NON-DIFFERENTIABLE CONSTRAINED SIGNAL RESTORATION BY SUBGRADIENT LEVEL METHODS

by

JIAN LUO

A dissertation submitted to the Graduate Faculty in Engineering in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

The City University of New York

2000

UMI Number: 9986354

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

This manuscript has been read and accepted for the Graduate Faculty in Engineering in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

8/24/2000
Date

Chair c

8-24-2000
Date

Executive Officer

PROFESSOR ROGER DORSINVILLE

PROFESSOR THAO NGUYEN

PROFESSOR KENNETH SOBEL

PROFESSOR FREDERICK E. THAU

DR. RICHARD TOLIMIERI

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

NON-DIFFERENTIABLE CONSTRAINED SIGNAL RESTORATION BY SUBGRADIENT LEVEL METHODS

by

Jian Luo

Advisor: Professor Patrick L. Combettes

The classical signal restoration problem is to estimate the original form of a signal from a degraded observation and some *a priori* information. Mathematically, a wide range of digital signal restoration problems can be formulated as minimizing a convex objective over a convex set representing the constraints derived from *a priori* knowledge and the observed signal. The goal of this dissertation is to develop numerical algorithms to solve this type of problems with nondifferentiable objectives. Such objectives arise for instance in constrained minimax, total variation, or L_1 norm problems. They have become popular in recent years due to their ability to capture certain features of signals such as sharp edges. However, the problem of developing reliable numerical schemes to solve the resulting constrained nondifferentiable optimization problems has received little attention. In the algorithms proposed in this dissertation, the potentially complex constraint set is disintegrated into an intersection of simpler sets defined by convex inequalities. At each iteration, the update is obtained through a combination of subgradient projections onto the individual constraint sets and a subgradient projection onto adaptively refined approximations to the unknown optimal level set of the objective. Various algorithms based on this variable target feasibility principle are developed and their convergence is established. The implementation of the algorithms is also discussed. Several numerical applications to signal and image restoration/denoising are demonstrated, with special emphasis on the total variation approach.

ACKNOWLEDGMENTS

I would like to thank my mentor, Professor P. L. Combettes, for his advice, guidance, and financial support. Thanks are also given to the Department of Electrical Engineering and in particular to Professor Sobel, Deputy Chair, for giving me many opportunities for teaching. They not only provided financial support for my Ph.D. studies but also added precious experience to it. I also thank the members of the examing committee, Professor R. Dorsinville, Professor T. Nguyen, Professor K. Sobel, Professor F. E. Thau, and Dr. R. Tolimieri.

I would like to give special thanks to Marry Jacson, former director of the international student office, for her help in finding funds to start my Ph.D. studies and her constant encouragements.

Finally, I would like to thank my parents and my sister for their long time support and constant encouragements. I should also thank my wife for being supportive and taking up all the household works during the period I prepared this dissertation and for making my family life cosy.

This work was partially supported by the National Science Foundation under grant MIP-9705504.

Contents

List of Figures

NOTATION

- $\mathbb{N}$ – the set of nonnegative integers.
- E – the standard real N-dimensional Euclidean space.
- $\langle \cdot \mid \cdot \rangle$ – the scalar product in E, i.e., for every $x = (x^1, \cdots, x^N) \in \mathsf{E}$ and $y = (y^1, \cdots, y^N) \in \mathsf{E}$, $\langle x \mid y \rangle = \sum_{i=1}^N x^i y^i$.
- $\|\cdot\|$ – the Euclidean norm in E, i.e., for every $x = (x^1, \cdots, x^N) \in \mathsf{E}$, $\|x\| = (\sum_{i=1}^N |x^i|^2)^{\frac{1}{2}}$.
- $\|\cdot\|_p$ – the p-norm in E, i.e., for every $x = (x^1, \cdots, x^N) \in \mathsf{E}$, $\|x\|_p = (\sum_{i=1}^N |x^i|^p)^{\frac{1}{p}}$.
- $\|\cdot\|_F$ – the Frobenius norm in $\mathbb{R}^{M \times M}$, i.e., for every $x \in \mathbb{R}^{M \times M}$,

$$\|x\|_F = \left[\sum_{i=1}^{M} \sum_{j=1}^{M} |x^{i,j}|^2 \right]^{\frac{1}{2}}.$$

- $d(x, S) = \inf_{y \in S} \|x - y\|$ – the distance from the point x to the set S.
- $\operatorname{diam}(S) = \sup_{(x,y) \in S^2} d(x, y)$ – the diameter of the set S.
- $\operatorname{conv}(S)$ – the convex hull of a set S, i.e., the smallest convex set containing S.
- Id – the identity operator on E, i.e., $(\forall x \in \mathsf{E})\ \operatorname{Id}(x) = x$.
- $\operatorname{Fix} T = \{x \in \mathsf{E} \mid T(x) = x\}$ – the set of fixed points of the operator $T \colon \mathsf{E} \to \mathsf{E}$.
- P_S – the projector onto a closed convex set S.
- $\alpha^+ = \max\{0, \alpha\}$, $(\alpha \in \mathbb{R})$.
- $\operatorname{lev}_{\leq \alpha} f = \{x \in \mathsf{E} \mid f(x) \leq \alpha\}$ – the lower level set of the function $f \colon \mathsf{E} \to \mathbb{R}$ at height $\alpha \in \mathbb{R}$.
- $\partial f(x) = \{t \in \mathsf{E} \mid (\forall y \in \mathsf{E})\ \ f(x) + \langle t \mid y - x \rangle \leq f(y)\}$ – the subdifferential of the function $f \colon \mathsf{E} \to \mathbb{R}$ at x.
- $\partial_\epsilon f(x) = \{t \in \mathsf{E} \mid (\forall y \in \mathsf{E})\ \ f(x) + \langle t \mid y - x \rangle \leq f(y) + \epsilon\}$ $(\epsilon > 0)$ – the ϵ-subdifferential of the function $f \colon \mathsf{E} \to \mathbb{R}$ at x.
- $\operatorname{div} f(x) = \sum_{i=1}^{N} \dfrac{\partial f(x)}{\partial x^i}$ – the divergence of the function $f \colon \mathsf{E} \to \mathbb{R}$ at $x = (x^1, \ldots, x^N) \in \mathsf{E}$.

Chapter 1

Introduction

1.1 The Signal Restoration Problem

Signal restoration refers to the problem of estimating the original form of a signal that has been recorded in the presence of sources of degradation [5, 12, 45]. Its applications cover a wide range of areas, including consumer and commercial imaging, geophysics, forensic science, medical imaging, surveillance, and astronomy. For example, an image obtained through an imaging system is usually degraded by two phenomena: blurring and noise corruption. The blurring may be caused by effects such as diffraction, transmission through a random medium, lens aberrations, out-of-focus lenses, or relative motion between the scene and the viewing device. On the other hand, the noise may be present in the image propagation medium and the recording device. In many situations the degradations are serious and need to be reduced so that useful information in the recorded signal can be retrieved.

To restore a signal some form of *a priori* knowledge concerning the degradation processes and/or the original signal is required. Examples of *a priori* information about the original signal include amplitude bounds, phase, region of support, energy bounds, and frequency band [12, 25, 36, 41, 43, 46, 45, 49]. A general vector space model for signal degradation is the additive noise model

$$y = L(x) + u, \tag{1.1}$$

where x, y and u represent respectively the original signal, the recorded signal, and additive system noise, and where L is an operator representing the degradation phenomenon. If L is a convolution operator, i.e., $L(x) = h * x$ for some impulse response h, (1.1) models various common situations [5].

1.2 Formulating Signal Restoration Problems

A priori knowledge is of paramount importance in signal restoration and should be used whenever possible [12, 26, 39, 41, 40, 43, 46, 45, 49]. Mathematically, *a priori* knowledge can be represented in the form of constraints and optimality criteria. In signal restoration, constraints arise from information about

1. The degradation mechanism;
2. The original signal;
3. The noise.

A signal restoration problem can typically be formulated as a constrained optimization problem of the form

$$\text{Find } x^* \in S = \bigcap_{0 \leq i \leq m} S_i \text{ such that } J(x^*) = \inf_{x \in S} J(x) \tag{1.2}$$

in a suitable vector space $\mathcal{B}$, where $(S_i)_{0 \leq i \leq m} \subset \mathcal{B}$ are sets describing signal constraints and $J \colon \mathcal{B} \to \mathbb{R}$ an optimality criterion. In practice, under suitable assumptions, it is often possible to approximate (1.2) by a finite dimensional problem after discretization. This will be the numerical framework adopted in this dissertation. More specifically, our assumptions regarding (1.2) will be following throughout the dissertation.

Assumption 1.1

(i) $J \colon \mathsf{E} \to \mathbb{R}$ is a convex function.

(ii) $(S_i)_{0 \leq i \leq m} \subset \mathsf{E}$ are closed convex sets and $\bigcap_{0 \leq i \leq m} S_i \neq \emptyset$.

(iii) There exists $\alpha \in \mathbb{R}$ such that $S \cap \operatorname{lev}_{\leq \alpha} J$ is nonempty and bounded.

(iv) S is an active set.

As a result, we have the following:

- (i)-(iii) imply that (1.2) has a solution (see Theorem 2.2).
- Denote by $\alpha^* = \inf_{x \in S} J(x)$ the optimal constrained value of J. Then (iv) implies that $\inf_{x \in \mathsf{E}} J(x) < \alpha^*$.

1.3 Nondifferentiable Convex Optimality Criteria

In many situations, one needs to minimize a convex function which is nonsmooth, i.e., not continuously differentiable [28, 33, 42]. In signal restoration, there are two general types of nondifferentiable cost functions: integral and pointwise maximum costs.

1.3.1 Integral Costs

A real-valued M-dimensional analog signal can be represented by a function

$$x : \Omega \subset \mathbb{R}^M \to \mathbb{R} \tag{1.3}$$

For such signals, a general cost function is

$$J : x \to \int_\Omega \phi(x(\omega), \nabla x(\omega)) d\omega, \tag{1.4}$$

where $\phi: \mathbb{R}^{M+1} \to \mathbb{R}$ is a convex function [9, 11, 26, 29, 32, 47] (equivalent costs can be defined for discrete time/space signal by proper discretization). It is convex, since ϕ is convex, and ∇ and $\int_\Omega$ (and their discrete approximation: summation and finite differences in a discrete setting) are linear operators. It is nondifferentiable if ϕ is nondifferentiable, e.g.,

$$\phi: (u, v) \quad \mapsto \quad |u| \tag{1.5}$$

$$\phi: (u, v) \quad \mapsto \quad \|v\|_p \text{ with } p \geq 1. \tag{1.6}$$

For instance, the total variation costs, which have been used in [29] and [38], are obtained if ϕ is given by (1.6) with $M = 2$ ($p = 1$ in [29]).

1.3.2 Minimax Problems

Another type of nondifferentiable convex costs which arise in signal restoration are those of the form

$$J = \max_{1 \le i \le p} J_i, \tag{1.7}$$

where each $J_i : \mathsf{E} \to \mathbb{R}$ is a convex function. Even if the functions $(J_i)_{1 \le i \le p}$ are differentiable, J may not be differentiable at those points at which two or more of the J_i's take the same value.

1.3.3 Remarks on Nondifferentiable Optimization

Differentiable optimization methods should not be used to minimize a nondifferentiable function as they may lead to serious failures [28, 42]. By way of warning, we now present a brief discussion based on that of [28].

1.3.3.1 Failure of Convergence

In $\mathbb{R}^2$, consider minimizing the function

$$J(x^1, x^2) = \begin{cases} \sqrt{3(|x^1|^2 + 2|x^2|^2)} & \text{if } 0 \le |x^2| \le 2x^1 \\ \dfrac{1}{\sqrt{3}}(x^1 + 4|x^2|) & \text{otherwise.} \end{cases} \tag{1.8}$$

It is convex but not differentiable on the ray $x^1 \le 0, x^2 = 0$. Moreover, $\lim_{u \to -\infty} J(x^1, 0) = -\infty$. As demonstrated in [28], starting from $x_0 = (2, 1)$ the sequence generated by the steepest descent method converges to the point $(0, 0)$, a non-optimal kink of J. Likewise, we can construct nondifferentiable functions for which Newton's method may fail. Thus, it is unwise to use differentiable optimization methods based on Taylor models to minimize nondifferentiable functions.

1.3.3.2 Lack of Optimality Test

Another problem associated with applying differentiable optimization methods to a nondifferentiable function is the lack of an implementable stopping rule. For a differentiable convex function $J\colon \mathsf{E} \to \mathbb{R}$, we have

$$J(x^*) = \inf_{x \in \mathsf{E}} J(x) \iff 0 = \nabla J(x^*). \tag{1.9}$$

Since J is continuously differentiable (Theorem 2.9), the norm of the gradient $\|\nabla J(x_n)\|$ will become small when x_n approaches some optimal point x^*. For a nondifferentiable function J, we have

$$J(x^*) = \inf_{x \in \mathsf{E}} J(x) \iff 0 \in \partial J(x^*), \tag{1.10}$$

and no stopping rule can be built based upon the norm of gradient, even if the gradient exists at all iterates. For example, for the absolute value function $J\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto |x|$, $|\nabla J(x_n)| = 1$ at each $x_n \neq 0$, no matter how close x_n is to the optimal kink $x^* = 0$.

1.3.3.3 Failure of Gradient Approximation

For a differentiable function $J\colon \mathsf{E} \to \mathbb{R}$

$$(\forall x \in \mathsf{E})(\forall h \in \mathsf{E})\ J(x+h) - J(x) = \langle \nabla J(x) \mid h \rangle + o(\|h\|). \tag{1.11}$$

Therefore, the gradient can be estimated by finite differences. For a nondifferentiable J, (1.11) no longer holds. The computation of subgradients becomes compulsory and cannot be done by a finite difference approximation scheme. For example, let $J\colon \mathbb{R}^3 \to \mathbb{R}\colon x \mapsto \max\{x^1, x^2, x^3\}$. Then the subdifferential at $(0,0,0)$ is

$$\partial J(0) = \operatorname{conv}\{(1,0,0), (0,1,0), (0,0,1)\}. \tag{1.12}$$

However, the forward, backward, and central differences of J at $(0,0,0)$ are $(1,1,1)$, $(0,0,0)$, and $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. None of them describes the behavior of J near $(0,0,0)$.

1.3.3.4 Nondifferentiable Functions and Their Subdifferential

Let $J\colon \mathsf{E} \to \mathbb{R}$ be a nondifferentiable convex function. Take an arbitrary $x \in \mathsf{E}$. If J is differentiable at x the subdifferential $\partial J(x)$ degenerates to the gradient $\nabla J(x)$, whose negative is always at a direction of descent of J at x, i.e.,

$$\nabla J(x) \neq 0 \Longrightarrow (\exists \epsilon > 0)\ J\left(x - \epsilon \nabla J(x)\right) < J(x). \tag{1.13}$$

If J is not differentiable at x, the negative of a subgradient $t \in \partial J(x)$ is not necessarily a direction of descent. As a particular example in $\mathbb{R}^2$, consider the nondifferentiable convex function

$$f(x^1, x^2) = \max\{-x^1 + |x^2|^2, |x^1|^2 - x^2\}. \tag{1.14}$$

At $(1,1)$ $\partial f(1,1) = \operatorname{conv}\{(-1,2), (2,-1)\}$. Take $t = (-0.7, 1.7) \in \partial f(1,1)$. Then

$$\begin{aligned} (\forall \epsilon > 0)\ J(1 + 0.7\epsilon, 1 - 1.7\epsilon) &= \max\{2.89\epsilon^2 - 4.1\epsilon,\ 0.49\epsilon^2 + 3.1\epsilon\} \\ &\geq 0.49\epsilon^2 + 3.1\epsilon > f(1,1) = 0. \end{aligned} \tag{1.15}$$

Since $\min_{(x^1,x^2) \in \mathsf{E}^2} f(x^1, x^2) < 0$, this shows that $-t$ is not at a direction of descent for f at $(1,1)$.

1.4 Proposed Work

1.4.1 State-of-the-Art in Nondifferentiable Signal Restoration

For digital signal restoration problems, oftentimes one piece of *a priori* knowledge is in form of the linear degradation model (1.1) [5]. With this model, nondifferentiable problems have shown up in one of the following two formats.

One is to formulate it as an unconstrained optimization problem

$$\text{Find } \arg\min_{x \in \mathsf{E}} \{J(x) + \lambda \|Lx - y\|^2\}, \tag{1.16}$$

where $J \colon \mathsf{E} \to \mathbb{R}$ is a convex objective and $\lambda > 0$ is used to balance the weight between the goodness of x in fitting the recorded signal y as required by the degradation model (1.1) and the penalty imposed by the cost J on x as required by *a priori* knowledge. The other is to formulate it as a constrained optimization problem

$$\begin{cases} \text{Find } x^\star \in S \text{ such that } J(x^\star) = \inf\limits_{x \in S} J(x) \\ S = \{x \in \mathsf{E} \mid \|Lx - y\|^2 \leq \delta\}, \end{cases} \tag{1.17}$$

where S is a noise constraint set, obtained through certain statistical hypotheses of the noise u, and $J \colon \mathsf{E} \to \mathbb{R}$ is a convex objective. Such formulations are limited to one constraint.

Formulations (1.16) and (1.17) are different but they are equivalent under certain conditions [30]. For formulation (1.16) one critical point is to find a suitable λ. Although Bayesian analysis provides a framework for obtaining a maximum likelihood estimate of λ for (1.16) under certain statistical hypothesis [6], the resulting problem is usually hard to solve. Oftentimes the parameter λ in (1.16) needs to be chosen by trial and error, usually, guided by some heuristic rules. In the end, the properties of the solution are often uncertain.

Formulations involving a nondifferentiable function have been found more appropriate in many applications [4, 3, 9, 29, 38, 48]. In [9, 48] signal restoration problems are formulated as (1.16), where J is the total variation cost. To solve it, it is transformed into a problem using auxiliary variables (in fact the total variation cost is slightly modified to make it differentiable) and the resulting problem is then solved by alternating two auxiliary minimization problems, one of which can be solved trivially in closed form and the other is a quadratic minimization problem and solved by quadratic programming. The method in [9, 48] has been reported to be robust and efficient [10].

In [3], a recursive method is proposed to solve the one-dimensional signal denoising problem

$$\text{Find } \arg\min \left\{ \|x - y\|_1 + \lambda \|\triangle x\|_1 \right\}, \tag{1.18}$$

where $\triangle$ represents the first order finite-difference operator, i.e, $(\triangle x)^i = x^i - x^{i-1}$. However, the method can be applied only when λ in (1.18) is of the form $\frac{n}{2}$ for some positive integer n. Therefore the solution of (1.18) cannot be guaranteed to satisfy the noise constraint in general. Moreover, how to generalize such a method to solve a two-dimensional signal denoising problem is still open.

In [38], signal denoising problems are formulated as (1.17), where J is the total variation cost. The method used to solve (1.17) is akin to a gradient projection method in which iterates are perturbed to avoid points of non-differentiability. Its numerical implementation is straightforward but it lacks a sound mathematical basis. In [29], signal restoration problems are formulated as (1.17), where the cost J is a variant of the total variation cost. The resulting problem is to minimize an ℓ_1 cost with a single quadratic constraint. It is solved by an affine scaling Newton method whose computation is expensive for large images.

All the above methods are limited to one signal constraint and a special type of cost. Moreover, using formulation (1.16), one does not solve problem (1.17) in general.

1.4.2 Proposed Work

In digital signal restoration, a wide range of signal constraints derived from *a priori* knowledge can be represented as closed convex sets $(S_i)_{0\leq i\leq m}$ in E. To use such constraints, signal restoration problems can be formulated as feasibility problem [12, 13, 43, 45, 49], i.e.,

$$\text{Find } x^* \in S \triangleq \bigcap_{0\leq i\leq m} S_i. \tag{1.19}$$

It is sometimes more advantageous to selectively pick a solution in S. For instance, in [11] algorithms are proposed to solve the best feasible approximation problem

$$\text{Find } x^* \in S = \bigcap_{0\leq i\leq m} S_i \ \text{ such that } \ \|x^* - r\| = \inf_{x\in S} \|x - r\|, \tag{1.20}$$

where r is a reference signal. Problem (1.20) is a multiple constraint quadratic convex minimization problem.

In many instances, signals have block features such as sharp edges which can be better captured by a nondifferentiable objective [3, 9, 29, 38, 48] and the resulting signal restoration problems should be formulated as a multiple-constraint, nondifferentiable convex optimization problems of the form (1.2), i.e.,

$$\text{Find } x^* \in S = \bigcap_{0\leq i\leq m} S_i \ \text{ such that } \ J(x^*) = \inf_{x\in S} J(x). \tag{1.21}$$

This is precisely the type of problems to be addressed in this dissertation. Our goal is to develop a general algorithm to solve (1.21) under mild assumptions.

1.4.3 Methodology

Consider our general problem (1.2) under Assumption 1.1 and let $\alpha^* = \inf_{x \in S} J(x)$. Then the solution set S^* of (1.2) is

$$S^* = S \cap \operatorname{lev}_{\leq \alpha^*} J. \tag{1.22}$$

If α^* is known, (1.2) is equivalent to the feasibility problem

$$\text{Find } x^* \in \left(\bigcap_{0 \leq i \leq m} S_i \right) \cap \operatorname{lev}_{\leq \alpha^*} J, \tag{1.23}$$

which can be solved by many methods [7, 8, 13, 14, 22, 24]. Unfortunately, in practice α^* is unknown.

Inspired by [20, 23], we shall use a level method to adaptively estimate the optimal constrained level α^* and treat (1.2) fundamentally as the feasibility problem shown in (1.23). Although the procedure for estimating α^* will be nontrivial, it will share with most other level methods the following basic principles:

(i) If $x \in \bigcap_{0 \leq i \leq m} S_i \cap \operatorname{lev}_{\leq \alpha} J \neq \emptyset$ can be found, then we infer $\alpha \geq \alpha^*$.

(ii) If $\bigcap_{0 \leq i \leq m} S_i \cap \operatorname{lev}_{\leq \alpha} J = \emptyset$ can be detected, then we infer $\alpha < \alpha^*$.

1.5 Contribution

The main contribution of this dissertation is the development of a general algorithm to solve multiple-constraint, nondifferentiable convex optimization problems (Chapter 4). Other significant new developments are the following.

- In Chapter 2, a new class of operators (nice operators) is introduced and some of their properties are derived.
- In Chapter 3, Polyak's projected subgradient method is extended (Theorem 3.4).
- In Chapter 4, a level subgradient method is proposed to solve the general signal restoration problem (1.2) under very mild assumptions. The convergence of the method is established and its (possibly parallel) implementation is discussed.
- In Chapter 5, applications of the proposed method to multiple-constraint nondifferentiable signal restoration and denoising problems are demonstrated.
- In Chapter 6, a special case of the subgradient level method proposed in Chapter 4 is investigated. Its application to the restoration/denoising of synthetic block signals is demonstrated.

Chapter 2

Mathematical Preliminaries

In this chapter, we provide some mathematical results that will be used in subsequent chapters.

2.1 Convex Analysis

For a detailed account of convex analysis and proofs of the following results, consult [18, 35, 37, 42].

Throughout, f is a real-valued function defined everywhere on E and is assumed to be convex, i.e.,

$$(\forall (x,y) \in \mathsf{E}^2)(\forall \lambda \in [0,1])\ f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y). \tag{2.1}$$

Theorem 2.1 *f is continuous on E and, for every $\alpha \in \mathbb{R}$, its lower level set*

$$\mathrm{lev}_{\leq\alpha} f = \{x \in \mathsf{E} \mid f(x) \leq \alpha\} \tag{2.2}$$

is closed and convex.

