Functionals for signal and image reconstruction: properties of their minimizers and applications

Research report

Mila NIKOLOVA

Centre de Mathématiques et de Leurs Applications (CMLA) - CNRS-UMR 8536 - ENS de Cachan
61 av. du Président Wilson, 94235 Cachan Cedex, France

Contents

Chapter 1

Introduction

1.1 Focus on inverse problems

We mainly address inverse problems where an original unknown function $u_o \in \mathcal{M}$—an image or a signal—has to be recovered based on data $v \in \mathcal{N}$, related to it like

$$v = A(u_o) \odot n \tag{1.1}$$

where $A : \mathcal{M} \to \mathcal{N}$ is a transform, n stands for perturbation, and $\mathcal{M}$ and $\mathcal{N}$ are suitable real spaces. Continuous signals and images are defined on subsets $\Omega \subset \mathbb{R}^d$, for $d = 1, 2, ...$, $\mathcal{M}$ is a functional space on Ω (e.g. $L^\infty(\Omega)$, $W^{1,2}(\Omega)$, $BV(\Omega)$ and so on). For discrete signals and images, $\mathcal{M}$ is $\mathbb{R}^p$ or a manifolds in $\mathbb{R}^p$ where p is the number of the samples composing u. Data v can be given either on a subset $\Omega' \subset \mathbb{R}^d$, $d = 1, 2, ...$, or on a discrete grid, say of cardinality q, in which case $\mathcal{N}$ is $\mathbb{R}^q$ or a subset of it.

In general, the *observation operator* A captures the main deterministic physical phenomena in the chain between the unknown u_o and the data v, while n models the random phenomena like perturbations and noise that contribute to data v via an operation symbolized by "$\odot$". Sharing data v between an ideal $A(u_o)$ and noise n is rarely unambiguous and is part of the modelling. Furthermore, the conception of a model (1.1) is subjected to the material constraint that it gives rise to realist analytical and numerical calculations. Partly in connection with the latter, the most widely used model is

$$v = Au_o + n \tag{1.2}$$

where A is a linear operator and n is white Gaussian noise. Classical (and yet unsolved) problems are the denoising and the segmentation of signals and images which corresponds to $A = I$ (the identity operator). In segmentation, however, n is composed out of noise, textures and insignificant objects inside the regions. Transmission through noisy channels, or faulty cells in camera sensors or memories corrupt u_o with impulse noise where some of its samples are replaced by random numbers. Radar and ultrasound images are corrupted with speckle noise: at each data point v_i the noise multiplies a function of $A(u_o)$. To model the blurring of signals and images, A represents the convolution of u_o with a kernel. Linear operators A arise in super-resolution problems where several low-resolution data sets are used to compute a high-resolution solution. In biomedical and geophysical imaging u_o is the distribution of the density of a material and v are measurements that are indirectly related to u_o, via a linear integral operator A.

In emission and transmission computed tomography the samples of v—observed photon counts—follow Poissonian distributions with means determined by $A(u_o)$. These and many other observation models are described in the literature; some classical textbooks are [49, 53, 44, 15, 81, 5].

We do not consider the rare cases when the noise is negligible (i.e. $v \approx A(u_o)$) and A admits a stable inverse, so that one can find a good solution by taking $A^{-1}(v)$... Except when A is the identity, A usually introduces a loss of useful information which means that A is ill-conditioned or even non-injective. The reconstruction of an estimate $\hat{u}$ of u_o based on (1.1) involves two stages, namely (a) the definition of a meaningful solution $\hat{u}$ and (b) the conception of realist numerical method for its computation. In the following we address both questions.

The author's publications are listed in chapter 9 and reference to them is of the form [$\star$ X (year)].

1.2 Solutions as minimizers of cost-functions (energies)

The noise in (1.1) can be described only by some statistics, such as its mean and its variance. Recovering an estimate $\hat{u}$ of u_o from noisy data v naturally appeals to use statistics. To present the ideas, and in accordance with the history of reconstruction methods, let us consider the linear case (1.2) where n is white, zero-mean Gaussian noise with variance σ^2. The likelihood of v—the distribution of v conditionally to u—reads $\pi(v|u) = \exp\{-\frac{1}{2\sigma^2}\|Au - v\|^2\}/Z$ where $Z > 0$ is the normalization constant. The maximum likelihood estimator proposes $\hat{u} = \arg\max_{u \in \mathcal{M}} \pi(v|u)$, which is actually the least-squares solution:

$$\hat{u} = \arg\min_{u \in \mathcal{M}} \|Au - v\|^2,$$

where $\|.\|$ is the Euclidian norm. The solution $\hat{u}$ necessarily satisfies

$$A^* A \hat{u} = A^* v, \tag{1.3}$$

where A^* is the adjoint of A (= the transposed when A is a matrix). Such a solution is not satisfactory. In the case of denoising (where $v = u + n$), the noise remains intact since (1.3) leads to $\hat{u} = v$. In typical applications, solving (1.3) is an ill-posed problem since A^*A is not invertible or is ill-conditioned. It is well known that the solutions of (1.3) are unstable with respect to the noise and the numerical errors, see for instance [77, 31, 82].

In order to stabilize ill-posed inverse problems, Tikhonov and Arsenin [79] proposed to define $\hat{u}$ as the (unique) minimizer of a cost-function $\mathcal{F}(., v) : \mathcal{M} \to \mathbb{R}$, called often an *energy*:

$$\mathcal{F}(u, v) = \|Au - v\|^2 + \beta\|u\|^2 \quad \text{for} \quad \beta > 0. \tag{1.4}$$

The first term in $\mathcal{F}(., v)$ measures the fidelity of $A\hat{u}$ to data v whereas the second encourages $\hat{u}$ to be smooth; the trade-off between these two goals is controlled by β. Since then, it is well known that smoothing the noise by (1.4) is reached at the expense of a large β which leads to solutions $\hat{u}$ that are too flattened with respect to u_o. Nevertheless, the cost-function in (1.4) is at the origin of the very popular *regularization approach* [31, 5]. The idea is to define a solution $\hat{u}$ to (1.1) as

$$\begin{aligned} \hat{u} &= \arg\min_{u \in \mathcal{M}} \mathcal{F}(u, v), && (1.5) \\ \mathcal{F}(u, v) &= \Psi(u, v) + \beta\Phi(u), && (1.6) \end{aligned}$$

where $\Psi : \mathcal{M} \times \mathcal{N} \to \mathbb{R}$ is a data-fidelity term and $\Phi : \mathcal{M} \to \mathbb{R}$ is a regularization term, and $\beta > 0$ is a parameter. The term Ψ ensures that $\hat{u}$ is close enough to satisfying (1.1). Usual models consider that A is linear—as in (1.2)—and that Ψ is either the L_2 or the ℓ_2 norm (according to $\mathcal{N}$) of the residual,

$$\Psi(u, v) = \|Au - v\|^2. \tag{1.7}$$

More specific terms Ψ arise when the noise n is non-Gaussian or non-additive. The formulation in (1.6) marks an important progress with respect to (1.4) since Φ in (1.6) is asked to partly compensate for the loss of information on u_o entailed by the observation system A and the noise n. To this end, Φ is required to push $\hat{u}$ to exhibit some *a priori* expected features like smoothness, presence of edges or textures. Such information is essentially contained in the derivatives, or the finite differences, of u, of different orders, denoted $D^k u$, $k \geq 1$. Hence faithful recovery of derivatives, and more generally of high frequency components, is a permanent concern in signal and image reconstruction methods. A straightforward improvement of (1.4) in this direction is

$$\Phi(u) = \|Du\|^2.$$

Even if this Φ allows $\hat{u}$ to have larger variations, edges are flattened. Regularization terms giving rise to a better recovery of edges and other fine features have been designed basically in a variational or a statistical framework. Although these approaches are fundamentally different, they have led to very similar, and even the same, cost-functions. Unifying frameworks are considered for instance in [31, 55]. If u is defined on a subset $\Omega \subset \mathbb{R}^d$, $d = 1, 2$,

$$\Phi(u) = \int_\Omega \phi(|Du|)dx, \tag{1.8}$$

whereas if it is defined on a discrete grid,

$$\Phi(u) = \sum_{i=1}^{r} \phi(|D_i u|), \tag{1.9}$$

where $\{D_i : 1 \leq i \leq r\}$ is a set of difference operators. In both expressions, $|.|$ is an appropriate norm and $\phi : \mathbb{R}_+ \to \mathbb{R}$, often called a potential function (PF). Several choices for ϕ are given in Table 1.1 [41, 9, 64, 42, 11, 69, 74, 39, 3, 18, 14, 10, 82]. The most popular function ϕ now is certainly

$$\phi(t) = |t|, \tag{1.10}$$

proposed in [9] to form a *median pixel prior* and introduced as *total variational* (TV) regularization in [74], which allows a better recovery of the edges in signals and images. The way Φ is chosen in a statistical or a variational setting is discussed next.

1.2.1 Bayesian approach

Statistical approaches deal with discrete signals and images, so $\mathcal{M} = \mathbb{R}^p$ and $\mathcal{N} = \mathbb{R}^q$ (or possibly subsets of them), and assimilate the samples u_i of u with particles and $\mathcal{F}(., v)$ with an energy. The Bayesian approach provides a rich framework to derive energies $\mathcal{F}$ that combine observation models (1.1) with

Convex PFs			
Smooth at zero PFs		Nonsmooth at zero PFs	
(f1)	$\phi(t) = \vert t\vert^\alpha,\ 1 < \alpha \leq 2$	(f5)	$\phi(t) = \vert t\vert$
(f2)	$\phi(t) = \sqrt{\alpha + t^2}$		
(f3)	$\phi(t) = \log(\cosh(\alpha t))$		
(f4)	$\phi(t) = \vert t\vert/\alpha - \log(1 + \vert t\vert/\alpha)$		
Nonconvex PFs			
Smooth at zero PFs		Nonsmooth at zero PFs	
(f6)	$\phi(t) = \min\{\alpha t^2, 1\}$	(f10)	$\phi(t) = \vert t\vert^\alpha,\ 0 < \alpha < 1$
(f7)	$\phi(t) = \dfrac{\alpha t^2}{1 + \alpha t^2}$	(f11)	$\phi(t) = \dfrac{\alpha\vert t\vert}{1 + \alpha\vert t\vert}$
(f8)	$\phi(t) = \log(\alpha t^2 + 1)$	(f12)	$\phi(t) = \log(\alpha\vert t\vert + 1)$
(f9)	$\phi(t) = 1 - \exp(-\alpha t^2)$	(f13)	$\phi(0) = 0,\ \phi(t) = 1$ if $t \neq 0$

Table 1.1: Commonly used PFs ϕ where $\alpha > 0$ is a parameter.

prior knowledge on the unknown u_o. It is based on the posterior distribution of the random variable u, namely

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

where $\pi(u)$ is the prior model for u_o and $Z = \pi(v)$ can be seen as a normalization constant. The Maximum a posteriori (MAP) estimator defines the optimal solution $\hat{u}$ as

$$\begin{aligned} \hat{u} &= \arg\max_{u\in\mathcal{M}} \pi(u|v) && (1.11) \\ &= \arg\min_{u\in\mathcal{M}} \left(-\ln \pi(v|u) - \ln \pi(u)\right) && (1.12) \end{aligned}$$

Comparing the last expression with (1.6) shows that MAP is equivalent to minimizing $\mathcal{F}(., v)$ as given in (1.5)-(1.6) if one takes

$$\begin{aligned} \Psi(u, v) &= -\ln \pi(v|u), && (1.13) \\ \Phi(u) &= -\frac{1}{\beta} \ln \pi(u). && (1.14) \end{aligned}$$

By (1.13), the term Ψ is completely determined by A and the distribution of the noise. There is some freedom in the choice of Φ. Since [8, 41, 9, 38], the sought-after u can be modelled as the realization of a Markov random field. By the famous Hammersley-Clifford theorem, $\pi(u)$ can be put into the form $\pi(u) = \exp\{-\Phi(u)\}/Z$ where Φ can convey the local interactions between each sample of u and its neighbors. For instance, Φ can be of the form (1.9) for a “general” function ϕ. It may be useful to notice that by (1.9), the subsets of neighbors (the cliques) are given by $\{j : D_i[j] \neq 0\}$, for all i (here $D_i[j]$ denotes the jth entry of $D_i \in \mathbb{R}^p$). With this interpretation, (1.9) provides a powerful tool for modelling. Indeed, a large variety of potential functions ϕ and difference operators D_i have been proposed in the Bayesian setting [42, 9, 48, 46, 57, 39, 14]. However, it is not difficult to establish that the distribution of $\hat{u}$ can be very different from the prior model (cf. chapter 6), which is misleading in the applications.

1.2.2 Variational approach

This approach is naturally developed for continuous signals and images, so that Φ is of the form (1.8). Suppose that $|Du| = \sqrt{u_{x_1}^2 + u_{x_2}^2}$ and that Ψ is of the form (1.7). At a minimizer point $\hat{u}$, the Euler-Lagrange equation formally reads

$$\frac{A^*(Au-v)}{\beta} = \operatorname{div}\left(\frac{\phi'(|Du|)}{2|Du|}Du\right). \tag{1.15}$$

In case $A = I$, a possible approximation is

$$u_t \approx \frac{u-v}{\beta}$$

in which case

$$u_t \approx \operatorname{div}\left(\frac{\phi'(|Du|)}{2|Du|}Du\right).$$

Under this approximation, we have anisotropic diffusion filtering with diffusivity $\frac{\phi'(s)}{2s}$, initial condition $u_0 = v$ and time step β. This approximation is usually rough, since it suppose a linear evolution for u_t. It is justified only in several special cases, according to the analysis of Steidl, Weickert, Mrázek and Welk in [76].

An important step to better understand regularization was the observation that (1.15) can be rewritten as [5]

$$A^*Au - \frac{\beta}{2}\left(\frac{\phi'(|Du|)}{|Du|}u_{tt} + \phi''(|Du|)u_{nn}\right) = A^*v, \tag{1.16}$$

where u_{tt} and u_{nn} are the second derivatives of u in the directions that are tangential and normal, respectively, to the isophote lines of u (i.e. the lines along which $u(x)$ is constant). Inside homogeneous regions we have $\nabla u \approx 0$ so smoothing is achieved if $\lim_{t\to 0^+}\phi''(t) = \phi''(0) > 0$, since then (1.16) is approximated by

$$A^*Au - \frac{\beta}{2}\phi''(0)\Delta u = A^*v.$$

This is a uniformly elliptic equation known to perform smoothing in all directions. In contrast, near an edge $|\nabla u|$ is large and it could be preserved large in $\hat{u}$ if $\lim_{t\to\infty}\phi''(t) = 0$ so that $\phi''(|\nabla u|)u_{nn} \approx 0$, and $\lim_{t\to\infty}\phi''(t)t/\phi'(t) = 0$ in which case the first term between the parentheses in (1.16) remains non-zero. As the limiting case one finds the TV regularization method corresponding to (1.10) which was proposed in [74]. Let us notice that this reasoning is basically qualitative since the contribution of A along u_{tt} and u_{nn} is ignored. Results on PDEs cannot be transported directly on regularized cost-functions.

1.3 Our approach : the minimizers as a function of the energy

We can remark that both approaches consider the effect of the regularization separately from the data-fidelity term and tend to evade the optimization of the whole functional. In [$\star$68(1994)] we did some preliminary studies that clearly showed that minimizers do not behave as expected by the prior model. The question that arise then is on the way in which the priors are *effectively* involved in $\hat{u}$—a local or a global minimizer of an energy $\mathcal{F}(.,v)$. It led us to formulate a different problem which is *the analysis of the relationship between the shape of* $\mathcal{F}$*—hence the shapes of* Ψ *and* Φ*—and the characteristic features* *exhibited by its (global or local) minimizers* $\hat{u}$. This problem, that we considered for the first time in [$\star$7(1997)], reflects a different point of view which, to the best of our knowledge, has never been studied in a systematic way before. It is completely general since it can be considered for any functional $\mathcal{F}$. The interest is that it leads to rigorous results on the solutions $\hat{u}$. Its ultimate ambition is to provide simple rules connecting the shape of $\mathcal{F}(.,v)$ to the characteristic features exhibited by its minimizers $\hat{u}$. Such a knowledge enables a real control on the regularization methods that are used. On the other hand, it allows the conception of specialized energies $\mathcal{F}$ whose minimizers correctly reflect the a priori expectations. Furthermore, it gives precious indication for a relevant choice of the parameters (β and others) involved in $\mathcal{F}$. Last, it can be used to conceive more economic numerical methods by taking a benefit from the features of the solution that are known in advance. The main difficulties are (a) to extract and to formalize pertinent properties of minimizers $\hat{u}$ and (b) to conduct the analysis.

Studying the properties of the minimizers of cost-functions as a function of the shape of the cost-function is a widely open problem. We have explored only several directions relevant to the (non)smoothness and the (non)convexity of $\mathcal{F}$. Given the intrinsic difficulty of the problem, we did not hesitate to focus on signals and images defined on discrete finite grids. In this practical framework, our results are quite general since we consider functions $\mathcal{F}$ which can be convex or non-convex, smooth or non-smooth, and results address local and global minimizers. Not without surprise it appeared that irregular cost-functions were easier to analyze and led us to stronger results than regular ones.

1.4 Preliminary notions

In our studies, we analyze the behavior of the minimizer points $\hat{u}$ of $\mathcal{F}(.,v)$ under variations of v. In words, we consider local minimizer functions.

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

The next lemma, which can be found in [36], addresses the regularity of the local minimizer functions when $\mathcal{F}$ is smooth. We use this result several times in the following in order to justify our assumptions.

Lemma 1 *Let $\mathcal{F}$ be $\mathcal{C}^m$, $m\geq 2$, on a neighborhood of $(\hat{u},v)\in\mathcal{M}\times\mathcal{N}$. Suppose that $\mathcal{F}(.,v)$ reaches at $\hat{u}$ a local minimum such that $D_1^2\mathcal{F}(\hat{u},v)$ is positive definite. Then there are a neighborhood $O\subset\mathcal{N}$ containing v and a unique $\mathcal{C}^{m-1}$ strict local minimizer function $\mathcal{U}:O\rightarrow\mathcal{M}$, such that $D_1^2\mathcal{F}(\mathcal{U}(\nu),\nu)$ is positive definite for every $\nu\in O$ and $\mathcal{U}(v)=\hat{u}$.*

A special attention being dedicated to non-smooth functions, we recall some basic facts. In what follows, M will denote the tangent of $\mathcal{M}$.

Definition 2 *Given $v\in\mathcal{N}$, the function $\mathcal{F}(.,v):\mathcal{M}\rightarrow\mathbb{R}$ admits at $\hat{u}\in\mathcal{M}$ a one-sided derivative in a direction $w\in M$, denoted $\delta_1\mathcal{F}(\hat{u},v)(w)$, if the following limit exists:*

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

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

In fact, $\delta_1\mathcal{F}(\hat{u}, v)(w)$ is a *right-side* derivative; the relevant *left-side* derivative is $-\delta_1\mathcal{F}(\hat{u}, v)(-w)$. If $\mathcal{F}(., v)$ is differentiable at $\hat{u}$ for w, then $\delta_1\mathcal{F}(\hat{u}, v)(w) = D_1\mathcal{F}(\hat{u}, v)w$. Next we recall the classical necessary condition for a local minimum of a possibly non-smooth function [72, 50, 26].

Lemma 2 *If $\mathcal{F}(., v)$ has a local minimum at $\hat{u} \in \mathcal{M}$, then $\delta_1\mathcal{F}(\hat{u}, v)(w) \geq 0$, for every $w \in M$.*

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

Chapter 2

Non-smooth regularization to recover strongly homogeneous regions

2.1 Main theoretical result

In [⋆2(2004)] we consider the general question of how, given a set of linear operators and vectors, denoted G_i and $\theta_i \in \mathbb{R}^s$ for $i = 1, \ldots, r$, the shape of $\mathcal{F}$ *favors*, or conversely *inhibits*, the possibility that for some $v \in \mathcal{N}$, the function $\mathcal{F}(., v)$ admits a (local) minimizer $\hat{u} \in \mathcal{M}$ satisfying

$$G_i \hat{u} = \theta_i, \quad \forall i \in \hat{J}, \quad \text{with} \quad \hat{J} \subset \{1, \ldots, r\} \quad \text{and} \quad \hat{J} \neq \emptyset, \tag{2.1}$$

where

$$\mathcal{M} = \{u \in \mathbb{R}^p : Cu = b\}, \quad C \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^{p_0}) \text{ and } b \in \mathbb{R}^{p_0}, \quad p_0 < p.$$

The samples $\hat{u}_i$ of $\hat{u}$ involved in (2.1) are said to form *strongly homogeneous regions*. Although [⋆2(2004)] addresses very general functions $\mathcal{F}$, we present our results for functions $\mathcal{F}$ of the form (1.6) along with

$$\Phi(u) = \sum_{i=1}^{r} \varphi_i(G_i u - \theta_i), \tag{2.2}$$

where for every $i = 1, \ldots, r$, the function $\varphi_i : \mathbb{R}^s \to \mathbb{R}$ is $\mathcal{C}^m$, $m \geq 2$ on $\mathbb{R}^s \setminus \{0\}$ while at zero φ_i is non-smooth, has a strict minimum and has a one-sided derivative application $w \to \delta\varphi_i(0)(w)$ which is uniform on the unit sphere S. The most typical form for the regularization term Φ in (2.2) corresponds with

$$\varphi_i(z) = \varphi(z) = \phi\left(\|z\|\right), \quad \forall i \in \{1, \ldots, r\}, \tag{2.3}$$

where $\phi : \mathbb{R}_+ \to \mathbb{R}$ is $\mathcal{C}^m$ for $m \geq 2$ and $\phi'(0) > 0$. Some examples of such functions ϕ are (f5) and (f10)-(f12) given in Table 1.1. The data-fidelity term $\Psi : \mathcal{M} \times \mathcal{N} \to \mathbb{R}$ in (1.6) is any explicit or implicit $\mathcal{C}^m$-function, $m \geq 2$. In particular, our formulation allows us to address cost-functions combining both smooth and non-smooth regularization terms.

Let us define the set-valued function $\mathcal{J}$ on $\mathcal{M}$ by

$$\mathcal{J}(u) = \Big\{ i \in \{1, \ldots, r\} \; : \; G_i u = \theta_i \Big\}. \tag{2.4}$$

Theorem 1 *For $v \in \mathbb{R}^q$ given, let $\hat{u} \in \mathcal{M}$ be a solution to (1.5). For $\hat{J} = \mathcal{J}(\hat{u})$, let $\mathcal{K}_{\hat{J}}$ be the subspace*

$$\mathcal{K}_{\hat{J}} = \left\{ u \in \mathcal{M} : G_i u = \theta_i, \forall i \in \hat{J} \right\} \neq \mathcal{M} \tag{2.5}$$

and $K_{\hat{J}}$ be its tangent. For $O \subset \mathbb{R}^q$ suppose that

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

(b) $\mathcal{F}|_{\mathcal{K}_{\hat{J}}}(., O)$ has a local minimizer function $\mathcal{U}_{\hat{J}} : O \to \mathcal{K}_{\hat{J}}$ which is continuous at v and $\hat{u} = \mathcal{U}_{\hat{J}}(v)$.

Then there is a neighborhood $O_{\hat{J}}$ of v such that $\mathcal{F}(., O_{\hat{J}})$ admits a local minimizer function $\mathcal{U} : O_{\hat{J}} \to \mathcal{M}$ which satisfies $\mathcal{U}(v) = \hat{u}$ and

$$\nu \in O_{\hat{J}} \quad \Rightarrow \quad \mathcal{U}(\nu) = \mathcal{U}_{\hat{J}}(\nu) \in \mathcal{K}_{\hat{J}} \quad \text{and} \quad G_i \mathcal{U}(\nu) = \theta_i, \ \text{for all } i \in \hat{J}.$$

If $\mathcal{U}_{\hat{J}}$ is a strict local minimizer function for $\mathcal{F}|_{\mathcal{K}_{\hat{J}}}(., O)$, then $\mathcal{U}$ is a strict local minimizer function for $\mathcal{F}(., O_{\hat{J}})$.

We show that the results of Theorem 1 holds also for irregular functions ϕ of the form (f13) in Table 1.1.

Commentary on the assumptions. Since $\mathcal{F}(., v)$ has a local minimum at $\hat{u}$, Lemma 2 shows that $\delta_1 \mathcal{F}(\hat{u}, v)(w) \geq 0$, for all $w \in M$ and this inequality cannot be strict unless $\mathcal{F}$ is non-smooth. When Φ is non-smooth as specified above, it is easy to see that (a) is not a strong requirement. By Lemma 1, condition (b) holds if $\mathcal{F}|_{\mathcal{K}_{\hat{J}}}$ is $\mathcal{C}^m$ on a neighborhood of $(\hat{u}, v)$ belonging to $\mathcal{K}_{\hat{J}} \times \mathcal{N}$, and if $D_1(\mathcal{F}|_{\mathcal{K}_{\hat{J}}})(\hat{u}, v) = 0$ and $D_1^2(\mathcal{F}|_{\mathcal{K}_{\hat{J}}})(\hat{u}, v)$ is positive definite, which is usually satisfied for the convex cost-functions $\mathcal{F}$ used in practice.

