7:370–375
- [31] Charbonnier P, Blanc-Féraud L, Aubert G, Barlaud M (1997) Deterministic edge-preserving regularization in computed imaging. *IEEE Trans Image Process* 6:298–311
- [32] Chellapa R, Jain A (1993) Markov random fields: theory and application. Academic, Boston
- [33] Chen X., M. Ng, Zhang C. (2012) Non-Lipschitz  $\ell_p$ -Regularization and Box Constrained Model for Image Restoration. *IEEE Trans Image Process* 21:4709–4721.
- [34] Chesneau C, Fadili J, Starck J-L (2010) Stein block thresholding for image denoising. *Appl. Comput. Harmon. Anal.* 28:67–88



- [35] Ciarlet PG (1989) Introduction to numerical linear algebra and optimization. Cambridge University Press, Cambridge
- [36] Coifman RR, Sowa A (2000) Combining the calculus of variations and wavelets for image enhancement. *Appl Comput Harmon Anal* 9:1–18
- [37] Demoment G (1989) Image reconstruction and restoration: overview of common estimation structure and problems. *IEEE Trans Acoust Speech Signal Process ASSP-37*:2024–2036
- [38] Do MN, Vetterli M (2005) The contourlet transform: an efficient directional multiresolution image representation. *IEEE Trans Image Process* 15:1916–1933
- [39] Dobson D, Santosa F (1996) Recovery of blocky images from noisy and blurred data. *SIAM J Appl Math* 56:1181–1199
- [40] Donoho DL, Johnstone IM (1994) Ideal spatial adaptation by wavelet shrinkage. *Biometrika* 81:425–455
- [41] Donoho DL, Johnstone IM (1995) Adapting to unknown smoothness via wavelet shrinkage. *J Acoust Soc Am* 90:1200–1224
- [42] Dontchev AL, Zollezi T (1993) Well-posed optimization problems. Springer, New York
- [43] Durand S, Froment J (2003) Reconstruction of wavelet coefficients using total variation minimization. *SIAM J Sci Comput* 24:1754–1767
- [44] Durand S, Nikolova M (2006) Stability of minimizers of regularized least squares objective functions I: study of the local behavior. *Appl Math Optim (Springer, New York)* 53:185–208
- [45] Durand S, Nikolova M (2006) Stability of minimizers of regularized least squares objective functions II: study of the global behaviour. *Appl Math Optim (Springer, New York)* 53:259–277
- [46] Durand S, Nikolova M (2007) Denoising of frame coefficients using  $\ell^1$  data-fidelity term and edge-preserving regularization. *SIAM J Multiscale Model Simulat* 6:547–576
- [47] Durand S, Fadili J, Nikolova M. (2010) Multiplicative noise removal using L1 fidelity on frame coefficients. *J. Math Imaging and Vision*, 36:201–226
- [48] Duval V, Aujol J-F, Gousseau Y (2009) The TVL1 model: a geometric point of view. *SIAM J Multiscale Model Simulat* 8:154–189
- [49] Ekeland I, Temam R (1976) Convex analysis and variational problems. North-Holland/SIAM, Amsterdam
- [50] Eldar, Y. C., Kutyniok, G.(2012) Compressed Sensing: Theory and Applications. Cambridge Univ. Press

- [51] Fessler F (1996) Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): applications to tomography. *IEEE Trans Image Process* 5:493–506
- [52] Fiacco A, McCormic G (1990) Nonlinear programming. *Classics in applied mathematics*. SIAM, Philadelphia
- [53] Geman D (1990) Random fields and inverse problems in imaging, vol 1427, *École d’Été de Probabilités de Saint-Flour XVIII - 1988*, Springer, *Lecture notes in mathematics*, pp 117–193
- [54] Geman D, Reynolds G (1992) Constrained restoration and recovery of discontinuities. *IEEE Trans Pattern Anal Mach Intell PAMI-14*:367–383
- [55] Geman D, Yang C (1995) Nonlinear image recovery with half-quadratic regularization. *IEEE Trans Image Process IP-4*:932–946
- [56] Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell PAMI-6*:721–741
- [57] Green PJ (1990) Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Trans Med Imaging MI-9*:84–93
- [58] Lustig M., Donoho D., Santos J. M. and Pauly L. M. (2008) Compressed Sensing MRI: a look how CS can improve our current imaging techniques: *IEEE Signal Proc. Magazine*. 25:72-82.
- [59] Haddad A, Meyer Y (2007) Variational methods in image processing, in “Perspective in Nonlinear Partial Differential equations in Honor of Haïm Brezis,” *Contemp Math (AMS)* 446:273–295
- [60] Hiriart-Urruty J-B, Lemaréchal C (1996) *Convex analysis and minimization algorithms*, vols I, II. Springer, Berlin
- [61] Hofmann B (1986) *Regularization for applied inverse and ill posed problems*. Teubner, Leipzig
- [62] Kak A, Slaney M (1987) *Principles of computerized tomographic imaging*. IEEE Press, New York
- [63] Keren D, Werman M (1993) Probabilistic analysis of regularization. *IEEE Trans Pattern Anal Mach Intell PAMI-15*:982–995
- [64] Li S (1995) *Markov random field modeling in computer vision*, 1st edn. Springer, New York
- [65] Li SZ (1995) On discontinuity-adaptive smoothness priors in computer vision. *IEEE Trans Pattern Anal Mach Intell PAMI-17*:576–586
- [66] Luisier F, Blu T (2008) SURE-LET multichannel image denoising: interscale orthonormal wavelet thresholding. *IEEE Trans Image Process* 17:482–492
- [67] Malgouyres F (2002) Minimizing the total variation under a general convex constraint for image restoration. *IEEE Trans Image Process* 11:1450–1456