Theorem 2.2 *Let $S \subset \mathsf{E}$ be closed. If, for some $\alpha \in \mathbb{R}$, $S \cap \mathrm{lev}_{\leq\alpha} f$ is nonempty and bounded then f achieves its infimum on S, i.e.,*

$$(\exists x^* \in S)\ f(x^*) = \inf_{x \in S} f(x). \tag{2.3}$$

Definition 2.3 A vector $t \in \mathsf{E}$ is a *subgradient* of f at a point $x \in \mathsf{E}$ if

$$(\forall y \in \mathsf{E})\ f(x) + \langle t \mid y - x\rangle \leq f(y). \tag{2.4}$$

The set of all subgradients of f at x is the *subdifferential* of f at x and is denoted by $\partial f(x)$.

Theorem 2.4 *For every $x \in \mathsf{E}$, the subdifferential $\partial f(x)$ is nonempty, bounded, convex, and closed. Moreover,*

$$(\forall x^* \in \mathsf{E})\ f(x^*) = \inf_{x \in \mathsf{E}} f(x) \Longleftrightarrow 0 \in \partial f(x^*). \tag{2.5}$$

Theorem 2.5 *∂f maps any bounded set and any nonempty lower level set of f into a bounded set.*

Theorem 2.6 *Let $f = \sum_{i=1}^{p} \lambda_i f_i$, where $(\lambda_i)_{1\leq i\leq p} \subset]0, +\infty[$ and $(f_i)_{1\leq i\leq p}\colon \mathsf{E} \to \mathbb{R}$ are convex. Then f is convex and*

$$(\forall x \in \mathsf{E})\ \partial f(x) = \sum_{i=1}^{p} \lambda_i \partial f_i(x) = \Big\{ \sum_{i=1}^{p} \lambda_i t_i \mid (\forall i \in \{1, \ldots, p\})\ t_i \in \partial f_i(x) \Big\}. \tag{2.6}$$

Theorem 2.7 *Let* $f = \max_{1\leq i\leq p} f_i$, *where* $(f_i)_{1\leq i\leq p}\colon \mathsf{E} \to \mathbb{R}$ *are convex. Then* f *is convex and*

$$(\forall x \in \mathsf{E})\ \partial f(x) = \operatorname{conv} \bigcup_{i\in I(x)} \partial f_i(x),\ \textit{where } I(x) = \big\{i \in \{1,\dots,p\} \mid f_i(x) = f(x)\big\}. \quad (2.7)$$

Definition 2.8 f is *differentiable* at x if there exists a vector $t \in \mathsf{E}$ (necessarily unique) such that

$$f(y) = f(x) + \langle t \mid y - x\rangle + o(\|y - x\|) \quad (2.8)$$

or, equivalently,

$$\lim_{y\to x} \frac{f(y) - f(x) - \langle t \mid y - x\rangle}{\|y - x\|} = 0. \quad (2.9)$$

Such a t, if it exists, is the *gradient* of f at x and is denoted by $\nabla f(x)$.

Theorem 2.9 f *is differentiable almost everywhere (Lebesgue) on* E *and, if* f *is differentiable at* x, *then* $\partial f(x) = \{\nabla f(x)\}$. *Moreover, if* f *is differentiable on* E *(i.e., differentiable at every point in* E*), then it is continuously differentiable on* E.

Theorem 2.10 *Let* $C \subset \mathsf{E}$ *be a nonempty closed convex set. Then, for every* $x \in \mathsf{E}$, *there exists a unique point* $P_C(x) \in C$ *such that*

$$\|x - P_C(x)\| = \min_{y\in C} \|x - y\|. \quad (2.10)$$

In addition, this point is characterized by

$$P_C(x) \in C \ \textit{ and } \ (\forall x \in \mathsf{E})(\forall y \in C)\ \langle x - P_C(x) \mid P_C(x) - y\rangle \geq 0. \quad (2.11)$$

Definition 2.11 The point $P_C(x)$ in Theorem 2.10 is the *projection* of x onto C and the operator $P_C\colon \mathsf{E} \to C$ is the *projector* onto C.

It follows from (2.11) that P_C is *firmly nonexpansive*, i.e.,

$$(\forall (x, y) \in \mathsf{E}^2)\ \|P_C(x) - P_C(y)\|^2 \leq \langle P_C(x) - P_C(y) \mid x - y \rangle. \tag{2.12}$$

and therefore *nonexpansive*, i.e.,

$$(\forall (x, y) \in \mathsf{E}^2)\ \|P_C(x) - P_C(y)\| \leq \|x - y\|. \tag{2.13}$$

2.2 Subgradient Projections

Definition 2.12 [7, 12, 13] Let η be a real number such that $\operatorname{lev}_{\leq \eta} f \neq \varnothing$. The *subgradient projection* of $x \in \mathsf{E}$ onto $\operatorname{lev}_{\leq \eta} f$ is

$$G_\eta^f(x) = \begin{cases} x - \dfrac{f(x) - \eta}{\|t\|^2} t & \text{if } f(x) > \eta, \text{ where } t \in \partial f(x) \\ x & \text{if } f(x) \leq \eta. \end{cases} \tag{2.14}$$

Note that $G_\eta^f(x)$ is the projection of x onto the half-space [7, 13]

$$H_\eta^f = \{ y \in \mathsf{E} \mid f(x) + \langle t \mid y - x \rangle \leq \eta \}, \tag{2.15}$$

and that by (2.4) and (2.14)

$$\operatorname{Fix} G_\eta^f = \operatorname{lev}_{\leq \eta} f \subset H_\eta^f. \tag{2.16}$$

It follows from (2.14) that

$$x \notin \operatorname{lev}_{\leq \eta} f \Longrightarrow \|G_\eta^f(x) - x\| = \frac{f(x) - \eta}{\|t\|} > 0. \tag{2.17}$$

Hence $G_\eta^f(x)$ can be regarded as an approximate projection of x onto $\mathrm{lev}_{\leq\eta} f$. Since it requires only a subgradient of f at x, it is significantly easier to implement than an exact projection and is used for solving a wide range of feasibility problems [7, 8, 12, 13, 24].

Lemma 2.13 $(\forall x \in \mathsf{E})(\forall y \in \mathrm{lev}_{\leq\eta} f)\ \|G_\eta^f(x) - y\|^2 \leq \|x - y\|^2 - \|G_\eta^f(x) - x\|^2.$

Proof. Fix $y \in \mathrm{lev}_{\leq\eta} f$, $x \in \mathsf{E}$, and $t \in \partial f(x)$. Then

$$\|G_\eta^f(x) - y\|^2 = \|G_\eta^f(x) - x\|^2 + \|x - y\|^2 - 2\langle x - G_\eta^f(x) \mid x - y\rangle. \tag{2.18}$$

By (2.14), (2.4), and $f(y) \leq \eta$

$$\begin{aligned} f(x) > \eta \Longrightarrow \langle x - G_\eta^f(x) \mid x - y\rangle &= \frac{f(x) - \eta}{\|t\|^2}\langle t \mid x - y\rangle \\ &\geq \frac{(f(x) - \eta)^2}{\|t\|^2} \\ &= \|G_\eta^f(x) - x\|^2. \end{aligned} \tag{2.19}$$

On the other hand, by (2.14), $f(x) \leq \eta \Rightarrow x - G_\eta^f(x) = 0$. Therefore

$$f(x) \leq \eta \Longrightarrow \langle x - G_\eta^f(x) \mid x - y\rangle = \|G_\eta^f(x) - x\|^2. \tag{2.20}$$

By (2.18)-(2.20), the lemma is proved. ☐

Proposition 2.14 *If $C \subset \mathsf{E}$ is bounded then*

$$(\forall (x_n)_{n\geq 0} \subset C)(\exists \tau \in]0, +\infty[)(\forall n \in \mathbb{N})(\forall t_n \in \partial f(x_n))\ \|t_n\| \leq \tau. \tag{2.21}$$

Proof. Apply Theorem 2.5. ☐

Remark 2.15 G_η^f is not continuous.

Proof. We prove this fact by providing a counterexample. Let $f\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto \max\{x, 4x\}$. Then f is nondifferentiable at $x = 0$ and $\mathrm{lev}_{\leq -1} f =]-\infty, -1]$. Now define $(\forall n \in \mathbb{N})$ $x_n = \dfrac{1}{n+1}$. Then $(x_n)_{n\geq 0} \subset]0, 1]$, $x_n \to 0$, and $(\forall n \in \mathbb{N})$ $\partial f(x_n) = \{4\}$ and $G^f_{-1}(x_n) = -\dfrac{1}{4}$. However by taking $t = 1 \in [1, 4] = \partial f(0)$, we obtain $G^f_{-1}(0) = -1$.

$$\lim_{n\to+\infty} G^f_{-1}(x_n) = -\frac{1}{4} \neq -1 = G^f_{-1}(0). \tag{2.22}$$

Therefore G^f_{-1} is not continuous at 0. □

Lemma 2.16 *Let $(x_n)_{n\geq 0} \subset \mathsf{E}$ and $(\eta_n)_{n\geq 0} \subset \mathbb{R}$. If $g\colon \mathsf{E} \to \mathbb{R}$ is continuous then*

$$\begin{cases} x_n \to x \\ \eta_n \to \eta \\ (g(x_n) - \eta_n)^+ \to 0 \end{cases} \implies g(x) \leq \eta. \tag{2.23}$$

Proof. Suppose $x_n \to x$, $\eta_n \to \eta$ and $(g(x_n) - \eta_n)^+ \to 0$. Since g is continuous, $g(x_n) \to g(x)$. Hence

$$\frac{|g(x_n) - \eta_n| + g(x_n) - \eta_n}{2} \to \frac{|g(x) - \eta| + g(x) - \eta}{2}. \tag{2.24}$$

On the other hand

$$\frac{|g(x_n) - \eta_n| + g(x_n) - \eta_n}{2} = (g(x_n) - \eta_n)^+ \to 0. \tag{2.25}$$

Therefore $|g(x) - \eta| = -(g(x) - \eta)$, which implies $g(x) - \eta \leq 0$ and $g(x) \leq \eta$. □

Proposition 2.17 *Let* $(\eta_n)_{n\geq 0} \subset \mathbb{R}$ *and* $(x_n)_{n\geq 0} \subset \mathsf{E}$. *Then*

$$\begin{cases} x_n \to x \\ \eta_n \to \eta \\ \|G^f_{\eta_n}(x_n) - x_n\| \to 0 \end{cases} \Longrightarrow x \in \operatorname{lev}_{\leq \eta} f. \tag{2.26}$$

Proof. For every $n \in \mathbb{N}$ we have

$$\|G^f_{\eta_n}(x_n) - x_n\| = \begin{cases} \dfrac{f(x_n) - \eta_n}{\|t_n\|} & \text{if } f(x_n) > \eta_n \ \text{ where } t_n \in \partial f(x_n) \\ 0 & \text{if } f(x_n) \leq \eta_n. \end{cases} \tag{2.27}$$

Since $(x_n)_{n\geq 0}$ converges, it is bounded and, by Proposition 2.14, there exists $\tau \in]0, +\infty[$ such that $\sup_{n\in I} \|t_n\| \leq \tau$, where $I = \{n \in \mathbb{N} \mid f(x_n) > \eta\}$. Therefore

$$(\forall n \in \mathbb{N}) \ \|G^f_{\eta_n}(x_n) - x_n\| \geq \frac{1}{\tau}\big(f(x_n) - \eta_n\big)^+. \tag{2.28}$$

It follows from $\|G^f_{\eta_n}(x_n) - x_n\| \to 0$ that $(f(x_n) - \eta_n)^+ \to 0$. Since $x_n \to x$ and $\eta_n \to \eta$, Lemma 2.16 yields $f(x) \leq \eta$ and therefore $x \in \operatorname{lev}_{\leq \eta} f$. □

2.3 Attracting Operators

Definition 2.18 [7] Let $T : \mathsf{E} \to \mathsf{E}$ be an operator such that Fix T is nonempty, closed, and convex, and let $\nu \in]0, +\infty[$. T is *ν-attracting* if

$$(\forall x \in \mathsf{E})(\forall w \in \operatorname{Fix} T) \ \|x - w\|^2 - \|T(x) - w\|^2 \geq \nu\|T(x) - x\|^2. \tag{2.29}$$

Proposition 2.19 *For any nonempty closed convex set* $S \subset \mathsf{E}$, $S = \operatorname{Fix} P_S$ *and* P_S *is 1-attracting.*

Proof. It follows from the definition of P_S that $S = \mathrm{Fix}\, P_S$. On the other hand, it follows from (2.11) that for any $x \in \mathsf{E}$ and $y \in S$

$$\begin{aligned} \|x-y\|^2 &= \|x - P_S(x)\|^2 + \|P_S(x) - y\|^2 + 2\langle x - P_S(x) \mid P_S(x) - y\rangle \\ &\geq \|x - P_S(x)\|^2 + \|P_S(x) - y\|^2. \end{aligned} \tag{2.30}$$

Thus P_S is 1-attracting. □

Proposition 2.20 *G_η^f is 1-attracting.*

Proof. By Lemma 2.13 and (2.16). □

Lemma 2.21 [7] *Let $(T_i)_{1\leq i\leq m}\colon \mathsf{E} \to \mathsf{E}$ be such that each T_i is ν_i-attracting and*

$$\bigcap_{1\leq i\leq m} \mathrm{Fix}\, T_i \neq \emptyset. \tag{2.31}$$

Then $T = T_1 \circ \cdots \circ T_m$ is $\min\{\nu_1, \cdots, \nu_m\}/2^{m-1}$-attracting and $\mathrm{Fix}\, T = \bigcap_{1\leq i\leq m} \mathrm{Fix}\, T_i$.

2.4 Nice Operators

Definition 2.22 $T\colon \mathsf{E} \to \mathsf{E}$ is *nice* if it is ν-attracting and, for every sequence $(x_n)_{n\geq 0} \subset \mathsf{E}$,

$$\begin{cases} x_n \to x \\ (T - \mathrm{Id})(x_n) \to 0 \end{cases} \quad \Longrightarrow \quad x \in \mathrm{Fix}\, T. \tag{2.32}$$

We now give some examples of nice operators.

Proposition 2.23 *Let C be a nonempty closed convex set. Then P_C is nice and* $\mathrm{Fix}\, P_C = C$.

Proof. P_C is 1-attracting by Proposition 2.19 and has fixed point set C. Moreover, since it is nonexpansive by (2.13), it is continuous. □

Proposition 2.24 G_η^f *is nice and* $\operatorname{Fix} G_\eta^f = \operatorname{lev}_{\leq \eta} f$.

Proof. Apply Propositions 2.17 and 2.20. □

Proposition 2.25 *Suppose* $(T_i)_{1 \leq i \leq m} \colon \mathsf{E} \to \mathsf{E}$ *are nice and* $\bigcap_{1 \leq i \leq m} \operatorname{Fix} T_i \neq \emptyset$. *Let* $T = T_1 \circ \cdots \circ T_m$. *Then* T *is nice and* $\operatorname{Fix} T = \bigcap_{1 \leq i \leq m} \operatorname{Fix} T_i$.

Proof. Since $(T_i)_{1 \leq i \leq m}$ are nice, suppose that each T_i is ν_i-attracting. By Lemma 2.21 T is $\min\{\nu_1, \ldots, \nu_m\}/2^{m-1}$-attracting and $\operatorname{Fix} T = \bigcap_{1 \leq i \leq m} \operatorname{Fix} T_i$. Now fix $w \in \operatorname{Fix} T$ and $(x_n)_{n \geq 0} \subset \mathsf{E}$. For every $n \in \mathbb{N}$, $T(x_n) = T_1 \circ \cdots \circ T_m(x_n)$ or, equivalently,

$$\begin{cases} x_{m+1,n} = x_n \\ x_{m,n} = T_m(x_{m+1,n}) \\ x_{m-1,n} = T_{m-1}(x_{m,n}) \\ \quad\vdots \\ x_{2,n} = T_2(x_{3,n}) \\ x_{1,n} = T_1(x_{2,n}) \\ T(x_n) = x_{1,n}. \end{cases} \tag{2.33}$$

Since each T_i is ν_i-attracting, we get from (2.33) that

$$\begin{cases} \|x_{m+1,n} - w\|^2 - \|x_{m,n} - w\|^2 \geq \nu_m \|x_{m,n} - x_{m+1,n}\|^2 \\ \|x_{m,n} - w\|^2 - \|x_{m-1,n} - w\|^2 \geq \nu_{m-1} \|x_{m-1,n} - x_{m,n}\|^2 \\ \qquad \vdots \\ \|x_{3,n} - w\|^2 - \|x_{2,n} - w\|^2 \geq \nu_2 \|x_{2,n} - x_{3,n}\|^2 \\ \|x_{2,n} - w\|^2 - \|x_{1,n} - w\|^2 \geq \nu_1 \|x_{1,n} - x_{2,n}\|^2. \end{cases} \tag{2.34}$$

It follows from (2.34) that

$$\|x_{m+1,n} - w\|^2 - \|x_{1,n} - w\|^2 \geq \sum_{i=1}^{m} \nu_i \|x_{i,n} - x_{i+1,n}\|^2. \tag{2.35}$$

Now suppose

$$x_n \to x \text{ and } (T - \mathrm{Id})(x_n) \to 0. \tag{2.36}$$

We must now show that $x \in \bigcap_{1 \leq i \leq m} \operatorname{Fix} T_i$, i.e., that (2.32) holds. Since $T(x_n) = x_{1,n}$ and $x_{m+1,n} = x_n$, we get

$$x_{1,n} - x_{m+1,n} \to 0 \text{ and } x_{1,n} - x_n \to 0. \tag{2.37}$$

Then, we get from $x_n \to x$ and (2.37) that

$$x_{m+1,n} \to x \text{ and } x_{1,n} \to x. \tag{2.38}$$

Therefore

$$\|x_{m+1,n} - w\| \to \|x - w\| \text{ and } \|x_{1,n} - w\| \to \|x - w\| \tag{2.39}$$

and

$$\|x_{m+1,n} - w\|^2 - \|x_{1,n} - w\|^2 \to 0. \tag{2.40}$$

From (2.40) and (2.35) we get $\sum_{i=1}^{m} \nu_i \|x_{i,n} - x_{i+1,n}\|^2 \to 0$ and therefore

$$x_{1,n} - x_{2,n} \to 0,\ x_{2,n} - x_{3,n} \to 0,\ \ldots,\ x_{m,n} - x_{m+1,n} \to 0. \tag{2.41}$$

Thus it follows from (2.38), (2.41), and (2.33) that

$$(\forall i \in \{1, \ldots, m\}) \quad \begin{cases} x_{i+1,n} \rightharpoonup x \\ (T_i - \mathrm{Id})(x_{i+1,n}) \to 0. \end{cases} \tag{2.42}$$

Since the operators $(T_i)_{1 \le i \le m}$ are nice, we conclude $x \in \bigcap_{1 \le i \le m} \operatorname{Fix} T_i$. □

Chapter 3

Polyak's Subgradient Projection Method

In this chapter, Polyak's subgradient projection method [34] is reviewed and some extensions are given. These algorithms will serve as a foundation for the nondifferentiable optimization algorithms developed in the following chapters.

3.1 Basic Principle

The basic signal recovery problem under consideration was formulated in (1.2) as

$$\text{Find } x^* \in S \text{ such that } J(x^*) = \inf_{x \in S} J(x). \tag{3.1}$$

Under Assumption 1.1 (3.1) has a solution. The solution is unique if J is strictly convex on S, i.e.,

$$(\forall (x,y) \in S^2)(\forall \lambda \in]0,1[)\ x \neq y \Longrightarrow J(\lambda x + (1-\lambda)y) < \lambda J(x) + (1-\lambda)J(y). \tag{3.2}$$

Now denote by $\alpha^* = \inf_{x \in S} J(x)$ the optimal constrained value of J. Then the solution set of (3.1) can be written as

$$S^* = S \cap \operatorname{lev}_{\leq \alpha^*} J. \tag{3.3}$$

When J is differentiable, (3.1) can be approached by the *projected gradient method*

$$(\forall n \in \mathbb{N})\ x_{n+1} = P_S(x_n - \gamma_n \nabla J(x_n)), \text{ where } \gamma_n \geq 0. \tag{3.4}$$

The step lengths $(\gamma_n)_{n \geq 0}$ can be chosen according to various criteria, e.g.,

$$(\forall n \in \mathbb{N})\ \gamma_n \in \arg\inf_{\gamma \geq 0} J(x_n - \gamma \nabla J(x_n)). \tag{3.5}$$

When J is nondifferentiable, (3.4) can be generalized to the *projected subgradient method*

$$(\forall n \in \mathbb{N})\ x_{n+1} = P_S(x_n - \gamma_n t_n), \text{ where } t_n \in \partial J(x_n). \tag{3.6}$$

Polyak's method [34] provides a scheme for choosing the sequence $(\gamma_n)_{n \geq 0}$ in (3.6) under the assumption that the optimal value α^* is known.

3.2 Polyak's Method and Its Convergence

Theorem 3.1 [34] *Given any $x_0 \in S$ and $\epsilon \in]0, 1]$, let $(x_n)_{n \geq 0}$ be a sequence generated by (3.6) with $(\gamma_n)_{n \geq 0}$ defined by*

$$(\forall n \in \mathbb{N})\ \gamma_n = \lambda_n \frac{J(x_n) - \alpha^*}{\|t_n\|^2}, \text{ where } \epsilon \leq \lambda_n \leq 2 - \epsilon. \tag{3.7}$$

Then $(x_n)_{n \geq 0}$ converges to a point in S^*.

Proof. It follows from Assumption 1.1 that S is an active constraint. Hence, by Theorem 2.4, no subgradient of J at any $x \in S$ is zero and (3.7) is well defined. It follows from

(2.13) that

$$
\begin{aligned}
(\forall w \in S)\ \|w - x_{n+1}\|^2 &= \|w - P_S(x_n - \gamma_n t_n)\|^2 \\
&= \|P_S(w) - P_S(x_n - \gamma_n t_n)\|^2 \\
&\leq \|w - x_n + \gamma_n t_n\|^2 \\
&= \|w - x_n\|^2 - 2\gamma_n \langle w - x_n \mid -t_n \rangle + \gamma_n^2 \|t_n\|^2. \qquad (3.8)
\end{aligned}
$$

Now take an arbitrary $x^* \in S^*$. From (3.7), (2.4), and $S^* \subset S$, we get

$$
\begin{aligned}
\|x^* - x_{n+1}\|^2 &\leq \|x^* - x_n\|^2 - \lambda_n(2 - \lambda_n)\frac{(J(x_n) - \alpha^*)^2}{\|t_n\|^2} \\
&\leq \|x^* - x_n\|^2 - \epsilon^2 \frac{(J(x_n) - \alpha^*)^2}{\|t_n\|^2}. \qquad (3.9)
\end{aligned}
$$

It follows from (3.9) that

$$
(\forall n \in \mathbb{N})\ \|x^* - x_{n+1}\|^2 \leq \|x^* - x_n\|^2 \leq \|x^* - x_0\|^2. \qquad (3.10)
$$

Therefore $(x_n)_{n \geq 0}$ is bounded and there exists a subsequence $(x_{n_k})_{k \geq 0}$ of $(x_n)_{n \geq 0}$ such that $x_{n_k} \rightharpoonup x$. As $(x_n)_{n \geq 0} \subset S$ and S is closed, we get $x \in S$. Moreover, since $(x_n)_{n \geq 0}$ is bounded, by Proposition 2.14, there exists $\tau \in]0, +\infty[$ such that $\sup_{n \geq 0} \|t_n\| \leq \tau$. Thus, (3.9) gives

$$
(\forall n \in \mathbb{N})\ \left(\frac{\epsilon}{\tau}\right)^2 \sum_{l=0}^{n} (J(x_l) - \alpha^*)^2 \leq \|x^* - x_0\|^2 - \|x^* - x_{n+1}\|^2 \leq \|x^* - x_0\|^2. \qquad (3.11)
$$

Hence $\sum_{n \geq 0} \left(J(x_n) - \alpha^*\right)^2 < +\infty$ and therefore $J(x_n) \to \alpha^*$. Since J is continuous, from $x_{n_k} \rightharpoonup x$ and $J(x_{n_k}) \to \alpha^*$, we get $J(x) = \alpha^*$ and $x \in \operatorname{lev}_{\leq \alpha^*} J$. Therefore $x \in S^*$ and, by (3.10), the whole sequence $(x_n)_{n \geq 0}$ converges to x. □

3.3 An Extension of Polyak's Method

The following variant of Theorem 3.1 was proposed in [2].