If Ψ is of the form (1.7) with A^*A invertible and if $\{\varphi_i\}$ are non-convex according to some non-restrictive assumptions, the analysis in [$\star$9(2005)] shows that for almost every v (except those contained in a negligible subset of $\mathbb{R}^q$), every local minimizer $\hat{u}$ of $\mathcal{F}(., v)$ satisfies both (a) and (b)—see Remark 3 in § 5.1.1 in this manuscript.

Significance of the results. The conclusion of the theorem can be reformulated as

$$v \in O_{\hat{J}} \quad \Rightarrow \quad \mathcal{J}\left(\mathcal{U}(v)\right) \supseteq \hat{J} \quad \Leftrightarrow \quad \mathcal{U}(v) \in \mathcal{K}_{\hat{J}}, \tag{2.6}$$

where $\mathcal{J}$ is defined in (2.4). The latter is a severe restriction since $\mathcal{K}_{\hat{J}}$ is a closed and negligible subset of $\mathcal{M}$ whereas data v vary on open subsets of $\mathcal{N}$. (The converse situation where a local minimizer $\hat{u}$ of $\mathcal{F}(., v)$ satisfies $G_i \hat{u} \neq \theta_i$, for all i seems quite natural.) Observe also that there is an open subset $\tilde{O}_{\hat{J}} \subset O_{\hat{J}}$ such that $\mathcal{J}\left(\mathcal{U}(v)\right) = \hat{J}$ for all $v \in \tilde{O}_{\hat{J}}$.

Focus on a minimizer function $\mathcal{U} : O \to \mathcal{M}$ for $\mathcal{F}(., O)$ and put $\hat{J} = \mathcal{J}(\mathcal{U}(v))$ for some $v \in O$. By Theorem 1 the sets $O_{\hat{J}}$ and $\tilde{O}_{\hat{J}}$ are of positive measure in $\mathcal{N}$. The chance that random points ν (*e.g.* noisy data) come across $O_{\hat{J}}$, or $\tilde{O}_{\hat{J}}$, is real. When data ν range over O, the set-valued function $(\mathcal{J} \circ \mathcal{U})$ generally takes several distinct values, say $\{J_j\}$. Thus, with a 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 $\mathcal{J}(\hat{u}) = J_j$. This is the reason why non-smooth cost-functions, as those

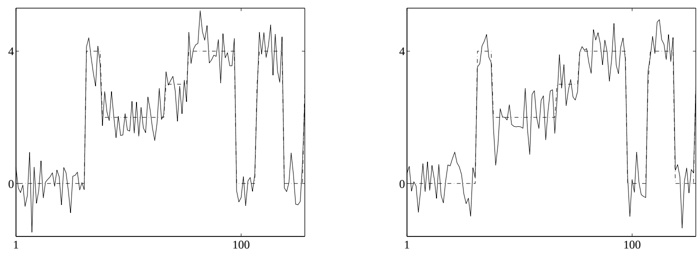
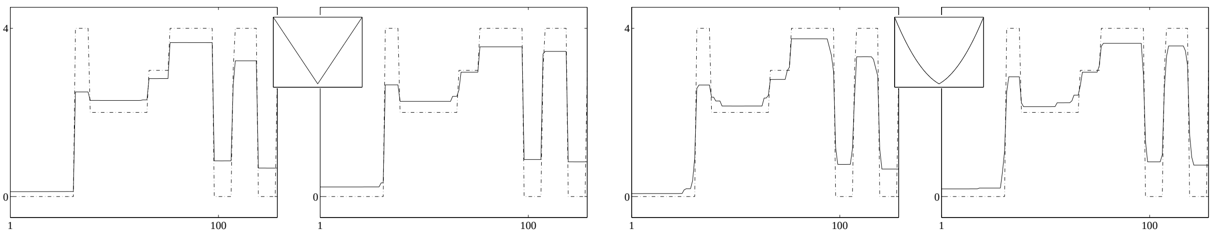


Figure 2.1: Data $v = u_o + n$ (—), corresponding to the original u_o (-.-.), contaminated with two different noise samples n in the left and in the right.

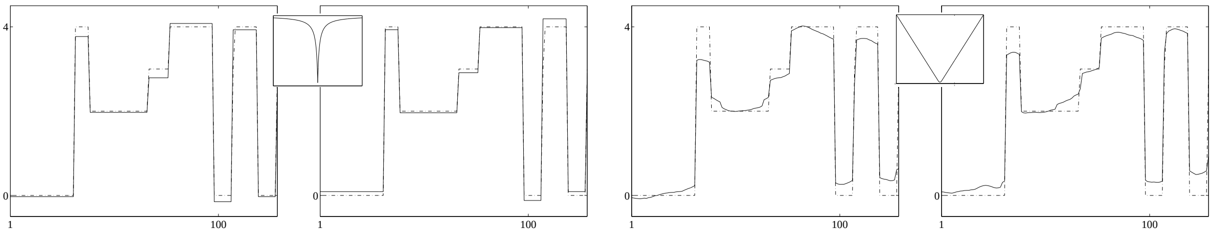
considered here, exhibit local minimizers which generically satisfy constraints of the form (2.1). For a regularized cost-function of the form defined by (1.6) and (2.2), with $\{\varphi_i\}$ as in (2.3), $\{G_i\}$ first-order difference operators and $\theta_i = 0$, for all i, minimizers $\hat{u}$ are typically constant on many regions. This explains in particular the *stair-casing effect* observed in total-variation (TV) methods [74, 32, 22]. Such minimizers can be seen for signals in Figs. 2.2 (a), (b) and (c), and for images in Fig. 2.4 (b) and in Fig. 2.5 (c) and (d). Initially we explained this phenomenon in a more restricted context in [$\star$7(1997)] and [$\star$5(2000)]. These papers provided the first mathematical explanation of stair-casing in the literature.

Restoration of a noisy signal. In Figs. 2.1 and 2.2 we consider the restoration of a piecewise constant signal u_o from noisy data $v = u_o + n$ by minimizing $\mathcal{F}(u,v) = \|u - v\|^2 + \beta \sum_{i=1}^{p-1} \phi(|u_i - u_{i+1}|)$. In this case, the strongly homogeneous regions are constant, $\left\{i \in \{1,\dots,p-1\} : \hat{u}_i = \hat{u}_{i+1}\right\}$. In order to evaluate the ability of different functions ϕ to recover, and to conserve, the strongly homogeneous zones yielded by minimizing the relevant $\mathcal{F}(.,v)$, we process in the same numerical conditions two data sets, contaminated by two very different noise realizations shown in Fig. 2.1. The minimizers shown in Figs. 2.2(a), (b) and (c) correspond to functions ϕ that are non-smooth at zero. In accordance with our theoretical results, they are constant on large segments. In each one of these figures, the reader is invited to compare the subsets where the minimizers corresponding to the two data sets in Fig 2.1, are constant. In contrast, the function ϕ in Fig. 2.2(d) is smooth and the resultant minimizers in are nowhere constant. This will be explained in § 2.3.

Deblurring of an image from noisy data. The original image u_o in Fig. 2.3(a) presents smoothly varying regions, constant regions and sharp edges. Data in Fig. 2.3(b) correspond to $v = a * u_o + n$, where a is a blur with entries $a_{i,j} = \exp\left(-(i^2 + j^2)/12.5\right)$ for $-4 \le i, j \le 4$, 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, $\{G_i : 1 \le i \le r\}$ correspond to the first-order differences of each pixel with its 8 nearest neighbors and $\theta_i = 0$; again, the strongly homogeneous regions are constant. In all figures, the obtained minimizers are displayed on the top. Below we give two sections of the restored images, corresponding to rows 54 and 90 where the relevant sections of the original image are plotted with a dotted line. The minimizers corresponding to non-convex functions ϕ are calculated using a generalized graduated non-convexity method developed in [$\star$8 (1999)] and briefly described in § 7.1.

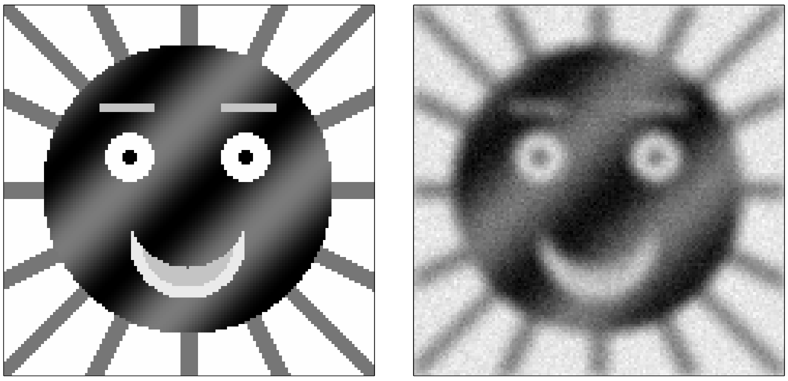


(a) $\phi(t) = |t|$ (b) $\phi(t) = \big(t - \alpha \mathrm{sign}(t)\big)^2$ (nonsmooth at 0)



(c) $\phi(t) = \alpha|t|/(1+\alpha|t|)$ (d)$\phi(t) = \begin{cases} \frac{t^2}{2} & \text{if } |t| \le \alpha \\ \alpha|t| - \frac{\alpha^2}{2} & \text{if } |t| > \alpha \end{cases}$ (smooth at 0)

Figure 2.2: Restoration using different functions ϕ. Original u_o (-.-.), minimizer $\hat{u}$ (—). Each figure from (a) to (d) shows the two minimizers $\hat{u}$ corresponding to the two data sets in Fig. 2.1 (left and right), while the shape of ϕ is plotted in the middle.



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

Figure 2.3: 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 Figs. 2.4 (b) and 2.5(c)-(d) correspond to non-smooth regularization and they involve large constant regions. No constant regions are observed in the other restorations which correspond to smooth regularization.

2.2 The 1D total variation regularization

The example below describes the sets $\tilde{O}_J$, for *every* $J \subset \{1, \ldots, r\}$, in the context of the one-dimensional discrete TV regularization. It provides a rich geometric interpretation of Theorem 1. Let $\mathcal{F} : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ be given by

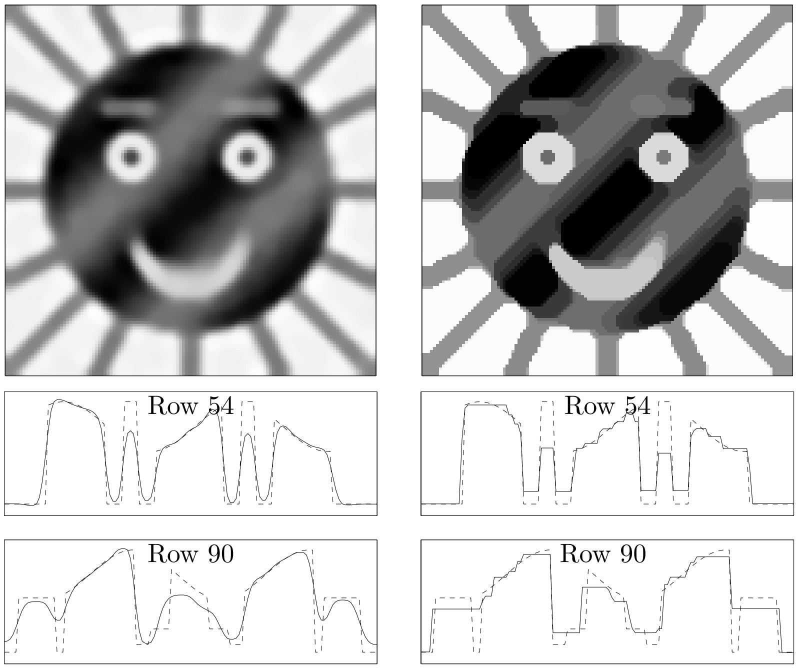
$$\mathcal{F}(u, v) = \|Au - v\|^2 + \beta \sum_{i=1}^{p-1} |u_i - u_{i+1}|, \tag{2.7}$$

where $A \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^p)$ is invertible and $\beta > 0$. It is easy to see that there is a unique minimizer function $\mathcal{U}$ for $\mathcal{F}(., \mathbb{R}^p)$. In [$\star$2(2004)] we exhibit two striking phenomena:

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 $\mathcal{F}(., v)$;

2. For every $J \subset \{1, \ldots, p-1\}$, there is a subset $\tilde{O}_J \subset \mathbb{R}^p$, composed of $2^{p-\#J-1}$ unbounded polyhedra (of dimension p) of $\mathbb{R}^p$, such that for every $v \in \tilde{O}_J$, the minimizer $\hat{u}$ of $\mathcal{F}(., v)$ satisfies $\hat{u}_i = \hat{u}_{i+1}$ for all $i \in J$ and $\hat{u}_i \neq \hat{u}_{i+1}$ for all $i \in J^c$. A description of these polyhedra is given in the appendix of [$\star$2(2004)]. It also shows how the parameter β works. Moreover, their closure forms a covering of $\mathbb{R}^p$.

Fast minimization method for (2.7) Since the discussion is on 1D TV, let us mention that in [$\star$3(2004)] we propose a fast numerical scheme to minimize (2.7). Consider the change of variables $z = Tu$ defined by $z_i = u_i - u_{i+1}$ for $1 \leq i \leq p-1$ and $z_p = \frac{1}{p}\sum_{i=1}^{p} u_i$, and put $B = AT^{-1}$ whose columns are denoted b_i, $1 \leq i \leq p$. At each iteration $k \in \mathbb{N}$, we do the following:



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

Figure 2.4: Restoration using convex PFs. Left: smooth at zero PF. Right: non-smooth at zero PF.

- for every $i \in \{1, \ldots, p-1\}$, calculate

$$\xi_i^{(k)} = 2\mathrm{b}_i^* B\left(z_1^{(k)}, z_2^{(k)}, \ldots, z_{i-1}^{(k)},\ 0,\ z_{i+1}^{(k-1)}, \ldots, z_p^{(k-1)}\right) - 2\mathrm{b}_i^* v,$$

then

$$\begin{array}{llll} \text{if} & \left|\xi_i^{(k)}\right| \leq \beta, & \text{set} & z_i^{(k)} = 0, \\ \text{if} & \xi_i^{(k)} < -\beta, & \text{set} & z_i^{(k)} = -\dfrac{\xi_i^{(k)} + \beta}{2\|\mathrm{b}_i\|^2} > 0, \\ \text{if} & \xi_i^{(k)} > \beta, & \text{set} & z_i^{(k)} = -\dfrac{\xi_i^{(k)} - \beta}{2\|\mathrm{b}_i\|^2} < 0; \end{array}$$

- for $i = p$,

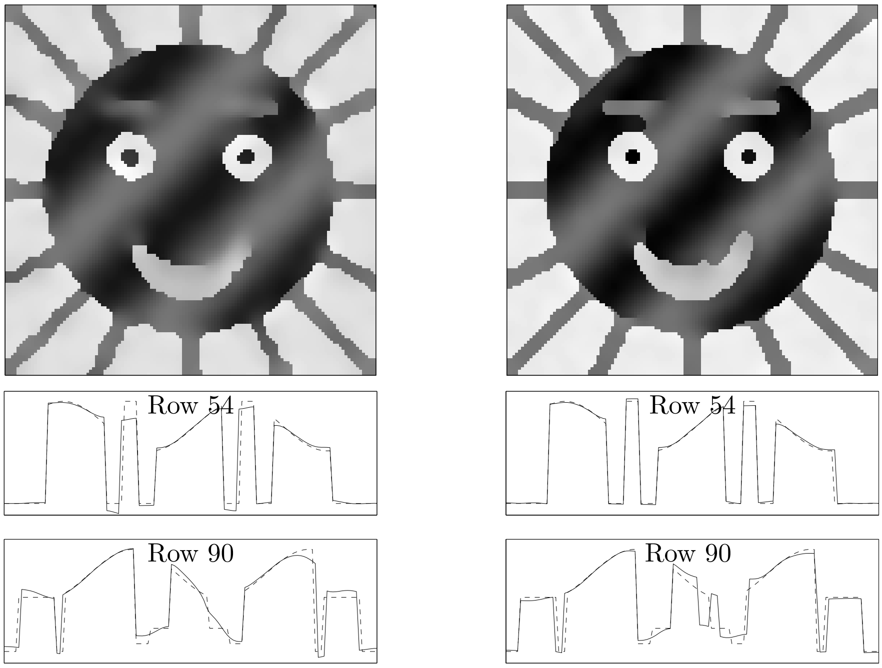
$$z_p^{(k)} = -\frac{\xi_p^{(k)}}{2\|\mathrm{b}_p\|^2}.$$

The convergence of $T^{-1}z^{(k)}$ towards the sought-after minimizer $\hat{u}$ of (2.7) is established in a more general context in [$\star$3(2004)]. Notice that this method cannot be extended to images.

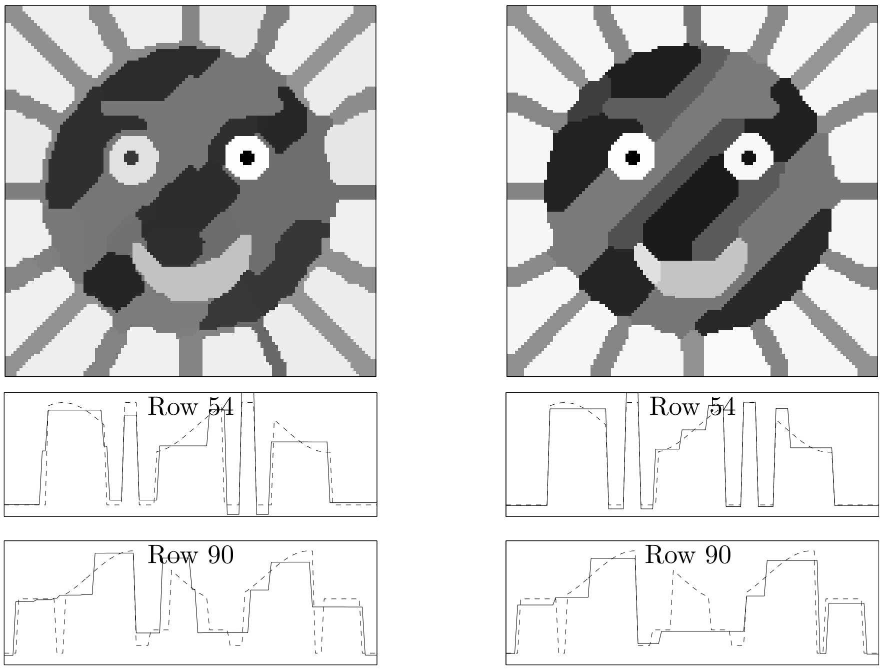
2.3 Comparison with smooth cost-functions

Here we explain that the special properties exhibited in § 2.1 do almost never occur if $\mathcal{F}$ is smooth.

Theorem 2 *Let $\mathcal{U} : O \to \mathcal{M}$ be a differentiable local minimizer function for $\mathcal{F}(., O)$ where $O \subset \mathcal{N}$ is open. Suppose that $\mathcal{F}$ is twice differentiable on $O_\mathcal{U} \times O$ where $O_\mathcal{U} \subset \mathcal{M}$ is open and $\mathcal{U}(O) \subset O_\mathcal{U}$. Let*



(a) $\varphi(t) = \dfrac{\alpha t^2}{1+\alpha t^2}$ for $\alpha = 25$, $\beta = 35$ (b) $\varphi(t) = \min\{\alpha t^2, 1\}$ for $\alpha = 60$, $\beta = 10$



(c) $\varphi(t) = \dfrac{\alpha |t|}{1+\alpha |t|}$ for $\alpha = 20$, $\beta = 100$ (d) $\varphi(t) = 1 - \mathbb{1}_{(t=0)}$ for $\beta = 25$

Figure 2.5: Restoration using non-convex PFs. First row ((a) and (b)): smooth at zero PFs. Second row ((c) and (d)): non-smooth at zero PFs.

$\theta \in \mathbb{R}^s$ and G be such that

$$\operatorname{rank} G > p - \operatorname{rank} D_{12}\mathcal{F}(u,v), \ \forall (u,v) \in O_{\mathcal{U}} \times O, \tag{2.8}$$

where p is the dimension of $\mathcal{M}$. For any $\mathcal{O} \subseteq O$ whose interior is nonempty define the subset $V_{\mathcal{O}}$ by

$$V_{\mathcal{O}} = \{v \in \mathcal{O} : G\mathcal{U}(v) = \theta\}. \tag{2.9}$$

Then

(i) the interior of $V_{\mathcal{O}}$ is empty;

(ii) if $\mathcal{U}$ is $\mathcal{C}^1$ on O, for any compact $\mathcal{O} \subset O$, the relevant $V_{\mathcal{O}}$ is included in a closed, negligible subset of $\mathcal{N}$.

If $\mathcal{F}$ is of the form (1.6)-(1.7) with A^*A invertible, the requirement in (2.8) is simplified to $\operatorname{rank} G > 0$.

We show that the same result holds if ϕ is non-smooth in such a way that $\phi'(\tau_i^-) > \phi'(\tau_i^+)$ for several nonzero points τ_i (see Proposition 2 and Theorem 4 in [$\star$2(2004)]). The most famous function ϕ of this kind is the truncated quadratic function,

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

This function was initially introduced in [41] for the restoration of images involving sharp edges. It can also be seen as the discrete version of the Mumford-Shah regularization [64] and was considered in a very large amount of papers [11, 19]. It is non-smooth at $1/\sqrt{\alpha}$ and $\mathcal{C}^\infty$ elsewhere.

Example 1 (Quadratic regularized cost-function) Consider the function $\mathcal{F} : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$,

$$\mathcal{F}(u,v) = \|Au - v\|^2 + \beta \|Gu\|^2, \tag{2.10}$$

where $\beta > 0$ and $G \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}^s)$. Under the trivial assumption $\ker(A^*A) \cap \ker(G^*G) = \{0\}$, for every $v \in \mathbb{R}^q$, the function $\mathcal{F}(\,.\,, v)$ is strictly convex and its unique minimizer function $\mathcal{U} : \mathbb{R}^q \to \mathbb{R}^p$ reads

$$\mathcal{U}(v) = (A^*A + \beta G^*G)^{-1} A^* v. \tag{2.11}$$

Let the rows of G be denoted $G_i \in \mathcal{L}(\mathbb{R}^p, \mathbb{R})$ for $i = 1, \ldots, r$. For a given $i \in \{1, \ldots, r\}$, the set $V_{\{i\}}$ of *all* data points $v \in \mathbb{R}^q$ for which $\mathcal{U}(v)$ satisfies exactly $G_i \mathcal{U}(v) = 0$ reads

$$\begin{aligned} V_{\{i\}} &= \{v \in \mathbb{R}^q : \langle v, p_i(\beta) \rangle = 0\} = \{p_i(\beta)\}^\perp, \\ p_i(\beta) &= A(A^*A + \beta G^*G)^{-1} G_i^*. \end{aligned}$$

We can have $p_i(\beta) = 0$ only if $\operatorname{rank} A < p$ and if β is such that $G_i^* \in \ker A(A^*A + \beta G^*G)^{-1}$: if there are $\beta > 0$ satisfying this system, they form a finite, discrete set of values. However, β in (2.10) will almost never belong to such a set, so in general, $p_i(\beta) \neq 0$. Then $V_{\{i\}} \subset \mathbb{R}^q$ is an affine subspace of dimension $q - 1$. More generally, we have the implication

$$\exists i \in \{1, \ldots, r\} \ \text{ such that } \ G_i \mathcal{U}_i(v) = 0 \quad \Rightarrow \quad v \in \bigcup_{i=1}^{r} V_{\{i\}}.$$

The union on the right side is of Lebesgue measure zero in $\mathbb{R}^q$. $\diamond$

2.4 Some applications

2.4.1 Segmentation of signals and images

When Φ is non-smooth as specified in §2.1, the function $\mathcal{J}$ defined in 2.4 naturally provides a classification rule if for any $u \in \mathcal{M}$ we consider the two classes $\hat{J} = \mathcal{J}(u)$ and $\hat{J}^c$, its complement in $\{1, \ldots, r\}$. If for every $i \in \{1, \ldots, r\}$ we have $\theta_i = 0$ and G_i yields either the first-order differences or the discrete approximation of the gradient at i, then $\mathcal{J}(\hat{u})$ addresses the constant regions in $\hat{u}$. Since $\mathcal{J}(\hat{u})$ is usually nonempty and large, $\hat{u}$ provides a segmentation of the original u_o. This is nicely illustrated in Figs. 2.4(b) and 2.5 (c)-(d). One can notice that this segmentation is neater if ϕ is non-convex rather then if it is convex—an explanation is provided by Theorem 9 below. By Theorem 1, this classification is *stable* with respect to small variations of the data (*e.g.* due to noise perturbations).

2.4.2 Restoration of binary images using convex energies

It is well known that no cost-function defined on the set of the binary images can be convex. The usual non-convex cost-functions for binary images are difficult to minimize while approximate solutions are very often of limited interest. On the other hand, general-purpose convex cost-functions yield continuous-valued smooth estimates which are far from being satisfying.