- [68] Morel J-M, Solimini S (1995) Variational methods in image segmentation. Birkhäuser, Basel
- [69] Morozov VA (1993) Regularization methods for ill posed problems. CRC Press, Boca Raton
- [70] Moulin P, Liu J (1999) Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors. *IEEE Trans Image Process* 45:909–919
- [71] Moulin P, Liu J (2000) Statistical imaging and complexity regularization. *IEEE Trans Inf Theory* 46:1762–1777
- [72] Mumford D, Shah J (1989) Optimal approximations by piecewise smooth functions and associated variational problems. *Commun Pure Appl Math* 42:577–684
- [73] Nashed M, Scherzer O (1998) Least squares and bounded variation regularization with nondifferentiable functional. *Numer Funct Anal Optim* 19:873–901
- [74] Nikolova M (1996) Regularisation functions and estimators. In *Proceedings of the IEEE international conference on image processing*, vol 2, pp 457–460
- [75] Nikolova M (1997) Estimées localement fortement homogènes. *Comptes-Rendus de l’Académie des Sciences* 325 (série 1):665–670
- [76] Nikolova M (2000) Thresholding implied by truncated quadratic regularization. *IEEE Trans Image Process* 48:3437–3450
- [77] Nikolova M (2001) Image restoration by minimizing objective functions with non-smooth data-fidelity terms. In *IEEE international conference on computer vision/workshop on variational and level-set methods*, pp 11–18
- [78] Nikolova M (2002) Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers. *SIAM J Num Anal* 40:965–994
- [79] Nikolova M (2004) A variational approach to remove outliers and impulse noise. *J Math Imaging Vis* 20:99–120
- [80] Nikolova M (2004) Weakly constrained minimization. Application to the estimation of images and signals involving constant regions. *J Math Imaging Vis* 21:155–175
- [81] Nikolova M (2005) Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares. *SIAM J Multiscale Model Simulat* 4:960–991
- [82] Nikolova M (2007) Analytical bounds on the minimizers of (nonconvex) regularized least-squares. *AIMS J Inverse Probl Imaging* 1:661–677
- [83] Nikolova M., Ng M., Tam C.P. (2013) On  $\ell_1$  Data Fitting and Concave Regularization for Image Recovery. *SIAM J. Sci. Comput* 35:397–430

- [84] Papadakis N., Yildizoglu R., Aujol J-F. and Caselles V. (2013) High-dimension multi-label problems: convex or non convex relaxation? *SIAM J. Imaging Sci.* 6:2603–2639
- [85] Perona P, Malik J (1990) Scale-space and edge detection using anisotropic diffusion. *IEEE Trans Pattern Anal Mach Intell PAMI*-12:629–639
- [86] Rockafellar RT, Wets JB (1997) *Variational analysis*. Springer, New York
- [87] Rudin L, Osher S, Fatemi C (1992) Nonlinear total variation based noise removal algorithm. *Physica* 60 D:259–268
- [88] Robini, M. and Magnin, I. (2010) Optimization by stochastic continuation. *SIAM J. Imaging Sci.*, 3:10961121
- [89] Robini, M. and Reissman, P.-J. (2013) From simulated annealing to stochastic continuation: a new trend in combinatorial optimization. *J. Global Optim* 56:185–215
- [90] Sauer K, Bouman C (1993) A local update strategy for iterative reconstruction from projections. *IEEE Trans Signal Process SP*-41:534–548
- [91] Scherzer O, Grasmair M, Grossauer H, Haltmeier M, Lenzen F (2009) *Variational problems in imaging*. Springer, New York
- [92] Tautenhahn U (1994) Error estimates for regularized solutions of non-linear ill posed problems. *Inverse Probl* 10:485–500
- [93] Tikhonov A, Arsenin V (1977) *Solutions of ill posed problems*, Winston, Washington
- [94] Vogel C (2002) *Computational methods for inverse problems*. *Frontiers in applied mathematics series*, vol 23. SIAM, New York
- [95] Welk M, Steidl G, Weickert J (2008) Locally analytic schemes: a link between diffusion filtering and wavelet shrinkage. *Appl Comput Harmon Anal* 24:195–224
- [96] Winkler G (2006) *Image analysis, random fields and Markov chain Monte Carlo methods. A mathematical introduction*. *Applications of mathematics*, 2nd edn, vol 27. Stochastic models and applied probability. Springer, Berlin