Theorem 3.2 [2] *Suppose* $\underline{\alpha} < \alpha^*$. *Given any* $x_0 \in S$ *and* $\epsilon \in]0,2[$, *let* $(x_n)_{n\geq 0}$ *be a sequence generated by*

$$(\forall n \in \mathbb{N})\ x_{n+1} = P_S(x_n - \gamma_n t_n),\ \textit{where} \begin{cases} \gamma_n = \lambda_n \dfrac{J(x_n) - \underline{\alpha}}{\|t_n\|^2} \\ t_n \in \partial J(x_n) \\ 0 < \lambda_n \leq 2 - \epsilon. \end{cases} \tag{3.12}$$

Then, if $\sum_{n\geq 0} \lambda_n = +\infty$,

$$(\forall \delta \in]0, +\infty[)(\exists m \in \mathbb{N})\ J(x_m) \leq \alpha^* + \frac{2-\epsilon}{\epsilon}(\alpha^* - \underline{\alpha}) + \delta. \tag{3.13}$$

Proof. It follows from Assumption 1.1 that S is an active constraint. Hence, by Theorem 2.4, no subgradient of J at any $x \in S$ is zero and (3.12) is well defined. Now fix $x^* \in S^*$. By (3.8)

$$(\forall n \in \mathbb{N})\ \|x^* - x_{n+1}\|^2 \leq \|x^* - x_n\|^2 - 2\gamma_n \langle x^* - x_n \mid -t_n\rangle + \gamma_n^2 \|t_n\|^2. \tag{3.14}$$

Thus, by (2.4) and (3.12), we get for every $n \in \mathbb{N}$

$$\begin{aligned} \|x^* - x_n\|^2 - \|x^* - x_{n+1}\|^2 &\geq \lambda_n \frac{J(x_n) - \underline{\alpha}}{\|t_n\|^2}\Big[2(J(x_n) - \alpha^*) - \lambda_n(J(x_n) - \underline{\alpha})\Big] \\ &= \lambda_n \frac{J(x_n) - \underline{\alpha}}{\|t_n\|^2}\Big[(2 - \lambda_n)(J(x_n) - \alpha^*) \\ &\qquad -\lambda_n(\alpha^* - \underline{\alpha})\Big]. \end{aligned} \tag{3.15}$$

Suppose that (3.13) is not true, i.e., there exists $\delta \in]0, +\infty[$ such that

$$(\forall n \in \mathbb{N})\ J(x_n) - \alpha^* > \frac{2-\epsilon}{\epsilon}(\alpha^* - \underline{\alpha}) + \delta. \tag{3.16}$$

Then

$$\begin{aligned}(\forall n \in \mathbb{N})\ \|x^* - x_n\|^2 - \|x^* - x_{n+1}\|^2 &> \lambda_n \frac{J(x_n) - \underline{\alpha}}{\|t_n\|^2}\Big[\frac{2-\epsilon}{\epsilon}(2-\lambda_n)(\alpha^* - \underline{\alpha}) \\ &\quad +(2-\lambda_n)\delta - \lambda_n(\alpha^* - \underline{\alpha})\Big] \\ &= \lambda_n \frac{J(x_n) - \underline{\alpha}}{\|t_n\|^2}\Big[\frac{4 - 2(\lambda_n + \epsilon)}{\epsilon}(\alpha^* - \underline{\alpha}) \\ &\quad +(2-\lambda_n)\delta\Big] \\ &\geq \lambda_n \epsilon\delta\, \frac{J(x_n) - \underline{\alpha}}{\|t_n\|^2} \\ &> \lambda_n \epsilon\delta\, \frac{J(x_n) - \alpha^*}{\|t_n\|^2} > \lambda_n \frac{\epsilon\delta^2}{\|t_n\|^2}. \end{aligned} \tag{3.17}$$

It follows from (3.17) that $(x_n)_{n\geq 0}$ is bounded. Hence, by Proposition 2.14, there exists $\tau \in]0, +\infty[$ such that $\sup_{n\geq 0} \|t_n\| \leq \tau$. Therefore, from (3.17), we get

$$(\forall n \in \mathbb{N})\ \|x^* - x_n\|^2 - \|x^* - x_{n+1}\|^2 > \lambda_n\left(\frac{\epsilon\delta^2}{\tau^2}\right). \tag{3.18}$$

Consequently

$$(\forall n \in \mathbb{N})\ \sum_{l=0}^{n} \lambda_l \leq \left(\frac{\tau^2}{\epsilon\delta^2}\right)\left(\|x^* - x_0\|^2 - \|x^* - x_{n+1}\|^2\right) \leq \left(\frac{\tau^2}{\epsilon\delta^2}\right)\|x^* - x_0\|^2. \tag{3.19}$$

Hence $\sum_{n\geq 0} \lambda_n < +\infty$, a contradiction. □

When a close underestimate $\underline{\alpha} < \alpha^*$ of α^* is available, although $\mathrm{lev}_{\leq\underline{\alpha}}\, J \cap S = \emptyset$, an approximate solution can still be found by the projected subgradient method (3.13) in the sense stated by Theorem 3.2.

3.4 A New Extension of Polyak's Method

Assumption 3.3 $T\colon \mathsf{E} \to \mathsf{E}$ is nice and $\operatorname{Fix} T = S$.

We now consider the recursion

$$(\forall n \in \mathbb{N})\ x_{n+1} = T(x_n - \gamma_n t_n), \text{ where } \gamma_n \geq 0 \text{ and } t_n \in \partial J(x_n). \tag{3.20}$$

Theorem 3.4 *Suppose $\overline{\alpha} \geq \alpha^*$. Given any $x_0 \in \mathsf{E}$ and $\epsilon \in]0,1]$, let $(x_n)_{n\geq 0}$ be a sequence generated by* (3.20) *under Assumption 3.3 with $(\gamma_n)_{n\geq 0}$ defined by*

$$(\forall n \in \mathbb{N})\ \gamma_n = \begin{cases} \lambda_n \dfrac{J(x_n) - \alpha_n}{\|t_n\|^2} & \text{if } J(x_n) > \alpha_n\text{, where} \begin{cases} \epsilon \leq \lambda_n \leq 2 - \epsilon \\ \alpha_n \geq \overline{\alpha} \end{cases} \\ 0 & \text{if } J(x_n) \leq \alpha_n. \end{cases} \tag{3.21}$$

Then, if $\alpha_n \to \overline{\alpha}$, $(x_n)_{n\geq 0}$ converges to a point in $S \cap \operatorname{lev}_{\leq \overline{\alpha}} J$.

Proof. Take an arbitrary $w \in \operatorname{Fix} T$. Since T is ν-attracting, by (2.29),

$$\begin{aligned} \|x_{n+1} - w\|^2 &= \|T(x_n - \gamma_n t_n) - w\|^2 \\ &\leq \|x_n - w - \gamma_n t_n\|^2 - \nu\|T(x_n - \gamma_n t_n) - (x_n - \gamma_n t_n)\|^2 \\ &= \|x_n - w\|^2 - 2\gamma_n\langle x_n - w \mid t_n\rangle + \gamma_n^2\|t_n\|^2 \\ &\quad -\nu\|T(x_n - \gamma_n t_n) - (x_n - \gamma_n t_n)\|^2. \end{aligned} \tag{3.22}$$

It follows from (2.4) and (3.22) that

$$\begin{aligned} \|x_{n+1} - w\|^2 &\leq \|x_n - w\|^2 - 2\gamma_n\left(J(x_n) - J(w)\right) + \gamma_n^2\|t_n\|^2 \\ &\quad -\nu\|T(x_n - \gamma_n t_n) - (x_n - \gamma_n t_n)\|^2. \end{aligned} \tag{3.23}$$

Since $S^* \subset \operatorname{lev}_{\leq \overline{\alpha}} J$ and $S^* \subset S = \operatorname{Fix} T$

$$\operatorname{lev}_{\leq \overline{\alpha}} J \cap \operatorname{Fix} T \neq \emptyset. \tag{3.24}$$

Now suppose $w \in \operatorname{lev}_{\leq \overline{\alpha}} J \cap \operatorname{Fix} T$ and note that $J(w) \leq \overline{\alpha}$. Then, from (3.23) and $\gamma_n \geq 0$, we get

$$\begin{aligned} \|x_{n+1} - w\|^2 &\leq \|x_n - w\|^2 - 2\gamma_n\left(J(x_n) - \overline{\alpha}\right) + \gamma_n^2\|t_n\|^2 \\ &\quad -\nu\|T(x_n - \gamma_n t_n) - (x_n - \gamma_n t_n)\|^2. \end{aligned} \tag{3.25}$$

If $J(x_n) > \alpha_n$, then $J(x_n) > \alpha_n \geq \overline{\alpha} \geq \alpha^* \geq \inf_{x \in E} J(x)$. Therefore $t_n \neq 0$ and by (3.21)

$$\begin{aligned} -2\gamma_n\left(J(x_n) - \overline{\alpha}\right) + \gamma_n^2\|t_n\|^2 &= -2\lambda_n \frac{(J(x_n) - \alpha_n)(J(x_n) - \overline{\alpha})}{\|t_n\|^2} + \lambda_n^2 \frac{(J(x_n) - \alpha_n)^2}{\|t_n\|^2} \\ &\leq -\lambda_n(2 - \lambda_n)\frac{(J(x_n) - \alpha_n)^2}{\|t_n\|^2} \\ &\leq -\epsilon^2 \frac{(J(x_n) - \alpha_n)^2}{\|t_n\|^2} \\ &= -\left(\frac{\epsilon}{\lambda_n}\right)^2 \|\gamma_n t_n\|^2 \\ &\leq -\left(\frac{\epsilon}{2 - \epsilon}\right)^2 \gamma_n^2\|t_n\|^2. \end{aligned} \tag{3.26}$$

Thus, from (3.25) and (3.26), we get

$$\begin{aligned} J(x_n) > \alpha_n \Longrightarrow \|x_{n+1} - w\|^2 &\leq \|x_n - w\|^2 - \left(\frac{\epsilon}{2 - \epsilon}\right)^2 \|\gamma_n t_n\|^2 \\ &\quad -\nu\|T(x_n - \gamma_n t_n) - (x_n - \gamma_n t_n)\|^2. \end{aligned} \tag{3.27}$$

If $J(x_n) \leq \alpha_n$, it follows from (3.21) that $\gamma_n = 0$ and (3.25) trivially gives

$$\begin{aligned} J(x_n) \leq \alpha_n \Longrightarrow \|x_{n+1} - w\|^2 &\leq \|x_n - w\|^2 - \left(\frac{\epsilon}{2 - \epsilon}\right)^2 \|\gamma_n t_n\|^2 \\ &\quad -\nu\|T(x_n - \gamma_n t_n) - (x_n - \gamma_n t_n)\|^2. \end{aligned} \tag{3.28}$$

It follows from (3.27) and (3.28) that in both cases

$$\begin{aligned}\|x_{n+1}-w\|^2 &\le \|x_n-w\|^2-\left(\frac{\epsilon}{2-\epsilon}\right)^2\|\gamma_n t_n\|^2\\ &\quad -\nu\|T(x_n-\gamma_n t_n)-(x_n-\gamma_n t_n)\|^2. \end{aligned} \tag{3.29}$$

Therefore

$$\begin{aligned}\left(\frac{\epsilon}{2-\epsilon}\right)^2\sum_{l=0}^{n}\|\gamma_l t_l\|^2+\nu\sum_{l=0}^{n}\|T(x_l-\gamma_l t_l)-(x_l-\gamma_l t_l)\|^2 &\le \|x_0-w\|^2-\|x_{n+1}-w\|^2\\ &\le \|x_0-w\|^2. \end{aligned} \tag{3.30}$$

Taking the limit as $n\to+\infty$, we get

$$\sum_{n\ge0}\|\gamma_n t_n\|^2<+\infty \text{ and } \sum_{n\ge0}\|(T-\mathrm{Id})(x_n-\gamma_n t_n)\|^2<+\infty. \tag{3.31}$$

Consequently

$$\gamma_n t_n\to0 \text{ and } (T-\mathrm{Id})(x_n-\gamma_n t_n)\to0. \tag{3.32}$$

It follows from (3.29) that

$$(\forall w\in \mathrm{lev}_{\le\bar\alpha} J\cap \mathrm{Fix}\, T)\ \|x_{n+1}-w\|\le\|x_n-w\|\le\|x_0-w\|. \tag{3.33}$$

Therefore $(x_n)_{n\ge0}$ is bounded and it has a subsequence $(x_{n_k})_{k\ge0}$ such that $x_{n_k}\rightharpoonup x$. Since $\gamma_{n_k}t_{n_k}\to0$, we get $x_{n_k}-\gamma_{n_k}t_{n_k}\rightharpoonup x$. Furthermore, since T is nice, (3.32) and Assumption 3.3 give

$$(T-\mathrm{Id})(x_{n_k}-\gamma_{n_k}t_{n_k})\to0\Longrightarrow x\in\mathrm{Fix}\,T\Longleftrightarrow x\in S. \tag{3.34}$$

From (3.21), we get

$$\|\gamma_n t_n\|=\begin{cases}\lambda_n\dfrac{J(x_n)-\alpha_n}{\|t_n\|} & \text{if } J(x_n)>\alpha_n\\[2ex] 0 & \text{if } J(x_n)\le\alpha_n.\end{cases} \tag{3.35}$$

Since $(x_n)_{n\geq 0}$ is bounded, by Proposition 2.14, there exists $\tau \in]0, +\infty[$ such that $\sup_{n\geq 0} \|t_n\| < \tau$. Thus

$$(\forall n \in \mathbb{N}) \begin{cases} \|\gamma_n t_n\| \geq \left(\frac{\epsilon}{\tau}\right) (J(x_n) - \alpha_n) & \text{if } J(x_n) > \alpha_n \\ \|\gamma_n t_n\| = 0 & \text{if } J(x_n) \leq \alpha_n. \end{cases} \tag{3.36}$$

Hence, in all cases,

$$(\forall n \in \mathbb{N})\ \|\gamma_n t_n\| \geq \left(\frac{\epsilon}{\tau}\right) (J(x_n) - \alpha_n)^+. \tag{3.37}$$

Consequently, it follows from (3.32) that $(J(x_{n_k}) - \alpha_{n_k})^+ \to 0$. Thus, since $x_{n_k} \to x$ and $\alpha_{n_k} \to \overline{\alpha}$, Lemma 2.16 yields $J(x) \leq \overline{\alpha}$ and therefore $x \in \operatorname{lev}_{\leq \overline{\alpha}} J$. Hence $x \in S \cap \operatorname{lev}_{\leq \overline{\alpha}} J$. Hence the whole sequence $(x_n)_{n\geq 0}$ converges to x by (3.33). □

Proposition 3.5 *Suppose $\overline{\alpha} \geq \alpha^*$. Given any $x_0 \in \mathsf{E}$ and $\epsilon \in]0, 1]$, let $(x_n)_{n\geq 0}$ be a sequence generated by*

$$(\forall n \in \mathbb{N})\ \ x_{n+1} = T\left(x_n + \lambda_n (G^J_{\alpha_n}(x_n) - x_n)\right), \ \textit{where} \begin{cases} \epsilon \leq \lambda_n \leq 2 - \epsilon \\ \alpha_n \geq \overline{\alpha} \end{cases} \tag{3.38}$$

under Assumption 3.3. Then, if $\alpha_n \to \overline{\alpha}$, $(x_n)_{n\geq 0}$ converges to a point in $S \cap \operatorname{lev}_{\leq \overline{\alpha}} J$.

Proof. Use (2.14) and apply Theorem 3.4. □

As seen in (3.3), the solution set of (3.1) can be written as $S^* = S \cap \operatorname{lev}_{\leq \alpha^*} J$. Then, by Theorem 3.4 and Proposition 3.5, if a close overestimate $\overline{\alpha} \geq \alpha^*$ and a nice operator $T \colon \mathsf{E} \to \mathsf{E}$ such that $\operatorname{Fix} T = S$ are available, an approximate solution of (3.1) can be obtained by recursion (3.20) and (3.38). In Chapter 4, Proposition 3.5 will be used to construct a general subgradient level method to solve our general problem (1.2) under a moderate assumption.

Chapter 4

A General Subgradient Level Method

4.1 Introduction

In (1.2) the restoration problem was posed as

$$\text{Find } x^* \in S = \bigcap_{0 \le i \le m} S_i \text{ such that } J(x^*) = \inf_{x \in S} J(x). \tag{4.1}$$

In this chapter, a general subgradient level method is proposed to solve (4.1) under Assumption 1.1 and

Assumption 4.1 S_0 is a simple compact convex set and, for every $i \in \{1, \dots, m\}$, $S_i = \operatorname{lev}_{\le 0} g_i$ where $g_i \colon \mathsf{E} \to \mathbb{R}$ is a convex function.

We recall that Assumption 1.1 guarantees that (4.1) has at least one solution.

Now let $\alpha^* = \inf_{x \in S} J(x)$. As seen in (3.3), (4.1) is equivalent to the convex feasibility problem

$$\text{Find } x^* \in S^* \triangleq \operatorname{lev}_{\leq \alpha^*} J \cap \bigcap_{0 \leq i \leq m} S_i. \tag{4.2}$$

When α^* is known, it follows from Propositions 3.5, 2.23, and 2.25, that (4.2) can be solved by the recursion

$$(\forall n \in \mathbb{N})\ x_{n+1} = P_{S_0} \circ T \circ G^J_{\alpha^*}(x_n), \tag{4.3}$$

where $T \colon \mathsf{E} \to \mathsf{E}$ is a nice operator (see Definition 2.22) with fixed point set $\bigcap_{1 \leq i \leq m} S_i$ and $G^J_{\alpha^*}(x_n)$ is the subgradient projection of x_n onto $\operatorname{lev}_{\leq \alpha^*} J$ given by (2.14). As seen in Section 2.4, a nice operator T with fixed point set $\bigcap_{1 \leq i \leq m} S_i$ can easily be constructed from (subgradient) projectors onto the individual sets $(S_i)_{1 \leq i \leq m}$. Alternative decomposition methods to solve (4.2) with approximate projectors onto the individual sets $\operatorname{lev}_{\leq \alpha^*} J$ and $(S_i)_{0 \leq i \leq m}$ can be found in [7, 13, 14, 24]. Unfortunately, α^* is usually unknown in practice and (4.3) cannot be implemented.

4.2 Description of the Main Algorithm

The idea behind the proposed algorithm is to replace (4.3) by the recursion

$$(\forall n \in \mathbb{N})\ x_{n+1} = P_{S_0} \circ T \circ G^J_{\alpha_n}(x_n), \tag{4.4}$$

where $T \colon \mathsf{E} \to \mathsf{E}$ is a nice operator with fixed point set $\bigcap_{1 \leq i \leq m} S_i$ and $(\alpha_n)_{n \geq 0} \subset \mathbb{R}$ is a sequence approaching the optimal level α^*. For the case $T = \mathrm{Id}$, i.e., the constraint set S in (4.1) is a simple constraint set S_0, the idea of replacing (4.3) by (4.4) can be found in [2, 17, 20, 23] (in [23] $G^J_{\alpha_n}$ is replaced by a projector onto a polyhedral approximation to $\operatorname{lev}_{\leq \alpha_n} J$ based on subgradients computed at the current and at previous iterations).

Theorem 4.2 *Let $T\colon \mathsf{E} \to \mathsf{E}$ be a nice operator with fixed point set $\bigcap_{1 \le i \le m} S_i$. Suppose that $(\forall n \in \mathbb{N})$ $\alpha_n \ge \overline{\alpha} \ge \alpha^*$ and $\alpha_n \to \overline{\alpha}$. Then any sequence $(x_n)_{n\ge 0}$ generated by (4.4) converges to a point in $S \cap \operatorname{lev}_{\le \overline{\alpha}} J$.*

Proof. Since P_{S_0} is nice with fixed point set S_0 and T is nice with fixed point set $\bigcap_{1 \le i \le m} S_i$, by Proposition 2.25, $P_{S_0} \circ T$ is nice with fixed point set S. Then, using Proposition 3.5 with $\lambda_n = 1$ in (3.38), we obtain the claim. □

Theorem 4.2 states that if an overestimate $\overline{\alpha}$ of the optimal constrained objective α^* is available, an approximate solution to (4.1) satisfying all the constraints and achieving at most the objective value $\overline{\alpha}$ can be found by (4.4).

Definition 4.3 A *penalty function* for (4.1) is a convex function $g\colon \mathsf{E} \to [0, +\infty[$ vanishing only on $\bigcap_{1 \le i \le m} S_i$.

Given a penalty function g for (4.1) and a real number $\beta > 0$, $S = S_0 \cap \operatorname{lev}_{\le 0} g$ and $S_0 \cap \operatorname{lev}_{\le \beta} g$ is the relaxed feasibility set at tolerance level β. By Theorem 2.1, this set is also closed and convex. We can define an approximate objective value of (4.1) as

$$\alpha^*_\beta = \inf_{x \in S_0 \cap \operatorname{lev}_{\le \beta} g} J(x). \tag{4.5}$$

Note that

$$(\forall (\beta_1, \beta_2) \in \mathbb{R}^2)\ 0 \le \beta_2 \le \beta_1 \Longrightarrow \alpha^* = \alpha^*_0 \ge \alpha^*_{\beta_2} \ge \alpha^*_{\beta_1}. \tag{4.6}$$

The main development of this Chapter is the following algorithm, where we give a precise description of the construction of the (possibly finite) sequence $(\alpha_n)_n$ in (4.4).

Algorithm 4.4 Given $v \in \mathsf{E}$, $(\epsilon, \beta) \in]0, +\infty[^2$, $\xi \in]0, 1[$, a nice operator $T\colon \mathsf{E} \to \mathsf{E}$ with fixed point set $\bigcap_{1 \le i \le m} S_i$, and a penalty function g for (4.1), a (possibly finite) sequence $(x_n)_n$ is constructed as follows.

Step 0. Set $x_0 = P_{S_0} \circ T(v)$, $\underline{\alpha}_0 < \alpha^*$, $\overline{\alpha}_0 = J(x_0)$, $\beta_0 = \max\{g(x_0), \beta\}$, and $n = 0$.

Step 1. If $\overline{\alpha}_n - \underline{\alpha}_n > \epsilon$, go to Step 5.

Step 2. If $\beta_n = \beta$, terminate.

Step 3. Do $x_n = P_{S_0} \circ T(x_n)$ while $g(x_n) > \max\{\xi\beta_n, \beta\}$.

Step 4. Set $\underline{\alpha}_{n+1} = \underline{\alpha}_n$, $\overline{\alpha}_{n+1} = J(x_n)$, $\beta_{n+1} = \max\{g(x_n), \beta\}$, $x_{n+1} = x_n$, and go to Step 10.

Step 5. Set $\alpha_n = (\underline{\alpha}_n + \overline{\alpha}_n)/2$.

Step 6. Set $x_{n+1} = P_{S_0} \circ T \circ G^J_{\alpha_n}(x_n)$.

Step 7. If $S \cap \operatorname{lev}_{\le \alpha_n} J = \emptyset$ is detected, go to Step 8; Otherwise, go to Step 9.