In [⋆60(1998)] we propose an alternative approach which is to construct *convex cost-functions* whose minimizers are continuous-valued but have a quasi-binary shape. For instance,

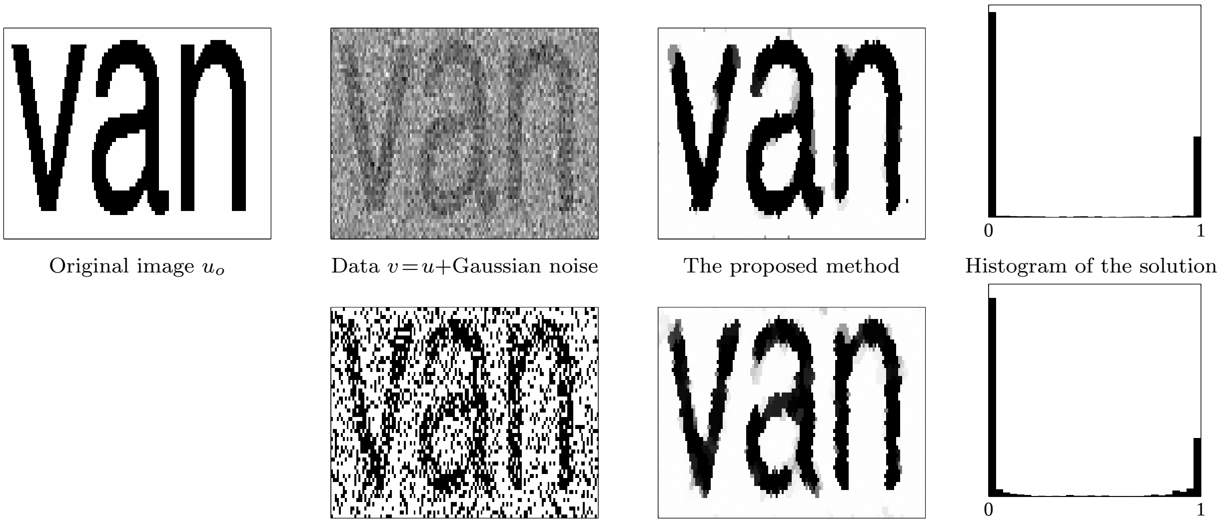
$$\mathcal{F}(u,v) = \|Au - v\|^2 - \alpha \sum_{i=1}^{p} \left(u_i - \frac{1}{2}\right)^2 + \beta \sum_{i \sim j} |u_i - u_j| \tag{2.12}$$

$$\text{subject to } u_i \in [0,1], \quad \forall i \in \{1, \ldots, p\} \tag{2.13}$$

where $i \sim j$ means that u_i and u_j are neighbors and $\alpha \geq 0$ and $\alpha \overset{<}{\approx} \min\{\lambda = \text{eigenvalue of } A^*A\}$, so that $\mathcal{F}(., v)$ is convex. By the concave term, pixels $\hat{u}_i$ are discouraged to be in the interior of $(0, 1)$ and by the constraint they cannot be outside $[0, 1]$, while via the regularization term, neighboring pixels are likely to be equal. The obtained solutions are almost binary, as seen in Fig. 2.6.

2.4.3 Blind deconvolution of binary signals

In collaboration with Florence Alberge and Pierre Duhamel, [⋆18(2002)] and [⋆19(2005)], we focus on different aspects of a blind deconvolution problem where the goal is to identify a channel and to estimate the binary symbols transmitted through it, based on the outputs of several arrays of censors. The main novelty of our approach is to introduce a continuous non-smooth cost-function which is convex with respect to the symbols and discourages them from taking non-binary values. The resultant method is robust and of low complexity.



Original image u_o | Data $v = u$+Gaussian noise | The proposed method | Histogram of the solution

Data with salt-and-pepper noise | The proposed method | Histogram of the solution

Figure 2.6: Restoration of quasi-binary images by minimizing (2.12).

Chapter 3

Non-smooth data-fidelity to fit exactly part of the data entries

In [$\star$4(2002)] and [$\star$3(2004)] we consider that in (1.6), the data-fidelity Ψ is non-smooth and that the regularization Φ is smooth. This situation is "dual" to the one considered in chapter 2. More specifically, let $\mathcal{M} = \mathbb{R}^p$ and $\mathcal{N} = \mathbb{R}^q$, and

$$\mathcal{F}(u,v) = \Psi(u,v) + \beta\Phi(u), \tag{3.1}$$

$$\Psi(u,v) = \sum_{i=1}^{q} \psi_i\left(\langle a_i, u\rangle - v_i\right), \tag{3.2}$$

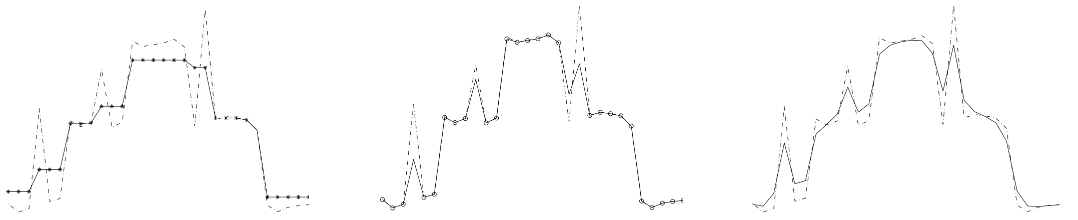
where $\psi_i : \mathbb{R} \to \mathbb{R}$, $i = 1, \ldots, q$, are continuous functions which decrease on $(-\infty, 0]$ and increase on $[0, +\infty)$, and a_i^* for $i = 1, \ldots, q$, are the rows of the observation matrix A. One usual choice is $\psi(t) = |t|^\rho$, for $\rho > 0$, which yields $\Psi(u,v) = \sum_{i=1}^{q} |\langle a_i, u\rangle - v_i|^\rho$ [71, 2]. We have already mentioned that the most often, $\Psi(x,y) = \|Ax - y\|^2$, that is, $\psi(t) = t^2$. Recall that many papers are dedicated to the minimization of this $\Psi(.,y)$ *alone* i.e., $\mathcal{F} = \Psi$, mainly for $\psi(t) = t^2$ [52], in some cases for $\psi(t) = |t|$ [12], but also $\psi(t) = |t|^\rho$ for $\rho \in (0,\infty]$ [71, 70]. Until our work (2002 and 2004), non-smooth data-fidelity terms in regularized cost-functions of the form (1.6) were very unusual and only Alliney [1], who considered $\mathcal{F}(u,v) = \sum_{i=1}^{p} |u_i - v_i| + \beta \sum_{i=1}^{p-1}(u_i - u_{i+1})^2$, made an exception. Nonsmooth data-fidelity terms Ψ were systematically avoided in image processing, for instance. Following our example, L_1-TV cost-functions were analyzed in [20] and in [$\star$22(2004)]. Many other papers considered cost-functions with L_1 data-fidelity terms later on.

3.1 Main theoretical results

We suppose that $\{\psi_i\}$ are $\mathcal{C}^m$-smooth on $\mathbb{R} \setminus \{0\}$, $m \geq 2$, whereas at zero their side derivatives satisfy

$$-\infty < \psi_i'(0^-) < \psi_i'(0^+) < +\infty \tag{3.3}$$

For simplicity, we consider that $\psi_i = \psi$ for all i. The term $\Phi : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ in (3.1) is any $\mathcal{C}^m$-function, which in particular may depend on data v as well. There is a striking distinction in the behavior of the minimizers relevant to non-smooth data-fidelity terms with respect to non-smooth regularization, as it



$\mathcal{F}(u,v) = \|u-v\|_2^2 + \beta\|Gu\|_1$ (a) Stair-casing $\quad$ $\mathcal{F}(u,v) = \|u-v\|_1 + \beta\|Gu\|_2^2$ (b) Exact data-fit $\quad$ $\mathcal{F}(u,v) = \|u-v\|_2^2 + \beta\|Gu\|_2^2$ (c)

Figure 3.1: G is a first-order difference operator, $[Gu]_i = u_i - u_{i+1}$, $i = 1,\ldots,p-1$. Data (- - -), Restored signal (—-).

can be seen in Fig. 3.1. It concerns the possibility of fitting *exactly* a certain number of the data entries, i.e., that for v given, a local minimizer $\hat{u}$ of $\mathcal{F}(.,v)$ satisfies $\langle a_i, \hat{u}\rangle = v_i$ for some, or even for many, indexes i. To analyze such a property we define $\mathcal{J}$ to be the following set-valued function:

$$(u,v) \quad \rightarrow \quad \mathcal{J}(u,v) = \Big\{ i \in \{1,\ldots,q\} : \langle a_i, u\rangle = v_i \Big\}. \tag{3.4}$$

Given v and a local minimizer $\hat{u}$ of $\mathcal{F}(.,v)$, the set of all data entries which are fitted exactly by $\hat{u}$ reads $\hat{J} = \mathcal{J}(\hat{u},v)$. Our main result, established in [$\star$4(2002)] is presented below.

Theorem 3 *Given $v \in \mathbb{R}^q$ and $\hat{u} \in \mathbb{R}^p$, for $\hat{J} = \mathcal{J}(\hat{u},y)$, let*

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

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

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

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

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

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

$$\nu \in O \quad \Rightarrow \quad \left\{ \begin{array}{lll} \langle a_i, \mathcal{U}(\nu)\rangle = \nu_i & \textit{if} & i \in \hat{J}, \\ \langle a_i, \mathcal{U}(\nu)\rangle \neq \nu_i & \textit{if} & i \in \hat{J}^c. \end{array} \right. \tag{3.5}$$

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

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.

Commentary on the assumptions. Assumption (a) does not require the independence of the whole set $\{a_i : i \in \{1,\ldots,q\}\}$. We show (Remark 6 in [$\star$4(2002)]) 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. As stated in § 5.1 hereafter, this assumption may fail to hold only for some v belonging to a negligible subset of $\mathbb{R}^q$.

If an arbitrary function $\mathcal{F}(.,v) : \mathbb{R}^p \to \mathbb{R}$ has a minimum at $\hat{u}$, then necessarily $\delta_1\mathcal{F}(\hat{u},v)(w) \geq 0$ for all $w \in T_{\hat{J}}^{\perp} \cap S$ [73]. In comparison, (c) requires only that the latter inequality be strict. One can easily conjecture that this inequality can be non-strict only for some data v included in a negligible subset of $\mathbb{R}^q$; the arguments are similar to those involved in the results summarized in Remark 3 in § 5.1.1.

Significance of the results. The result in (3.5) means that the set-valued function $v \to \mathcal{J}(\mathcal{U}(v), v)$ is constant on O, 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. Intuitively, this may seem unlikely, especially for noisy data.

Theorem 3 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 (e.g., for A invertible, $\beta = 0$ yields $\hat{J} = \{1, \ldots, q\}$ and the data volume relevant to this $\hat{J}$ is $\mathbb{R}^q$). In general, there are volumes corresponding to various $\hat{J}$ so that noisy data come across them. That is why in practice, non-smooth data-fidelity terms yield minimizers fitting exactly a certain number of the data entries. The resultant numerical effect is observed in Fig. 3.1(b) as well as in Figs. 3.3 and 3.4.

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

Numerical experiment. The original image u_o in Fig. 3.2(a) can be supposed to be a noisy version of an ideal piecewise constant image. Data v in Fig. 3.2(b) are obtained by replacing some pixels of u_o, whose locations are seen in Fig. 3.5-left, by aberrant impulsions. The image in Fig. 3.3(a) corresponds to an ℓ_1 data-fidelity term for $\beta = 0.14$. The outliers are well visible although their amplitudes are considerably reduced. The image of the residuals $v - \hat{u}$, shown in Fig. 3.3(b), is null everywhere except at the positions of the outliers in v. The pixels corresponding to non-zero residuals (*i.e.* the elements of $\hat{J}^c$) provide a faithful estimate of the locations of the outliers in v, as seen in Fig. 3.5-middle. Next, in Fig. 3.4(a) we show a minimizer $\hat{u}$ of the same $\mathcal{F}(.,v)$ obtained for $\alpha = 0.25$. 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. 3.4(b) is null only on restricted areas, but has a very small magnitude everywhere beyond the outliers. However, applying the above detection rule now leads to numerous false detections, as seen in Fig. 3.5-right.

The minimizers of two different cost-function $\mathcal{F}$ involving a *smooth* data-fidelity term Ψ, shwon in Fig. 3.6, do not fit any data entry. The mathematical explanation will be presented in § 3.3.

3.2 Minimization method

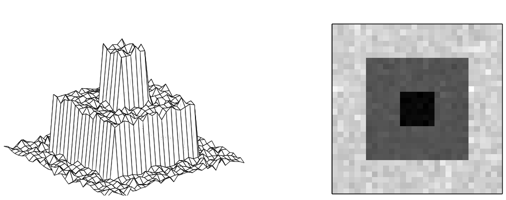
In [$\star$3(2004)] we propose a numerical method to minimize convex non-smooth energies $\mathcal{F}$ where Ψ is defined by (3.2)-(3.3), Φ is convex and $\mathcal{C}^1$ continuous, and $\{a_i : 1 \leq i \leq q\}$ are linearly independent.

Using the change of variables $z = T(u)$ defined by $z_i = \langle a_i, u\rangle - v_i$ for $1 \le i \le q$ and if $p > q$, $z_i = u_i$, $q+1 \le i \le p$, we consider equivalently the minimization of $F(z) = \sum_{i=1}^{q} \psi_i(z_i) + \beta\Phi(T^{-1}(z))$. We focus on relaxation-based minimization and generalize a method proposed by Glowinski, Lions and Trémolières in [43]. Let $z^{(0)} \in \mathbb{R}^p$ be a starting point. At every iteration $k = 1, 2, \ldots$, the new iterate $z^{(k)}$ is obtained from $z^{(k-1)}$ by calculating successively each one of its entries $z_i^{(k)}$ using one-dimensional minimization:

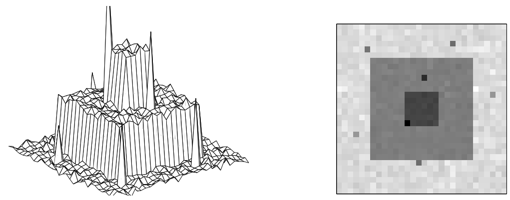
$$\text{for any} \quad i = 1, \ldots, p, \quad \text{find} \quad z_i^{(k)} \quad \text{such that}$$

$$F(z_1^{(k)}, \ldots, z_{i-1}^{(k)},\ z_i^{(k)},\ z_{i+1}^{(k-1)}, \ldots, z_p^{(k-1)}) \le F(z_1^{(k)}, \ldots, z_{i-1}^{(k)},\ t,\ z_{i+1}^{(k-1)}, \ldots, z_p^{(k-1)}), \quad \forall t \in \mathbb{R}. \tag{3.6}$$

The details of the method and its convergence under mild assumptions are given in section 2.2 in [$\star$3(2004)]. Let us notice that the points where ψ_i is non-smooth—which are the most difficult to reach using standard minimization methods—are easily solved by checking an inequality. The convergence of the method to a minimizer $\hat{u}$ of $\mathcal{F}(., v)$ is established under mild assumptions.

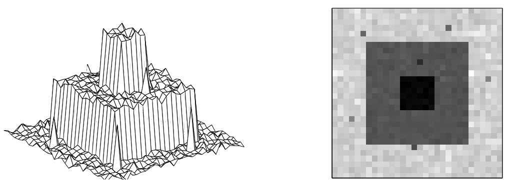


(a) Original u_o.

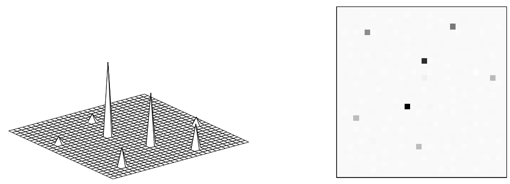


(b) Data $v = u{\cdot}$outliers.

Figure 3.2: Original u_o and data v degraded by outliers.

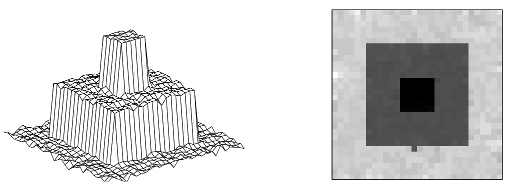


(a) Restoration $\hat{u}$ for $\beta = 0.14$.

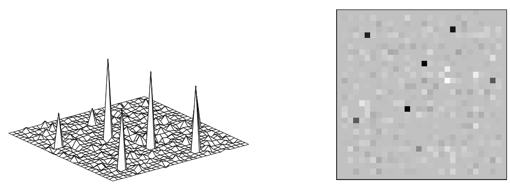


(b) Residual $v - \hat{u}$.

Figure 3.3: Restoration using $\mathcal{F}(u, v) = \sum_i |u_i - v_i| + \beta \sum_{i\sim j} |u_i - u_j|^\alpha$ $\alpha = 1.1$ and $\beta = 0.14$.



(a) Restoration $\hat{u}$ for $\beta = 0.25$.



(b) Residual $v - \hat{u}$.

Figure 3.4: Restoration using $\mathcal{F}(u, v) = \sum_i |u_i - v_i| + \beta \sum_{i\sim j} |u_i - u_j|^\alpha$ for $\alpha = 1.1$ and $\beta = 0.25$.

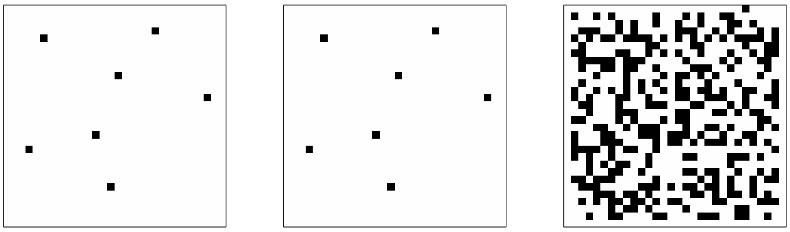
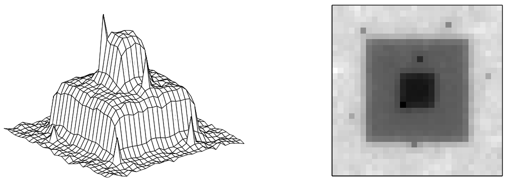
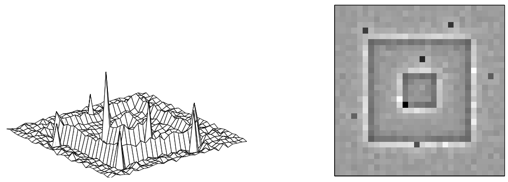


Figure 3.5: Left: the locations of the outliers in v. Next—the locations of the pixels of a minimizer $\hat{u}$ at which $\hat{u}_i \neq v_i$. Middle: these locations for the minimizer obtained for $\beta = 0.14$, Fig. 3.3. Right: the same locations for the minimizer relevant to $\beta = 0.25$, see Fig. 3.4.



(a) Restoration $\hat{u}$ for $\beta = 0.2$.



(b) Residual $v - \hat{u}$.

Figure 3.6: Restoration using a smooth cost-function, $\mathcal{F}(u,v) = \sum_i (u_i - v_i)^2 + \beta \sum_{i \sim j} (u_i - u_j)^2$, $\beta = 0.2$.

3.3 Comparison with smooth cost-functions

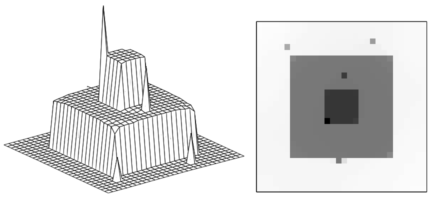
The special properties exhibited in § 3.1 are due to the non-smoothness of Ψ and they do almost never occur if $\mathcal{F}$ is smooth. The theorem below shows this under strong sufficient conditions.

Theorem 4 *Let $\mathcal{F}$ be $\mathcal{C}^m$, $m \geq 2$, and let $J \subset \{1 \ldots, q\}$ be nonempty. Assume the following:*

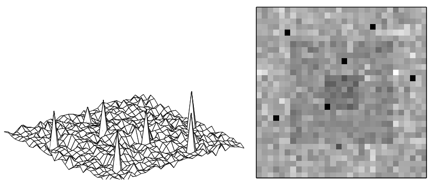
(a) $\psi_i : \mathbb{R} \to \mathbb{R}$ satisfy $\psi_i''(t) > 0$ for all $t \in \mathbb{R}$, $\forall i = 1, \ldots, q$;

(b) A is invertible (recall that for every $i = 1, \ldots, q$, the ith row of A is a_i^*);

(c) there is an open domain $O \subset \mathbb{R}^q$ so that $\mathcal{F}(., O)$ admits a $\mathcal{C}^{m-1}$ local minimizer function $\mathcal{U} : O \to \mathbb{R}^p$, such that $D_1^2 \mathcal{F}(\mathcal{U}(v), v)$ is positive definite, for all $v \in O$;



(a) Restoration $\hat{u}$ for $\beta = 0.2$.



(b) Residual $v - \hat{u}$.

Figure 3.7: Restoration using non-smooth regularization $\mathcal{F}(u,v) = \sum_i |u_i - v_i| + \beta \sum_{i \sim j} |u_i - u_j|$, $\beta = 0.2$.

(d) for every $u \in \mathcal{U}(O) \subset \mathbb{R}^p$ and for every $i \in J$ we have $D^2\Phi(u)[A^{-1}]_i \neq 0$, where $[A^{-1}]_i$ denotes the ith column of A^{-1}, for $i = 1, \ldots, q$.

For any $\mathcal{O} \subset O$ a closed subset, put

$$V_{\mathcal{O}} = \{v \in \mathcal{O} : \langle a_i, \mathcal{U}(v)\rangle = v_i \ \forall i \in J\}. \tag{3.7}$$

Then $V_{\mathcal{O}}$ is a closed subset of $\mathbb{R}^q$ whose interior is empty.

Hence, there is no chance that noisy v give rise to a minimizer $\hat{u}$ such that $\langle a_i, \hat{u}\rangle = v_i$ for some $i \in \{1, \ldots, q\}$.

Assumption (a) holds for the usual data-fidelity terms. Assumption (b) is easy to extend to the case when A^*A invertible. Assumption (c) here is the same as (b) in Theorem 3 and was commented there. Assumption (d) holds for almost all A, as explained in Remark 11 in [⋆4(2002)]. Proposition 2 in [⋆4(2002)] states the same conclusions but under different assumptions.

Example 1, continued. We now determine the set of *all* data points $v \in \mathbb{R}^q$ for which $\hat{u} = \mathcal{U}(v)$ fits exactly the ith data entry v_i. To this end, we solve with respect to $v \in \mathbb{R}^q$ the equation $\langle a_i, \mathcal{U}(v)\rangle = v_i$ Using (2.11), this is equivalent to

$$\langle p_i(\beta), v\rangle = 0, \tag{3.8}$$
$$p_i(\beta) = A(A^*A + \beta\, G^*G)^{-1}a_i - e_i, \tag{3.9}$$

where e_i is the ith vector of the canonical basis of $\mathbb{R}^q$. We can have $p_i(\beta) = 0$ only if β belongs to the set of several values which satisfy (3.9). However, β will almost never belong to such a set, so $p_i(\beta) \neq 0$ in general. Then (3.8) implies $v \in \{p_i(\beta)\}^\perp$. More generally, we have the implication

$$\exists i \in \{1, \ldots, q\} \ \text{ such that } \ \mathcal{U}_i(v) = v_i \quad \Rightarrow \quad v \in \bigcup_{j=1}^{q} \{p_j(\beta)\}^\perp.$$

Since every $\{p_i(\beta)\}^\perp$ is a subspace of $\mathbb{R}^q$ of dimension $q-1$, the union on the right-hand side is a *closed, negligible subset of* $\mathbb{R}^q$. The chance that noisy data v yield a minimizer $\mathcal{U}(v)$ that fits even one data entry, is null.

3.4 ℓ_1 data-fidelity to detect and remove outliers

3.4.1 Theoretical method

The properties established in [⋆4(2002)] suggest that non-smooth data-fidelity terms can be used to process “spiky” data. In [⋆3(2004)] we develop a method to detect and to remove outliers and impulse noise by minimizing $\mathcal{F}$ of the form (3.1)-(3.2) where

$$\Psi(u, v) = \sum_{i=1}^{p} |u_i - v_i|, \tag{3.10}$$

$$\Phi(u) = \frac{1}{2}\sum_{i=1}^{p}\sum_{j\in\mathcal{N}_i} \phi(u_i - u_j). \tag{3.11}$$

Here $\mathcal{N}_i$ denotes the set of the neighbors of i, for every $i = 1, \ldots, p$. If u is a signal, $\mathcal{N}_i = \{i, i+1\}$, if $i = 2, \ldots, p-1$; for a 2D image, $\mathcal{N}_i$ is the set of the 4, or the 8 pixels adjacent to i. The function ϕ is $\mathcal{C}^1$, convex, symmetric, and edge-preserving [14, 24, 46], as (f1)-(f4) given in Table 1.1. Suppose that $\mathcal{F}(., v)$ reaches its minimum at $\hat{u}$ and put $\hat{J} = \{i \ : \ \hat{u}_i = v_i\}$. The idea of our method is that *every v_i, for $i \in \hat{J}$, is uncorrupted (i.e. regular), since $\hat{u}_i = v_i$; in contrast, every v_i, for $i \in \hat{J}^c$, can be an outlier since $\hat{u}_i \neq v_i$, in which case $\hat{u}_i$ is an estimate of the original entry.* In particular, we define the *outlier detector* function

$$v \to \hat{J}^c = \{i \in \{1, \ldots, p\} : \hat{u}_i \neq v_i\}, \quad \text{where } \hat{u} \text{ is such that } \ \mathcal{F}(\hat{u}, v) \leq \mathcal{F}(u, v), \ \forall u \in \mathbb{R}^p. \tag{3.12}$$

The rationale of this method relies on the properties of the minimizers $\hat{u}$ of $\mathcal{F}(., v)$ when v involves outliers. Below we give a flavor of these properties.