Step 8. Set $\underline{\alpha}_{n+1} = \alpha_n$, $\overline{\alpha}_{n+1} = \overline{\alpha}_n$, $\beta_{n+1} = \beta_n$, $x_{n+1} = x_n$, and go to Step 10.

Step 9. Set $\underline{\alpha}_{n+1} = \underline{\alpha}_n$. If $g(x_{n+1}) \le \beta_n$ and $J(x_{n+1}) < \overline{\alpha}_n$, set $\overline{\alpha}_{n+1} = J(x_{n+1})$ and $\beta_{n+1} = \max\{g(x_{n+1}), \beta\}$; Otherwise, set $\overline{\alpha}_{n+1} = \overline{\alpha}_n$ and $\beta_{n+1} = \beta_n$.

Step 10. Set $n = n + 1$ and go to Step 1.

In above algorithm, a nonincreasing auxiliary sequence $(\beta_n)_{n \ge 0}$ is generated. It is used primarily in the construction of sequence $(\overline{\alpha}_n)_{n \ge 0}$ (see Proposition 4.7(v)). Let us emphasize that

- At each iteration n, β_n specifies the relaxed feasibility set $S_0 \cap \mathrm{lev}_{\leq \beta_n} g$ at level β_n and, by (4.5), the lower bound of J on $S_0 \cap \mathrm{lev}_{\leq \beta_n} g$.

- $\alpha_J^* \leq \alpha^*$ and since $(\beta_n)_{n \geq 0}$ is nonincreasing, by (4.6), $(\alpha_{\beta_n}^*)_{n \geq 0}$ is nondecreasing.

Proposition 4.5 *Let $T \colon \mathsf{E} \to \mathsf{E}$ be a nice operator with fixed point set $\bigcap_{1 \leq i \leq m} S_i$. Given $y_0 \in \mathsf{E}$, let $(y_n)_{n \geq 0}$ be any sequence generated by $y_{n+1} = P_{S_0} \circ T(y_n)$. Then $(y_n)_{n \geq 0}$ converges to a point in $S_0 \cap \mathrm{lev}_{\leq 0} g$.*

Proof. Apply Theorem 4.2 by letting J be a constant function on E, e.g., $J \colon x \mapsto c$ and take $\overline{\alpha} = c$. □

Remark 4.6 Since g is continuous, by Proposition 4.5, Step 3 terminates in a finite number of steps and produces $x_n \in S_0 \cap \mathrm{lev}_{\leq \max\{\xi\beta_n, \beta\}} g$.

Let us now describe the algorithm in detail. At Step 0, x_0, $\underline{\alpha}_0$, $\overline{\alpha}_0$, and β_0 are initialized as

$$\begin{cases} x_0 = P_{S_0} \circ T(v) \\ \underline{\alpha}_0 < \alpha^* \\ \overline{\alpha}_0 = J(x_0) \\ \beta_0 = \max\{g(x_0), \beta\}. \end{cases} \tag{4.7}$$

If $\overline{\alpha}_0 - \underline{\alpha}_0 \leq \epsilon$ and $g(x_0) \leq \beta$, Algorithm 4.4 terminates at Step 2 and we get from (4.7) that

$$J(x_0) \leq \underline{\alpha}_0 + \epsilon < \alpha^* + \epsilon \ \text{ and } \ x_0 \in S_0 \cap \mathrm{lev}_{\leq \beta} g. \tag{4.8}$$

Otherwise, the iterations start and, at each iteration $n \in \mathbb{N}$, the update is produced by one of the following three loops.

Loop 1: Step 1 $\to$ Step 5 $\to$ Step 6 $\to$ Step 9 $\to$ Step 10 $\to$ Step 1.

Loop 2: Step 1 $\to$ Step 5 $\to$ Step 6 $\to$ Step 8 $\to$ Step 10 $\to$ Step 1.

Loop 3: Step 1 $\to$ Step 3 $\to$ Step 4 $\to$ Step 10 $\to$ Step 1.

More specifically:

- Loop 1 is executed when $\overline{\alpha}_n - \underline{\alpha}_n > \epsilon$ and $S \cap \operatorname{lev}_{\leq \alpha_n} J = \varnothing$ is not detected at Step 7;
- Loop 2 is executed when $\overline{\alpha}_n - \underline{\alpha}_n > \epsilon$ and $S \cap \operatorname{lev}_{\leq \alpha_n} J = \varnothing$ is detected at Step 7;
- Loop 3 is executed when $\overline{\alpha}_n - \underline{\alpha}_n \leq \epsilon$.

The updating operation (4.4) on x_n takes place at Step 6, which appears only in Loops 1 and 2 with

$$\alpha_n = \frac{\underline{\alpha}_n + \overline{\alpha}_n}{2} \tag{4.9}$$

set by Step 5. Since the updating of β_n takes place only at Steps 4 and 9, it is easy to see that

$$(\forall n \in \mathbb{N})\ \beta_n \geq \beta. \tag{4.10}$$

In view of Steps 9 and 8 and of (4.9), we then get

$$\overline{\alpha}_n - \underline{\alpha}_n > \epsilon \Longrightarrow \begin{cases} \underline{\alpha}_{n+1} \geq \underline{\alpha}_n \\ \overline{\alpha}_{n+1} \leq \overline{\alpha}_n \\ \beta \leq \beta_{n+1} \leq \max\{\beta_n, \beta\} = \beta_n \end{cases} \tag{4.11}$$

and, by Remark 4.6, we get from Step 4 that

$$\overline{\alpha}_n - \underline{\alpha}_n \leq \epsilon \Longrightarrow \begin{cases} \underline{\alpha}_{n+1} = \underline{\alpha}_n \\ \overline{\alpha}_{n+1} = J(x_n) \text{ with } x_n \in S_0 \cap \operatorname{lev}_{\leq \max\{\xi\beta_n, \beta\}} g \\ \beta \leq \beta_{n+1} \leq \max\{\xi\beta_n, \beta\} \leq \beta_n. \end{cases} \tag{4.12}$$

It follows from (4.11) that

$$(\forall n \in \mathbb{N})\ \overline{\alpha}_n - \underline{\alpha}_n > \epsilon \Longrightarrow \overline{\alpha}_{n+1} \leq \overline{\alpha}_n. \tag{4.13}$$

Moreover, $\overline{\alpha}_{n+1} > \overline{\alpha}_n$ may happen only for those $n \in \mathbb{N}$ such that $\overline{\alpha}_n - \underline{\alpha}_n \leq \epsilon$ since $\overline{\alpha}_{n+1} = J(x_n)$ in (4.12) and we may have $J(x_n) > \overline{\alpha}_n$. Note also that Algorithm 4.4 terminates at iteration n if and only if

$$\overline{\alpha}_n - \underline{\alpha}_n \leq \epsilon \text{ and } \beta_n = \beta. \tag{4.14}$$

Proposition 4.7 *At every iteration n we have*

(i) $x_n \in S_0$.

(ii) $\underline{\alpha}_0 \leq \underline{\alpha}_n \leq \underline{\alpha}_{n+1} < \alpha^*$.

(iii) $\beta \leq \beta_{n+1} \leq \beta_n$.

(iv) *If $\overline{\alpha}_n - \underline{\alpha}_n > \epsilon$, then α_n is defined and $\underline{\alpha}_n + \frac{\epsilon}{2} < \alpha_n < \overline{\alpha}_n - \frac{\epsilon}{2}$.*

(v) $\overline{\alpha}_n = \min\{J(x_k) \mid 0 \le k \le n \text{ and } g(x_k) \le \beta_n\}$.

Proof. (i) follows from Steps 0, 3, 4, 6 and 8. (ii): We get from Step 0 that $\underline{\alpha}_0 < \alpha^*$. Moreover, the value of $\underline{\alpha}_n$ is updated only in Loop 2 as

$$\underline{\alpha}_{n+1} = \alpha_n = \underline{\alpha}_n + \frac{\overline{\alpha}_n - \underline{\alpha}_n}{2} > \underline{\alpha}_n + \frac{\epsilon}{2} \tag{4.15}$$

when $S \cap \operatorname{lev}_{\le \alpha_n} J = \emptyset$ ($\alpha_n < \alpha^*$) is detected. From Step 0 we get $\beta_0 \ge \beta$ and, by using (4.11) and (4.12), we get (iii). (iv) follows from the fact that α_n is defined only when Step 5 is executed at iteration n, which occurs only in Loops 1 and 2. In this case it is set to

$$\alpha_n = \frac{\underline{\alpha}_n + \overline{\alpha}_n}{2} = \underline{\alpha}_n + \frac{\overline{\alpha}_n - \underline{\alpha}_n}{2} = \overline{\alpha}_n - \frac{\overline{\alpha}_n - \underline{\alpha}_n}{2}. \tag{4.16}$$

To prove (v), let us first note that $\overline{\alpha}_n$ and β_n are updated only at Steps 8, 9, and 4. When $\overline{\alpha}_n - \underline{\alpha}_n > \epsilon$, Loop 1 or 2 is executed. Note that $\overline{\alpha}_{n+1} < \overline{\alpha}_n$ can take place only at Step 9 when

$$g(x_{n+1}) \le \beta_n \text{ and } J(x_{n+1}) < \overline{\alpha}_n. \tag{4.17}$$

In connection with (4.17), note that

$$\begin{cases} g(x_{n+1}) \le \beta_n \\ J(x_{n+1}) < \overline{\alpha}_n \end{cases} \implies \begin{cases} \beta_{n+1} = \max\{g(x_{n+1}), \beta\} \le \max\{\beta_n, \beta\} = \beta_n \\ \overline{\alpha}_{n+1} = J(x_{n+1}) < \overline{\alpha}_n. \end{cases} \tag{4.18}$$

On the other hand, when $\overline{\alpha}_n - \underline{\alpha}_n > \epsilon$, we have

$$\big(g(x_{n+1}) > \beta_n \text{ or } J(x_{n+1}) \ge \overline{\alpha}_n\big) \implies \begin{cases} \beta_{n+1} = \beta_n \\ \overline{\alpha}_{n+1} = \overline{\alpha}_n. \end{cases} \tag{4.19}$$

Therefore, if $(\forall n \in \mathbb{N})\ \overline{\alpha}_n - \underline{\alpha}_n > \epsilon$, then

$$(\forall n \in \mathbb{N})\ \overline{\alpha}_n = \min\big\{J(x_k) \mid 0 \le k \le n \ \text{ and } \ g(x_k) \le \beta_n\big\}. \tag{4.20}$$

On the other hand, when $\overline{\alpha}_n - \underline{\alpha}_n \le \epsilon$, we get $\overline{\alpha}_n \le \underline{\alpha}_n + \epsilon < \alpha^* + \epsilon$. If $\beta_n = \beta$, Algorithm 4.4 terminates. Otherwise, we get $\beta_n > \beta$ and, by Remark 4.6, β_n is then updated at Step 4 as (see Steps 3 and 4)

$$\beta_{n+1} = \max\{g(x_n), \beta\} \le \max\{\xi\beta_n, \beta\} < \beta_n, \tag{4.21}$$

and $\overline{\alpha}_{n+1}$ and x_{n+1} are set at the same time as $\overline{\alpha}_{n+1} = J(x_n)$ and $x_{n+1} = x_n$. Therefore

$$\overline{\alpha}_{n+1} = J(x_{n+1}) = J(x_n) \ \text{ and } \ g(x_{n+1}) = g(x_n) = \beta_{n+1} < \beta_n. \tag{4.22}$$

Since $(\forall k \in \{0, \dots, n-1\})\ g(k) > \beta_{n+1}$, we get

$$\overline{\alpha}_{n+1} = \min\big\{J(x_k) \mid 0 \le k \le n+1 \ \text{ and } \ g(x_k) \le \beta_{n+1}\big\}. \tag{4.23}$$

□

4.3 Asymptotic Properties

In this section, ϵ and β are the tolerances fixed in Algorithm 4.4. Define

$$\mathbb{N}_\epsilon = \big\{k \in \mathbb{N} \mid \overline{\alpha}_k - \underline{\alpha}_k > \epsilon\big\} \tag{4.24}$$

and, for every $n \in \mathbb{N}$, let $\mathbb{N}_n$ be the largest interval in $\mathbb{N}$ of the form

$$\big\{k \in \mathbb{N} \mid \underline{\alpha}_k = \underline{\alpha}_n \text{ and } \overline{\alpha}_k \le \overline{\alpha}_{k-1}\big\} \tag{4.25}$$

containing n. $\mathbb{N}_\epsilon$ is the set of iteration indices at which Loop 1 or 2 is executed and $\mathbb{N}_n$ is the largest set of iteration indices containing n such that $(\underline{\alpha}_k)_{k\in\mathbb{N}_n}$ is a constant sequence and $(\overline{\alpha}_k)_{k\in\mathbb{N}_n}$ is a nonincreasing sequence.

Proposition 4.8 *Let $n \in \mathbb{N}$. If $\mathbb{N}_n \subset \mathbb{N}_\epsilon$, then $(\exists \bar{n} \in \mathbb{N}_n)\ \alpha_{\bar{n}} < \alpha^*$.*

Proof. Let $n \in \mathbb{N}$ be such that $\mathbb{N}_n \subset \mathbb{N}_\epsilon$ and suppose $(\forall k \in \mathbb{N}_n)\ \alpha_k \geq \alpha^*$. Then

$$(\forall k \in \mathbb{N}_n)\ S \cap \operatorname{lev}_{\leq \alpha_k} \neq \emptyset \text{ and therefore } \underline{\alpha}_{k+1} = \underline{\alpha}_k. \tag{4.26}$$

Since $\mathbb{N}_n \subset \mathbb{N}_\epsilon$, $(\forall k \in \mathbb{N}_n)\ \overline{\alpha}_k - \underline{\alpha}_k > \epsilon$ and, by (4.11), $(\forall k \in \mathbb{N}_n)\ \overline{\alpha}_{k+1} \leq \overline{\alpha}_k$. Therefore

$$(\forall k \in \mathbb{N}_n)\ \big(\alpha_k \geq \alpha^* \text{ and } k \in \mathbb{N}_n \subset \mathbb{N}_\epsilon\big) \Longrightarrow k+1 \in \mathbb{N}_n. \tag{4.27}$$

Hence, if $\mathbb{N}_n \subset \mathbb{N}_\epsilon$ and $(\forall k \in \mathbb{N}_n)\ \alpha_n \geq \alpha^*$, then $\mathbb{N}_n$ is an infinite interval in $\mathbb{N}$. Since $(\forall k \in \mathbb{N}_n \subset \mathbb{N}_\epsilon)\ \alpha_k = \dfrac{\underline{\alpha}_k + \overline{\alpha}_k}{2}$ and $\alpha_k \geq \alpha^*$, $(\alpha_k)_{k \in \mathbb{N}_n}$ is nonincreasing and bounded from below by α^* and therefore

$$(\exists \overline{\alpha} \in [\alpha^*, +\infty[)\ \lim_{k \to +\infty} \alpha_k = \overline{\alpha}. \tag{4.28}$$

Since for every $k \in \mathbb{N}_n$ Loop 1 is executed, it follows from Theorem 4.2 that

$$\lim_{k \to +\infty} x_k = x \in S \cap \operatorname{lev}_{\leq \overline{\alpha}} J. \tag{4.29}$$

However, since g and J are continuous and $S = S_0 \cap \operatorname{lev}_{\leq 0} g$, we get from (4.29) that

$$\lim_{k \to +\infty} g(x_k) = g(x) \leq 0 \text{ and } \lim_{k \to +\infty} J(x_k) = J(x) \leq \overline{\alpha}. \tag{4.30}$$

Since by Proposition 4.7(v) $\overline{\alpha}_k = \min\{J(x_l) \mid 0 \leq l \leq k \text{ and } g(x_l) \leq \beta_k\}$, where $\beta_k \geq \beta > 0$, we get

$$(\exists K \in \mathbb{N})(\forall k \in \mathbb{N}_n, k \geq K)\ \overline{\alpha}_k = \min\{J(x_l) \mid K \leq l \leq k\} \leq J(x_k). \tag{4.31}$$

Then, by Proposition 4.7(iv),

$$(\forall k \in \mathbb{N}_n, k \geq K)\ \alpha_k < \overline{\alpha}_k - \frac{\epsilon}{2} \leq J(x_k) - \frac{\epsilon}{2}. \tag{4.32}$$

Thus, by (4.28) and (4.30), we get $\overline{\alpha} \leq \overline{\alpha} - \frac{\epsilon}{2}$, a contradiction. Hence $(\exists \tilde{n} \in \mathbb{N}_n)\ \alpha_{\tilde{n}} < \alpha^*$.

□

Proposition 4.9 *For every $n \in \mathbb{N}$ such that $\mathbb{N}_n \subset \mathbb{N}_\epsilon$, if $S \cap \operatorname{lev}_{\leq \alpha_{\tilde{n}}} J = \emptyset$ is detected for some $\tilde{n} \in \mathbb{N}_n$, then $(\exists\, l \in \mathbb{N})\ x_l \in S_0 \cap \operatorname{lev}_{\leq \beta} g$ and $J(x_l) < \alpha^* + \epsilon$.*

Proof. For every $n \in \mathbb{N}$ such that $\mathbb{N}_n \subset \mathbb{N}_\epsilon$, by Proposition 4.8, there exists $\tilde{n} \in \mathbb{N}_n$ such that $S \cap \operatorname{lev}_{\leq \alpha_{\tilde{n}}} J = \emptyset$. Now assume that for every $n \in \mathbb{N}$ such that $\mathbb{N}_n \subset \mathbb{N}_\epsilon$,

$$(\exists \tilde{n} \in \mathbb{N}_n)\ S \cap \operatorname{lev}_{\leq \tilde{n}} J = \emptyset \text{ is detected.} \tag{4.33}$$

Let $q \in \mathbb{N}$ and suppose $(\forall n \in \mathbb{N}, n \geq q)\ \overline{\alpha}_n - \underline{\alpha}_n > \epsilon$. Then Loop 2 will be executed an infinite number of times and, by Steps 8 and 5, the update

$$\underline{\alpha}_{n+1} = \alpha_n = \underline{\alpha}_n + \frac{\overline{\alpha}_n - \underline{\alpha}_n}{2} > \underline{\alpha}_n + \frac{\epsilon}{2} \tag{4.34}$$

will be executed an infinite number of times. As a result $\underline{\alpha}_n$ will become arbitrarily large as n increases, which contradicts Proposition 4.7(ii). Thus, for any $n \in \mathbb{N}$,

$$\text{if } \overline{\alpha}_n - \underline{\alpha}_n > \epsilon \text{ then } (\exists p \in \mathbb{N}, p > n)\ \overline{\alpha}_p - \underline{\alpha}_p \leq \epsilon. \tag{4.35}$$

Now let $n \in \mathbb{N}$ be such that $\overline{\alpha}_n - \underline{\alpha}_n \leq \epsilon$. Then $\overline{\alpha}_n \leq \underline{\alpha}_n + \epsilon < \alpha^* + \epsilon$. If $\beta_n = \beta$, Algorithm 4.4 terminates and, since by Proposition 4.7(v)

$$\overline{\alpha}_n = \min\{J(x_k) \mid 0 \leq k \leq n \text{ and } g(x_k) \leq \beta_n\}, \tag{4.36}$$

we get

$$(\exists\, l \in \{1, \ldots, n\})\ x_l \in S_0 \cap \operatorname{lev}_{\leq \beta} g \text{ and } J(x_l) < \alpha^* + \epsilon. \tag{4.37}$$

Otherwise ($\beta_n > \beta$), by construction, Loop 3 is executed. Therefore, x_n, $\overline{\alpha}_n$, $\underline{\alpha}_n$ and β_n are updated as

$$\begin{cases} x_{n+1} \in S_0 \cap \operatorname{lev}_{\leq \max\{\xi\beta_n, \beta\}} g \\ \overline{\alpha}_{n+1} = J(x_{n+1}) \\ \underline{\alpha}_{n+1} = \underline{\alpha}_n \\ \beta_{n+1} = \max\{\xi\beta_n, \beta\} < \beta_n \end{cases} \tag{4.38}$$

and we go back to Step 1. Since Loop 3 is executed only when $\overline{\alpha}_n - \underline{\alpha}_n \leq \epsilon$ and produces $\beta_{n+1} < \beta_n$, it follows from (4.35) that, by iterating this argument, we get

$$(\exists l \in \mathbb{N})\ \overline{\alpha}_l - \underline{\alpha}_l \leq \epsilon \text{ and } \beta_l = \beta. \tag{4.39}$$

Consequently, by Proposition 4.7(ii)(v), we get

$$(\exists l \in \mathbb{N})\ J(x_l) \leq \alpha^\star + \epsilon \text{ and } x_l \in S_0 \cap \operatorname{lev}_{\leq \beta} g. \tag{4.40}$$

□

We now present the main convergence result.

Theorem 4.10 *Suppose that infeasibility can be detected at Step 7 and let $(x_n)_n$ be an arbitrary sequence generated by Algorithm 4.4. Then $(\exists n \in \mathbb{N})$ $x_n \in S_0 \cap \operatorname{lev}_{\leq \beta} g$ and $J(x_n) < \alpha^\star + \epsilon$.*

Proof. Apply Proposition 4.9. □

Theorem 4.10 states that for any nice operator $T \colon \mathsf{E} \to \mathsf{E}$ with fixed point set $\bigcap_{1 \leq i \leq m} S_i$, if infeasibility detection at Step 7 can be implemented, Algorithm 4.4 produces a signal that achieves any preset tolerance values on the constrained optimal value of the objective and on the joint enforcement of the constraints. In other words, if infeasibility detection can be achieved in a finite number of steps, Algorithm 4.4 produces an approximate solution to our basic problem 1.2 whose accuracy can be made arbitrarily good.

4.4 Implementation of the Main Algorithm

To implement Algorithm 4.4 it is necessary to:

1. Select a nice operator $T\colon \mathsf{E} \to \mathsf{E}$ such that $\operatorname{Fix} T = \bigcap_{1\leq i\leq m} S_i$.

2. Select a penalty function $g\colon \mathsf{E} \to [0, +\infty[$ for (4.1).

3. Devise a scheme to detect infeasibility in a finite number of steps at Step 7.

4.4.1 Construction of the Penalty Function g

Proposition 4.11 *Let* $(w_i)_{1\leq i\leq m} \subset]0, +\infty[$. *Then*

$$\sum_{1\leq i\leq m} w_i {g_i}^+ \quad \textit{and} \quad \max_{1\leq i\leq m} g_i^+ \tag{4.41}$$

are penalty functions for (4.1).

Proof. Since $(g_i)_{1\leq i\leq m}\colon \mathsf{E} \to \mathbb{R}$ are convex, by Theorems 2.6 and 2.7, $g = \sum_{1\leq i\leq m} w_i {g_i}^+$ and $g = \max_{1\leq i\leq m} g_i^+$ are convex. On the other hand, for either g

$$x \in \operatorname{lev}_{\leq 0} g \Leftrightarrow [(\forall i \in \{1, \dots, m\})\ g_i(x) \leq 0] \Leftrightarrow x \in \bigcap_{1\leq i\leq m} S_i. \tag{4.42}$$

□

Algorithm 4.4 terminates with a signal $x_n \in S_0 \cap \operatorname{lev}_{\leq \beta} g \cap \operatorname{lev}_{\leq \alpha^* + \epsilon} J$. Therefore, if $g = \sum_{1 \leq i \leq m} w_i g_i{}^+$, then for every $i \in \{1, \ldots, m\}$ $g_i(x_n) \leq \frac{\beta}{w_i}$. On the other hand, if g is given by $g = \max_{1 \leq i \leq m} g_i^+$, then $\max_{1 \leq i \leq m} g_i(x_n) \leq \beta$.

4.4.2 Construction of the Nice Operator T

Proposition 4.12 *Let $T = P_{S_1} \circ \cdots \circ P_{S_m}$. Then T is nice and* Fix $T = \bigcap_{1 \leq i \leq m} S_i$.