First, the p-dimensional subset given below is the set of all outlier-free data,

$$\left\{ v \in \mathbb{R}^p : \left| \sum_{j \in \mathcal{N}_i} \phi'(v_i - v_j) \right| \leq \frac{1}{\beta}, \ \ \forall i = 1, \ldots, p \right\}, \tag{3.13}$$

since $\mathcal{F}(., v)$ reaches its minimum at $\hat{u} = v$ for every v belonging to this set. Observe that the latter contains signals or images with smoothly varying and textured areas and edges.

- *Detection of outliers.* A datum v_i is detected to be an outlier if $|v_i|$ is much larger than its neighbors, $\sum_{j \in \mathcal{N}_i} \phi'(v_i - \hat{u}_j) > \frac{1}{\beta}$. Notice that v_i is compared only with *faithful neighbors*—regular entries v_j for $j \in \mathcal{N}_i \cap \hat{J}$ and estimates of outliers $\hat{u}_j$ for $j \in \mathcal{N}_i \cap \hat{J}^c$. This is crucial for the reliability of the detection of outliers and allows very restricted neighborhoods $\{\mathcal{N}_i\}$ to be used (e.g. the four nearest neighbors in the case of images). Since ϕ' is increasing on $\mathbb{R}$ and $\phi'(0) = 0$, we see that if v_i is too dissimilar with respect to its neighbors, v_i is replaced by an $\hat{u}_i$ such that $|\hat{u}_i| < |v_i|$ and whose value is independent of the magnitude $|v_i|$ of the outlier (Lemma 3 in [$\star$3(2004)]). Hence the robustness of $\hat{u}$. Moreover, we show that the detection of outliers is *stable*: the set $\hat{J}^c$ is constant under small perturbations of regular data entries v_i for $i \in \hat{J}$ and under arbitrary deviations of outliers v_i for $i \in \hat{J}^c$ (Theorem 3 in [$\star$3(2004)]).

- *Restoration of sets of neighboring outliers.* Let ζ be a connected component of $\hat{J}^c$ and let $\hat{u}_\zeta$ be the restriction of $\hat{u}$ to ζ. We show that $\hat{u}_\zeta$ is the minimizer of a *regularized cost-function* $f(., v)$ of the form (1.6). Its first term encourages every boundary entry for ζ, namely $\hat{u}_i$ for an $i \in \zeta$ such that $\mathcal{N}_i \cap \hat{J} \neq \emptyset$, to fit neighboring regular data entries v_j for $j \in \mathcal{N}_i \cap \hat{J}$. Its second term is a smoothness constraint on $\hat{u}_\zeta$ since it favors neighboring entries for ζ, say $\hat{u}_i$ and $\hat{u}_j$, with $i, j \in \zeta$ and $j \in \mathcal{N}_i$, to have similar values. Thus we can expect that edges in $\hat{u}_\zeta$ are well restored if ϕ is a good edge-preserving function. We show that $\hat{u}_\zeta$ results from a continuous minimizer function $\mathcal{U}_\zeta$ which depends only on the neighboring regular data entries v_i for $i \in \mathcal{N}_\zeta$ and is independent of the value of the outliers v_i for $i \in \zeta$ (see Lemma 4 in [$\star$3(2004)]).

  We also show that this restoration introduces a small bias.

Furthermore, we derive bounds for β and show that there is a compromise between detection of outliers of small amplitude and preservation of large edges. Last, we justify the choice of $\psi_i(t) = |t|$ in (3.10).

(a) Original image u_o (size 300×300).

(b) Data v with 10% random-valued noise.

Figure 3.8: Original picture and data with random-valued noise.

The computation of minimizers $\hat{u}$ is done by adapting the relaxation method sketched in §3.2.

Picture with 10% random-valued impulse noise. The original image u_o is shown in Fig. 3.8 (a). In Fig. 3.8(b), 10% of the pixels have random values uniformly distributed on $\left[\min_i [u_o]_i,\ \max_i [u_o]_i\right]$. Denoising results using different methods are displayed in Fig. 3.9, where all parameters are finely tuned. The restoration in Fig. 3.9 (a) corresponds to one iteration of a 3×3 window median filter. The image in (b) is calculated using a 3×3 window recursive CWM for $\alpha = 3$. The result in (c) corresponds to a 3×3 window PWM filter for $\alpha = 4$. These images are slightly blurred, the texture of the sea is deformed, and several outliers still remain. The image $\hat{u}$ in (d) is the minimizer of $\mathcal{F}$ as given in 3.10)-(3.11), with $\phi(t) = |t|^{1.1}$, $\mathcal{N}_i$ the set of the 4 adjacent neighbors and $\beta = 0.3$. All details are well preserved and the image is difficult to distinguish from the original. Indeed, for 85% of the pixels, $|\hat{u}_i - [u_o]_i| / \Delta \leq 2\%$, where $\Delta = \max_i [u_o]_i - \min_i [u_o]_i$. Based on previous, one can expect that a smaller β can reduce the number of regular data entries erroneously detected as outliers, but that detected outliers are not smoothed enough.

Picture with 45% salt-and-pepper noise. 45% of the entries in Fig. 3.10 (a), with locations uniformly distributed over the grid of the image, are equal either to $\min_i [u_o]_i$, of to $\max_i [u_o]_i$, with probability 1/2. The image in Fig. 3.10 (b) is obtained after 2 iterations of a 3×3 window recursive median filter. It has a poor resolution and exhibits a stair-case effect. The result in (c) results from a 7×7 window PWM filter for $\alpha = 14$. Although better than (b), the resolution is poor, there are artifacts along the edges and the texture of the sea is destroyed. The images in (d) is the minimizer of $\mathcal{F}$ as given in (3.10)-(3.11) where $\phi(t) = |t|^{1.3}$ and $\mathcal{N}_i$ is the set of the 4 adjacent neighbors. for $\beta = 0.18$. The quality of the restoration is clearly improved: the contours are neater, the texture of the sea in better preserved and some details on the boat can be distinguished.

(a) Median filtering ($\|\hat{u}-u\|_2 = 4155$).

(b) Recursive CWM ($\|\hat{u}-u\|_2 = 3566$).

(c) PWM ($\|\hat{u}-u\|_2 = 3984$).

(d) The proposed method ($\|\hat{u}-u\|_2 = 2934$).

Figure 3.9: Restoration from a picture with 10% random-valued noise.

(a) Data v with 45% salt-and-pepper noise.

(b) Recursive median filter ($\|\hat{u}-u\|_2 = 7825$).

(c) PWM ($\|\hat{u}-u\|_2 = 6265$).

(d) The proposed method ($\|\hat{u}-u\|_2 = 6064$)

Figure 3.10: Picture with 45% salt-and-pepper noise.

3.4.2 The $\ell_1 - \ell_2$ cost-function

In [⋆41(2003)] we specialize the method in § 3.4.1 to quadratic regularization terms Φ, namely

$$\Phi(u)\frac{1}{2}\sum_{i=1}^{p}\sum_{j\in\mathcal{N}_i}(u_i-u_j)^2.$$

This regularization is not edge-preserving but the method is still quite good. The main interest is that the computation is very fast. Once again, we use a relaxation minimization scheme. Let the intermediate solution at step $i-1$ of iteration k be denoted $u^{(k,i-1)} = (u_1^{(k)}, u_2^{(k)}, \ldots, u_{i-1}^{(k)}, u_i^{(k-1)}, u_{i+1}^{(k-1)}, \ldots, u_p^{(k-1)})$. At the next step we calculate $u_i^{(k)}$ according to the rule:

$$\begin{array}{rcl} \xi_i^{(k)} & = & v_i - \chi_i^{(k)} \ \text{ where } \ \chi_i^{(k)} = \dfrac{1}{\#N_i}\displaystyle\sum_{j\in N_i} u_j^{(k,i-1)}, \\ \text{if } \left|\xi_i^{(k)}\right| & \leq & \dfrac{1}{2\beta\#N_i} \ \Rightarrow \ u_i^{(k)} = v_i, \\ \text{if } \left|\xi_i^{(k)}\right| & > & \dfrac{1}{2\beta\#N_i} \ \Rightarrow \ u_i^{(k)} = \chi_i^{(k)} + \dfrac{\text{sign}(\xi_i^{(k)})}{2\beta\#N_i}. \end{array}$$

Notice that updating each entry $u_i^{(k)}$ involves only the samples belonging to its neighborhood N_i, so computation can be done in a parallel way. Based on previous results, we recommend to start with $u^{(0)} = v$.

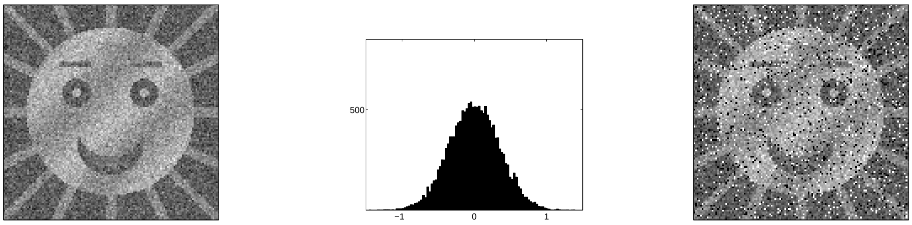
Another important point of interest is that the minimizers $\hat{u}$ can be characterized almost explicitly. Some results are given in [⋆41(2003)] but we did not yet complete the full paper.

In our experiment, the sought image u_o, shown in Fig. 3.11(a), is related to u^* in Fig. 2.3(a) by $u_o = u^* + n$, where n is white Gaussian noise with 20 dB SNR. The histogram of n is plotted in Fig. 3.11(b). Our goal is to restore u_o in Fig. 3.11 (a) which contains Gaussian noise based on the data v in Fig. 3.11(c) which contain 10% salt-and-pepper noise. Restoring u_o is a challenge since the white noise n there must be preserved. The histogram of the estimated noise, $\hat{n} = \hat{u} - u^*$, must be close to the initial noise n. The image in Fig. 3.12 (a) corresponds to one iteration of median filter over a 3×3 window. The histogram of the resultant estimate of the Guassian noise in u is given in (b). The image in Fig. 3.12 (c) is calculated using CWM with a 5×5 window and multiplicity parameter 14. The relevant noise estimate is given in (d). In these estimates, the distribution of the noise estimate $\hat{n}$ is quite different from the distribution of n. Fig. 3.13 displays the issue of the proposed method, for $\beta = 1.3$. It achieves a good preservation of the statistics of the noise in u, as seen from the histogram of the estimated noise $\hat{n}$. The proposed method accurately suppresses the outliers and preserves the Gaussian noise.

3.4.3 Restoration of wavelet coefficients

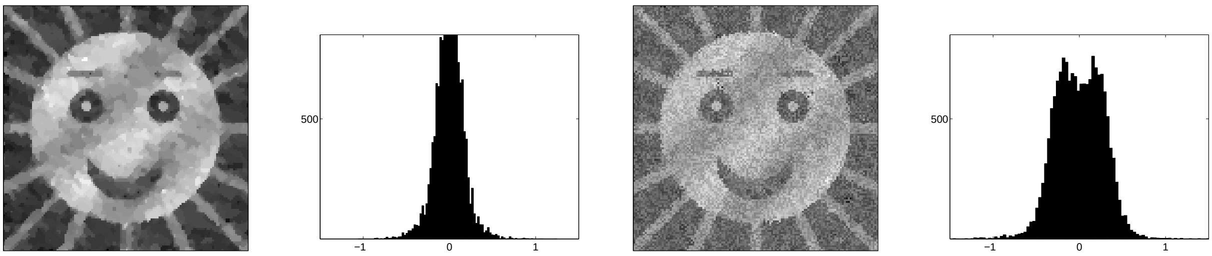
In collaboration with Sylvain Durand [⋆42(2003)] we consider hybrid methods to denoise a signal or an image $v = u_o + n \in \mathbb{R}^p$, where n is white Gaussian noise, based on the coefficients y of a frame transform of the data

$$y = \widetilde{W}v = \sum_i \widetilde{w}_i v.$$



(a) $u_o = u^* + n$ to recover (b) Histogram of n to recover (c) Data=Image $\odot$ SP

Figure 3.11: The goal is to restore u_o in (a) which contains Gaussian noise n, see (b) from the data in (c).



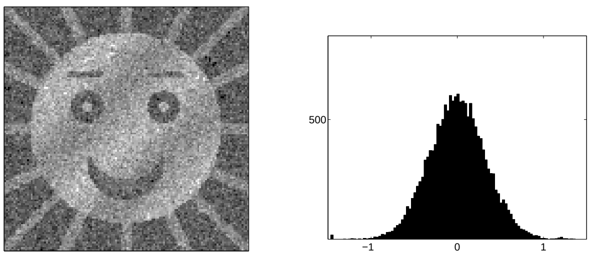
(a) Median Filter
$\|\hat{u}-u_o\|_1 = 43$, $\|\hat{u}-u_o\|_2 = 46$

(b) Noise estimate $\hat{n}$

(c) Center-Weighted Median
$\|\hat{u}-u_o\|_1 = 22$, $\|\hat{u}-u_o\|_2 = 35$

(d) Noise estimate $\hat{n}$

Figure 3.12: Classical methods to clean the outliers.



(a) $\mathcal{F} = \ell_1 + \ell_2$
$\|\hat{u}-u_o\|_1 = 14$, $\|\hat{u}-u_o\|_2 = 26$

(b) Noise estimate $\hat{n} = \hat{u} - u^*$

Figure 3.13: The proposed method.

Since the inaugural work of Donoho and Johnstone [33], shrinkage estimators are a popular and fast tool to denoise images and signals. Let I denote the set of the indexes of all noisy coefficients that must be restored. Hard-thresholding defines the denoised coefficients by

$$\hat{x}[i] = \begin{cases} y[i] & \text{if} \quad i \in I_1, \\ 0 & \text{if} \quad i \in I_0, \end{cases} \tag{3.14}$$

where

$$I_0 = \{i \in I : \, |y[i]| < T\} \;\text{ and }\; I_1 = I \setminus I_0, \tag{3.15}$$

whereas soft-thresholding corresponds to replace $y[i]$ in (3.14) by $y[i] - T\text{sign}(y[i])$ if $i \in I_1$. These are asymptotically optimal in the minimax sense if $\widetilde{W}$ is an orthogonal wavelet transform and

$$T = \sigma\sqrt{2\log_e p}, \tag{3.16}$$

where σ is the standard deviation of the noise. An intrinsic difficulty is that T increases along with the size p of u which entails a loss of useful information. Refinements of these methods have been proposed in order to adapt thresholding to the scale of the coefficients [34]. Other shrinkage methods are based on a priori models for the distribution of the coefficients $x = \widetilde{W}u$ and MAP estimation, thus leading to minimize energies of the form $\|x - y\|^2 + \sum_i \lambda_i \phi(x_i)$ [84, 75, 63, 7, 4]. The denoised signal or image is then

$$\hat{u} = W\hat{x} = \sum_i \hat{x}_i w_i \tag{3.17}$$

where W is a left inverse of $\widetilde{W}$ and $\{w_i\}$ the associated dual frame. The major problems with these methods is that shrinking large coefficients leads to oversmoothing of edges, while shrinking small coefficients towards zero yield Gibbs-type oscillations in the vicinity of edges. On the other hand, if shrinkage is weak, some coefficients bearing mainly noise will remain (almost) unchanged and (3.17) shows that they generate in $\hat{u}$ artifacts with the shape of the functions w_i. Furthermore, priors on the coefficients x cannot adequately address the presence of edges and smooth regions in u_o.

Instead, several authors [13, 28, 37, 60, 17, 35] combined the information contained in the large coefficients $y[i]$ with pertinent priors directly on the sought-after function u. Although based on different motives, these "hybrid" methods amount to define the restored function $\hat{u}$ as

$$\text{minimize} \;\; \Phi(u) = \int_\Omega \phi(|\nabla u(s)|)\, ds \;\; \text{subject to} \;\; \hat{u} \in \left\{u : \; \left|\left(\widetilde{W}(u - v)\right)[i]\right| \leq \mu_i, \, \forall i \in I\right\}, \tag{3.18}$$

where $\{\mu_i\}$ depend on y. In the first such method, introduced in [13], $\phi(t) = t^2$. General functions ϕ are considered in [28]. In order to remove pseudo-Gibbs oscillations, the authors of [37, 60, 17, 35] focused on $\phi(t) = |t|$ and did various choices for the operator W—orthogonal bases in [13, 28, 37, 35], curvlets transforms in [17], unions of wavelet bases in [60]. These methods differ also in the choice of parameters $\{\mu_i\}_{i\in J}$. If the use of an edge-preserving function for ϕ is clearly a pertinent choice, the strategy for the selection of parameters $\{\mu_i\}_{i\in J}$ remains an open question. In our paper we provide a critical analysis of the strategies adopted by these authors.

Our approach is to determine $\{\mu_i\}_{i\in J}$ based both on the data and on a prior regularization term. To this end, restored coefficients $\hat{x}$ are defined to minimize a cost-function of the form

$$F(x, y) = \sum_{i \in I_1} \lambda_i \, |(u - v)[i]| + \sum_{i \in I_0} \lambda_i \, |x[i]| + \int_\Omega \phi(|\nabla W x|)\, ds.$$

where I_0 and I_1 are defined by (3.15) for a T smaller than the value prescribed in (3.16), and ϕ is a convex edge preserving potential function (e.g. a possibly smooth approximation of $\phi(t) = t$). In such a case, I_1 is composed out of **(a)** large coefficients which bear the main features of u and **(b)** coefficients which are highly contaminated by noise, called *outliers*. Based on [$\star$4(2002)], we can expect that the first are preserved intact whereas the second are restored by the regularization term so that no wavelet-shaped artifacts appear in $\hat{u}$. Furthermore, I_0 is composed out of **(c)** noise coefficients and **(d)** coefficients $y[i]$ which correspond to edges and other details in u. The coefficients relevant to (c) are likely to remain zero, which is certainly the best one can do, whereas those in (d) will be restored according to the regularization term which prevents $\hat{u}$ from Gibbs-like oscillations. These conjectures are demonstrated for simple input signals and corroborated by the experiments. Parameters $\{\lambda_i\}$ can be determined in a semi-automatic way as a function of the frame $\widetilde{W}$. Notice that the method is robust with respect to T since both outliers (b) and erroneously thresholded coefficients (d) are restored.

Numerical experiment. We consider the restoration of the 512-length original signal in Fig. 3.14(a) from the data shown there, contaminated with white Guassian noise with $\sigma = 10$. The restoration in (b) is obtained using the sure-shrink method [34] and the toolbox WaveLab. The result displayed in Fig. 3.14 (d) is the minimizer of $\mathcal{F} = \|Au - v\|^2 + \beta \sum_i \phi(\|D_i u\|)$ where $\phi(t) = \sqrt{\alpha + t^2}$ for $\alpha = 0.1$ and $\beta = 100$. Smooth zones are rough, edges are slightly smoothed and spikes are eroded, while some diffused noise is still visible on the signal.

The other restorations are based on thresholded wavelet coefficients where $\widetilde{W}$ is an orthogonal basis of Daubechies wavelets with 8 vanishing moments. The optimal T, as given in (3.16), reads $T = 35$. The wavelet-thresholding estimate $\sum_{i \in I_1} y[i]w$ is shown in Fig. 3.14 (c). It involves important Gibbs artifacts, as well as wavelet-shaped oscillations due to aberrant coefficients. In Fig. 3.14 (e) we present the result obtained with the proposed method which corresponds to $T = 23$, $\alpha = 0.05$, $\lambda_{j,\kappa} = 0.5 \times 2^{j/2}$ if $(j,\kappa) \in I_0$ and $\lambda_{j,\kappa} = 1.5 \times 2^{j/2}$ if $(j,\kappa) \in I_1$. In this restoration, edges are neat and polynomial parts are well recovered. Fig. 3.14(f) illustrates how restored coefficients $\hat{x}$ are placed with respect to thresholded and original coefficients. In particular, we observe how erroneously thresholded coefficients are restored and how outliers are smoothed. More details are given in [$\star$42(2003)].

3.5 Practical two-stage methods for impulse noise removal

Median-based filters are well-known to locate outliers accurately but to replace them inaccurately, and to be faster than variational methods. On the other hand, the variational methods of the form presented in § 3.4.1 have the default to possibly detect some noise-free pixels as noisy if these are placed near to edges (see (3.13)), and the quality to give rise to accurate restoration of the noisy pixels. Hence the idea to approximate the outlier-detection stage in the variational method in § 3.4.1 by a median-based filtering method. In collaboration with Raymond Chan and several students, we investigate how to combine the advantages of filtering and variational methods for salt-and-paper and random-valued impulse noise [$\star$13(2004)], [$\star$14(2005)], [$\star$15(2005)] and [$\star$38(2005)]. Following this direction, we develop efficient two-stage methods whose principal steps are

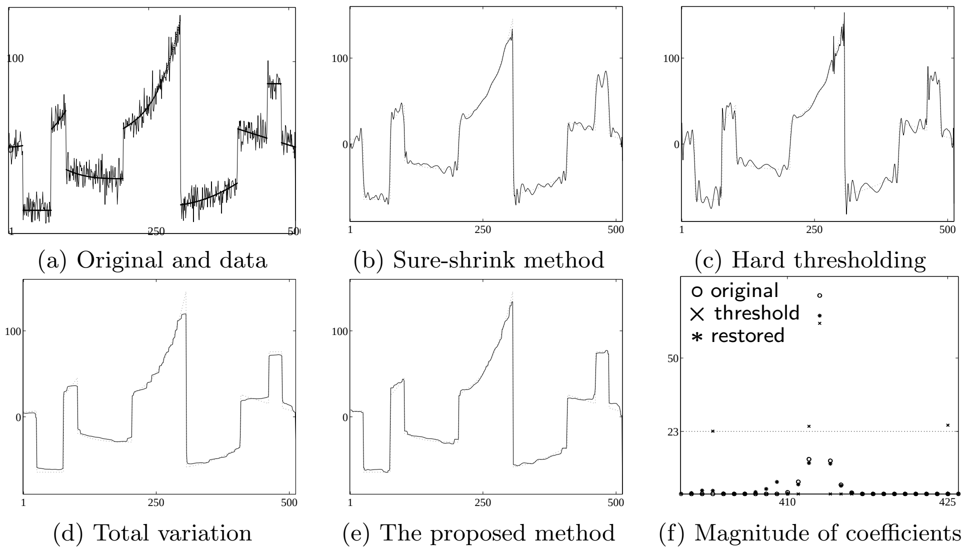


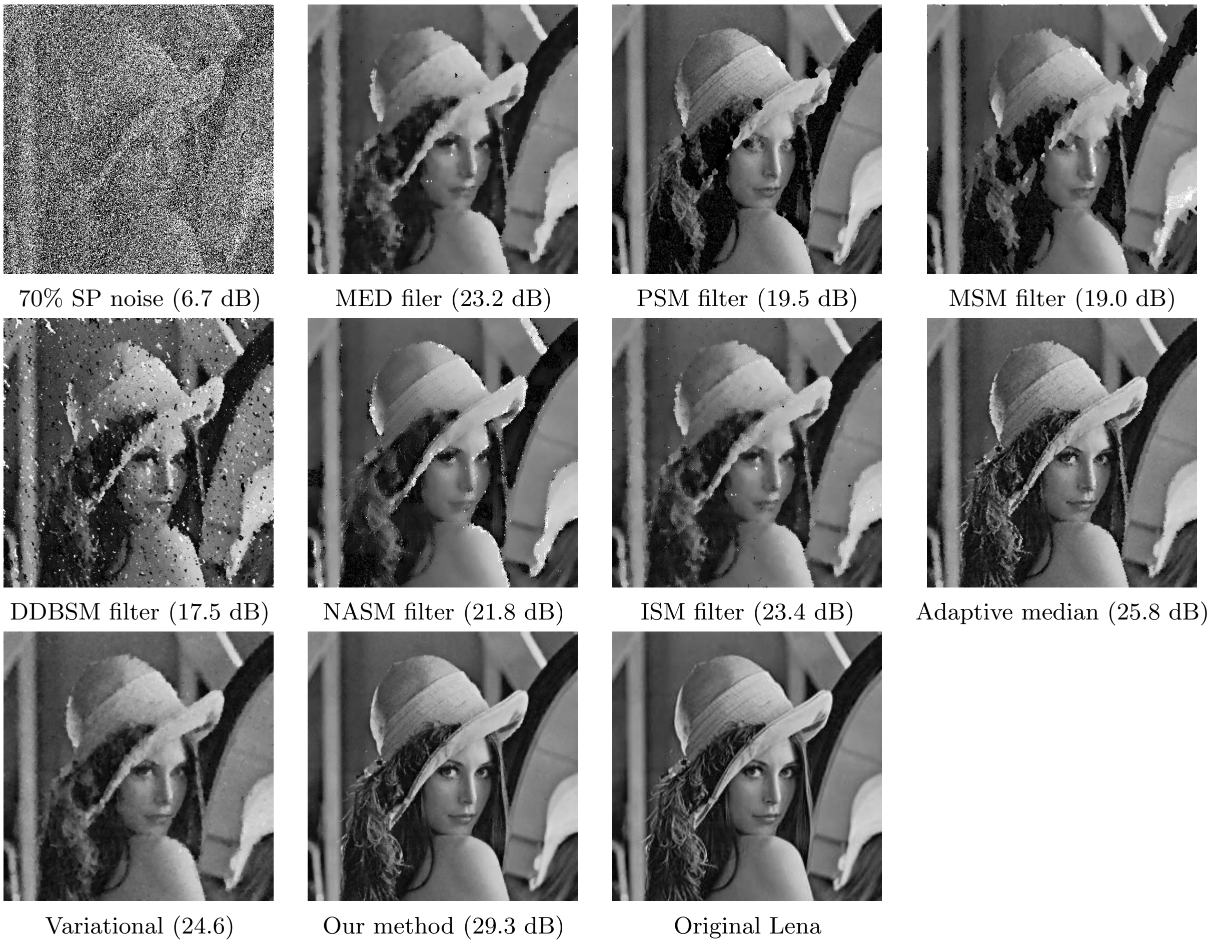
Figure 3.14: Various methods to restore the noisy signal in (a). Restored signal (—), original signal (- -).

(a) Find $\hat{v}$ by applying an appropriate median-based filter to v and consider that the corrupted pixels are

$\hat{J}^c = \{i : \hat{v}_i \neq v_i\}$;

(b) Restore the noise candidates $\left\{\hat{u}_i : i \in \hat{J}^c\right\}$ using a variational method of the form § 3.4.1, restricted only to $\hat{J}^c$, and keep $\hat{u}_i = v_i$ for all $i \in \hat{J}$.

After extended experiments, we found out that the adaptive median filter is well suited for salt-and-pepper noise, while the adapted center-weight median filter is better for random-valued impulse noise. Random impulse noise being much more difficult to clean out, we designed a refining iterative scheme. We do not present here the details of these methods. In Fig. 3.15 we present restoration results from 70% salt-and-pepper impulse noise. In Figs. 3.16 and 3.17 we show the results from 30% and 50% random impulse noise, respectively. Even at very high noise ratios, we could obtain accurate restorations at a low numerical cost.

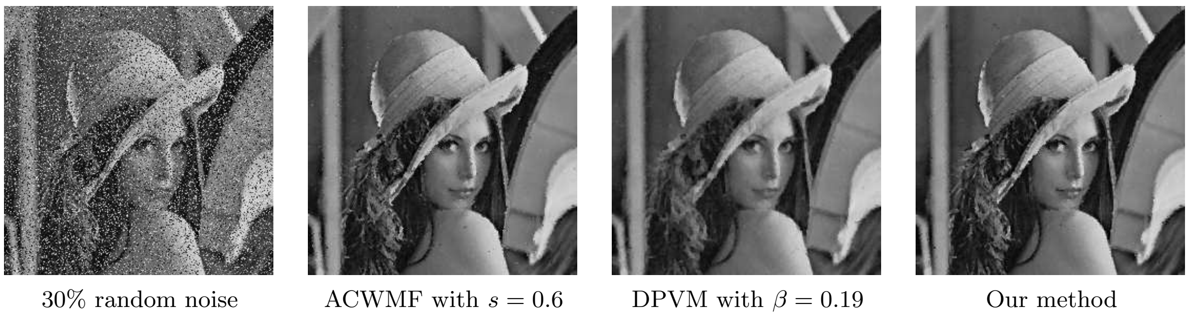


70% SP noise (6.7 dB) | MED filer (23.2 dB) | PSM filter (19.5 dB) | MSM filter (19.0 dB)

DDBSM filter (17.5 dB) | NASM filter (21.8 dB) | ISM filter (23.4 dB) | Adaptive median (25.8 dB)

Variational (24.6) | Our method (29.3 dB) | Original Lena

Figure 3.15: Restoration from 70% salt-and-pepper noise using different methods.



30% random noise | ACWMF with $s = 0.6$ | DPVM with $\beta = 0.19$ | Our method

Figure 3.16: Restoration from 30% random-valued impulse noise using different methods.

50% random noise | ACWMF with $s = 0.3$ | DPVM with $\beta = 0.19$ | Our method

Figure 3.17: Restoration from 50% random-valued impulse noise using different methods

Chapter 4

Nonsmooth data-fidelity and regularization

4.1 A corollary of chapters 2 and 3

If both terms Φ and Ψ are non-smooth as required in § 2.1 and in § 3.1, respectively, Theorems 1 and 3 indicate that minimizers $\hat{u}$ are likely to be such that

$$\begin{array}{lll} G_i\hat{u} = 0 & \text{for} & i \in \hat{J}_\phi \neq \emptyset, \\ \langle a_i, \hat{u}\rangle = v_i & \text{for} & i \in \hat{J}_\psi \neq \emptyset \end{array} \tag{4.1}$$

Minor refinements in the conditions and the proofs are necessary for a proper proof of this conjecture. We underestimated the potential importance of such a result. Inspired by our non-smooth data-fidelity terms in [$\star$4(2002)] and [$\star$3(2004)], T. Chan and S. Esedoglou did in [20] a very interesting analysis of the minimizers of cost-functions composed of an L_1 data-fidelity term and TV regularization. Even though their work concerned images on an open domain of $\mathbb{R}^2$, the conjecture in (4.1) was confirmed. Some new reports in this direction are for instance [47, 16].

4.2 Restoration of binary images using an $L_1 - TV$ energy

Based on (4.1), one can see that if v is binary (e.g. with values in $\{0, 1\}$), then minimizers $\hat{u}$ are likely to be binary as well.

In a joint work with Toni Chan and Selim Esedoglu, [$\star$22(2004)] and [$\star$37(2005)] we started by exploring the possibilities to restore a binary image on $\mathbb{R}^d$, $d \geq 2$, based on binary noisy data v, by minimizing a *convex* non-smooth energy. Binary data v can be expressed as

$$v(x) = 1\!\!1_\Omega(x),$$

where Ω is a bounded domain of $\mathbb{R}^d$ whose boundary $\partial\Omega$ can be very rough because of the noise (one can think of a noisy Fax document, for instance). A straightforward variational method, suggested in [80], is to consider a TV energy constrained to the set of the binary images $u(x) = 1\!\!1_\Sigma(x)$:

$$\min_{\substack{\Sigma\subset\mathcal{R}^N \\ u(x)=1\!\!1_\Sigma(x)}} \int_{\mathbb{R}^d} \Big(u(x) - v(x)\Big)^2 dx + \beta \int_{\mathcal{R}^N} \phi(|\nabla u|). \tag{4.2}$$

This problem is non-convex because the minimization is over a non-convex set of functions. It is equivalent to the following geometry problem:

$$\min_{\Sigma \subset \mathbb{R}^d} \Big(|\Sigma \bigtriangleup \Omega| + \beta \mathrm{Per}(\Sigma) \Big) \tag{4.3}$$

where Per stands for perimeter, $|\cdot|$ is the d-dimensional Lebesgue measure, and $S_1 \bigtriangleup S_2$ denotes the symmetric difference between the two sets S_1 and S_2. The unknown set Σ in (4.3) can be described by its boundary $\partial\Sigma$. So a common approach of solving (4.3) has been to use some *curve evolution* process, sometimes referred to as *active contours*, where $\partial\Sigma$ is updated iteratively according to gradient flow for the energy involved. The usual numerical methods, such as the explicit curve representation [54], the level set method of Osher and Sethian [68], or the Gamma convergence method [29], are prone to get stuck in spurious local minima, thus leading to images with wrong level of detail.

The crux of our approach is to consider minimization of the following convex energy, defined for any given observed image $v(x) \in L^1(\mathbb{R}^d)$ and $\beta > 0$:

$$\mathcal{F}(u, v) = \int_{\mathbb{R}^d} |u(x) - f(x)| dx + \beta \int_{\mathbb{R}^d} |\nabla u| \tag{4.4}$$

The relevance of (4.4) for our purposes comes from the fact that $\mathcal{F}(., v)$ is *convex*, hence its minimum can practically be reached, and from the equivalence theorem stated below.

Theorem 5 *We have the following:*

(i) If $\hat{u} = \mathbb{1}_\Sigma$ is a (global) solution to (4.2), then $\mathcal{F}(., v)$ reaches its minimum at $\hat{u}$.

(ii) If $\mathcal{F}(., v)$ reaches its minimum at $\hat{w}$, then for almost every $\mu \in (0, 1)$ the function $\hat{u} = \mathbb{1}_\Sigma$ where $\Sigma = \{x \in \mathbb{R}^d : \hat{w}(x) > \mu\}$ is a global solution to (4.2).

These statements come from the obvious fact that the energies in (4.2) and (4.4) agree on binary images, and from the theory developed in [20].

Algorithm. To find a solution (i.e. a global minimizer) $\hat{u}$ of the non-convex variational problem (4.2), it is sufficient to carry out the following three steps:

1. Find any minimizer $\hat{w}(x)$ of the *convex* energy (4.4).
2. Determine $\Sigma = \{x \in \mathbb{R}^d : \hat{w}(x) > \mu\}$ for some $\mu \in (0, 1)$.
3. Set $\hat{u}(x) = \mathbb{1}_\Sigma(x)$: then $\hat{u}$ is a global minimizer of (4.2) for almost every choice of μ.

Next, we extended this approach to solve the piecewise constant Mumford-Shah segmentation energy [21] where data v are no longer binary. This is presented in section 4 in [$\star$22(2004)].

Chapter 5

Non-convex regularization

5.1 Stability results

In [⋆48(2001)], [⋆9(2005)] and [⋆10(2005)] in collaboration with Sylvain Durand, we explore the difficult question of the stability of the minimizers—local and global—of non-convex and possibly non-smooth cost-functions $\mathcal{F}$. More precisely, we focus on $\mathcal{F} : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ of the form

$$\mathcal{F}(u, v) = \|Au - v\|^2 + \beta \sum_{i=1}^{r} \varphi_i(G_i u),$$

where for every $i \in \{1, \dots, r\}$, the function $\varphi_i : \mathbb{R}^s \to \mathbb{R}$ is continuous on $\mathbb{R}^s$ and $\mathcal{C}^m$, $m \geq 2$, everywhere except possibly at a given $\theta_i \in \mathbb{R}^s$, and $G_i : \mathbb{R}^p \to \mathbb{R}^s$ is a linear operator. We systematically assume that A is injective, i.e. that $\operatorname{rank} A = p$.

5.1.1 Local minimizers

The first part of our work [⋆9(2005)] focuses on the local minimizers of $F(., v)$. Studying the stability of local minimizers (rather than global minimizers only) is a matter of critical importance in its own right for several reasons. In many applications, smoothing is performed by only locally minimizing a non-convex cost-function in the vicinity of some initial solution. Second, it is worth recalling that no minimization algorithm guarantees the finding of the global minimum of a general non-convex cost-function. The practically obtained solutions are frequently only local minimizers.

To this end, we analyze the extent in $\mathbb{R}^q$ of the subset $\Omega \subset \mathbb{R}^q$ of all data leading to minimizers which have good regularity properties as specified below:

Definition 3 *Let $\mathcal{F}(., v)$ be $\mathcal{C}^m$ with $m \geq 2$ almost everywhere on $\mathbb{R}^p$, for every $y \in \mathbb{R}^q$. Denote*

$$\Omega = \left\{ v \in \mathbb{R}^q : \begin{array}{l} \textit{if } \hat{u} \textit{ is a strict local minimizer of } \mathcal{F}(., v) \textit{ then there} \\ \textit{is a } \mathcal{C}^{m-1} \textit{ strict (local) minimizer function } \mathcal{U} : O \to \mathbb{R}^p \\ \textit{such that } v \in O \subset \mathbb{R}^q \textit{ and } \hat{u} = \mathcal{U}(v) \end{array} \right\}.$$

When Φ is $\mathcal{C}^m$, no special assumptions are taken. In the non-smooth case, we have a few non-restrictive assumptions (H4-H6 in [⋆9(2005)]) which are shown to hold in the most important case when

$$\varphi_i(z) = \phi(\|z - \theta_i\|) \text{ for } \phi \in \mathcal{C}^m(\mathbb{R}_+) \text{ and } \phi'(0) > 0, \quad 1 \leq i \leq r. \tag{5.1}$$

In particular, ϕ can be any one of the functions (f10)-(f12) in Table 1.1.

Theorem 6 *The set Ω^c—the complement of Ω in $\mathbb{R}^q$—has Lebesgue measure zero in $\mathbb{R}^q$. Moreover, if*

$$\forall i,\ \forall t \in \mathbb{R}, \quad \lim_{t\to\infty} \frac{\nabla\varphi_i(tz)}{t} \to 0 \text{ uniformly with } z \in S, \tag{5.2}$$

then $\overline{\Omega^c}$ has Lebesgue measure zero in $\mathbb{R}^q$.

When Φ is $\mathcal{C}^m$, $m \geq 2$ we can formally write that $r = 1$ and $\Phi = \varphi_1$, so Φ can be arbitrary. Let us mention some interesting auxiliary results.

Remark 2 When $\mathcal{F}$ is $\mathcal{C}^m$, $m \geq 2$, define the subset $\Omega_0 \subset \Omega$ via

$$\Omega_0^c = \left\{ v \in \mathbb{R}^q : \exists \hat{u} \in \mathbb{R}^p \ \text{ such that } \ D\mathcal{F}(\hat{u}, v) = 0 \text{ and } D^2\mathcal{F}(\hat{u}, v) = 0 \right\}$$

The proof of Theorem 6 consists in showing that Ω_0^c *if of Lebesgue measure zero in* $\mathbb{R}^q$ *and that if (5.2) holds, then* $\overline{\Omega_0^c}$ *has Lebesgue measure zero in* $\mathbb{R}^q$ *as well.* It follows that for almost every $v \in \mathbb{R}^q$, all local minimizers $\hat{u}$ of $\mathcal{F}(., v)$ are such that $D^2\mathcal{F}(\hat{u}, v)$ is positive definite. This is an important result. In particular it shows that the conditions of Lemma 1 hold for almost every $v \in \mathbb{R}^q$ (except those contained in a negligible subset of $\mathbb{R}^q$).

Consider now that Φ piecewise smooth only. For any nonempty $J \subset \{1, \ldots, r\}$, define

$$\mathcal{K}_J = \{u \in \mathbb{R}^p : G_i u = \theta_i, \forall i \in J \text{ and } G_i u \neq \theta_i, \forall i \in J^c\}.$$

Notice that on $\mathcal{K}_J$, every φ_i for $i \in J^c$ is smooth, while $\varphi_i = 0$ for every $i \in J$. Define the following subsets:

$$\begin{aligned}
A_J &= \{v \in \mathbb{R}^q : \exists u_J \in \mathcal{K}_J \ \text{ such that } \ D\mathcal{F}|_{\mathcal{K}_J}(u_J, v) = 0 \text{ and } D^2\mathcal{F}|_{\mathcal{K}_J}(u_J, v) = 0\}; \\
B_J &= \{v \in \mathbb{R}^q : \exists \hat{u} \in \mathcal{K}_J \text{ local minimizer of } \mathcal{F}(., v) \text{ and } \exists w \in \mathcal{K}_J^\perp \cap S \text{ such that } \delta_1\mathcal{F}(\hat{u}, v)(w) = 0\}.
\end{aligned}$$

Remark 3 *We show in Propositions 2 and 3 in [⋆9(2005)] that*

(a) the sets A_J and B_J are of Lebesgue measure zero in $\mathbb{R}^q$;

(b) if (5.2) holds, then $\overline{A_J}$ and $\overline{B_J}$ are of Lebesgue measure zero in $\mathbb{R}^q$ as well.

Several interesting results on local minimizers established in [⋆10(2005)] are given below.

- Local minimizer functions never cross on Ω_0. Let us consider two minimizer functions $\mathcal{U}_1$ and $\mathcal{U}_2$ defined on an open and connected domain $O \subset \Omega_0$. We show that either $\mathcal{U}_1 \equiv \mathcal{U}_2$ on O, or

$$\mathcal{U}_1(v) \neq \mathcal{U}_2(v), \quad \forall\, v \in O. \tag{5.3}$$

- For every bounded $O \subset \mathbb{R}^q$, there is a compact subset $Q \subset \mathbb{R}^p$ such that for every $v \in O$, if $\mathcal{F}(., v)$ has a (local) minimum at $\hat{u}$, then $\hat{u} \in Q$. This is very important for practical applications.

- Every open set of $\mathbb{R}^q$ contains an open subset O on which $\mathcal{F}$ admits exactly n local minimizer functions $\mathcal{U}_i : O \to \mathbb{R}^p$, $i = 1, \ldots, n$, which are $\mathcal{C}^{m-1}$ and are such that for all $v \in O$, all the local minimizers of $\mathcal{F}(., v)$ read $\mathcal{U}_i(v), i = 1, \ldots, n$ and satisfy $\mathcal{F}(\mathcal{U}_i(v), v) \neq \mathcal{F}(\mathcal{U}_j(v), v)$, $\forall i, j \in \{1, \ldots, n\}$ with $i \neq j$.

Even though the statements concerning non-smooth functions $\mathcal{F}$ are basically the same as those corresponding to smooth functions $\mathcal{F}$, their proofs when $\mathcal{F}$ is non-smooth are much more technical.

5.1.2 Global minimizers

The global minimizers of $\mathcal{F}$ are analyzed in the second part of our work [$\star$10(2005)]. Under the assumptions mentioned in § 5.1.1, function $\mathcal{F}(.,v)$ is coercive for every $v \in \mathbb{R}^q$, hence it admits minimizers [25, 73]. However, $\mathcal{F}(.,v)$ may have several global minimizers which may be misleading in applications. So we focus on the subset

$$\Gamma = \{v \in \mathbb{R}^q : \mathcal{F}(.,v) \text{ has a unique global minimizer}\}.$$

On Γ, we consider the *global minimizer function* $\hat{\mathcal{U}} : \Gamma \to \mathbb{R}^p$—the function which for every $v \in \Gamma$ yields $\hat{\mathcal{U}}(v)$, the unique global minimizer of $\mathcal{F}(.,v)$.

If Φ is $\mathcal{C}^m$, no specific assumptions are taken. When Φ is piecewise smooth, we add two non-restrictive assumptions to those given in § 5.1.1. These are H7 and H8 in section 3 in [$\star$10(2005)] and we show they hold in the case of (5.1).

Theorem 7 *Let (5.2) hold. Then*

(i) Γ^c *has Lebesgue measure zero in* $\mathbb{R}^q$ *and the interior of* Γ *is dense in* $\mathbb{R}^q$.

(ii) The global minimizer function $\hat{\mathcal{U}} : \Gamma \to \mathbb{R}^p$ *is* $\mathcal{C}^{m-1}$ *on an open subset of* Γ *which is dense in* $\mathbb{R}^q$.

A crucial consequence of Theorem 7 is that in a real-world problem *there is no chance of getting data v leading to a cost-function having more than one global minimizers.* Equivalently, the optimal solution $\hat{u}$ is almost surely unique.

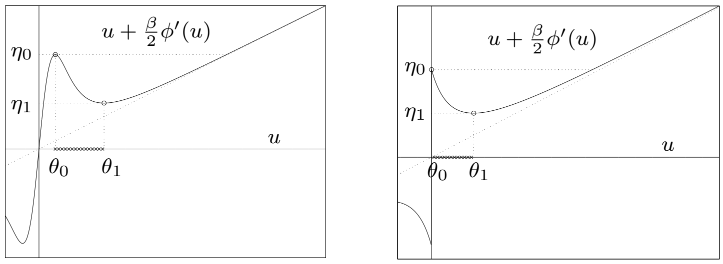
5.2 Edge enhancement

Edges in images and breaking points in signals concentrate critical information. Hence the requirement that the regularization term Φ in (1.6) leads to minimizers $\hat{u}$ involving large gradients $|Du|$ or differences $|\langle g_i, \hat{u}\rangle|$ at the location of edges in the original u, and smooth differences elsewhere. Since the pioneering work of Geman & Geman [41], different non-convex functions Φ have been considered [64, 42, 9, 69, 39, 40, 58, 5]. The relevant minimizers exhibit neat edges and well smoothed homogeneous regions. However, they are awkward to compute, to control and to analyze... In order to alleviate these intricacies, a considerable effort has been done to derive *convex edge-preserving* functions Φ, see for instance [10, 14, 24, 46, 58]. Nevertheless, possibilities are limited with respect to non-convex regularization. Various non-convex regularization functions are currently used in engineering problems. Research is mainly focused on the Mumford-Shah model, see e.g. [65, 62, 61]. For non-convex regularization of the form (1.8) or (1.9), some necessary conditions and heuristics have been suggested in [39, 58, 24].

In [$\star$6(2000)] and [$\star$1(2005)] we address the question of non-convex regularization from the point of view proposed in § 1.3. We consider that $\mathcal{F}(.,v)$ of the form (1.6)-(1.7) along with

$$\Phi(u) = \sum_{i=1}^{r} \phi(\langle g_i, u\rangle), \tag{5.4}$$

where $\{g_i\}$ are difference operators on $\mathbb{R}^p$; in the following, G will denote the $r \times p$ matrix whose rows are g_i^* for $1 \le i \le r$. In these papers, we derive formal results that explain how edges are enhanced and small differences smoothed out when ϕ is non-convex. The potential function



$\phi(u) = \frac{\alpha u^2}{(1+\alpha u^2)}$ $\qquad$ $\phi(u) = \frac{\alpha |u|}{(1+\alpha |u|)}$

Figure 5.1: Plots of $\dfrac{1}{2}D_1\mathcal{F}(u,v) - v = u + \dfrac{\beta}{2}\phi'(u)$ on $\mathbb{R}\setminus\{0\}$ for a PF satisfying H2 on the left and H3 on the right. These plots suggest how to find the local minimizers $\hat{u}$ of $\mathcal{F}(.,v)$ graphically.

H1 *$\phi : \mathbb{R} \to \mathbb{R}$ is $\mathcal{C}^2$ on $\mathbb{R}\setminus\{0\}$, satisfies $\phi(t) = \phi(-t)$ and $\phi'(t) \geq 0$ for all $t > 0$, has a strict minimum at 0 and is non-convex in the following sense: there is $\theta > 0$ such that $\phi''(\theta) < 0$ while $\lim_{t\to\infty} \phi''(t) = 0$.*

According to the smoothness of ϕ at zero, we consider either H2 or H3 as given below.

H2 *ϕ is $\mathcal{C}^2$ and there are $\tau > 0$ and $\mathcal{T} \in (\tau, \infty)$ such that $\phi''(t) \geq 0$ if $t \in [0, \tau]$ and $\phi''(t) \leq 0$ if $t \geq \tau$, where ϕ'' is decreasing on $(\tau, \mathcal{T})$ and increasing on $(\mathcal{T}, \infty)$.*

H3 *$\phi'(0^+) > 0$ and ϕ'' is increasing on $(0, \infty)$ with $\phi''(t) \leq 0$, for all $t > 0$.*

These assumptions are satisfied for almost all non-convex PFs used in practice (check Table 1.1). They can be extended to other classes of functions, too.

5.2.1 An instructive illustration on $\mathbb{R}$

For $v \in \mathbb{R}_+$, let $\mathcal{F} : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ read

$$\mathcal{F}(u,v) = (u-v)^2 + \beta\phi(u), \tag{5.5}$$

where ϕ is as specified above. Consider that $\beta > -\frac{2}{\phi''(\mathcal{T})}$ under H2 and $\beta > -\frac{2}{\phi''(0^+)}$ under H3. Define

$$\theta_0 = \inf C_\beta \quad \text{and} \quad \theta_1 = \sup C_\beta,$$

where $C_\beta = \left\{ t \in (0,\infty) : \phi''(t) < -\dfrac{2}{\beta} \right\}$. Notice that $\theta_0 = 0$ if H3 holds and that $\mathcal{T} \in (\theta_0, \theta_1)$ if H2 holds. In both cases,

$$D_1^2\mathcal{F}(u,v) = 2 + \beta\phi''(u) < 0 \quad \text{if} \quad \theta_0 < |u| < \theta_1. \tag{5.6}$$

This shows that *for any $v \in \mathbb{R}_+$, no local minimizer $\hat{u}$ of $\mathcal{F}(.,v)$ lies in $(-\theta_1, -\theta_0) \cup (\theta_0, \theta_1)$*. Conversely, minimizers $\hat{u}$ satisfy either $|\hat{u}| \in [0, \theta_0]$ or $|\hat{u}| \in [\theta_1, \infty)$. This observation underlies the property of recovering either shrunk or enhanced differences at the (local) minimizers $\hat{u}$ of $\mathcal{F}(.,v)$, discussed next. It is worth noticing that θ_0 decreases with β while θ_1 increases with β.