Proof. By Proposition 2.23, for every $i \in \{1, \ldots, m\}$, P_{S_i} is nice and Fix $P_{S_i} = S_i$. Since $\bigcap_{1 \leq i \leq m} S_i = S \neq \emptyset$, by Proposition 2.25, T is nice and Fix $T = \bigcap_{1 \leq i \leq m} S_i$. □

Proposition 4.13 *Let $T = G_0^{g_1} \circ G_0^{g_2} \circ \cdots \circ G_0^{g_m}$. Then T is nice and* Fix $T = \bigcap_{1 \leq i \leq m} S_i$.

Proof. By Proposition 2.24, for each $i \in \{1, \ldots, m\}$, $G_0^{g_i}$ is nice and Fix $G_0^{g_i} = \operatorname{lev}_{\leq 0} g_i$. Since $\bigcap_{1 \leq i \leq m} \operatorname{lev}_{\leq 0} g_i = \bigcap_{1 \leq i \leq m} S_i \neq \emptyset$, by Proposition 2.25, T is nice and Fix $T = \bigcap_{1 \leq i \leq m} S_i$.
□

Proposition 4.14 *Let $T = G_0^g$, where $g \colon \mathsf{E} \to]0, +\infty[$ is a penalty function for* (4.1). *Then T is nice and* Fix $T = \bigcap_{1 \leq i \leq m} S_i$.

Proof. Since g is a penalty function for (4.1), $\operatorname{lev}_{\leq 0} g = \bigcap_{1 \leq i \leq m} S_i$. Moreover, by Proposition 2.24, G_0^g is nice and Fix $G_0^g = \operatorname{lev}_{\leq 0} g = \bigcap_{1 \leq i \leq m} S_i$. □

If the penalty function $g = \max_{1 \leq i \leq m} g_i^+$ is used, the computation of $G_0^g(x_n)$ in Proposition 4.14 can be parallelized when computing the independent values $(g_i(x_n))_{1 \leq i \leq m}$. Indeed, to get $G_0^g(x_n)$ we need to compute each $g_i(x_n)$ so as to obtain $g(x_n)$ and then a subgradient $t \in \partial g(x_n)$, which can be obtained by Theorem 2.7.

If the penalty function $g = \sum_{1 \leq i \leq m} w_i {g_i}^+$, where $(w_i)_{1 \leq i \leq m} \subset]0, +\infty[$, is used, the computation of $G_0^g(x_n)$ in Proposition 4.14 can be parallelized when computing $(g_i(x_n))_{1 \leq i \leq m}$ and a subgradient $t \in \partial g(x_n)$. Indeed, by Theorems 2.6 and 2.7,

$$\left(\sum_{i \in I(x_n)} w_i t_i \right) \in \partial g(x_n), \tag{4.43}$$

where $I(x_n) = \{i \in \{1, \ldots, m\} \mid g_i(x_n) > 0\}$ and $t_i \in \partial g_i(x_n)$. Therefore, $(t_i)_{i \in I(x_n)}$ can be computed in parallel to obtain a subgradient in (4.43) and $(g_i^+(x_n))_{1 \leq i \leq m}$ can be computed in parallel to obtain $g(x_n)$.

4.4.3 Implementation of Infeasibility Detection at Step 7

At any iteration n at which $\varnothing \neq \mathbb{N}_n \subset \mathbb{N}_\epsilon$, we have, by Proposition 4.8,

$$(\exists \bar{n} \in \mathbb{N}_n) \; S \cap \operatorname{lev}_{\leq \alpha_{\bar{n}}} J = \varnothing. \tag{4.44}$$

There is no general scheme to identify such a $\bar{n}$ at which infeasibility occurs directly unless the sets $(S_i)_{0 \leq i \leq m}$ and $\operatorname{lev}_{\leq \alpha_n} J$ are extremely simple. However, if either an upper bound on $\operatorname{diam}(S_0)$ or on $(d(x_k, S^\bullet))_{k \in \mathbb{N}_n}$ is known. the following results provide a sufficient condition for identifying $\bar{n}$ in (4.44).

4.4.3.1 General Analysis

Assumption 4.15 For every $i \in \{1, \ldots, m\}$, $S_i = \operatorname{Fix} T_i$ and $T_i \colon \mathsf{E} \to \mathsf{E}$ is a 1-attracting nice operator.

Now consider the recursion (4.4) with $T = T_1 \circ \cdots \circ T_m$, or equivalently

$$\begin{cases} x_{m+1,k} = G^J_{\alpha_k}(x_k) \\ x_{m,k} = T_m(x_{m+1,k}) \\ x_{m-1,k} = T_{m-1}(x_{m,k}) \\ \qquad \vdots \\ x_{1,k} = T_1(x_{2,k}) \\ x_{k+1} = P_{S_0}(x_{1,k}). \end{cases} \tag{4.45}$$

Proposition 4.16 *For an arbitrary $x_0 \in \mathsf{E}$, define at every iteration k*

$$\begin{cases} \rho_{m+1,k} = \|G^J_{\alpha_k}(x_k) - x_k\|^2 \\ (\forall i \in \{1, \cdots, m\})\ \rho_{i,k} = \|T_i(x_{i+1,k}) - x_{i+1,k}\|^2 \\ \rho_{0,k} = \|P_{S_0}(x_{1,k}) - x_{1,k}\|^2 \\ \rho_k = \sum_{i=0}^{m+1} \rho_{i,k}. \end{cases} \tag{4.46}$$

If for some $q \in \mathbb{N}$ and $x^* \in S^*$

$$\sum_{k=0}^{q} \rho_k > \|x_0 - x^*\|^2 - \|x_{q+1} - x^*\|^2, \tag{4.47}$$

then

$$(\exists \bar{n} \in \{0, \ldots, q\})\ S \cap \operatorname{lev}_{\leq \alpha_{\bar{n}}} J = \varnothing. \tag{4.48}$$

Proof. For an arbitrary $k \in \mathbb{N}$ suppose $S \cap \mathrm{lev}_{\leq \alpha_k} J \neq \emptyset$ and take an arbitrary point $x^* \in S^*$. Then $x^* \in S_0 \cap \left(\bigcap_{1 \leq i \leq m} \mathrm{Fix}\ T_i \right) \cap \mathrm{lev}_{\leq \alpha_k} J$. Since $\mathrm{Fix}\ P_{S_0} = S_0$ and $\mathrm{Fix}\ G^J_{\alpha_k} = \mathrm{lev}_{\leq \alpha_k} J$

$$x^* \in \mathrm{Fix}\ P_{S_0} \cap \left(\bigcap_{1 \leq i \leq m} \mathrm{Fix}\ T_i \right) \cap \mathrm{Fix}\ G^J_{\alpha_k}. \tag{4.49}$$

Since P_{S_0}, $G^J_{\alpha_k}$, and $(T_i)_{1 \leq i \leq m}$ are 1-attracting by Assumption 4.15 we then get

$$\begin{cases} \|x_k - x^*\|^2 - \|x_{m+1,k} - x^*\|^2 \geq \rho_{m+1,k} \\ (\forall i \in \{1, \ldots, m\})\ \|x_{i+1,k} - x^*\|^2 - \|x_{i,k} - x^*\|^2 \geq \rho_{i,k} \\ \|x_{1,k} - x^*\|^2 - \|x_{k+1} - x^*\|^2 \geq \rho_{0,k} \end{cases} \tag{4.50}$$

and consequently

$$S \cap \mathrm{lev}_{\leq \alpha_k} J \neq \emptyset \Longrightarrow \|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2 \geq \rho_k. \tag{4.51}$$

Now suppose that, for every $k \in \{0, \ldots, q\}$, $S \cap \mathrm{lev}_{\leq \alpha_k} J \neq \emptyset$. Then (4.51) yields

$$\|x_0 - x^*\|^2 - \|x_{q+1} - x^*\|^2 \geq \sum_{k=0}^{q} \rho_k. \tag{4.52}$$

Thus for any $q \in \mathbb{N}$, if $\|x_0 - x^*\|^2 - \|x_{q+1} - x^*\|^2 < \sum_{k=0}^{q} \rho_k$ then there exists $\bar{n} \in \{0, \ldots, q\}$ such that $\mathrm{lev}_{\leq \alpha_{\bar{n}}} J \cap S = \emptyset$. □

4.4.3.2 Special Cases

As seen in Section 4.4.2, the general form of a nice operator is

$$T = T_1 \circ \cdots \circ T_m. \tag{4.53}$$

For such operators, infeasibility detection at Step 7 can be implemented as follows.

Proposition 4.17 *Set* $\kappa \geq \operatorname{diam}(S_0)$. *At each iteration* n *such that* $\overline{\alpha}_n - \underline{\alpha}_n > \epsilon$ *define for every* $k \in \mathbb{N}_n$ ρ_k *as in* (4.46). *Let* l *be the smallest integer in* $\mathbb{N}_n$. *Then, for every* $\gamma \geq d(x_l, S^*)$,

$$\text{if } \sum_{k \in \mathbb{N}_n} \rho_k > \min\left\{\kappa^2, \gamma^2, 2\gamma\|x_l - x_{n+1}\| - \|x_l - x_{n+1}\|^2\right\} \text{ then } S \cap \operatorname{lev}_{\leq \alpha_n} J = \emptyset. \tag{4.54}$$

Proof. Take an arbitrary $x^* \in S^*$ and n such that $\overline{\alpha}_n - \underline{\alpha}_n > \epsilon$. Apply Proposition 4.16, with $x_0 = x_l$, we get

$$\sum_{k \in \mathbb{N}_n} \rho_k > \|x_l - x^*\|^2 - \|x_{n+1} - x^*\|^2 \Longrightarrow (\exists \tilde{n} \in \mathbb{N}_n) \operatorname{lev}_{\leq \alpha_{\tilde{n}}} J \cap S = \emptyset. \tag{4.55}$$

Since $(\alpha_k)_{k \in \mathbb{N}_n}$ is nonincreasing, we get then

$$\sum_{k \in \mathbb{N}_n} \rho_k > \|x_l - x^*\|^2 - \|x_{n+1} - x^*\|^2 \Longrightarrow \operatorname{lev}_{\leq \alpha_n} J \cap S = \emptyset. \tag{4.56}$$

Now let $x^* = P_{S^*}(x_l)$. Then the inequalities

$$\|x_l - x^*\|^2 - \|x_{n+1} - x^*\|^2 \leq \|x_l - x^*\|^2 \leq (\operatorname{diam}(S_0))^2 \leq \kappa^2, \tag{4.57}$$

$$\|x_l - x^*\|^2 = d^2(x_l, S^*) \leq \gamma^2, \tag{4.58}$$

and

$$\begin{aligned} \|x_l - x^*\|^2 - \|x_{n+1} - x^*\|^2 &= 2\langle x_l - x^* \mid x_l - x_{n+1}\rangle - \|x_l - x_{n+1}\|^2 \\ &\leq 2\|x_l - x^*\| \cdot \|x_l - x_{n+1}\| - \|x_l - x_{n+1}\|^2 \\ &\leq 2\gamma\|x_l - x_{n+1}\| - \|x_l - x_{n+1}\|^2 \end{aligned} \tag{4.59}$$

give the result. □

Although the condition stated in Proposition 4.17 is only sufficient, it can be used to detect infeasibility in (4.44) by virtue of the following proposition.

Proposition 4.18 *Let $n \in \mathbb{N}$. If $\mathbb{N}_n \subset \mathbb{N}_\epsilon$ and $\mathbb{N}_n$ is infinite, then* $\sum_{k \in \mathbb{N}_n} \rho_k = +\infty$.

Proof. Suppose that, for some $n \in \mathbb{N}$, $\mathbb{N}_n \subset \mathbb{N}_\epsilon$, $\mathbb{N}_n$ is infinite, and $\sum_{k \in \mathbb{N}_n} \rho_k < +\infty$. By Proposition 4.8 $(\exists \bar{n} \in \mathbb{N}_n)\ \alpha_{\bar{n}} < \alpha^*$. Since $(\alpha_k)_{k \in \mathbb{N}_n}$ is nonincreasing and, by Proposition 4.7(ii)(iv), bounded from below by $\underline{\alpha}_0$, we get $\alpha_k \downarrow \alpha < \alpha^*$. It follows from $\sum_{k \in \mathbb{N}_n} \rho_k < +\infty$ that $\rho_k \to 0$. Therefore, by (4.46),

$$\begin{cases} \rho_{m+1,k} = \|G^J_{\alpha_k}(x_k) - x_k\|^2 = \|x_{m+1,k} - x_k\|^2 \to 0 \\ (\forall i \in \{1, \dots, m\})\ \rho_{i,k} = \|T_i(x_{i+1,k}) - x_{i+1,k}\|^2 = \|x_{i,k} - x_{i+1,k}\|^2 \to 0 \\ \rho_{0,k} = \|P_{S_0}(x_{1,k}) - x_{1,k}\|^2 = \|x_{k+1} - x_{1,k}\|^2 \to 0. \end{cases} \tag{4.60}$$

Since by Proposition 4.7(i) $(x_k)_{k \in \mathbb{N}_n} \subset S_0$, and S_0 is a compact set, there exists a subsequence $(x_{k_l})_{l \geq 0}$ of $(x_k)_{k \in \mathbb{N}_n}$ such that $x_{k_l} \to x \in S_0$. It follows from (4.60) that

$$(\forall i \in \{1, \dots, m\})\ x_{i+1,k_l} \to x. \tag{4.61}$$

Since $(T_i)_{1 \leq i \leq m}$ are nice operators, (4.60) yields

$$(\forall i \in \{1, \dots, m\})\ x \in \operatorname{Fix} T_i \text{ and therefore } x \in \bigcap_{1 \leq i \leq m} \operatorname{Fix} T_i. \tag{4.62}$$

Since, for each $i \in \{1, \dots, m\}$ $\operatorname{Fix} T_i = S_i$, $x \in \bigcap_{0 \leq i \leq m} S_i$ and therefore $x \in S$. On the other hand, since $\alpha_k \downarrow \alpha$, $x_{k_l} \to x$ and $\|G^J_{\alpha_k}(x_{k_l}) - x_{k_l}\| \to 0$, $x \in \operatorname{lev}_{\leq \alpha} J$ by Proposition 2.17. However, since $\alpha < \alpha^*$, $S \cap \operatorname{lev}_{\leq \alpha} J = \emptyset$. We reach a contradiction. □

By Proposition 4.18 if infeasibility detection at Step 7 is implemented by Proposition 4.17 then every $\mathbb{N}_n \subset \mathbb{N}_\epsilon$ is finite and therefore infeasibility (4.44) can be detected. Moreover, in view of Proposition 4.17, infeasibility (4.44) will be detected rapidly if a tight bound γ is available.

4.5 Variants of the Main Algorithm

Algorithm 4.4 is a specific implementation of the level set method (see Section 1.4.3) for adaptive estimation of the constrained optimal objective. Several variants can be devised. Since these variants can be derived directly from Algorithm 4.4, in this section, we simply describe a few of them without proof. The chief purpose here is merely to shed more light on the proposed level set method for solving (4.1).

4.5.1 Variant I

At Step 5 of Algorithm 4.4, instead of taking $\alpha_n = (\underline{\alpha}_n + \overline{\alpha}_n)/2$, we can take more generally

$$\alpha_n = \underline{\alpha}_n + \lambda(\overline{\alpha}_n - \underline{\alpha}_n), \quad \text{where} \quad \lambda \in]0, 1[. \tag{4.63}$$

This added flexibility may help may improve the convergence behavior of the algorithm in some problems.

4.5.2 Variant II

Given a penalty function g for (4.1) and a preset tolerance $\beta > 0$, if a point $v \in S_0 \cap \operatorname{lev}_{\leq \beta} g$ is available then, it can easily be checked that, for every $n \in \mathbb{N}$, Algorithm 4.4 produces $\beta_n = \beta$. In this case, Algorithm 4.4 takes the following form.

Algorithm 4.19 Given $\epsilon \in]0, +\infty[$ and $v \in S_0 \cap \operatorname{lev}_{\leq \beta} g$, a sequence $(x_n)_n$ is constructed as follows.

Step 0. Set $x_0 = v$, $\underline{\alpha}_0 < \alpha^*$, $\overline{\alpha}_0 = J(x_0)$, and $n = 0$.

Step 1. If $\overline{\alpha}_n - \underline{\alpha}_n \leq \epsilon$ terminate.

Step 2. Set $\alpha_n = (\underline{\alpha}_n + \overline{\alpha}_n)/2$.

Step 3. Set $x_{n+1} = P_{S_0} \circ T \circ G^J_{\alpha_n}(x_n)$.

Step 4. If $S \cap \operatorname{lev}_{\leq \alpha_n} J = \emptyset$ is detected, go to Step 5; Otherwise, go to Step 6.

Step 5. Set $\underline{\alpha}_{n+1} = \alpha_n$, $\overline{\alpha}_{n+1} = \overline{\alpha}_n$, $x_{n+1} = x_n$, and go to Step 7.

Step 6. Set $\underline{\alpha}_{n+1} = \underline{\alpha}_n$. If $g(x_{n+1}) \leq \beta$ and $J(x_{n+1}) < \overline{\alpha}_n$, set $\overline{\alpha}_{n+1} = J(x_{n+1})$; Otherwise, set $\overline{\alpha}_{n+1} = \overline{\alpha}_n$.

Step 7. Set $n = n + 1$ and go to Step 1.

Remark 4.20 A point $v \in S_0 \cap \operatorname{lev}_{\leq \beta} g$ can be obtained by iterating a finite number of times the recursion $x_{n+1} = P_{S_0} \circ T(x_n)$ (see Proposition 4.5).

4.5.3 Variant III

At Step 5 of Algorithm 4.4, instead of taking $\alpha_n = (\underline{\alpha}_n + \overline{\alpha}_n)/2$, we can take

$$\alpha_n = \overline{\alpha}_n - d_n, \quad \text{where } d_n > 0. \tag{4.64}$$

With this approach, one does not need to construct the lower levels $(\underline{\alpha}_n)_n$ and only the last infeasibility detection needs to be strictly implemented (see [27] for further discussion). This leads to the following algorithm.

Algorithm 4.21 Given $v \in \mathsf{E}$, $(\epsilon, \beta) \in]0, +\infty[^2$, $(\xi, \lambda) \in]0, 1[^2$, and a penalty function g for (4.1), a sequence $(x_n)_n$ is constructed as follows.

Step 0. Set $x_0 = P_{S_0}(v)$, $\overline{\alpha}_0 = J(x_0)$, $\beta_0 = \max\{g(x_0), \beta\}$, $d_0 > \epsilon$, and $n = 0$.

Step 1. If $d_n > \epsilon$, go to Step 5.

Step 2. If $\beta_n = \beta$, terminate.

Step 3. Do $x_n = P_{S_0} \circ T(x_n)$ while $g(x_n) > \max\{\xi\beta_n, \beta\}$.

Step 4. Set $d_{n+1} > \epsilon$, $\overline{\alpha}_{n+1} = J(x_n)$, $\beta_{n+1} = \max\{g(x_n), \beta\}$, $x_{n+1} = x_n$, and go to Step 10.

Step 5. Set $\alpha_n = \overline{\alpha}_n - d_n$.

Step 6. Set $x_{n+1} = P_{S_0} \circ T \circ G^J_{\alpha_n}(x_n)$.

Step 7. If $S \cap \mathrm{lev}_{\leq \alpha_n} J = \emptyset$ is detected, go to Step 8; Otherwise, go to Step 9.

Step 8. Set $d_{n+1} = \lambda d_n$, $\overline{\alpha}_{n+1} = \overline{\alpha}_n$, $\beta_{n+1} = \beta_n$, $x_{n+1} = x_n$, and go to Step 10.

Step 9. Set $d_{n+1} = d_n$. If $g(x_{n+1}) \leq \beta_n$ and $J(x_{n+1}) < \overline{\alpha}_n$, set $\overline{\alpha}_{n+1} = J(x_{n+1})$ and $\beta_{n+1} = \max\{g(x_{n+1}), \beta\}$; Otherwise, set $\overline{\alpha}_{n+1} = \overline{\alpha}_n$ and $\beta_{n+1} = \beta_n$.

Step 10. Set $n = n + 1$ and go to Step 1.

When a reasonable $\underline{\alpha}_n < \alpha^*$ is hard to estimate, Algorithm 4.21 may be preferred over Algorithm 4.4.

Chapter 5

Application to Constrained Minimum Total Variation Image Restoration and Denoising

5.1 Introduction

A two-dimensional image can be represented discretely by an array of pixels. For a gray-level image each pixel is represented by a number which indicates the intensity of the image at that relative position. In this chapter, the images are square images having M^2 ($M = 128$) pixels and the range of the pixel values is [0,255]. The signal space is the matrix space $\mathsf{E} = \mathbb{R}^{M\times M}$ endowed with the Frobenius norm $\|\cdot\|_F$. For simplicity, we use $\|x\|$ to denote the Frobenius norm of $x \in \mathbb{R}^{M\times M}$ and $|\cdot|_2$ to denote the 2-norm in $\mathbb{R}^2$.

We consider image restoration problems in which the image degradation model (1.1) is available and is of the form

$$y = Lx + u, \tag{5.1}$$

where x, y, and u are elements of E representing respectively the original image, the recorded image, and the additive noise, and where $L\colon \mathsf{E} \to \mathsf{E}$ is a linear operator. We conduct two experiments. In experiment I, the components of u are normally distributed i.i.d zero mean random variables. In experiment II, the components of u are uniformly distributed i.i.d zero mean random variables. In each experiment, using available *a priori* knowledge, the image restoration problem will be formulated as a constrained total variation minimization problem of the form (1.2) and solved by Algorithm 4.4. The issue of selecting the nice operator T and the penalty function g will also be addressed.

5.2 Total Variation

Denote by $\mathcal{C}^1(\Omega)$ the space of real-valued continuously differentiable functions on an open set $\Omega \subset \mathbb{R}^2$ and by $\mathcal{C}_0^1(\Omega)$ the subset of $\mathcal{C}^1(\Omega)$ consisting of the functions in $\mathcal{C}^1(\Omega)$ having compact support. The total variation of a real-valued function $\mathrm{x}\colon \Omega \mapsto \mathbb{R}$ is [16]

$$J_{\mathrm{tv}}(\mathrm{x}) = \sup\left\{\int_\Omega \mathrm{x}(\omega)\,\mathrm{div}\,\varphi(\omega)d\omega \mid \varphi \in \mathcal{C}_0^1(\Omega) \text{ and } (\forall\omega\in\Omega)\ |\varphi(\omega)| \leq 1\right\}, \tag{5.2}$$

where div is the divergence operator. If $\mathrm{x} \in \mathcal{C}^1(\Omega)$, one can show using integration by parts that [16]

$$J_{\mathrm{tv}}(\mathrm{x}) = \int_\Omega |\nabla\mathrm{x}(\omega)|_2 d\omega. \tag{5.3}$$

The cost (5.3) has been proposed in [38] as an optimality criterion for signal denoising and then used as an optimality criterion for signal restoration (see [9, 47] and the references therein). It is in the form of (1.4) with ϕ given by (1.6) ($p = 2$) and is nondifferentiable if x is a constant function on Ω. In general, (5.3) will be interpreted in a distributional sense.