Let us now focus on the global minimization of $\mathcal{F}(.,v)$. Without loss of generality, suppose that the equation $\phi''(t) = -\frac{2}{\beta}$ is solved on $\mathbb{R}_+^*$ only for $\{\theta_0\}\cup\{\theta_1\}$ when H2 holds, and only for $\{\theta_1\}$ if H3 holds

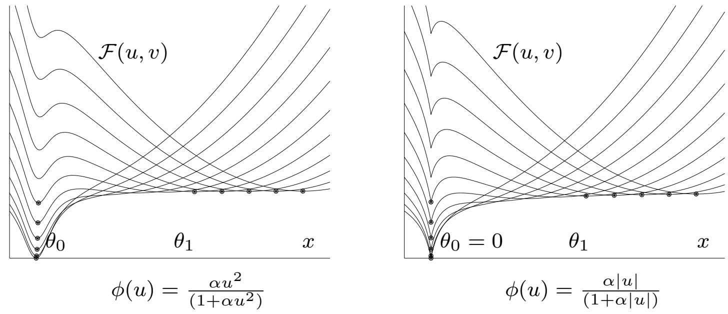


Figure 5.2: Each curve represents $\mathcal{F}(u,v) = (u-v)^2 + \beta\phi(u)$ for a different $v \in (\eta_1, \eta_0)$. The global minimizer of each $\mathcal{F}(.,v)$ is marked with "o". Observe also that no local minimizer belongs to (θ_0, θ_1).

(reader is invited to check Fig. 5.1). Some simple calculations show that for any $v \geq 0$, any (local) minimizer $\hat{u}$ of $\mathcal{F}(.,v)$ satisfies $0 \leq \hat{u} \leq v$. Define η_1 and η_0 by (see Fig. 5.1 again)

$$\eta_1 = \theta_1 + \frac{\beta}{2}\phi'(\theta_1) \quad \text{and} \quad \eta_0 = \begin{cases} \theta_0 + \dfrac{\beta}{2}\phi'(\theta_0) & \text{under H2,} \\ \dfrac{\beta}{2}\phi'(0^+) & \text{under H3.} \end{cases} \tag{5.7}$$

One can then see that if $v \in (\eta_1, \eta_0)$, there are two strict local minimizers, $\hat{u}_0 \in [0, \theta_0]$ and $\hat{u}_1 > \theta_1$. Let $\omega_0 : [0, \eta_0) \to [0, \theta_0]$ and $\omega_1 : (\eta_1, \infty) \to (\theta_1, \infty)$ denote the relevant minimizer functions. Notice that these functions are $\mathcal{C}^1$, that $\omega_0 = 0$ if H3 holds, and that

$$\big(v - \omega_0(v)\big) > \big(v - \omega_1(v)\big) + (\theta_1 - \theta_0), \quad \forall v \in (\eta_1, \eta_0).$$

Since $v - \omega_0(v)$ and $v - \omega_1(v)$ are the bias on the relevant minimizer, we can say that $\hat{u}_0 = \omega_0(v)$ incurs strong smoothing while smoothing for $\hat{u}_1 = \omega_1(v)$ is weak. If in addition $\lim_{t\to\infty} \phi'(t) = 0$ (which is a current assumption), then

$$\lim_{v\to\infty} |v - \omega_1(v)| = 0.$$

Hence smoothing for $\hat{u}_1 = \omega_1(v)$ is vanishing which corresponds to the recovery of sharp edges.

Furthermore, one can show that there is a unique $\overline{\eta} \in (\eta_1, \eta_0)$ such that

- if $v \in [0, \overline{\eta})$, the *global* minimizer $\hat{u}$ of $\mathcal{F}(.,v)$ satisfies $\hat{u} = \omega_0(v) \in [0, \theta_0]$;
- if $v > \overline{\eta}$, the *global* minimizer $\hat{u}$ of $\mathcal{F}(.,v)$ is such that $\hat{u} = \omega_1(v) > \theta_1$,

whereas $\mathcal{F}_{\overline{\eta}}$ has two global minimizers, $\omega_0(\overline{\eta})$ and $\omega_1(\overline{\eta})$. This behavior is illustrated in Fig. 5.2. Clearly, the *global minimizer function* is discontinuous at $v = \overline{\eta}$. The critical value $v = \overline{\eta}$ can be seen as a threshold to decide whether or not the global minimizer $\hat{u}$ of $\mathcal{F}(.,v)$ incurs strong smoothing. In the context of signal and image reconstruction, this corresponds to the decision on whether a difference belongs to a homogeneous region or to an edge.

5.2.2 Either shrinkage or enhancement of the differences

Consider first that the regularization is smooth, i.e. ϕ satisfies H1 and H2.

Theorem 8 *Assume the set* $\{g_i : 1 \le i \le r\}$ *is linearly independent and put* $\mu = \max_{1\le i\le r} \|G^*(GG^*)^{-1}e_i\|$.

$$\text{If} \quad \beta > \frac{2\mu^2\, \|A^*A\|}{|\phi''(\mathcal{T})|}, \tag{5.8}$$

there exist $\theta_0 \in (\tau, \mathcal{T})$ *and* $\theta_1 \in (\mathcal{T}, \infty)$ *such that for every* $v \in \mathbb{R}^q$, *every local minimizer* $\hat{u}$ *of* $\mathcal{F}(., v)$ *satisfies*

$$\text{either} \quad |\,\langle g_i, \hat{u}\rangle\,| \le \theta_0, \quad \text{or} \quad |\,\langle g_i, \hat{u}\rangle\,| \ge \theta_1, \quad \forall i \in \{1, \dots, r\}. \tag{5.9}$$

A reciprocal statement says that if we fix either θ_0 or θ_1, we can find a suitable β such that (5.9) holds. The thresholds θ_0 and θ_1 in the proof delimit only the regions in $\mathbb{R}^p$ where $D^2\mathcal{F}(u, v)$ is not non-negative definite. We can expect that they are pessimistic.

The assumption that $\{g_i\}$ is linearly independent fails in usual image restoration problems. Nevertheless, the analysis above is easy to extend to all situations where a (local) minimizer $\hat{u}$ is homogeneous on some connected regions. Details are given in [$\star$1(2005)].

Truncated quadratic PF. This PF $\phi(t) = \min\{\alpha t^2, 1\}$ is the discrete version of the Mumford-Shah model [11]. It fails to satisfy the assumptions above. We consider that $\{g_i : 1 \le i \le r\}$ is linearly independent.

Proposition 1 *If* $\mathcal{F}(., v)$ *reaches its global minimum at* $\hat{u}$, *then for every* $i \in \{1, \dots, r\}$ *we have:*

(i) if $Be_i = 0$, *then* $\langle g_i, \hat{u}\rangle = 0$*;*

(ii) if $Be_i \ne 0$, *then*

$$\text{either} \quad |\,\langle g_i, \hat{u}\rangle\,| \le \frac{1}{\sqrt{\alpha}}\,\Gamma_i \quad \text{or} \quad |\,\langle g_i, \hat{u}\rangle\,| \ge \frac{1}{\sqrt{\alpha}\,\Gamma_i}, \tag{5.10}$$

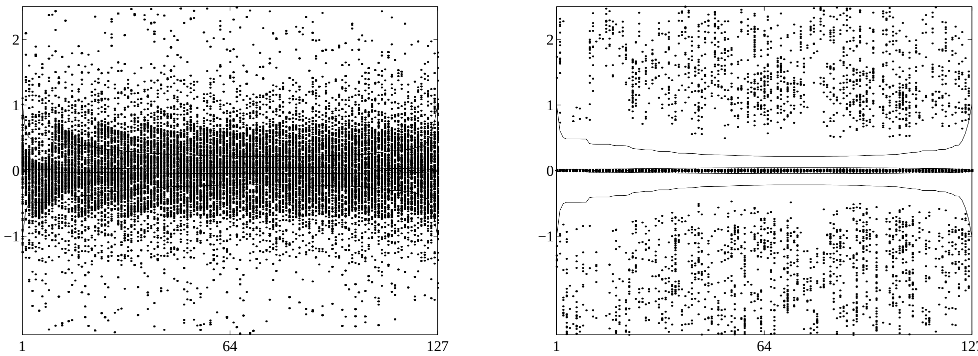
where B *is a matrix that depends only on* A *and* G, *and*

$$\Gamma_i = \sqrt{\frac{\|Be_i\|^2}{\|Be_i\|^2 + \alpha\beta}} < 1.$$

Moreover, the inequalities in (5.10) are strict if $\mathcal{F}(., v)$ *has a unique global minimizer.*

The matrix B is easy to compute but we skip the details. Since the result holds for global minima exclusively, this proposition furnishes a necessary condition for a global minimum of $\mathcal{F}(., v)$. The result is pretty fine since the thresholds are adapted to each difference. In particular, (5.9) holds for $\theta_0 = \frac{\gamma}{\sqrt{\alpha}}$ and $\theta_1 = \frac{1}{\sqrt{\alpha}\gamma}$, where $\gamma = \max_{1\le i\le r} \Gamma_i < 1$.

Numerical illustration of Proposition 1. We generate 100 random signals u^k, $1 \le k \le 100$, of length 128. At each position $i \in \{1, \dots, 127\}$ of the x-axis of Fig. 5.3 (a) we plot (with dots) the set $\{u_i^k - u_{i+1}^k : 1 \le k \le 100\}$. For a given A, we compute the thresholds in (5.10). Namely, the curves with a solid line in Fig. 5.3 (b), from the top to the bottom, represent $(\sqrt{\alpha}\Gamma_i)^{-1}$, $\Gamma_i/\sqrt{\alpha}$, $-\Gamma_i/\sqrt{\alpha}$ and $-(\sqrt{\alpha}\Gamma_i)^{-1}$. For each $k \in \{1, \dots, 100\}$ we generate data $v^k = Au^k + n^k$ where n^k is noise and then compute the global minimizer $\hat{u}^k$ of $\mathcal{F}(u, v^k)$ where ϕ is the truncated quadratic function. At each position



(a) The differences for the original signals: for $1 \le k \le 100$, each dot at position $i \in \{1, \dots, 127\}$ on the x-axis represents $u_i^k - u_{i+1}^k$

(b) The differences for the global minimizers for $1 \le k \le 100$, each dot at position $i \in \{1, \dots, 127\}$ on the x-axis represents $\hat{u}_i^k - \hat{u}_{i+1}^k$

Figure 5.3: Distribution of the differences of the original signals (left) and of the global minimizers (right). The thresholds $\pm\Gamma_i/\sqrt{\alpha}$, $\pm 1/\sqrt{\alpha}\Gamma_i$ for $i = 1, \dots, 127$ are plotted with solid lines (—).

$i = 1, \dots, 127$ of the x-axis in Fig. 5.3 (b) we represent (with dots) the set $\{\hat{u}_i^k - \hat{u}_{i+1}^k : 1 \le k \le 100\}$. For many k, many differences satisfy $|\hat{u}_i^k - \hat{u}_{i+1}^k| \le \Gamma_i/\sqrt{\alpha}$. As predicted by Proposition 1, for no i, no difference has its magnitude $|\hat{u}_i^k - \hat{u}_{i+1}^k|$ in $(\Gamma_i/\sqrt{\alpha}, (\sqrt{\alpha}\Gamma_i)^{-1}$, for any k.

Next we focus on functions ϕ that are non-smooth at zero. In this case, all formulae are more complicated but the results are stronger. Notice that no assumption on $\{g_i : 1 \le i \le r\}$ is done.

Theorem 9 *Let H1 and H3 hold. If*

$$\beta > \frac{2\mu^2 \, \|A^* A\|}{|\phi''(0^+)|}, \tag{5.11}$$

where $\mu > 0$ is a constant that depends only on $\{g_i : 1 \le i \le r\}$, then there exists $\theta_1 > 0$ such that for every $v \in \mathbb{R}^q$, every (local) minimizer $\hat{u}$ of $\mathcal{F}(., v)$ satisfies

$$\textit{either} \quad |\langle g_i, \hat{u}\rangle| = 0, \quad \textit{or} \quad |\langle g_i, \hat{u}\rangle| \ge \theta_1, \quad \forall i \in \{1, \dots, r\}. \tag{5.12}$$

If $|\phi''(0^+)| = \infty$, we find $\beta_0 = 0$ in (5.11). We also prove a reciprocal statement saying that if we fix $\theta_1 > 0$, there is $\beta > 0$ such that (5.12) holds.

Remark 4 By (5.12), *all non-zero differences are necessarily larger then θ_1*. Thus $\hat{u}$ is piecewise constant with large edges. We are hence faced with an enhanced stair-casing effect.

"0-1" PF. This function reads $\varphi(0) = 0$, $\varphi(t) = 1$ if $t \ne 0$: it is discontinuous at 0 and does not satisfy the assumptions given above. As in Proposition 1, we suppose that $\{g_i : i \in J\}$ is linearly independent and focus on the global minimizers of $\mathcal{F}(., v)$.

Proposition 2 *If $\mathcal{F}(., v)$ has a global minimum at $\hat{u}$, then for every $i \in \{1, \dots, r\}$,*

(i) if $Be_i = 0$, then $\langle g_i, \hat{u}\rangle = 0$;

(ii) if $Be_i \neq 0$, then

$$\textit{either} \quad \langle g_i, \hat{u}\rangle = 0 \quad \textit{or} \quad |\,\langle g_i, \hat{u}\rangle\,| \geq \frac{\sqrt{\beta}}{\|Be_i\|}, \tag{5.13}$$

where B is the matrix mentioned in Proposition 1. The last inequality is strict if $\mathcal{F}(.,v)$ has a unique global minimizer.

This proposition provides a simple necessary condition for a global minimum of $\mathcal{F}(.,v)$. Notice that (5.13) is finely adapted to each difference $\langle g_i, \hat{u}\rangle$, and that (5.12) holds if we put $\theta_1 = \min\limits_{1\leq i\leq r} \dfrac{\sqrt{\beta}}{\|Be_i\|}$.

5.3 Selection for the global minimum

In section 4 in [$\star$1(2005)] we study how an original image or signal of the form $\eta\mathbb{1}_\Sigma$, where $\eta > 0$, the sets $\Sigma \subset \{1,\ldots,p\}$ and Σ^c are nonempty, and $\mathbb{1}_\Sigma \in \mathbb{R}^p$ reads

$$\mathbb{1}_\Sigma[i] = \begin{cases} 1 & \text{if } \ i \in \Sigma, \\ 0 & \text{if } \ i \in \Sigma^c, \end{cases} \tag{5.14}$$

is recovered at the *global* minimizer $\hat{u}$ of $\mathcal{F}(.,v)$ when $v = A\ \eta\mathbb{1}_\Sigma$ and $\{g_i\}$ are first-order difference operators. Additional assumptions are that $\phi(t) \leq 1$ for all $t \in \mathbb{R}$ (it is easily justified by Table 1.1) and that A^*A is invertible. By the latter, $\mathcal{F}(.,v)$ reaches its global minimum on a bounded subset of $\mathbb{R}^p$. From now on, we systematically denote

$$J = \Big\{ i \in \{1,\ldots,r\} : \langle g_i, \mathbb{1}_\Sigma\rangle \neq 0 \Big\} \tag{5.15}$$

Observe that J address the edges in $\mathbb{1}_\Sigma$. It will be convenient to denote by $\hat{u}_\eta$ a *global minimizer* of the function $u \to \mathcal{F}(u, \eta A\mathbb{1}_\Sigma)$. Below we resume the main results demonstrated in section 4 in [$\star$1(2005)].

- Considering smooth regularization in the context of Theorem 8, we exhibit two constants $\eta_0 > 0$ and $\eta_1 > \eta_0$ such that

  $$\eta \in [0, \eta_0) \quad \Rightarrow \quad |\,\langle g_i, \hat{u}_\eta\rangle\,| \leq \theta_0, \quad \forall i \in \{1,\ldots,r\} \tag{5.16}$$

  whereas

  $$\eta \geq \eta_1 \quad \Rightarrow \quad \begin{array}{ll} |\,\langle g_i, \hat{u}_\eta\rangle\,| \leq \theta_0, & \forall i \in J^c, \\ |\,\langle g_i, \hat{u}_\eta\rangle\,| \geq \theta_1, & \forall i \in J. \end{array}$$

  This result corroborates the interpretation of θ_0 and θ_1 as thresholds for the detection of smooth differences and edges, respectively.

- For the truncated quadratic function, define $\omega_\Sigma \in \mathbb{R}^p$ by

  $$\omega_\Sigma = (A^*A + \beta\alpha G^*G)^{-1}\, A^*A\mathbb{1}_\Sigma. \tag{5.17}$$

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

  $$\eta \in [0, \eta_0) \quad \Rightarrow \quad \hat{u}_\eta = \eta\, \omega_\Sigma, \tag{5.18}$$

  $$\eta \geq \eta_1 \quad \Rightarrow \quad \hat{u}_\eta = \eta\ \mathbb{1}_\Sigma. \tag{5.19}$$

  Moreover, $\hat{u}_\eta$ in (5.18) and (5.19) is the unique global minimizer of the relevant $\mathcal{F}(., \eta A\mathbb{1}_\Sigma)$. Observe that $\eta\omega_\Sigma$ in (5.18) is the regularized least-squares solution, and it does not involve edges. For $\eta \geq \eta_1$ the global minimizer $\hat{u}_\eta$ is equal to the original u.

- When ϕ is non-smooth at zero, we have $\eta_0 > 0$ and $\eta_1 > \eta_0$ such that

$$\eta \in [0, \eta_0) \quad \Rightarrow \quad \hat{u}_\eta = \eta\zeta \, \mathbb{1}, \quad \text{where} \quad \zeta = \frac{(A\mathbb{1})^* A\mathbb{1}_\Sigma}{\|A\mathbb{1}\|^2} \tag{5.20}$$

whereas

$$\eta > \eta_1 \quad \Rightarrow \quad \begin{array}{ll} \langle g_i, \hat{u}_\eta \rangle = 0, & \forall i \in J^c, \\ |\langle g_i, \hat{u}_\eta \rangle| \geq \theta_1, & \forall i \in J. \end{array}$$

So, if η is small, $\hat{u}_\eta$ is constant, while for η large enough, $\hat{u}_\eta$ has the same edges and the same constant regions as the original $\eta\mathbb{1}_\Sigma$. Moreover, if Σ and Σ^c are connected with respect to $\{g_i : 1 \leq i \leq r\}$, there are $\hat{s}_\eta \in (0, \eta]$ and $\hat{c}_\eta \in \mathbb{R}$ such that

$$\hat{u}_\eta = \hat{s}_\eta \mathbb{1}_\Sigma + \hat{c}_\eta \mathbb{1}, \tag{5.21}$$

and $\hat{s}_\eta \to \eta$ and $\hat{c}_\eta \to 0$ as $\eta \to \infty$. Hence $\hat{u}_\eta$ provides a faithful restoration of the original $\eta\mathbb{1}_\Sigma$.

- Very light assumptions are needed to show that when ϕ is the "0-1" potential function, there are $\eta_0 > 0$ and $\eta_1 > \eta_0$ such that

$$\eta \in [0, \eta_0) \quad \Rightarrow \quad \hat{u}_\eta = \eta\zeta \, \mathbb{1}, \tag{5.22}$$

$$\eta > \eta_1 \quad \Rightarrow \quad \hat{u}_\eta = \eta \, \mathbb{1}_\Sigma, \tag{5.23}$$

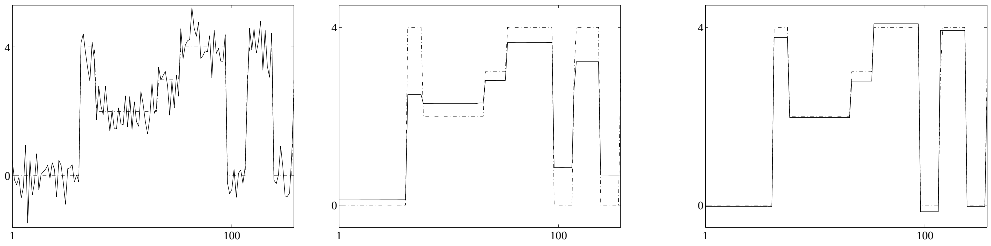
where ζ is given in (5.20). Moreover, $\hat{u}_\eta$ in (5.22) and (5.23) is the unique global minimizer of $F(., \eta A\mathbb{1}_\Sigma)$.

Experiments on signal denoising. Let us come back to the experiment presented in Figs. 2.1 and 2.2 in § 2.1. The function ϕ used in Fig. 2.2(c) is non-convex and non-smooth at zero. The two minimizers plotted there, corresponding to two different realizations of the noise corrupting the data, involve sharp edges whose height is close to the original and they are neatly segmented in the same way. In contrast, the edges in Figs. 2.2(a)-(b) are underestimated which can be explained by the convexity of ϕ.

Experiments on image deblurring. We interpret the experiments in § 2.1 in the light of the new theoretical results on edge-enhancement. The restorations in Fig. 2.4 (a) and (b) correspond to convex functions ϕ. The edges in (a) are slightly blurred and underestimated while (b) shows a strong stair-caising effect. The restorations in Fig. 2.5 are calculated using non-convex functions ϕ. Those in the first row correspond to smooth at zero functions ϕ while those in the second row correspond to non-smooth at zero functions ϕ. On the average, the important edges are very neat and their amplitude is correct. In addition, the images corresponding to smooth functions ϕ—Fig. 2.5 (a) and (b)— have smoothly varying homogeneous regions whereas those corresponding to a non-smooth at zero ϕ—(c) and (d)—are piecewise constant with high edges, as suggested by (5.12) in Theorem 9.

5.4 Convex edge-preserving regularization versus non-convex regularization

Each local minimizer $\hat{u}$ of $\mathcal{F}(., v)$ can be seen as resulting from a continuous local minimizer function $v \to \omega(v)$ defined on a subset of $O \subset \mathbb{R}^q$, i.e. $\hat{u} = \omega(v)$. From the continuity, $\omega : O \to \mathbb{R}^p$ produces



Data $v = u + n$ (solid line) Original u_o (dashed line) | Total-variation regularization $\varphi(t) = t$ | Nonconvex non-smooth regularization $\varphi(t) = \alpha t/(1 + \alpha t)$

Figure 5.4: Convex edge-preserving regularization versus non-convex regularization.

minimizers $\hat{u} = \omega(v)$, for $v \in O$, that have the same set of edges $J = \{i \in \{1, \ldots, r\} : \langle g_i, \omega(v)\rangle \geq \theta_1\}$ for all $v \in O$.

Given $v \in \mathbb{R}^q$, let $\mathcal{F}(., v)$ reach its global minimum at $\hat{u} = \omega(v)$ with edges J and homogeneous regions J^c. When data vary in a neighborhood of v in such a way that noticeable edges either appear or disappear in the original u, the *global* minimum jumps from the (local) minimizer function ω with edges J to another (local) minimizer function ω' whose edges are $J' \neq J$. This discontinuity of the *global minimizer function* allows edges to be detected or removed at the global minimum of $\mathcal{F}(., v)$. Using the results in § 5.1, such discontinuities occur only at data points included in a negligible subset of $\mathbb{R}^q$.

In contrast, if $\mathcal{F}(., v)$ is strictly convex for every $v \in \mathbb{R}^q$, there is a unique minimizer function $\omega : \mathbb{R}^q \to \mathbb{R}^p$ and the latter is continuous [14]. In particular, differences $\langle g_i, \hat{u}\rangle$ can take any value on $\mathbb{R}$. The edge-preservation properties of $\phi(t) = |t|$—the TV regularization—have been extensively discussed in the literature. We should emphasize that they are based on a totally different property. As explained in § 2.1, the relevant minimizers $\hat{u}$ exhibit *stair-casing*: for many differences, $\langle g_i, \hat{u}\rangle = 0$, so $\hat{u}$ contains constant regions. The non-zero differences that separate the constant regions in $\hat{u}$ then naturally appear as edges. This effect is observed in Fig. 2.4(b) where numerous spurious edges appear on planar-shaped regions. A nice illustration is given in Fig. 5.4.

Thus, image and signal restoration using non-convex regularization is fundamentally different from restoration using convex regularization. The main difference is related to the (dis)continuity of the global minimizers with respect to the data.

Chapter 6

Critical remarks on Bayesian MAP

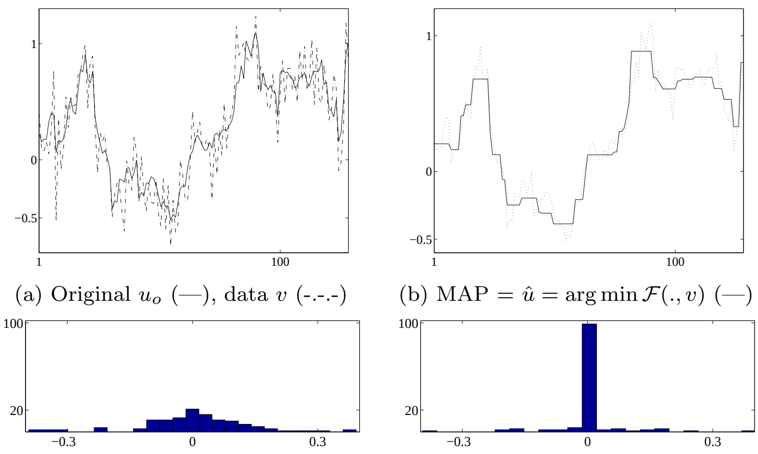
Let us consider the simple example presented in Fig. 6.1. The original u, plotted with a solid line in Fig. 6.1 (a), is generated according to the law $\pi(u_i - u_{i+1}) = \exp\{-\beta|u_i - u_{i+1}|\}/Z$, $1 \leq i \leq p-1$, for $p = 128$ and $\beta = 10$. The histogram of the differences $u_i - u_{i+1}$ is seen below in (c). Data v, plotted in (a) with a dashed line, are generated as $v = u+n$ where n is white Gaussian noise with $\sigma^2 = 0.04$. The true Bayesian MAP solution is hence the minimizer $\hat{u}$ of $\mathcal{F}(u,v) = \frac{1}{2\sigma^2}\|u-v\|^2 + \beta\sum_{i=1}^{p-1}|u_i - u_{i+1}|$. This solution is plotted in (b) with a solid line (data v are recalled there with a dashed line) and the histogram of the differences $\hat{u}_i - \hat{u}_{i+1}$ is presented in (d). In particular, $\hat{u}$ involves 96 zero-valued differences. Obviously the obtained $\hat{u}$ is very far from the prior model!

Our studies on the properties of minimizers exhibited several points of disagreement between prior models and MAP solutions. Some of them were discussed in [$\star$52(2000)]. We briefly describe the essential points and skip the technical details.

- Let Φ be of the form (2.2) with φ continuous and non-smooth at zero, as considered in that section. The prior $\pi(u)$ relevant to (1.14) is a continuous function, so $\Pr\big(G_i u = \theta_i\big) = 0$. (Here Pr stands for probability.) Consider that $\mathcal{F}$ is of the form (1.6) where Ψ is a smooth data-fidelity function. According to Theorem 1, that the minimizers $\hat{u}$ of $\mathcal{F}(.,v)$, when v ranges over some open domain, are such that $\Pr\big(G_i\hat{u} = \theta_i\big) > 0$, hence the law of $\hat{u}$ contains Dirac distributions. Such a solution $\hat{u}$ cannot be in agreement with the prior conveyed by Φ.

- Consider that Ψ is of the form (3.2) with ψ non-smooth at zero and continuous, and that the assumptions given in § 3.1 hold. For definiteness, let ψ be even and increasing on $\mathbb{R}_+$, and let Φ be a smooth regularization function. Using (1.13), the likelihood $\pi(v|u)$ is a continuous function, hence $\Pr\big(\langle a_i, u\rangle = v_i\big) = \Pr(\text{noise}_i = 0) = 0$. However, by Theorem 3, the minimizer $\hat{u}$ is such that $\Pr\big(\langle a_i, \hat{u}\rangle = v_i\big) > 0$. The law of the residuals corresponding to $\hat{u}$ contains Dirac distributions. There is a gap between the model for the noise and the behavior of the residuals.

- Now let Φ be continuous and non-convex as considered in § 5.2. Consider that u belongs to a ball $B(0,\rho)$ where the radius $\rho > 0$ is large enough. The prior defined by (1.14) clearly satisfies $\pi(u) > 0$ for all $u \in B(0,\rho)$. Let θ_0 and θ_1 be the thresholds exhibited in § 5.2. In particular, $\Pr\big(|\langle g_i, u\rangle| \in$



(a) Original u_o (—), data v (-.-.-)

(b) MAP $= \hat{u} = \arg\min \mathcal{F}(., v)$ (—)

(c) Histogram of $|u_i - u_{i+1}|$

(d) Histogram of $|\hat{u}_i - \hat{u}_{i+1}|$ at $\hat{u}$.

Figure 6.1: MAP signal reconstruction using the true prior law and the true parameters.

$(\theta_0, \theta_1)\big) > 0$. Furthermore, let Ψ be a smooth data-fidelity term. From Theorems 8 and 9, and Propositions 1 and 2, it follows that the minimizers $\hat{u}$ of $\mathcal{F}(., v)$ are such that $\Pr\big(|\langle g_i, \hat{u}\rangle| \in (\theta_0, \theta_1)\big) = 0$. The latter is a contradiction with the prior model for u. By the way, it is nicely illustrated in Fig. 5.3 for the case of a truncated quadratic regularization.

- The two methods for restoration of binary images, described in § 2.4.2 and § 4.2, are a blatant illustration of the disagreement between model and solution. In a Bayesian setting, these models corresponds to Gaussian or Laplace noise, respectively, on the data v, and to Laplace prior on the differences between neighboring pixels. In spite of this, the solutions are either quasi-binary, or binary.

MAP estimation comes from the minimization of a 0-1 cost and expresses a risk diminishing. In particular, it does not require that the obtained solutions are in agreement with the likelihood or the prior models. However, the disagreements between the models and the relevant solutions is misleading when one has to solve signal and image reconstruction problems.

Chapter 7

Computational issues

7.1 Non-convex minimization using continuation ideas

The minimization of non-convex energies of the form (1.6) especially when $\Psi = \sum_i \psi_i$ is such that the calculation of each ψ_i involves more than a few samples of u, is still an open problem. The reasons explained in [$\star$8(1999)] and [$\star$20(1998)] remain currently in force. We briefly sketch the approach developed in these papers which was inspired by the work of Blake and Zisserman [11] in the context of visual reconstruction.

We consider that Ψ in (1.6) is convex but possibly only nonstrictly so, as is the case in usual inverse problems. The regularization term is of the form

$$\Phi(u) = \sum_k \gamma_k \sum_i \phi(|\langle g_i^k, u\rangle|),$$

where $\{g_i^k : i \geq 1\}$ are difference operators of the kth order, γ_k are weighting constants and ϕ are general non-convex and possibly non-smooth potential functions (including all non-convex PFs given in Table 1.1). Global minimizer $\hat{u}$ is approximated by a sequence of local minimizers $\hat{u}_\ell$ for $\ell = 1, \ldots, L$,

$$\begin{aligned} \hat{u}^\ell &= \arg \min_{V(\hat{u}^{\ell-1})} \mathcal{F}_{\rho_\ell}(u, v) \\ \mathcal{F}_{\rho_\ell}(u, v) &= \Psi(u, v) + \beta \Phi_{\rho_\ell}(u) \end{aligned}$$

where $V(u)$ denotes the smaller neighborhood of u containing a local minimizer of $\mathcal{F}_{\rho_\ell}(., v)$, so that the minimization of each $\mathcal{F}_{\rho_\ell}(., v)$ is realized using standard descent methods. The sequence ρ_ℓ is increasing in $[0, 1]$ and approximated energies $\mathcal{F}_{\rho_\ell}(., v)$ are smooth and such that $\mathcal{F}_{\rho_0}(., v)$ is strictly convex and $\mathcal{F}_{\rho_L}(., v) \approx \mathcal{F}(., v)$. Even though convergence of $\hat{u}^\ell$ towards a global minimizer cannot be ensured, a careful construction of $\mathcal{F}_\rho$ can improve the chance to get a better approximation. In this respect, approximations $\mathcal{F}_\rho$ are as close as possible to the original $\mathcal{F}$ with a special control on the points where $D_1^2 \mathcal{F}_\rho(u, v)$ reaches its minimum (which is < 0 for $\ell > 1$). Examples and counter-examples are provided. Two numerical experiments—an image deblurring and an emission tomography reconstruction—illustrate the performance of the proposed technique with respect to the main alternative approaches.

7.2 Convergence rate for the main forms of half-quadratic regularization

Consider $\mathcal{F}$ of the form defined by (1.6)-(1.7) and (5.4) where $\phi : \mathbb{R}_+ \to \mathbb{R}$ is convex, smooth and *edge-preserving*. Examples of such functions are (f1)-(f4) in Table 1.1. Suppose also that A^*A is invertible and/or $\phi''(t) > 0$, $\forall t \in \mathbb{R}_+$. Then for every $v \in \mathbb{R}^p$, $\mathcal{F}(.,v)$ has a unique minimum and the latter is strict. In spite of this well-posedness, $\mathcal{F}(.,v)$ may exhibit nearly flat regions where minimization methods progress very slowly. In order to cope with the computation, *half-quadratic* (HQ) reformulation of $\mathcal{F}$ was pioneered, in two different ways, in [39] and [40]. For simplicity, we will write $\mathcal{F}(u)$ for $\mathcal{F}(u,v)$. The idea is to construct an *augmented cost-function* $F : \mathbb{R}^p \times \mathbb{R}^r \to \mathbb{R}$ which involves an auxiliary variable $b \in \mathbb{R}^r$,

$$F(u,b) = \|Au - y\|^2 + \beta \sum_{i=1}^{r} \Big(Q(\langle g_i, u\rangle \, , \, b_i) + \psi(b_i) \Big), \tag{7.1}$$

where $\psi : \mathbb{R} \to \mathbb{R}$ satisfies $\phi(t) = \min_{s\in\mathbb{R}} \{Q(t, \ s) + \psi(s)\}$, $\forall t \in \mathbb{R}$, so that $\mathcal{F}(u) = \min_{b\in\mathbb{R}^r} F(u,b)$. The main forms for Q were introduced in [39] and [40] and read:

$$\text{multiplicative form} \qquad Q(t,s) = \frac{1}{2}t^2 s, \quad \text{for } \ t \in \mathbb{R}, \quad s \in \mathbb{R}_+ \tag{7.2}$$

$$\text{additive form} \qquad Q(t,s) = (t-s)^2, \quad \text{for } \quad t \in \mathbb{R}, \quad s \in \mathbb{R}. \tag{7.3}$$

The minimizer $(\hat{u}, \hat{b})$ of F in (7.1) is found using *alternate minimization.* So at iteration k we calculate

$$b^{(k)} = \arg\min_b F(u^{(k-1)}, b) \tag{7.4}$$

$$u^{(k)} = \arg\min_u F(u, b^{(k)}). \tag{7.5}$$

The key points are that the minimization with respect to b is done in a componentwise way while the function on the right is quadratic with respect to u. We formally write these iterations in the form

$$b_i^{(k)} = \sigma\left(\left\langle g_i, u^{(k-1)}\right\rangle\right), \quad 1 \le i \le r, \tag{7.6}$$

$$u^{(k)} = \omega(b^{(k)}), \tag{7.7}$$

where $\sigma : \mathbb{R} \to \mathbb{R}$ and $\omega : \mathbb{R}^r \to \mathbb{R}^p$ are the relevant minimizer functions.

The multiplicative form was considered in [23, 24, 30, 51, 56, 78] and the additive form in [6, 23, 27, 51]. In [27], the auxiliary variable is introduced in a non-convex data-fidelity term. Extensions of the multiplicative form were proposed in [78, 51]. The convergence rate of (7.4)-(7.5) for the multiplicative form is considered in [24, 30, 51] and for the additive form in [6]. The numerical results have shown that both forms of HQ regularization can speed up computation. However, their convergence rates has never been analyzed and compared in a systematic way.

In a joint work with Michael Ng [$\star$11(2005)], we show that both the multiplicative and the additive forms of HQ regularization can be put into the form of quasi-Newton minimization and we analyze their convergence rates.

Multiplicative form. The standard assumptions are that $t \to \phi(|t|)$ is convex and $\mathcal{C}^1$ on $\mathbb{R}$, $t \to \phi(\sqrt{t})$ is concave on $\mathbb{R}_+$, $\phi''(0^+) > 0$ and $\lim_{t\to\infty} \phi(t)/t^2 = 0$. The minimizer function σ introduced in (7.4) and (7.6) has an explicit form that was originally determined in [24]:

$$\sigma(t) = \begin{cases} \phi''(0^+) & \text{if} \quad t = 0, \\ \dfrac{\phi'(t)}{t} & \text{if} \quad t \neq 0. \end{cases} \tag{7.8}$$

It satisfies $\sigma(t) > 0$ for all $t \in \mathbb{R}$. The minimizer function with respect to u is

$$\begin{aligned} \omega(b) &= (H(b))^{-1} \, 2A^*v \\ H(b) &= 2A^*A + \beta G^* \text{diag}(b) G \succ 0, \ \forall b > 0. \end{aligned} \tag{7.9}$$

For our convergence rate analysis, we suppose also that ϕ is $\mathcal{C}^2$ on $\mathbb{R}$ and $\mathcal{C}^3$ near zero, which is not a practical assumption.

Additive form. The basic assumptions are that ϕ is convex and $\exists c > 0$ such that $t \to ct^2/2 - \phi(t)$ is convex, ϕ is continuous on $\mathbb{R}_+$, $\lim\limits_{|t|\to\infty} \phi(t)/t^2 < c/2$. The minimizer function σ admits an explicit form [23, 27, 6]:

$$\sigma(t) = ct - \phi'(t). \tag{7.10}$$

The minimizer function with respect to u reads

$$\begin{aligned} \omega(b) &= \mathcal{H}^{-1} \left(2A^*y + \beta G^* b \right), \\ \mathcal{H} &= 2A^*A + \beta c \, G^*G \succ 0. \end{aligned} \tag{7.11}$$

For each form, let $\mathcal{T} : \mathbb{R}^p \to \mathbb{R}^p$ be defined as

$$\mathcal{T}(u) = \omega\Big([\sigma(\langle g_i, u \rangle)]_{i=1}^r \Big). \tag{7.12}$$

Half-quadratic iterations (7.4)-(7.5) are then equivalent to

$$u^{(k)} = \mathcal{T}(u^{(k-1)}), \quad \text{for all} \ \ k \in \mathbb{N}. \tag{7.13}$$

More precisely, we show that

$$\mathcal{T}(u) = u - (\mathcal{H}(u))^{-1} D\mathcal{F}(u), \tag{7.14}$$

where and $D\mathcal{F}$ denotes the differential of $\mathcal{F}$ and

$$\begin{aligned} &\text{multiplicative form} & \mathcal{H}(u) &= 2A^*A + \beta G^* \text{diag}\left([\sigma(\langle g_i, u \rangle)]_{i=1}^r \right) G \succ 0, & (7.15) \\ &\text{additive form} & \mathcal{H}(u) = \mathcal{H} &= 2A^*A + \beta c G^*G \succ 0. & (7.16) \end{aligned}$$

Hence each form of HQ regularization is a variant of Newton minimization where the matrix $\mathcal{H}(u)$ provides a correction of the steepest descent direction $-D\mathcal{F}(u)$. Recall that the classical Newton method corresponds to $\mathcal{H}(u) = D^2\mathcal{F}(u)$, where

$$D^2\mathcal{F}(u) = 2A^*A + \beta G^* \text{diag}\left([\phi''(\langle g_i, u \rangle)]_{i=1}^r \right) G.$$

More details are given in section 4 in [$\star$11(2005)]. Furthermore, HQ minimization amounts to find the fixed point $\hat{u}$ of $\mathcal{T}$. Given a matrix $H \in \mathbb{R}^{p\times p}$, we write $\rho_{\max}(H)$ for its largest-in-magnitude eigenvalue and $\rho_{\min}(H)$ for its smallest-in-magnitude eigenvalue. In order to analyze the contraction properties of $\mathcal{T}$, we focus on the *root-convergence factor* [25, 67]

$$\begin{aligned} C(\mathcal{T}, \hat{u}) &= \sup\left\{ \limsup_{k\to\infty} \left\| u^{(k)} - \hat{u} \right\|^{\frac{1}{k}} \ : \ u^{(0)} \in \mathbb{R}^p \right\}, && (7.17) \\ &= \rho_{\max}(D\mathcal{T}(\hat{u})), && (7.18) \end{aligned}$$

where the second equality comes from the linear convergence theorem [67, p. 301].

Theorem 10 *For each form of HQ regularization the spectral radius of $D\mathcal{T}$ at $\hat{u}$ satisfies*

$$\rho_{\max}\left(D\mathcal{T}(\hat{u})\right) \leq K \max_{1\leq i\leq r} \mathcal{R}(|\langle g_i, \hat{u}\rangle|) \ < 1, \qquad (7.19)$$

where $K \in [0,1]$ and $\mathcal{R} : \mathbb{R}_+ \to [0,1]$ is continuous and

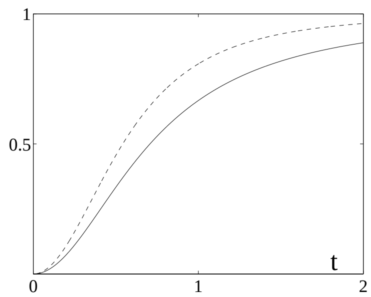
$$\text{multiplicative form} \quad \left\{ \begin{aligned} K &= \frac{\beta\phi''(0)\rho_{\max}(G^*G)}{2\rho_{\min}(A^*A) + \beta\phi''(0)\rho_{\max}(G^*G)}, \\ \mathcal{R}(t) &= 1 - \frac{\phi''(t)}{\sigma(t)}, \end{aligned} \right. \qquad (7.20)$$

$$\text{additive form} \quad \left\{ \begin{aligned} K &= \frac{\beta c\rho_{\max}(G^*G)}{2\rho_{\min}(A^*A) + \beta c\rho_{\max}(G^*G)}, \\ \mathcal{R}(t) &= 1 - \frac{\phi''(t)}{c} \end{aligned} \right. \qquad (7.21)$$

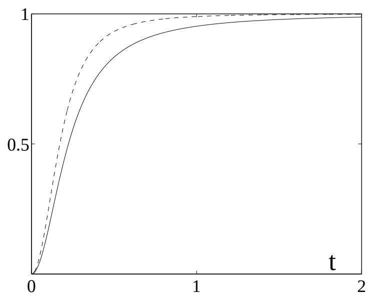
The result in (7.21) recommend to choose $c = \phi''(0)$ for the constant c in the additive form § **??** In such a case, the constant K is the same for both forms. If A^*A is singular, $K = 1$. Furthermore, for all edge-preserving functions ϕ used in practice, the function $\mathcal{R}$ is monotone increasing on $\mathbb{R}_+$ and

$$\underset{\text{multipl. form}}{\mathcal{R}(t)} \quad < \quad \underset{\text{additive form}}{\mathcal{R}(t),} \qquad \forall t \in \mathbb{R} \setminus \{0\},$$

with $\mathcal{R}(0) = 0$ for both forms. An illustration is seen in Fig. 7.1. This suggests that the multiplicative form needs fewer iterations than the additive form in order to reach the minimizer $\hat{u}$. This is corroborated by the experiments presented in Section 5 in [$\star$11(2005)].



(a) $\alpha = 0.5$



(b) $\alpha = 0.05$

Figure 7.1: The shape of $\mathcal{R}$ as a function of α for $\phi(t) = \sqrt{\alpha + t^2}$. Multiplicative form "——", Additive form "- - - -" for $c = \phi''(0)$.

In both the multiplicative and the additive forms of HQ regularization, the calculation of $b^{(k)}$ has basically the same complexity. In the multiplicative form, the matrix $H(b^{(k)})$ in (7.9) is changing at each iteration. Although invertible, it can be ill-conditioned so that finding $\omega(b^{(k)})$ can need a large number of iterations. In the additive form, H is fixed—see (7.11). If p is small, we can compute H^{-1} before to start iterations. When p is large, the conjugate gradient method can be used to find $\omega(b^{(k)})$ at each iteration. The convergence of the additive form can be improved using preconditioning techniques. So, the computational cost of each iteration is much smaller for the additive form than for the multiplicative form. In order to verify these claims in practice, we provide the average results from extensive numerical experiments where we compare the convergence properties (number of iterations, computational times, conditioning) of the two forms of HQ regularization and then compare them with standard minimization methods (see section 5 in [$\star$11(2005)]). In conclusion, the computational cost of each iteration for the additive form of HQ regularization is smaller than for the multiplicative form and it can substantially be improved using fast solvers and preconditioning.

7.3 Equivalence result for the multiplicative half-quadratic regularization

The contribution of our joint paper with Raymond Chan [$\star$23(2005)] is to show that the multiplicative form of half-quadratic regularization is equivalent to the very classical *gradient linearization* approach, known also as the *fixed point iteration*. Connection between both approaches has been mentioned by Vogel in [83] for the particular case when $\phi(t) = \sqrt{\alpha + t^2}$. As we show, equivalence holds in general.

With only a little loss of generality, we keep here the framework of the previous section § 7.2. In order to solve the equation $D\mathcal{F}(u) = 0$, at each iteration k of the fixed point iteration method one finds $u^{(k)}$ by solving a linear problem,

$$\mathcal{L}\left(u^{(k-1)}\right)\, u^{(k)} = z, \tag{7.22}$$

where $z \in \mathbb{R}^p$ is independent of u, and for any $u \in \mathbb{R}^p$ given, z and $\mathcal{L}(u) \in \mathbb{R}^{p\times p}$ are uniquely defined by

$$D\mathcal{F}(u) = \mathcal{L}(u)\, u - z. \tag{7.23}$$

Let us remind that

$$\begin{aligned} D\mathcal{F}(u) &= 2A^*Au + \beta G^*\left[\phi'(\langle g_i, u\rangle)\right]_{i=1}^r - 2A^*v \\ &= \mathcal{H}(u)u - 2A^*v \end{aligned}$$

where $\mathcal{H}(u)$ is the matrix given in (7.15). The identification with (7.22) clearly reads

$$\mathcal{L}(u) = \mathcal{H}(u) \quad \text{and} \quad z = 2A^*v.$$

On the other hand, inserting (7.8) and (7.9) into (7.12) yields

$$\mathcal{T}(u) = (\mathcal{H}(u))^{-1}A^*v$$

By (7.13), the kth iteration for the multiplicative HQ regularization reads

$$u^{(k)} = \left(\mathcal{H}(u^{(k-1)})\right)^{-1} z.$$

Hence the result.

7.4 Fast minimization for cost-functions of the form $\ell_2 - \ell_1$ and $\ell_1 - \ell_1$

The work in [$\star$12(2005)] is aimed at providing efficient methods to minimize non-smooth cost-functions $\mathcal{F}$ of the form

$$\ell_2 - \ell_1 \qquad \mathcal{F}(u) = \|Au - v\|_2^2 + \beta\|Gu\|_1 \tag{7.24}$$

$$\ell_1 - \ell_1 \qquad \mathcal{F}(u) = \|Au - v\|_1^1 + \beta\|Gu\|_1 \tag{7.25}$$

under the constraint that $u \geq 0$. Here, $u \in \mathbb{R}^p$ is an image and G is a difference operator, e.g. the operator that yields the differences between each pixel and its 4 or 8 adjacent neighbors. Instead of directly minimizing these functions, we restate them as the minimization of differentiable cost-functions under linear constraints.

The $\ell_1 - \ell_1$ cost-function. Put $x = Au - v$ and $y = \beta Gu$. Let us decompose x as $x = x^+ - x^-$ where $x_i^+ = \max\{x_i, 0\}$ and $x_i^- = \max\{-x_i, 0\}$, $1 \leq i \leq p$. In a similar way, let $y = y^+ - y^-$. Minimizing $\mathcal{F}$ in (7.25) is then equivalent to

$$\begin{aligned} \text{minimize} \quad & \left\langle \mathbb{1}, x^+ \right\rangle + \left\langle \mathbb{1}, x^- \right\rangle + \left\langle \mathbb{1}, y^+ \right\rangle + \left\langle \mathbb{1}, y^+ \right\rangle \\ \text{subject to} \quad & Au - v = x^+ - x^- \\ & \beta Gu = y^+ - y^- \\ & x^+ \geq 0,\ x^- \geq 0,\ y^+ \geq 0,\ y^- \geq 0,\ u \geq 0. \end{aligned}$$

This is a linear programming (LP) problem that can be put into the standard form

$$\text{minimize } \langle c, z \rangle \text{ subject to } Hz = b \text{ and } z \geq 0.$$

The $\ell_2 - \ell_1$ cost-function. Now put $y = \beta Gu$ and decompose y as above. Minimizing $\mathcal{F}$ in (7.24) is equivalent to

$$\begin{aligned} \text{minimize} \quad & \|Au - v\|_2^2 + \left\langle \mathbb{1}, y^+ \right\rangle + \left\langle \mathbb{1}, y^+ \right\rangle \\ \text{subject to} \quad & \beta Gu = y^+ - y^- \\ & y^+ \geq 0,\ y^- \geq 0,\ u \geq 0. \end{aligned}$$

This is a quadratic programming (QP) problem whose standard form reads

$$\text{minimize } \frac{1}{2} z^* Q z + \langle c, z \rangle \text{ subject to } Hz = b \text{ and } z \geq 0.$$

The steps that follow are to define the optimal solution using generalized Lagrange multipliers and then to eliminate a certain number of variables. The resultant constrained minimization problem is solved using an *interior point method* [66] (chapter 1) and [85] (chapter 14), combined with factorized sparse inverse preconditioners [59]. All details are explained in [$\star$12(2005)].