For a compactly supported two-dimensional image $\mathrm{x} \in C^1(\Omega)$, we can use the discrete approximations

$$\begin{cases} \mathrm{x}(\omega) \to x^{i,j} \\ |\nabla \mathrm{x}(\omega)|_2 \to \sqrt{|x^{i+1,j} - x^{i,j}|^2 + |x^{i,j+1} - x^{i,j}|^2} \\ \int \to \sum. \end{cases} \tag{5.4}$$

Here, $x^{i,j}$ denotes the (i,j)th component of $x \in \mathsf{E}$. Then, considering boundary effects, the total variation of a discrete image $x \in \mathsf{E}$ will be defined as

$$\begin{aligned} J_{\mathrm{tv}}(x) &= \sum_{i=0}^{M-2}\sum_{j=0}^{M-2} \sqrt{|x^{i+1,j} - x^{i,j}|^2 + |x^{i,j+1} - x^{i,j}|^2} \\ &\quad + \sum_{i=0}^{M-2} \sqrt{|x^{i+1,M-1} - x^{i,M-1}|^2} + \sum_{j=0}^{M-2} \sqrt{|x^{M-1,j+1} - x^{M-1,j}|^2}. \end{aligned} \tag{5.5}$$

Proposition 5.1 *$J_{\mathrm{tv}}\colon \mathsf{E} \to \mathbb{R}$ is convex and nondifferentiable.*

Proof. Since each term in the summations is a convex function on E, i.e.,

$$\begin{cases} \sqrt{|x^{i+1,j} - x^{i,j}|^2 + |x^{i,j+1} - x^{i,j}|^2} = \|M_{i,j}x\| \\ \sqrt{|x^{i+1,M-1} - x^{i,M-1}|^2} = \|M_i x\| \\ \sqrt{|x^{M-1,j+1} - x^{M-1,j}|^2} = \|M_j x\| \end{cases} \tag{5.6}$$

and $M_{i,j}\colon x \mapsto (x^{i+1,j} - x^{i,j}, x^{i,j+1} - x^{i,j})$, $M_i\colon x \mapsto x^{i+1,M-1} - x^{i,M-1}$, and $M_j\colon x \mapsto x^{M-1,j+1} - x^{M-1,j}$ are linear operator on E, by Theorem 2.6, J_{tv} is convex on E. It follows from (5.5) that, for any $0 \le i \le M-2$ and $0 \le j \le M-2$, $\partial J_{\mathrm{tv}}(x)/\partial x^{i+1,j}$, $\partial J_{\mathrm{tv}}(x)/\partial x^{i,j}$ and $\partial J_{\mathrm{tv}}(x)/\partial x^{i,j+1}$ do not exist if $x^{i+1,j} = x^{i,j} = x^{i,j+1}$; for any $0 \le i \le M-2$,

$\partial J_{\text{tv}}(x)/\partial x^{i+1,M-1}$ and $\partial J_{\text{tv}}(x)/\partial x^{i,M-1}$ do not exist if $x^{i+1,M-1} = x^{i,M-1}$; and, for any $0 \leq j \leq M-2$, $\partial J_{\text{tv}}(x)/\partial x^{M-1,j+1}$ and $\partial J_{\text{tv}}(x)/\partial x^{M-1,j}$ do not exist if $x^{M-1,j+1} = x^{M-1,j}$. Thus J_{tv} is nondifferentiable. □

The motivation for minimizing the total variation cost (5.3) or its variants [1, 9, 29, 48] in signal denoising and restoration lies in that it does not penalize discontinuities, i.e., it has no particular bias toward a discontinuous or continuous solution. Further, minimizing (5.3) tends to preserve the location of the discontinuities of the original signal and is appropriate for signals which have block features [9, 38, 48].

5.3 Experiment I

5.3.1 Image Restoration

5.3.1.1 Description

In the following example, the degraded image y shown in Fig. 5.3 is obtained by convolving the original image x shown in Fig. 5.1 with the point spread function shown in Fig. 5.2 and adding a random noise image u whose components are i.i.d zero mean Gaussian random variables. Using these statistical hypotheses on the components of u, we can define the constraint set [12, 46]

$$S_0 = \left\{x \in \mathsf{E} \mid \|Lx - y\|^2 \leq \delta\right\}. \tag{5.7}$$

Since x is an image, it is nonnegative and we can therefore define the constraint set

$$S_1 = [0, +\infty[^{M \times M}. \tag{5.8}$$

Next, knowing *a priori* that the original image has block features, the total variation cost (5.5) is chosen as the criterion to be minimized. One is thus led to the constrained nondifferentiable convex optimization problem

$$\text{Find } x^* \in S = S_0 \cap S_1 \text{ such that } J_{tv}(x^*) = \inf_{x \in S} J_{tv}(x). \tag{5.9}$$

Problem (5.9) is solved by Algorithm 4.4. It should be pointed out that, with the nondifferentiable cost J_{tv}, the multiple constraints optimization problem (5.9) cannot be solved by the methods found in literature (see Section 1.4.1).

5.3.1.2 Implementation

The projector P_{S_0} is implemented by the method described in [46] and the penalty function g for (5.9) is chosen as

$$g : x \mapsto \max_{0 \leq i,j \leq M-1} (x^{i,j^+} - x^{i,j}). \tag{5.10}$$

This function penalizes the most negative pixel of the image x. The nice operator is $T = P_{S_1}$.

Infeasibility detection at Step 7 is implemented by Proposition 4.17. For signal restoration problems, since the diameter of S_0 defined by (5.7) can be very large, the infeasibility detection test of Proposition 4.17 relies on the estimation of $\gamma \geq d(x_l, S^*)$. For the above image restoration problem, since

$$(\forall x^* \in S^*) \; d(x_l, S) \leq \|x_l - x^*\| \leq \|x_l - P_{S_0}(0)\| + \|P_{S_0}(0) - x^*\|, \tag{5.11}$$

we obtain $\gamma \geq d(x_l, S^*)$ as

$$\gamma = \|x_l - P_{S_0}(0)\| + \zeta, \tag{5.12}$$

where $\zeta \geq d(P_{S_0}(0), S^*)$ is estimated heuristically as $\zeta = 0.5\|P_{S_0}(0)\|$. Finally, we set $v = 0$.

The restored image is shown in Fig. 5.4. As we can see, using only the two signal constraints (5.7) and (5.8), most features of the original image are recovered by minimizing the total variation cost.

Next, we replace the total variation cost by the energy cost

$$J_e : x \mapsto \|x\|^2 \tag{5.13}$$

and solve

$$\text{Find } x^* \in S = S_0 \cap S_1 \text{ such that } J_e(x^*) = \inf_{x \in S} J_e(x). \tag{5.14}$$

This quadratic problem is a special case of (1.20) and it is solved by the minimum norm algorithm [12]

$$x_0 = 0 \text{ and } (\forall n \in \mathbb{N}) \;\; x_{n+1} = \frac{n}{2(n+1)} \sum_{i=0}^{1} P_{S_i}(x_n). \tag{5.15}$$

The degraded image Fig. 5.3 is shown again in Fig. 5.5 and the restored image is shown in Fig. 5.6. It is clearly not as sharp as the image obtained with the total variation cost.

5.3.2 Image Denoising

Total variation works especially well as a minimization criterion for denoising signals [38]. In denoising problems, the operator L in (5.1) becomes the identity operator and the constraint set S_0 in (5.7) is a ball centered at y.

In the following examples, the noisy images shown in Fig. 5.7/5.9 and Fig. 5.11/5.13 are obtained by adding to the original image shown in Fig. 5.1 a random noise image whose components are i.i.d zero mean white Gaussian random variables. The signal-to-noise ratios are 11.66 dB and 5.64 dB respectively. The denoising problems are formulated as (5.9) and solved by Algorithm 4.4.

In these two denoising problems, the implementation of Algorithm 4.4 is the same as in Section 5.3.1, except that the diameter of S_0 defined by (5.7) can be used directly in Proposition 4.17 to detect infeasibility.

The denoised images are shown in Fig. 5.8 and Fig. 5.12. As we can see, when the level of the noise is high, patches start to appear in the denoised image. For comparison, the energy cost (5.13) is used next. The denoised images are shown respectively in Fig. 5.10 and Fig. 5.14.

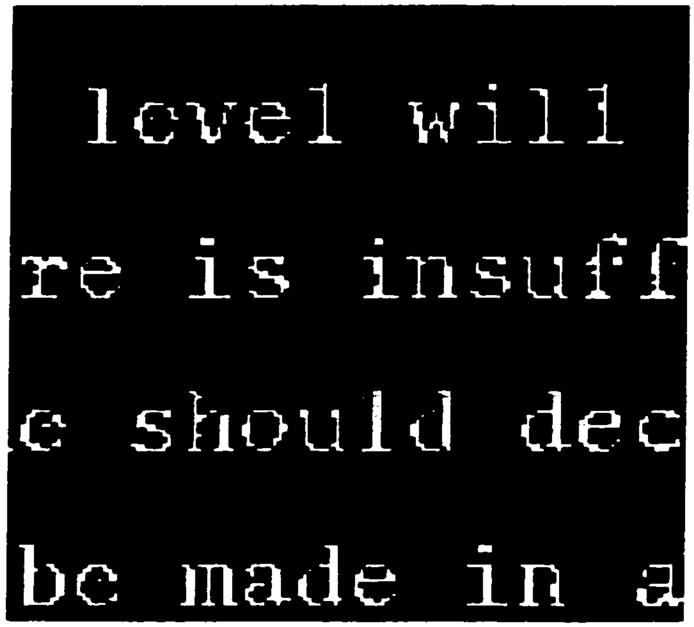

Figure 5.1: Original image.

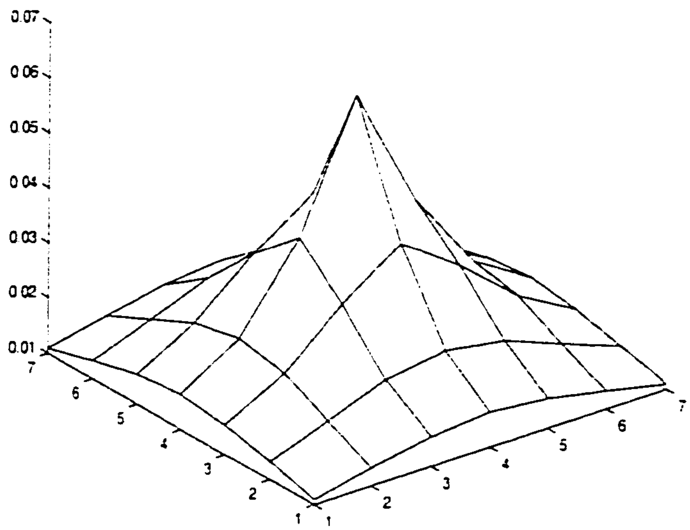

Figure 5.2: Point spread function.

Figure 5.3: Degraded image. blurred SNR=30.23 dB.

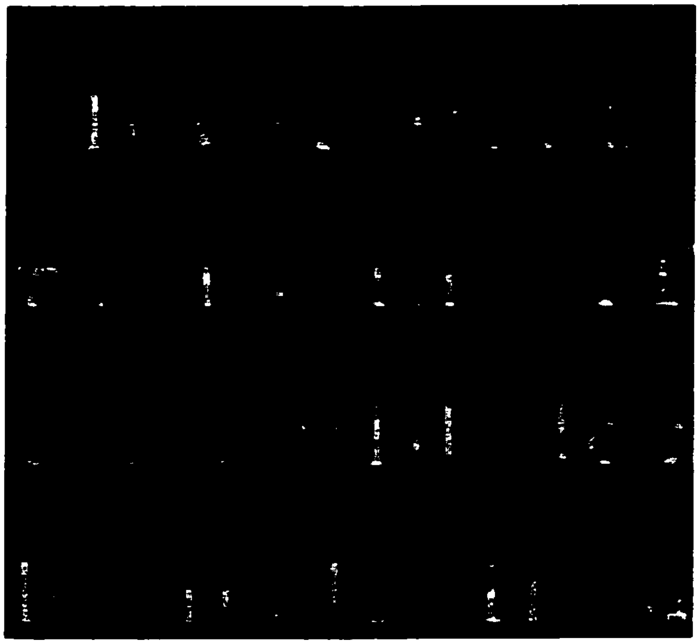

Figure 5.4: Image restored with the total variation cost.

Figure 5.5: Degraded image, blurred SNR=30.23 dB.

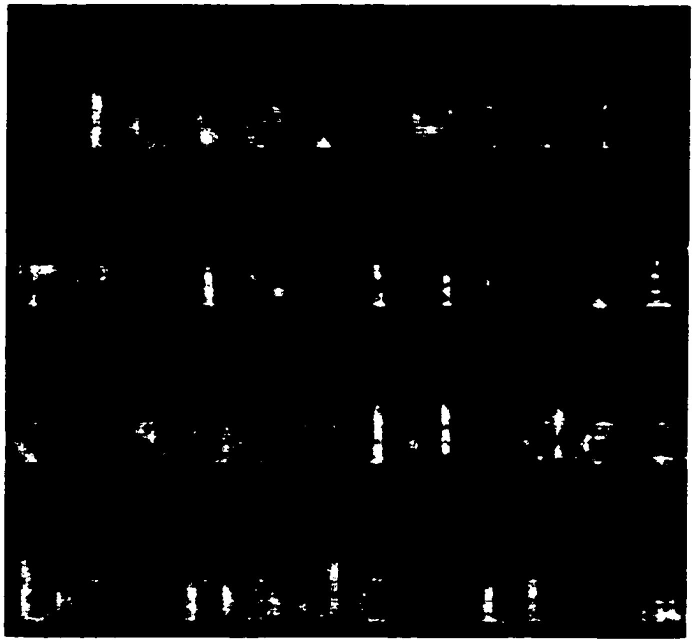

Figure 5.6: Image restored with the energy cost.

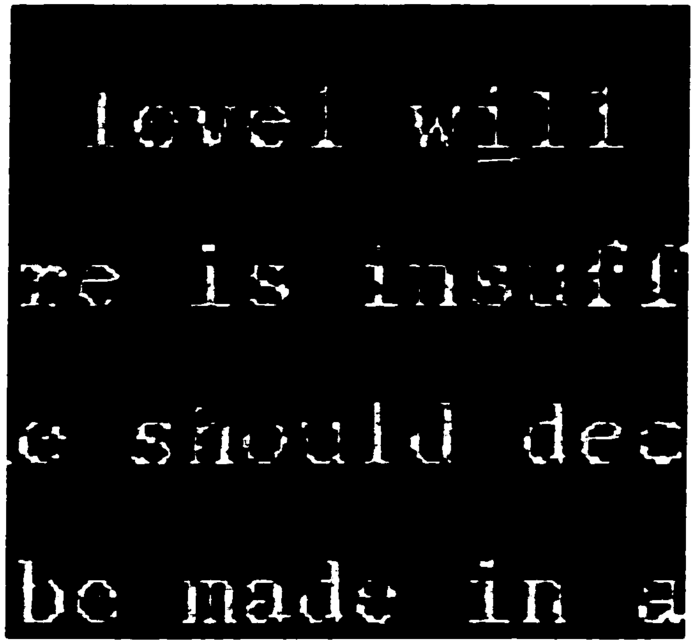

Figure 5.7: Noisy image, SNR=11.66 dB.

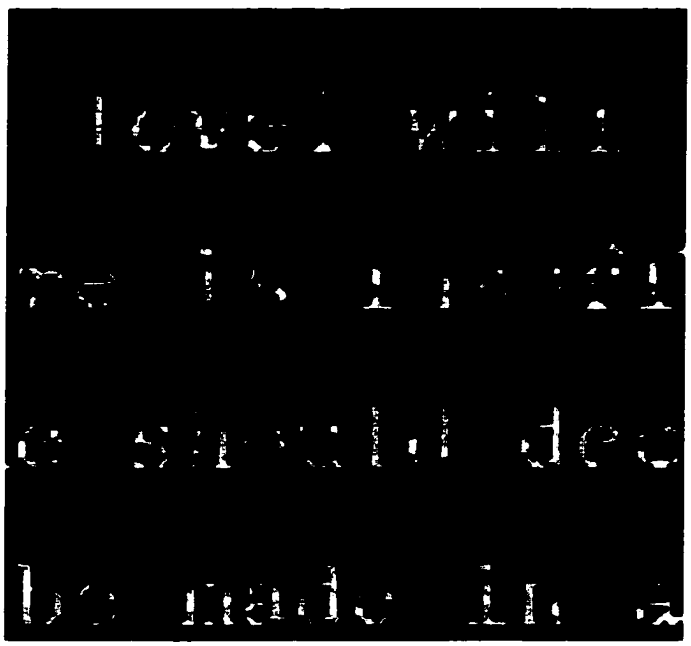

Figure 5.8: Image denoised with the total variation cost.

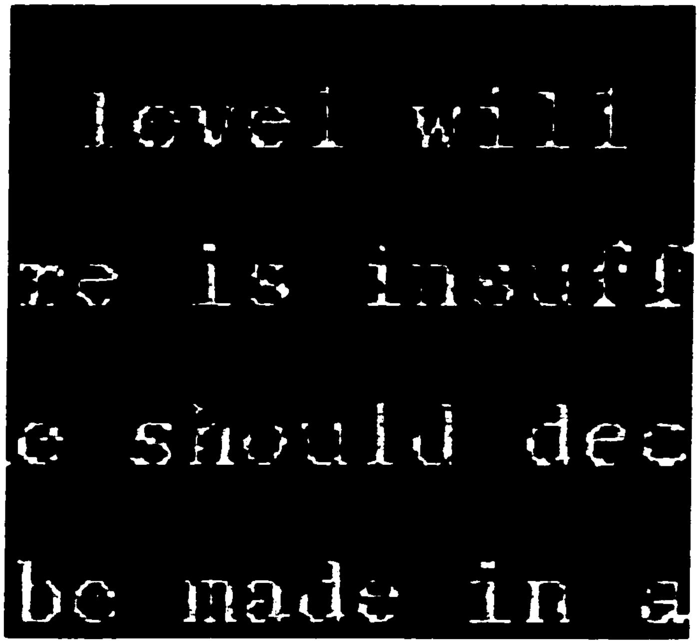

Figure 5.9: Noisy image, SNR=11.66 dB.

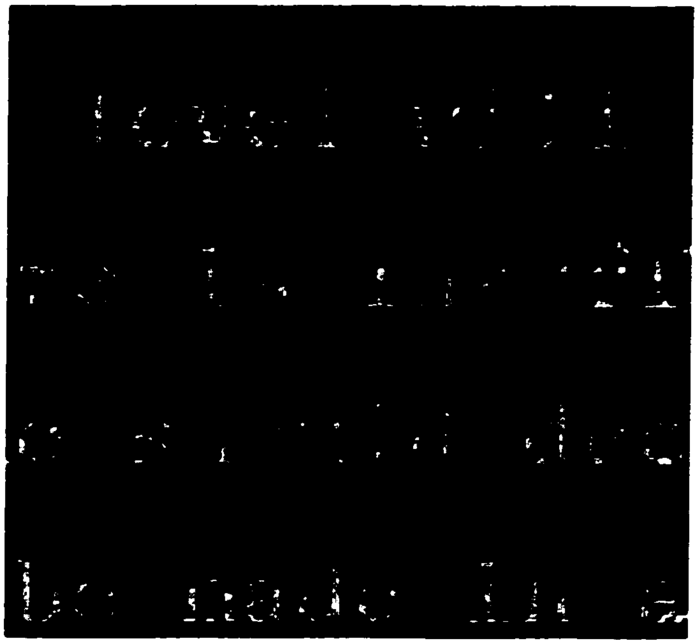

Figure 5.10: Image denoised with the energy cost.

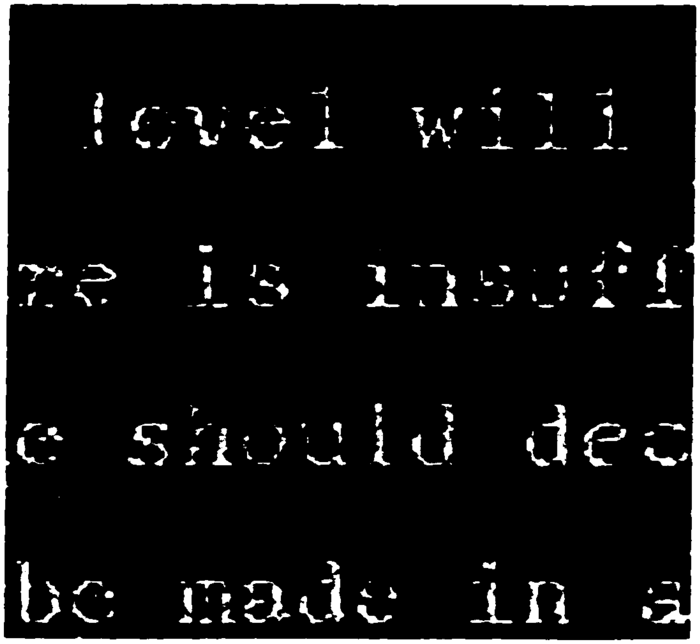

Figure 5.11: Noisy image. SNR=5.64 dB.

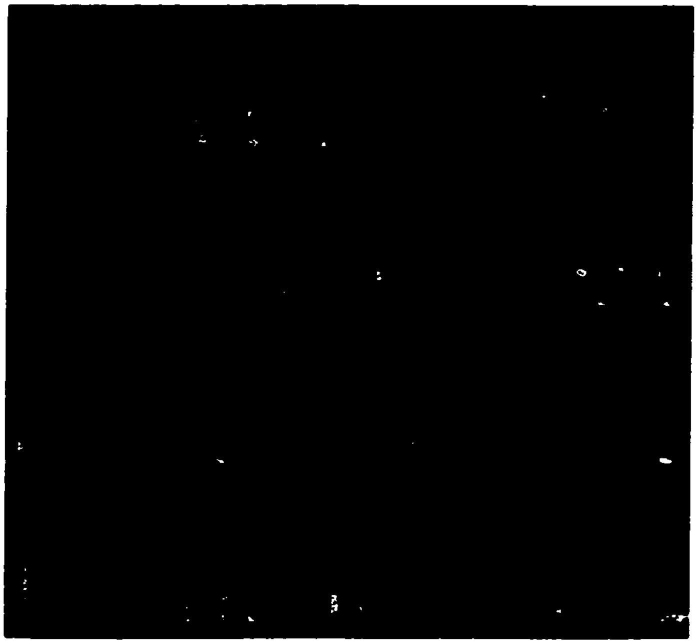

Figure 5.12: Image denoised with the total variation cost.

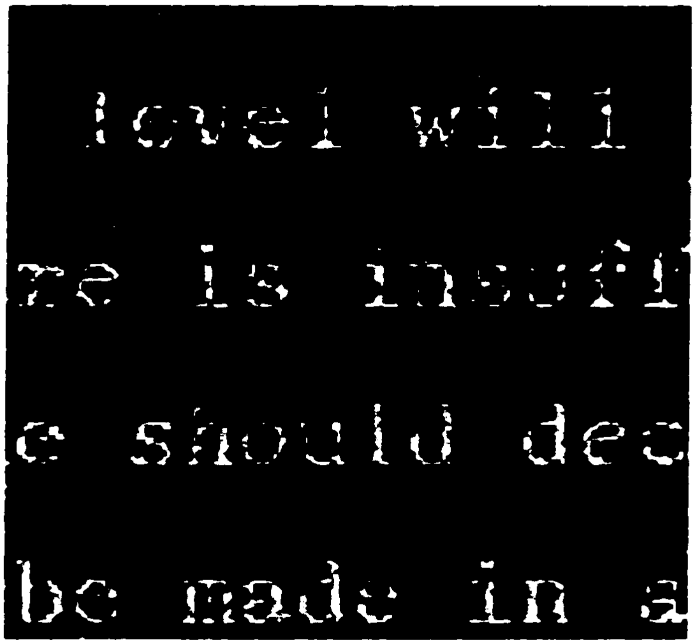

Figure 5.13: Noisy image, SNR=5.64 dB.