Chapter 8

Perspective and future research

The approach proposed in section 1.3 is new and important. Even if a few results could been obtained, it is widely open for future research and arises a lot of new questions. The slogan of my future research can be formulated as *the conception of feasible signal and image reconstruction methods that make a correct use of all available information, i.e. that respect the data acquisition and the perturbation models as well as the priors.* Several directions that I feel can be fruitful to explore are presented next.

(a) The contradictions relevant to Bayesian methods exhibited in chapter 6 need a deep analysis. Hence the foundations of statistical methods combining knowledge of different origins, have to be revisited from the point of view of signal and image reconstruction.

We belief that this can help to determine ways to construct solutions of inverse problems that really do respect the models.

Will we be happy with solutions that respect the models? In many cases probably not. The reason is that current models, and especially priors, are usually taken in an ad hoc manner. Let us recall the work of Gousseau and Morel [45] demonstrating that natural images do not satisfy the Bounded Variation model which underlies one of the most popular image reconstruction methods. In the recent years, researchers become aware that little is known on natural images and even less on digital images. Realist modelling of images will certainly be the center of important research in the next years. Once one has right models, it will be crucial to have reconstruction methods that respect the models.

(b) I will pursue studying the properties of the minimizers relevant to various cost-functions. This is important for several reasons. It allows to obtain rigorous results on the solutions that are currently used, and hence a real control on them and on the parameters involved there. This gives an access to the models that these solutions follow *effectively*. The obtained results can provide a reversed way to perform modelling by providing a "dictionary" where different properties of the minimizers are related to different features of the underlying cost-function.

Using the elements of this "dictionary", one can think of creating specialized cost-functions adapted to solve particular problems. Relevant examples in our previous research are the binary image reconstruction using convex cost-functions and the outliers suppression methods based on ℓ_1 data-fidelity.

This is a challenge for the intellect and for the applications as far as it can lead to simple solutions of complicate problems.

An important extension of the analytical results on the properties of the minimizers will be to put them in a statistical context. More precisely, the idea is to give a statistical description of these properties as a function of the randomness of the data. This provides an elegant way to get an access to the statistics of the solutions, which remains an unsolved problem.

(c) The computational aspects are crucial for the success of a signal or image reconstruction method. The knowledge on some properties of the minimizers points can in principle be used in the optimization schemes in order to reduce the search space and thus simplify the optimization. Furthermore, it may be possible to find (partial) equivalences with PDE filtering methods and thus improve the convergence rates.

The problems briefly described above are likely to yield various ramifications. I will be happy to have the opportunity to initiate young researchers to this research. Hence my motivation to apply for the grade *habilité à dirigier des recherches*.

Chapter 9

Complete list of publications

All publications are available at http://www.cmla.ens-cachan.fr/∼nikolova/

Refereed Journal Articles

[⋆1] M. Nikolova, *Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares*, SIAM Journal on Multiscale Model. Simul., Vol. 4, No. 3, 2005, pp. 960-991.

[⋆2] M. Nikolova, *Weakly constrained minimization. Application to the estimation of images and signals involving constant regions*, Journal of Mathematical Imaging and Vision, no. 2, vol. 21, Sep. 2004, pp. 155-175.

[⋆3] M. Nikolova, *A variational approach to remove outliers and impulse noise*, Journal of Mathematical Imaging and Vision, vol. 20, no. 1-2, 2004, pp. 99-120.

[⋆4] M. Nikolova, *Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers*, SIAM Journal on Numerical Analysis, vol. 40, no. 3, (2002), 965-994.

[⋆5] M. Nikolova, *Local strong homogeneity of a regularized estimator*, SIAM Journal on Applied Mathematics, vol. 61, no. 2, (2000), pp. 633-658.

[⋆6] M. Nikolova, *Thresholding implied by truncated quadratic regularization*, IEEE Transactions on Signal Processing, vol. 48, December (2000), pp. 3437–3450.

[⋆7] M. Nikolova, "Homogénéité forte d'une solution régularisée", *Comptes-rendus de l'Académie des sciences*, Paris, t. 325, Série I (Mathématiques), pp. 665-670, 1997.

[⋆8] M. Nikolova, "Markovian reconstruction using a GNC approach", *IEEE Image Processing*, vol. 8, n. 9, Sept. 1999, pp. 1204-1220.

[⋆9] S. Durand and M. Nikolova, *Stability of minimizers of regularized least squares objective functions I: study of the local behaviour*, Journal of Applied Mathematics and Optimization, Vol. 53, n. 2, March 2006, pp. 185-208.

[⋆10] S. Durand and M. Nikolova, *Stability of minimizers of regularized least squares objective functions II: study of the global behavior*, Journal of Applied Mathematics and Optimization (2006).

[⋆11] M. Nikolova and M. Ng, *Analysis of Half-Quadratic Minimization Methods for Signal and Image Recovery*, SIAM Journ. on Scientific computing, vol. 27, No. 3, pp. 937-966, 2005.

[⋆12] Haoying Fu, Michael K. Ng, Mila Nikolova and Jesse L. Barlow *Efficient minimization methods of mixed $\ell 1$-$\ell 1$ and $\ell 2$-$\ell 1$ norms for image restoration*, SIAM Journ. on Scientific computing, Vol. 27, No 6, 2006, pp. 1881-1902.

[⋆13] Chan R., C.W. Ho and M. Nikolova, *Convergence of Newton's Method for a Minimization Problem in Impulse Noise Removal*, J.Comput.Math., vol. 22, 2004, pp. 168-177.

[⋆14] Chan R., C. Hu and M. Nikolova, *An Iterative Procedure for Removing Random-Valued Impulse Noise*, IEEE Signal Proc. Letters, 11 (2004), 921-924.

[⋆15] Chan R., Chung-Wa Ho and M. Nikolova, *Salt-and-Pepper Noise Removal by Median-type Noise Detector and Detail-preserving Regularization*, IEEE Trans. on Image Processing, Vol. 14, No. 10, Oct. 2005, pp. 1479-1485.

[⋆16] R. Peeters, P. Kornprobst, M. Nikolova, S. Sunaert, T. Vieville, G. Malandain, R. Deriche, O. Fougeras, M. Ng and P. Hecke *The use of superresolution techniques to reduce slice thickness in functional MRI*, International Journal of Imaging Systems and Technology, Vol. 14, No. 3, 2004.

[⋆17] Roullot E., A. Herment, I. Bloch, A. Cesare, M. Nikolova and E. Mousseaux, *Modeling anisotropic undersampling of magnetic resonance angiographies and reconstruction of a high-resolution isotropic volume using half-quadratic regularization techniques*, Signal Processing, vol. 84, 2004, pp. 743-762.

[⋆18] Alberge F., M. Nikolova and P. Duhamel, *Blind Identification / Equalization using Deterministic Maximum Likelihood and a partial prior on the input*, IEEE Trans. on Signal Processing, Vol. 54, Issue 2, Feb. 2006, pp. 724- 737

[⋆19] Alberge F., P. Duhamel and M. Nikolova, *Adaptive solution for blind identification / equalization using deterministic maximum likelihood*, IEEE Trans. on Signal Processing, vol. 50, no 4, April 2002, pp. 923-936.

[⋆20] Nikolova M., Idier J. and Mohammad-Djafari A., *Inversion of large-support ill-posed linear operators using a piecewise Gaussian MRF*, IEEE Image Processing, vol. 8, no. 4, pp. 457-460, April 1998.

[⋆21] Nikolova M. & A. Mohammad-Djafari, *Eddy Current Tomography Using a Markov model*, Signal Processing, vol. 49, no. 2, pp. 119–132, 1996.

Reports

[⋆22] Chan Tony, Selim Esedoglu and Mila Nikolova, *Algorithms for Finding Global Minimizers of Image Segmentation and Denoising Models*, Rapport CAM (UCLA) no. 04-54, Septembre 2004.

[⋆23] Nikolova Mila and Raymond Chan, *Half-Quadratic Minimization of Regularized Objectives as the Fixed Point Iteration for the Linearized Gradient*, Report CMLA N. 2005-22, Septembre 2005.

Conference Invited Presentations

[⋆24] M. Nikolova, *How to choose our objective functions?*, International Conference for Super-resolution Imaging, The University of Hong Kong, 29-31 August 2005.

[⋆25] M. Nikolova, *Essential Features of the (Local) Minimizers of Non-Smooth Regularized Cost-Functions*, AMS-SIAM Annual Meeting, Special Session on Nonsmooth Analysis in Variational and Imaging Problems, Atlanta, January 5 - 8, 2005.

[⋆26] M. Nikolova, *Recovery of edges in signals and images by minimizing nonconvex regularized least-squares*, Workshop - Mathematical Image Analysis and Processing, Banff Research Station, Octobre 2004.

[⋆27] M. Nikolova, *Non-smooth cost-functions in image and signal reconstruction*, International Conference on Numerical Methods in Imaging Science and Information Processing, Institute of Mathematical Sciences, Singapore, 15-19 Dec. 2003.

[⋆28] M. Nikolova and M. Ng, *Analysis of half-quadratic minimization methods for signal and image recovery*, Workshop on Mathematics in Image Processing, Institute of Mathematical Sciences, Singapore, 8-9 Dec. 2003.

[⋆29] S. Durand and M. Nikolova, *Restoration of wavelet coefficients by Minimizing a specially designed objective function*, Special Session on Multiscale and Oscillatory Phenomena, Joint Mathematics Meetings of the American Mathematical Society, Phoenix, Jan. 2004.

[⋆30] M. Nikolova, *Minimizers of cost-functions involving non-smooth data-fidelity terms. Application to the processing of outliers*, Colloque Mathematics and Image Analysis, 10-14 Septembre 2002, Paris.

[⋆31] M. Nikolova, *Image restoration by minimizing objective functions with nonsmooth data-fidelity terms*, The Fisrt SIAM Conference on Imaging Science, Boston, March 2002.

[⋆32] M. Nikolova, *Local homogeneity of images using regularization*, SIAM annual meeting, July 2000.

[⋆33] S. Durand and M. Nikolova, *Stability of minimizers of objective functions*, SIAM annual meeting, July 2000.

[⋆34] M. Nikolova, *Local features of images for different families of objective functions*, Workshop on Mathematics in Image Processing, University of Hong Kong, Dec. 2000.

[⋆35] M. Nikolova, *Locally homogeneous images as minimizers of an objective function*, in Proceedings of the IEEE International Conference on Image Processing, Oct. 1999.

Peer Reviewed Conference Papers

[⋆36] M. Nikolova, *Restoration of edges by minimizing non-convex cost-functions*, IEEE Int. Conf. on Image Processing, Sept. 2005, vol. II, pp. 786-789.

[⋆37] T. Chan, S. Esedoglu and M. Nikolova, *Finding the Global Minimum for Binary Image Restoration*, IEEE Int. Conf. on Image Processing, Sept. 2005, vol. I, pp. 121-124.

[⋆38] R. Chan, C. W. Ho, C. Y. Leung and M. Nikolova *Minimization of Detail-preserving Regularization Functional by Newton's Method with Continuation*, IEEE Int. Conf. on Image Processing, Sept. 2005, vol I, pp. 125-128.

[⋆39] Haoying Fu, Michael K. Ng, Mila Nikolova, Jesse L. Barlow, Wai-Ki Ching, *Fast Algorithms for ℓ_1 Norm/Mixed ℓ_1 and ℓ_2 Norms for Image Restoration.* ICCSA (4) 2005: 843-851

[⋆40] M. Nikolova, *Minimization of cost-functions with non-smooth data-fidelity terms to clean impulsive noise*, Int. workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition, Lecture Notes in Computer Science, Springer-Verlag, pp. 391-406, 2003.

[⋆41] M. Nikolova, *Efficient removing of impulsive noise based on an L1-L2 cost-function*, IEEE Int. Conf. on Image Processing, Sept. 2003.

[⋆42] S. Durand and M. Nikolova, *Restoration of wavelet coefficients by minimizing a specially designed objective function*, IEEE Int. Conf. on Computer Vision / Workshop on Variational and Level-Set Methods, Oct. 2003.

[⋆43] P. Kornprobst, R. Peeters, M. Nikolova, R. Deriche, M. Ng and P. Van Hecke, *A superresolution framework for fMRI sequences and its impact on resulting activation maps*, Medical Image Computing and Computer-Assisted Intervention-MICCAI2003, Lecture Notes in Computer Science, Springer-Verlag.

[⋆44] R. Deriche, P. Kornprobst, M. Nikolova and M. Ng, *Half-quadratic regularization for MRI image restoration*, in Proceedings of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing, vol. VI, pp. 585-588, 2003.

[⋆45] M. Nikolova and M. Ng, *Comparison of the main forms of half-quadratic regularization*, Proceedings of the IEEE Int. Conf. on Image Processing, vol. 1, pp. 349–352, 2002.

[⋆46] S. Zinger, M. Nikolova, M. Roux and H. Maitre, *3D resampling for airborne Laser data of urban areas*, Proceedings of ISPRS, vol. XXXIV, n. 3A, pp. 418–423, 2002.

[⋆47] S. Zinger, M. Nikolova, M. Roux and H. Maitre, *Rééchantillonnage de données 3D laser aéroporté en milieu urbain*, Congrès Francophone de Vision par Ordinateur ORASIS 2003, Mai 2003, pp. 75-82.

[⋆48] S. Durand and M. Nikolova, *Stability of image restoration by minimizing regularized objective functions*, Proceedings of the IEEE Int. Conf. on Computer Vision / Workshop on Variational and Level-Set Methods, 2001, pp. 73-80, Vancouver, Canada.

[⋆49] M. Nikolova, *Image restoration by minimizing objective functions with non-smooth data-fidelity terms*, Proceedings of the IEEE Int. Conf. on Computer Vision / Workshop on Variational and Level-Set Methods, 2001, pp. 11-18, Vancouver, Canada.

[⋆50] M. Nikolova, *Smoothing of outliers in image restoration by minimizing regularized objective functions with non-smooth data-fidelity terms*, Proceedings of the IEEE Int. Conf. on Image Processing, 2001, Thessaloniki, Greece.

[⋆51] M. Nikolova and M. Ng, *Fast image reconstruction algorithms combining half-quadratic regularization and preconditioning*, Proceedings of the IEEE Int. Conf. on Image Processing, 2001, Thessaloniki, Greece.

[⋆52] M. Nikolova, *Assumed and effective priors in Bayesian MAP estimation*, in Proceedings of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing, June 2000.

[⋆53] F. Alberge, M. Nikolova and P. Duhamel, *Adaptive deterministic maximum likelihood using a quasi-discrete prior*, in Proceedings of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing, June 2000.

[⋆54] F. Alberge, P. Duhamel and M. Nikolova, *Blind Identification/Equalization using Deterministic Maximum Likelihood and a partial information on the input.* IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Annapolis (Maryland), May 99.

[⋆55] F. Alberge, P. Duhamel and M. Nikolova, *Low cost adaptive algorithm for blind channel identification and symbol estimation.* EUSIPCO 2000. Tampere (Finland).

[⋆56] E. Roullot, A. Herment, I. Bloch, M. Nikolova and E. Mousseaux, *Regularized reconstruction of 3D high resolution Magnetic Resonance Images from acquisitions of anisotropically degraded resolutions*, in ICPR 2000, vol. 3, p. 350-353.

[⋆57] E. Roullot, A. Herment, I. Bloch, M. Nikolova and E. Mousseaux, *Reconstruction régularisée d'images de résonance magnétique 3D de haute résolution à partir d'acquisitions de résolutions anistropes*, RFIA 2000 (Paris, France), vol. II, p. 59-68.

[⋆58] Nikolova, M. and A. Hero, *Noisy word recognition using denoising and moment matrix discriminants*, IEEE Workshop on Higher Order Statistics, 1999.

[⋆59] Nikolova, M., *Local continuity and thresholding using truncated quadratic regularization*, IEEE Workshop on Higher Order Statistics, 1999.

[⋆60] Nikolova, M., *Estimation of binary images using convex criteria*, Proc. of IEEE Int. Conf. on Image Processing, Oct. 1998, Vol. II, pp. 108–112.

[⋆61] Nikolova, M., *Reconstruction of locally homogeneous images*, European Signal Proc. Conf., Sept. 1998, Vol. II, pp. 833–836.

[⋆62] Nikolova, M., *Estimation of signals containing strongly homogeneous zones*, Proc. of IEEE Stat. Signal and Array Proc., Sept. 1998, pp. 76–79.

[⋆63] Nikolova M. and A. Hero, *Segmentation of road edges from a vehicle-mounted imaging radar*, Proc. of IEEE Stat. Signal and Array Proc., Sept. 1998, pp. 212–215.

[⋆64] Nikolova M., *Regularisation functions and estimators*, Proc. of IEEE Int. Conf. on Image Processing, Nov. 1996, pp. 457–460.

[⋆65] Nikolova, M., *Non convex regularization and the recovery of edges*, Proc. IEEE Workshop on Nonlinear Signal and Image Processing, Greece, June. 1995, pp. 1042–1045.

[⋆66] Nikolova, M., *Markovian reconstruction in computed imaging and Fourier synthesis*, Int. Conf. on Image Processing, Nov. 1994, pp. 690-694.

[⋆67] Nikolova M., A. Mohammad-Djafari and J. Idier, *Inversion of large-support ill-conditioned linear operators using a Markov model with a line process*, Proc. IEEE Int. Conf. Acoust. Speech Signal Process., Adelaide, Apr. 1994, vol. V, pp. 357-360.

[⋆68] Nikolova M. and A. Mohammad-Djafari, *Discontinuity reconstruction from linear attenuating operators using the weak-string model*, European Signal Proc. Conf., Sept. 1994, pp. 1062–1065.

Book chapters

[⋆69] Nikolova M. & A. Mohammad-Djafari, *Maximum Entropy Image Reconstruction in Eddy Current Tomography*, in Maximum entropy and Bayesian methods, A. Mohammad-Djafari & G. Demoment eds. Kluwer Academic Publ., 1993, pp.273-278.

[⋆70] Zorgati R. & M. Nikolova, *Eddy Current Imaging: An Overview*, in *Studies in Applied Electromagnetics and Magnetics 9, Non-Destructive Testing of Materials*, Kenzomiya et al. eds., IOS Press., 1996, 8 p.

[⋆71] Nikolova M., *Inversion de données pour le contrôle non destructif : une synthèse des travaux du groupe P21 – Direction des Études et Recherches, EDF*, Notes de la DER – EDF, Rapport EDF/DER/HP-21/96/013, Sept. 1996, 112 p., diffusion externe.

[⋆72] Mohammad-Djafari A., H. Carfantan and M. Nikolova, *New advances in Bayesian calculation for linear and non linear inverses problems*, in *Maximum entropy and Bayesian methods*, Berg-en-Dal, Kluwer Academic Publ., 1996.

Bibliography

[1] S. Alliney, *A property of the minimum vectors of a regularizing functional defined by means of absolute norm*, IEEE Transactions on Signal Processing, 45 (1997), pp. 913–917.

[2] S. Alliney and S. A. Ruzinsky, *An algorithm for the minimization of mixed l_1 and l_2 norms with application to Bayesian estimation*, IEEE Transactions on Signal Processing, SP-42 (1994), pp. 618–627.

[3] L. Alvarez, P. L. Lions, and J. M. Morel, *Image selective smoothing and edge detection by nonlinear diffusion (II)*, SIAM Journal on Numerical Analysis, 29 (1992), pp. 845–866.

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

[5] G. Aubert and P. Kornprobst, *Mathematical problems in images processing*, Springer-Verlag, Berlin, 2002.

[6] G. Aubert and L. Vese, *A variational method in image recovery*, SIAM Journal on Numerical Analysis, 34 (1997), pp. 1948–1979.

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

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

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

[10] M. Black and A. Rangarajan, *On the unification of line processes, outlier rejection, and robust statistics with applications to early vision*, International Journal of Computer Vision, 19 (1996), pp. 57–91.

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

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

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

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

[15] A. C. Bovik, *Handbook of image and video processing*, Academic Press, New York, 2000.

[16] X. Bresson, S. Esedoglu, P. Vandergheynst, J. Thiran, and S. Osher, *Global minimizers of the active contour/snake model*, tech. report, UCLA, CAM Report (05-04), 2005.

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

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

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

[20] T. Chan and S. Esedoglu, *Aspects of total variation regularized l^1 function approximation*, SIAM Journal on Applied Mathematics(to appear).

[21] T. Chan and L. Vese, *Active contours without edges*, IEEE Transactions on Image Processing, 10 (2001), pp. 266–277.

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

[23] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, *Two deterministic half-quadratic regularization algorithms for computed imaging*, in Proceedings of IEEE ICIP, vol. 2, 1994, pp. 168–172.

[24] ——, *Deterministic edge-preserving regularization in computed imaging*, IEEE Transactions on Image Processing, 6 (1997), pp. 298–311.

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

[26] F. Clarke, *Optimization and Nonsmooth Analysis*, SIAM, Philadelphia, 1990.

[27] L. D. Cohen, *Auxiliary variables and 2-step iterative algorithms in computer vision problems*, Journal of Mathematical Imaging and Vision, 6 (1996), pp. 376–386.

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

[29] G. Dal Maso, *An introduction to Γ-convergence*, Progress in Nonlinear Differential Equations and their Applications, Birkhauser Boston Inc., Boston, MA, 1993.

[30] A. H. Delaney and Y. Bresler, *Globally convergent edge-preserving regularized reconstruction: an application to limited-angle tomography*, IEEE Transactions on Image Processing, 7 (1998), pp. 204–221.

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

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

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

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

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

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

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

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

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

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

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

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

[43] R. Glowinski, J. Lions, and R. Trémolières, *Analyse numérique des inéquations variationnelles*, vol. 1, Dunod, Paris, 1 ed., 1976.

[44] R. C. Gonzales and P. A. Wintz, *Digital Image Processing*, Addison-Wesley, MA, 1987.

[45] Y. Gousseau and J.-M. Morel, *Are natural images of bounded variation?*, SIAM Journal on Mathematical Analysis, (2001).

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

[47] L. He, M. Burger, and S. Osher, *Iterative total variation regularization with non-quadratic fidelity*, tech. report, UCLA, CAM Report (05-38), 2005.

[48] T. Hebert and R. Leahy, *A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors*, IEEE Transactions on Medical Imaging, 8 (1989), pp. 194–202.

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

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

[51] J. Idier, *Convex half-quadratic criteria and auxiliary interacting variables for image restoration*, IEEE Transactions on Image Processing, 10 (2001), pp. 1001–1009.

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

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

[54] M. Kass, A. Witkin, and D. Terzopoulos, *Snakes: Active contour models*, International Journal of Computer Vision, (1987), pp. 321–331.

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

[56] P. Kornprobst, R. Deriche, and G. Aubert, *Image sequence analysis via partial differential equations*, Journal of Mathematical Imaging and Vision, 11 (1999), pp. 5–26.

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

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

[59] F.-R. Lin, M. K. Ng, and W.-K. Ching, *Factorized banded inverse preconditionners for matrices with toeplitz structure*, SIAM Journal on Scientific Computing, 26 (2005).

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

[61] J.-M. Morel and S. Solimini, *Segmentation of images by variational methods: A constructive approach*, Rev. Math. Univ. Comput., (1988), pp. 169–182.

[62] ——, *Variational Methods in Image Segmentation*, Birkhäuser, Basel, 1995.

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

[64] D. Mumford and J. Shah, *Boundary detection by minimizing functionals*, in Proceedings of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 1985, pp. 22–26.

[65] ——, *Optimal approximations by piecewise smooth functions and associated variational problems*, Communications on Pure and Applied Mathematics, (1989), pp. 577–684.

[66] J. Nocedal and S. Wright, *Numerical Optimization*, Springer, New York, 1999.

[67] J. Ortega and W. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

[68] S. Osher and J. Sethian, *Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations*, Journal of Computational Physics, 79 (1988), pp. 12–49.

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

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

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

[72] R. T. Rockafellar, *The Theory of Subgradients and its Application to Problems of Optimization: Convex and Nonconvex Functions*, Helderman Verlag, Berlin, 1981.

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

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

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

[76] G. Steidl, J. Weickert, P. Mrázek, and M. Welk, *On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and sides*, SIAM Journal on Numerical Analysis, 42 (2004), pp. 686–713.

[77] A. Tarantola, *Inverse problem theory : Methods for data fitting and model parameter estimation*, Elsevier Science Publishers, Amsterdam, 1987.

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

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

[80] L. Vese and S. Osher, *The level set method links active contours, mumford-shah segmentation, and total variation restoration*, tech. report, UCLA CAM Report (02-05), 2002.

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

[82] C. R. Vogel and M. E. Oman, *Iterative method for total variation denoising*, SIAM Journal on Scientific Computing, 17 (1996), pp. 227–238.

[83] C. R. Vogel and M. E. Oman, *Fast, robust total variation-based reconstruction of noisy, blurred images*, IEEE Transactions on Image Processing, 7 (1998), pp. 813–824.

[84] G. Wang, J. Zhang, and G.-W. Pan, *Solution of inverse problems in image processing by wavelet expansion*, IEEE Transactions on Image Processing, 4 (1995), pp. 579–593.

[85] S. Wright, *Primal-Dual Interior-Point Methods*, SIAM Publications, Philadelphia, 1997.