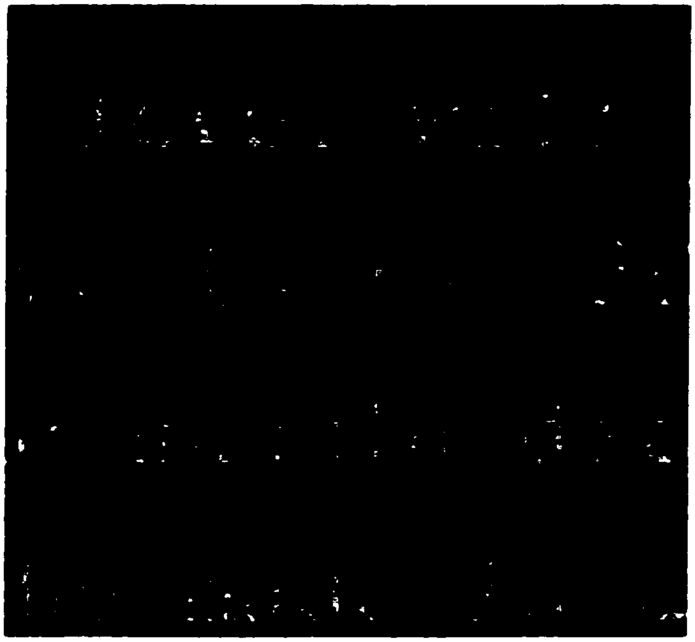

Figure 5.14: Image denoised with the energy cost.

5.4 Experiment II

5.4.1 Image Restoration

In this example, we use different constraints. The degraded image shown in Fig. 5.15/5.17 is obtained by convolving the original image shown in Fig. 5.1 with a 7×7 uniform blur and adding a random noise image u whose components are i.i.d. and uniformly distributed in $[-\zeta, \zeta]$, where $\zeta = 8$. The blurred image-to-noise ratio is 22.98 dB.

Using an *a priori* bound on the amplitude of the pixels of the original image, we define S_0 as

$$S_0 = [0, \zeta_0]^{M \times M}. \tag{5.16}$$

Then, following [12], we use the above statistical hypotheses on the noise u to form the M^2 sets

$$\begin{cases} S_{i \times M + j + 1} = \{x \in \mathsf{E} \mid |y^{i,j} - (Lx)^{i,j}| \leq \zeta\} \\ 0 \leq i, j \leq M - 1, \end{cases} \tag{5.17}$$

as well as the set

$$S_{M^2+1} = \{x \in \mathsf{E} \mid \|y - Lx\|^2 \leq \delta\}. \tag{5.18}$$

Thus, we have a total of $M^2 + 2 = 16386$ sets.

Again, the total variation cost (5.5) is chosen as the minimizing criterion. This leads to the constrained nondifferentiable convex optimization problem

$$\text{Find } x^* \in S = \bigcap_{0 \leq i \leq M^2+1} S_i \ \text{ such that } \ J_{\text{tv}}(x^*) = \inf_{x \in S} J_{\text{tv}}(x), \tag{5.19}$$

which is solved by Algorithm 4.4.

The penalty function g for (5.19) is chosen as $g = \max\{g_1^+, g_2^+\}$ where

$$\begin{cases} g_1(x) = \|y - Lx\|^2 - \delta \\ g_2(x) = \max\left\{|y^{i,j} - (Lx)^{i,j}| - \zeta \mid 0 \leq i, j \leq M-1\right\}, \end{cases} \tag{5.20}$$

and the nice operator as

$$T = P_{S_1} \circ \cdots \circ P_{S_{M^2}} \circ G_0^{g_1}, \tag{5.21}$$

where P_{S_i} $(1 \leq i \leq M^2)$ is the projector onto the hyperslab S_i of (5.17) and is implemented as in [13].

Infeasibility detection at Step 7 is implemented by Proposition 4.17. Since

$$(\forall x^* \in S^*)\ d(x_l, S) \leq \|x_l - x^*\| \leq \|x_l - P_{S_{M^2+1}}(0)\| + \|P_{S_{M^2+1}}(0) - x^*\|, \tag{5.22}$$

we obtain $\gamma \geq d(x_l, S^*)$ as

$$\gamma = \|x_l - P_{S_{M^2+1}}(0)\| + \zeta, \tag{5.23}$$

where $\zeta \geq d(P_{S_{M^2+1}}(0), S^*)$ is estimated heuristically as $\zeta = 0.5\|P_{S_{M^2+1}}(0)\|$. The projector $P_{S_{M^2+1}}$ is implemented by the method described in [46].

The restored image is shown in Fig. 5.16. We see that the edges are restored in reasonably good form. For comparison, as in experiment I, the energy cost J_e of (5.13) is also used in this problem. The restored image is shown in Fig. 5.18. This image was obtained by the minimum norm algorithm [12]

$$x_0 = 0 \text{ and } (\forall n \in \mathbb{N})\ x_{n+1} = \frac{n}{(M^2+2)(n+1)} \sum_{i=0}^{M^2+1} P_{S_i}(x_n). \tag{5.24}$$

5.4.2 Image Denoising

We now consider a denoising problem. The noisy image $y = x+u$ is shown in Fig. 5.19/5.21, where the components of u are uniformly distributed in $[-70, 70]$. The signal-to-noise ratio is 5.56 dB. The denoising problem is formulated as (5.19) and solved by Algorithm 4.4. The denoised image is shown in Fig. 5.20. Although the images shown in Fig. 5.19 and Fig. 5.11 have the similar signal-to-noise ratios, the denoised image shown in Fig. 5.20 has a better quality than the denoised image shown in Fig. 5.12.

For comparison, the energy cost J_e is also used for above image denoising problem. The denoised image is shown in Fig. 5.22.

Let us note that constraints of type (5.17) are not very effective for Gaussian noise. Indeed, in (5.17), a statistical constraint is obtained for each of the 16384 pixels. Since the constraints are independent, even if we let ζ in (5.17) be such that each constraint has a confidence level of 0.99, we obtain a joint confidence level of $0.99^{16384} \approx 0$ [15]. Moreover, choosing a tighter bound ζ may result in $\bigcap_{1 \le i \le 16384} S_i = \emptyset$. For these reasons, the constraints (5.17) were not used in experiment I.

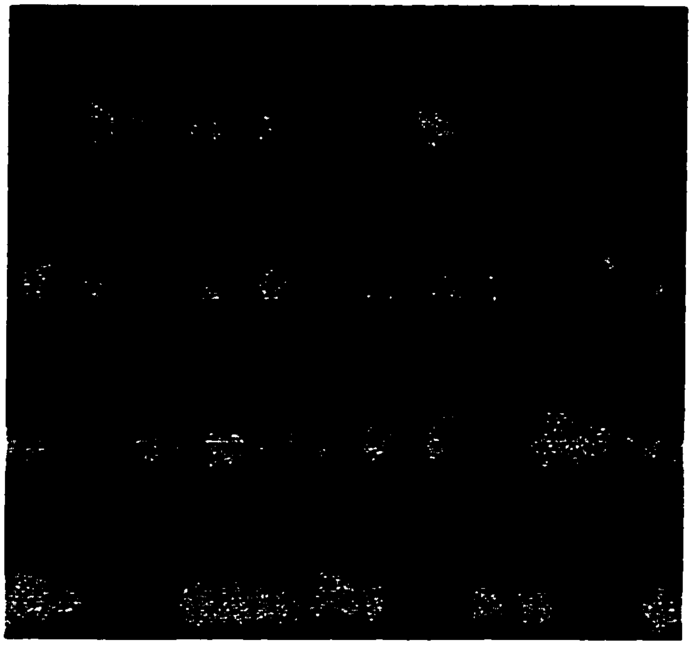

Figure 5.15: Degraded image, blurred SNR=22.98 dB.

Figure 5.16: Image restored with the total variation cost.

Figure 5.17: Degraded image. blurred SNR=22.98 dB.

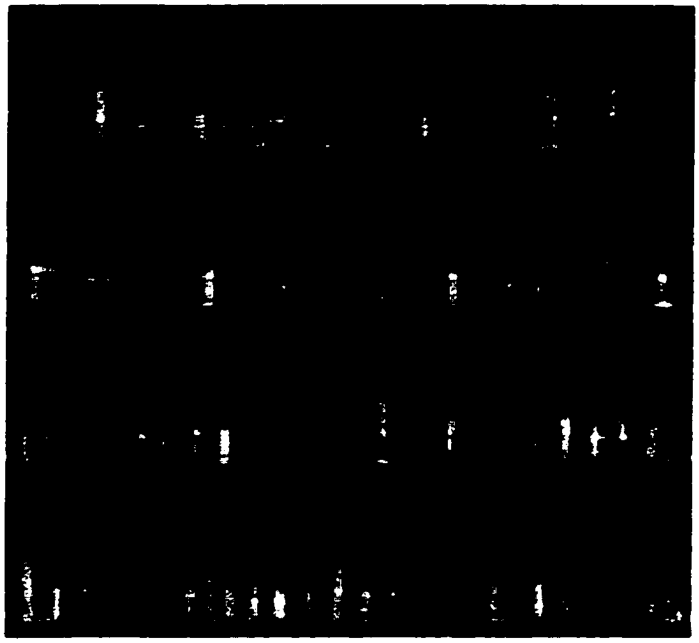

Figure 5.18: Image restored with the energy cost.

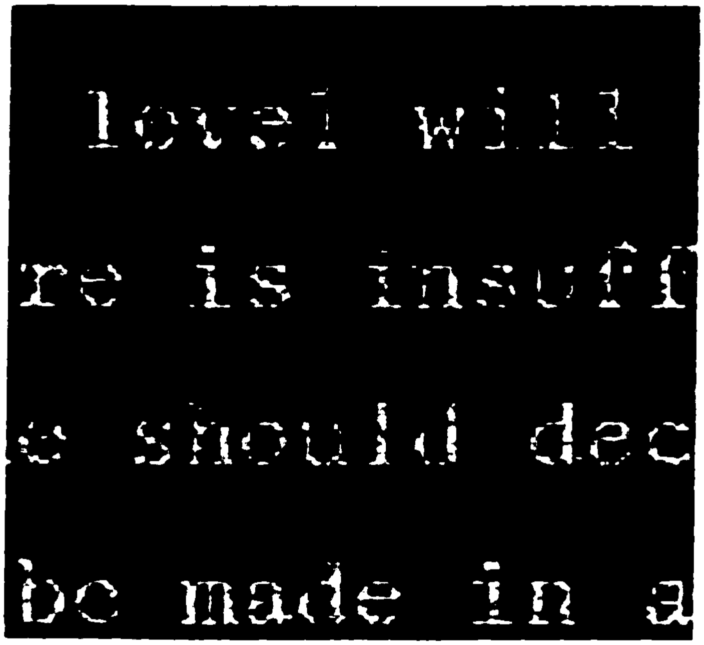

Figure 5.19: Noisy image, SNR=5.56 dB.

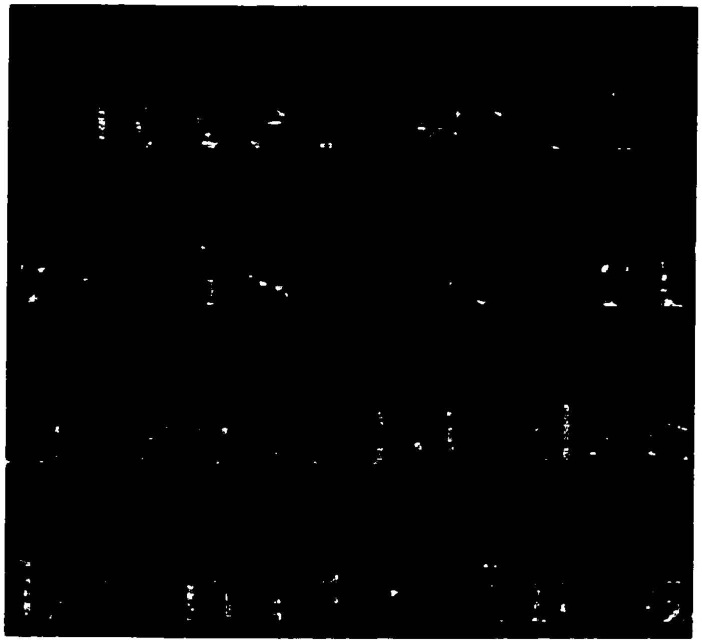

Figure 5.20: Image denoised with the total variation cost.

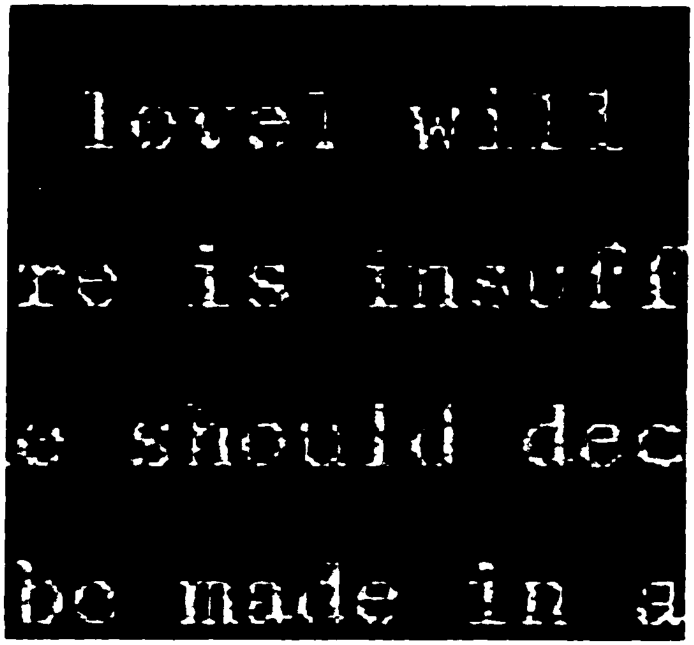

Figure 5.21: Noisy image. SNR=5.56 dB.

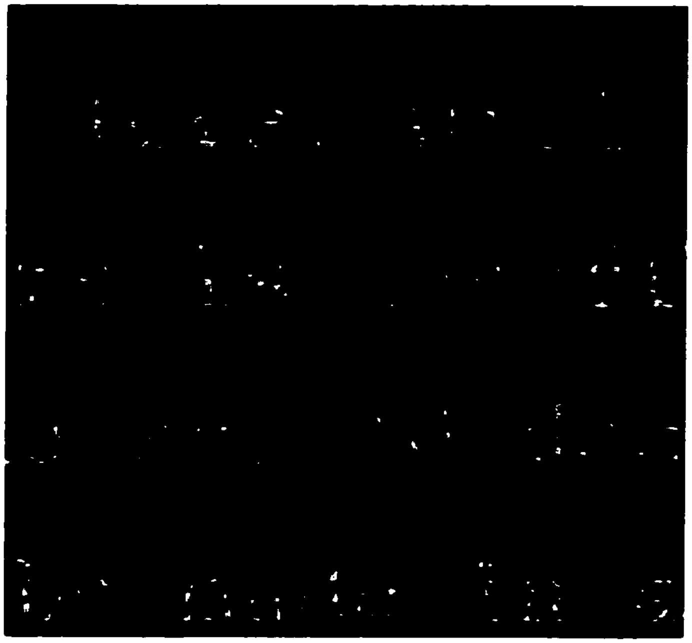

Figure 5.22: Image denoised with the energy cost.

Chapter 6

A Simple Subgradient Level Algorithm

6.1 Introduction

Our basic problem was formulated in (1.2) as follows: Given a convex function $J\colon \mathsf{E} \to \mathbb{R}$ and a nonempty closed convex set $S \subset \mathsf{E}$,

$$\text{Find } x \in S \text{ such that } J(x^*) = \inf_{x \in S} J(x) \triangleq \alpha^*. \tag{6.1}$$

In Chapter 4, a general level subgradient method was proposed for solving (6.1) under Assumption 4.1, i.e.,

$$S = \bigcap_{0 \le i \le m} S_i, \tag{6.2}$$

where S_0 is a simple compact convex set and, for every $i \in \{1, \dots, m\}$, $S_i = \operatorname{lev}_{\le 0} g_i$ where $g_i\colon \mathsf{E} \to \mathbb{R}$ is a convex function.

In this chapter, we assume that $S = S_0$ is a simple compact convex set and present a version of the subgradient level method developed in Chapter 4 adapted to this assumption. We obtain a method akin to the algorithms proposed in [17, 20] for solving the convex optimization problem

$$\text{Find } x^* \in S_0 \text{ such that } J(x^*) = \inf_{x \in S_0} J(x). \tag{6.3}$$

6.2 Description of the Algorithm and Convergence

Consider the special case of the subgradient level method proposed in Chapter 4 in which the nice operator is $T = \mathrm{Id}$. Then $\mathrm{Fix}\, T = \mathsf{E}$ and therefore we can let the penalty function $g\colon \mathsf{E} \to \mathbb{R}$ for (4.1) be

$$g\colon x \mapsto 0. \tag{6.4}$$

For this special case, Algorithm 4.4 becomes the following simple algorithm due to the fact that the constraint represented by S_0 is automatically enforced at each iteration.

Algorithm 6.1 Given $v \in \mathsf{E}$, $\epsilon \in]0, +\infty[$, a (possibly finite) sequence $(x_n)_n$ is constructed as follows.

Step 0. Set $x_0 = P_{S_0}(v)$, $\underline{\alpha}_0 < \alpha^*$, $\overline{\alpha}_0 = J(x_0)$ and $n = 0$.

Step 1. If $\overline{\alpha}_n - \underline{\alpha}_n \leq \epsilon$, terminate.

Step 3. Set $\alpha_n = (\underline{\alpha}_n + \overline{\alpha}_n)/2$.

Step 4. Set $x_{n+1} = P_{S_0} \circ G^J_{\alpha_n}(x_n)$.

Step 5. If $S \cap \mathrm{lev}_{\leq \alpha_n} J = \emptyset$ is detected, go to Step 6; Otherwise, go to Step 7.

Step 6. Set $\underline{\alpha}_{n+1} = \alpha_n$, $\overline{\alpha}_{n+1} = \overline{\alpha}_n$, $x_{n+1} = x_n$, $n = n+1$, and go to Step 1.

Step 7. Set $\underline{\alpha}_{n+1} = \underline{\alpha}_n$, $\overline{\alpha}_{n+1} = \min\{J(x_{n+1}), \overline{\alpha}_n\}$, $n = n+1$, and go to Step 1.

Since $(\forall \beta \in]0, +\infty[)$ $\mathrm{lev}_{\leq \beta}\, g = \mathsf{E}$, Theorem 4.10 becomes

Theorem 6.2 *Let $(x_n)_n$ be an arbitrary sequence generated by Algorithm 6.1. Then*

$$(\exists n \in \mathbb{N}) \;\; x_n \in S_0 \;\; \text{and} \;\; J(x_n) < \alpha^* + \epsilon. \tag{6.5}$$

Theorem 6.2 states that Algorithm 6.1 produces a signal that satisfies the constraint S_0 and achieves any preset tolerance value on the optimal value of the objective.

6.3 Implementation

The implementation of Algorithm 6.1 is straightforward except for infeasibility detection at Step 5. Since Algorithm 6.1 is a special case of Algorithm 4.4, infeasibility detection can be implemented by Proposition 4.17. For Algorithm 6.1, by construction, at every iteration n we have:

- $\overline{\alpha}_n \geq \overline{\alpha}_{n+1}$.
- Termination when $\overline{\alpha}_n - \underline{\alpha}_n \leq \epsilon$.

Therefore, the set $\mathbb{N}_n$ defined by (4.25) becomes $\mathbb{N}_n = \{k \in \mathbb{N} \mid \underline{\alpha}_k = \underline{\alpha}_n\}$ and $(\forall n \in \mathbb{N})$ $\mathbb{N}_n \subset \mathbb{N}_\epsilon$. Thus, for Algorithm 6.1, Proposition 4.17 can be stated as

Proposition 6.3 *Let* $\kappa \geq \operatorname{diam}(S_0)$. *At each iteration* n, *define*

$$(\forall k \in \mathbb{N}_n)\ \rho_k = \|G^J_{\alpha_k}(x_k) - x_k\|^2 + \|P_{S_0} \circ G^J_{\alpha_k}(x_k) - G^J_{\alpha_k}(x_k)\|^2 \tag{6.6}$$

and let l *be the smallest integer in* $\mathbb{N}_n$. *Then, for every* $\gamma \geq d(x_l, S^*)$,

$$\textit{if } \sum_{k \in \mathbb{N}_n} \rho_k > \min\left\{\kappa^2, \gamma^2, 2\gamma\|x_l - x_{n+1}\| - \|x_l - x_{n+1}\|^2\right\} \textit{ then } S \cap \operatorname{lev}_{\leq \alpha_n} J = \varnothing. \tag{6.7}$$

Proposition 6.3 can be implemented as in Section 5.3.1.

6.4 Applications

As in Chapter 5, we consider signal restoration and denoising problems in which the signal degradation model is available and given respectively by

$$y = Lx + u \tag{6.8}$$

and

$$y = x + u. \tag{6.9}$$

Recall that in this model x, y, and u are in $\mathsf{E} = \mathbb{R}^{M \times M}$ and represent respectively the original signal, the recorded signal, the additive noise, while $L\colon \mathsf{E} \to \mathsf{E}$ is a linear operator. For each instances, we use the statistical hypotheses on the components of noise u to form the constraint set [12, 46]

$$S_0 = \left\{x \in \mathsf{E} \mid \|Lx - y\|^2 \leq \delta\right\} \text{ or } S_0 = \left\{x \in \mathsf{E} \mid \|x - y\|^2 \leq \delta\right\}. \tag{6.10}$$

Knowing that the original signal has block features, the total variation cost (5.5) is chosen as the optimality criterion. Signal restoration/denoising problems are then formulated as the constrained total variation minimization problem

$$\text{Find } x^* \in S_0 \text{ such that } J_{tv}(x^*) = \inf_{x \in S_0} J_{tv}(x) \tag{6.11}$$

and solved by Algorithm 6.1.

Below, several examples are given to show that the exact constrained minimization of the total variation cost, as posed in (6.11) and carried out with Algorithm 6.1, leads to satisfactory restoration and denoising of blocky signals, even when the level of the noise is high.

In the signal restoration examples, the degraded signals are obtained by convolving the original signal shown in Fig. 6.1 with the point spread function shown in Fig. 6.2 and adding zero mean Gaussian white noise.

In the signal denoising examples, the noisy signals are obtained by adding a zero mean Gaussian white noise to the original signal shown in Fig. 6.1. The level of the additive noise is increased gradually and the degraded/noisy signal and restored/denoised are shown in pairs. Furthermore, as in Chapter 5, the energy cost $J_e\colon x \mapsto \|x\|^2$ is also used for comparison purposes.

The degraded signals appear in Fig. 6.3/6.5, Fig. 6.7/6.9, Fig. 6.11, and Fig. 6.13. The restored signals are shown in Fig. 6.4, Fig. 6.6, Fig. 6.8, Fig. 6.10, Fig. 6.12, and Fig. 6.14. Finally, the noisy signals are shown in Fig. 6.15/6.17 Fig. 6.19/6.21, and Fig. 6.23 and the denoised signals are shown in Fig. 6.16, Fig. 6.18, Fig. 6.20, Fig. 6.22, and Fig. 6.24.

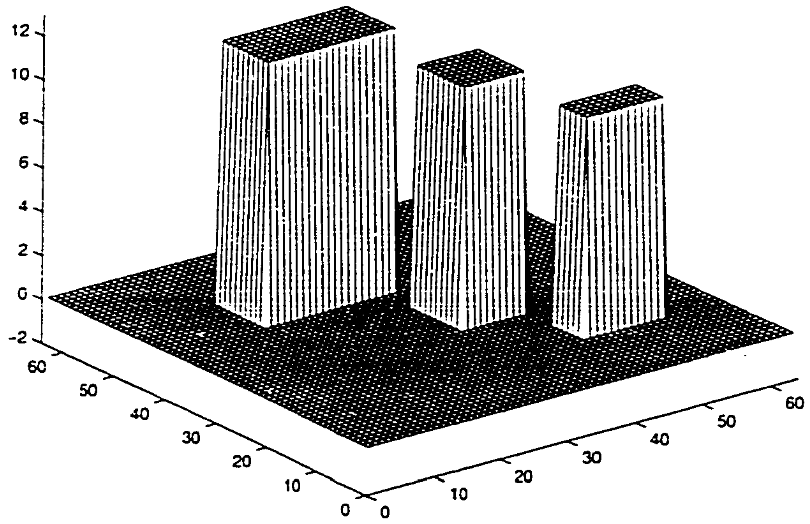

Figure 6.1: Original signal.

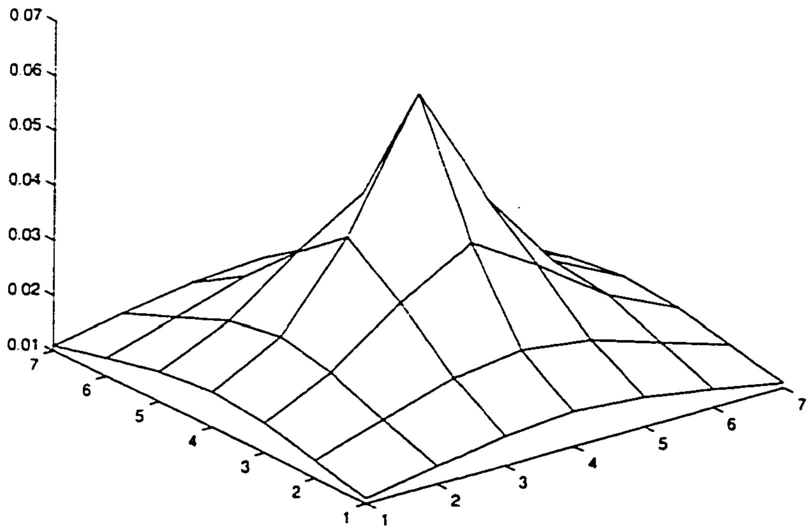

Figure 6.2: Impulse response.

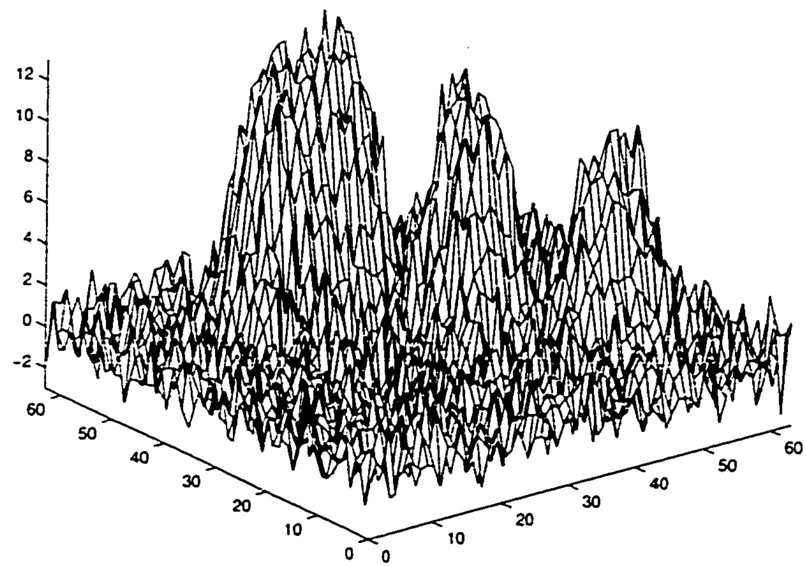

Figure 6.3: Degraded signal, blurred SNR=9.02 dB.

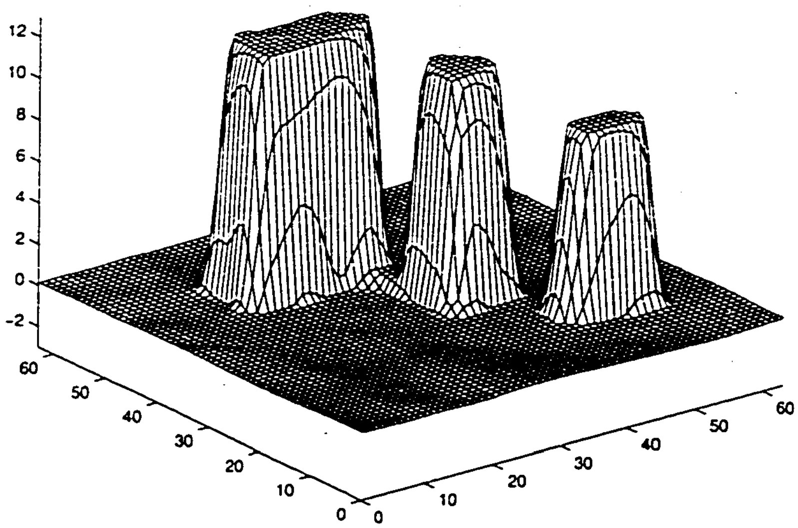

Figure 6.4: Signal restored with the total variation cost.

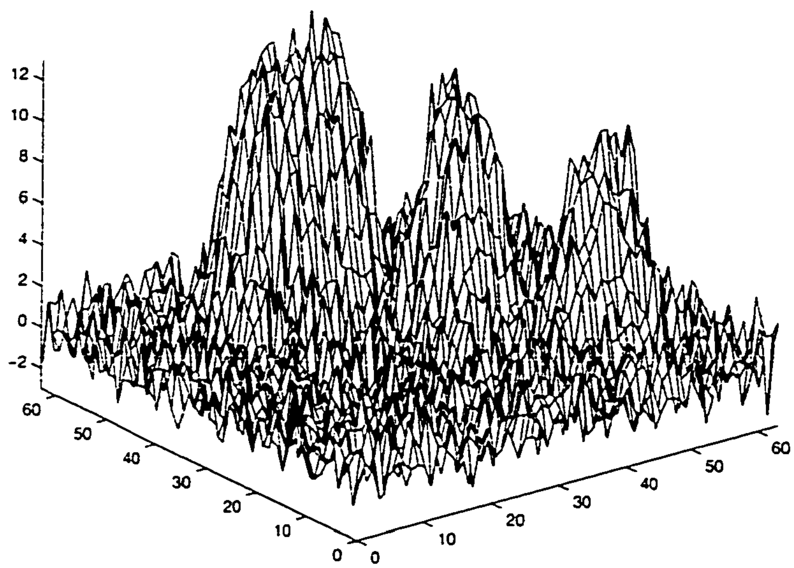

Figure 6.5: Degraded signal, blurred SNR=9.02 dB.

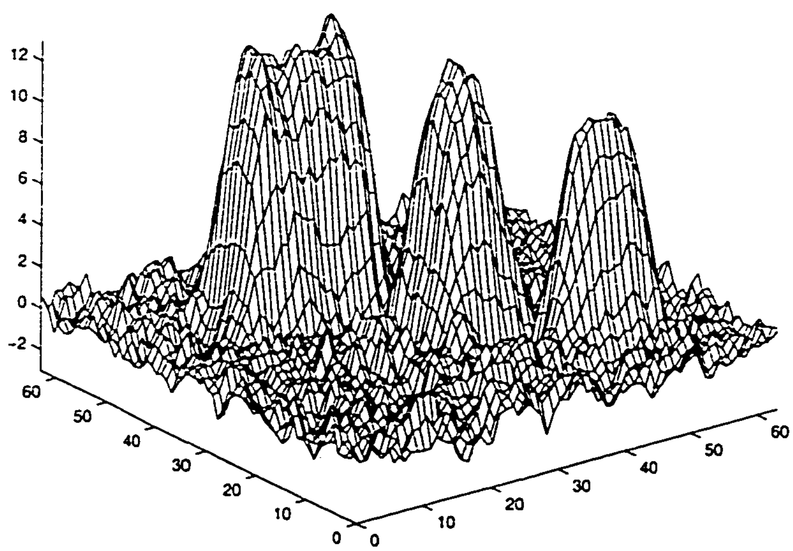

Figure 6.6: Signal restored with the energy cost.

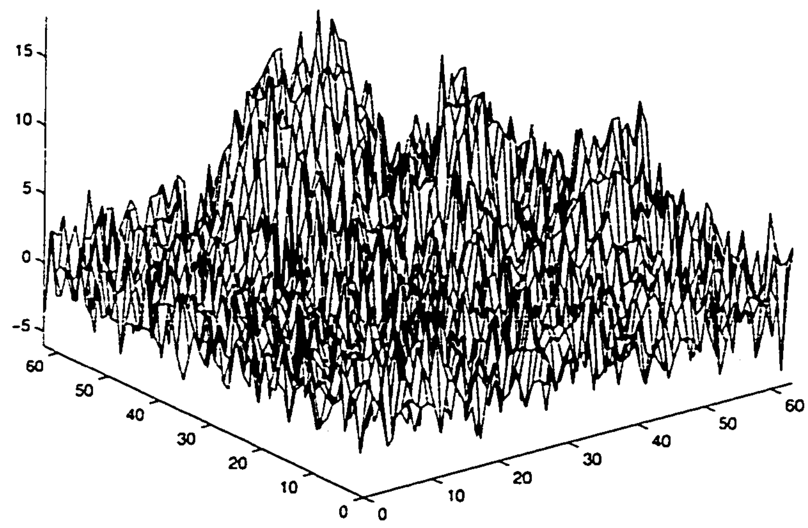

Figure 6.7: Degraded signal, blurred SNR=3.00 dB.

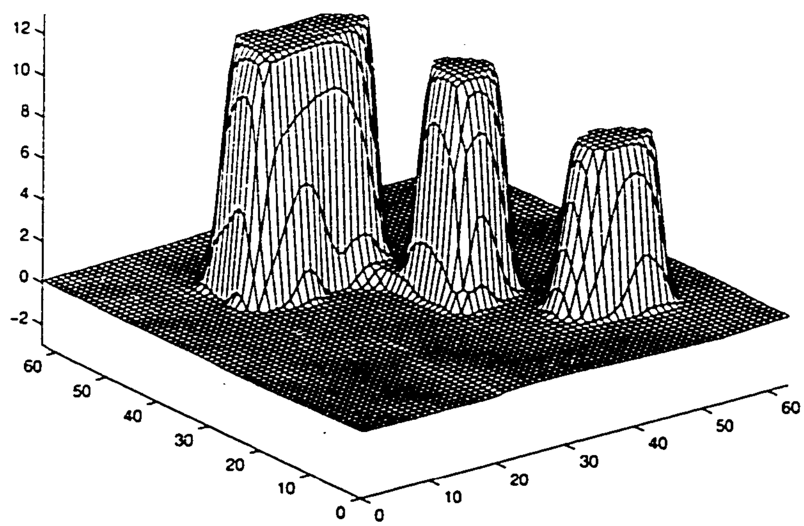

Figure 6.8: Signal restored with the total variation cost.

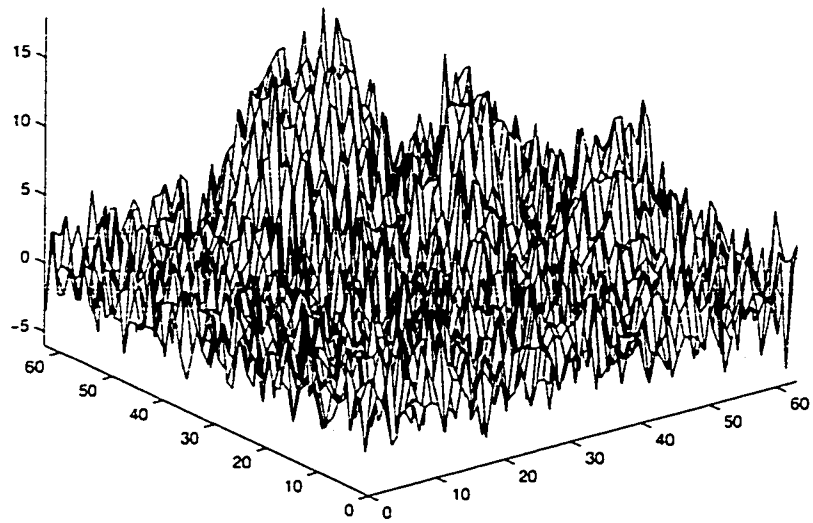

Figure 6.9: Degraded signal, blurred SNR=3.00 dB.

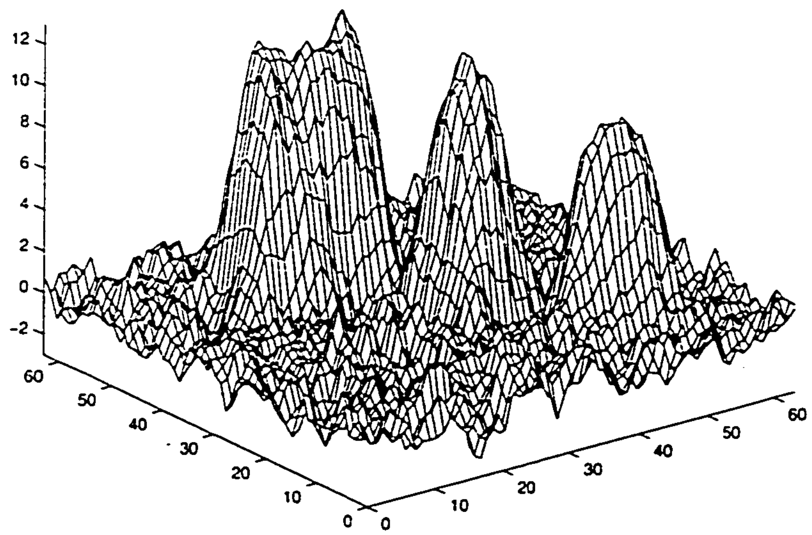

Figure 6.10: Signal restored with the energy cost.

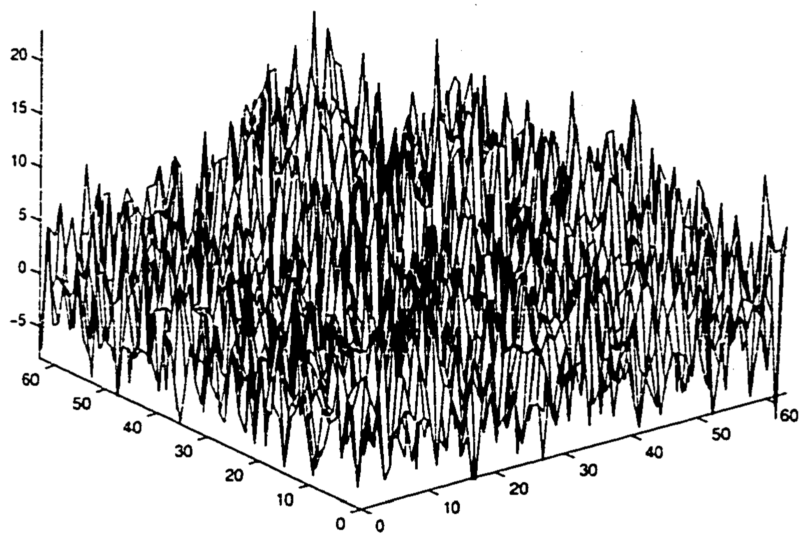

Figure 6.11: Degraded signal, blurred SNR=-3.01 dB.

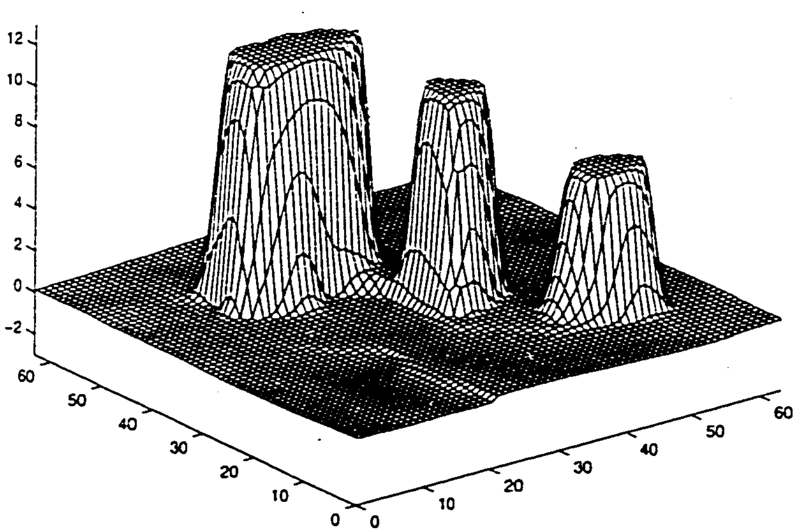

Figure 6.12: Signal restored with the total variation cost.

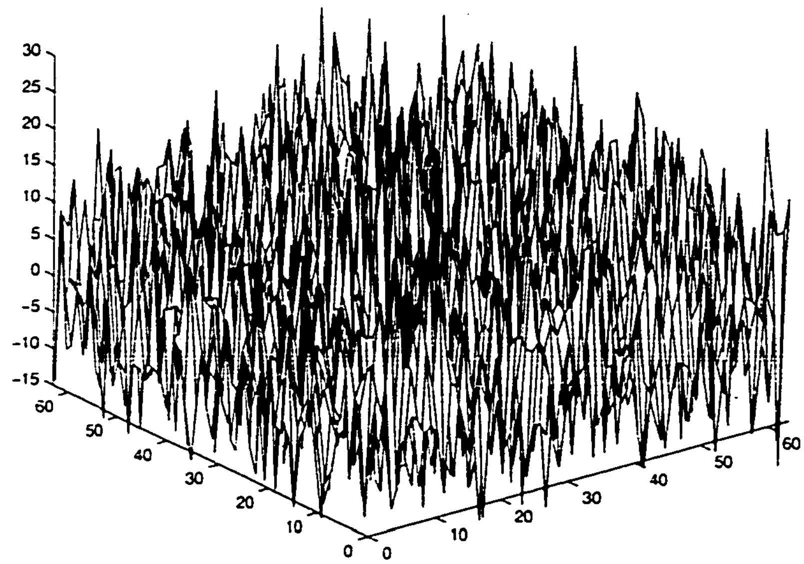

Figure 6.13: Degraded signal, blurred SNR=-9.04 dB.

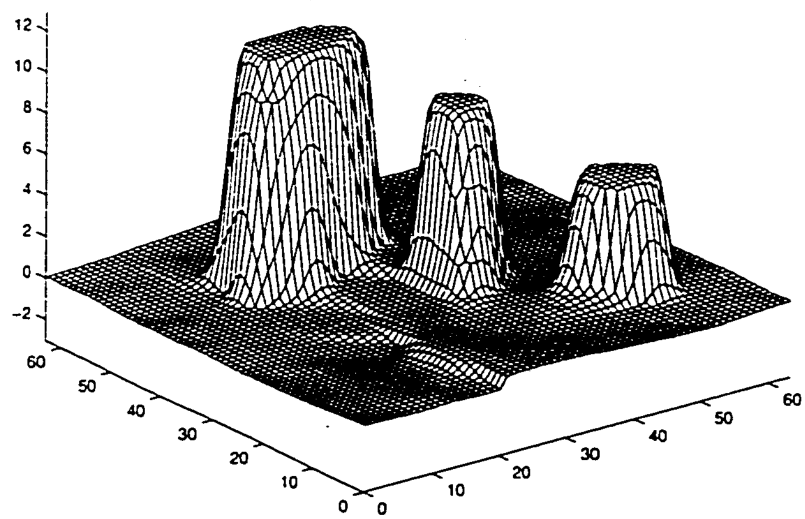

Figure 6.14: Signal restored with the total variation cost.

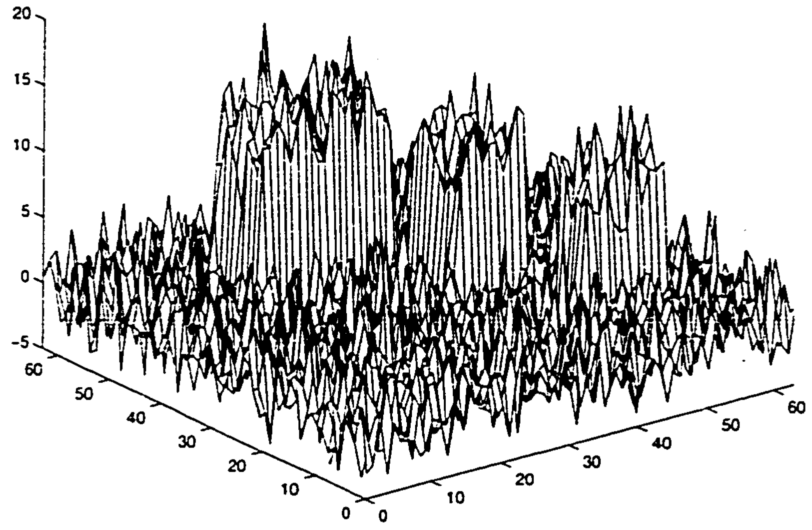

Figure 6.15: Noisy signal, SNR=4.62 dB.

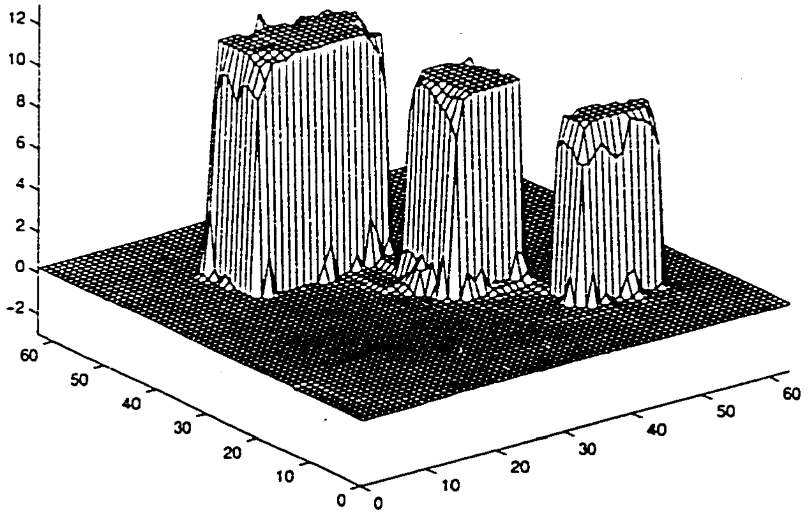

Figure 6.16: Signal denoised with the total variation cost.

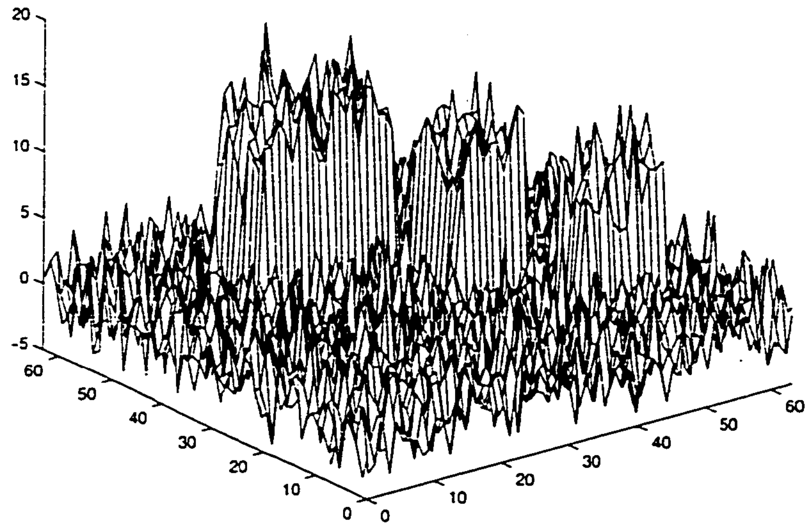

Figure 6.17: Noisy signal, SNR=4.62 dB.

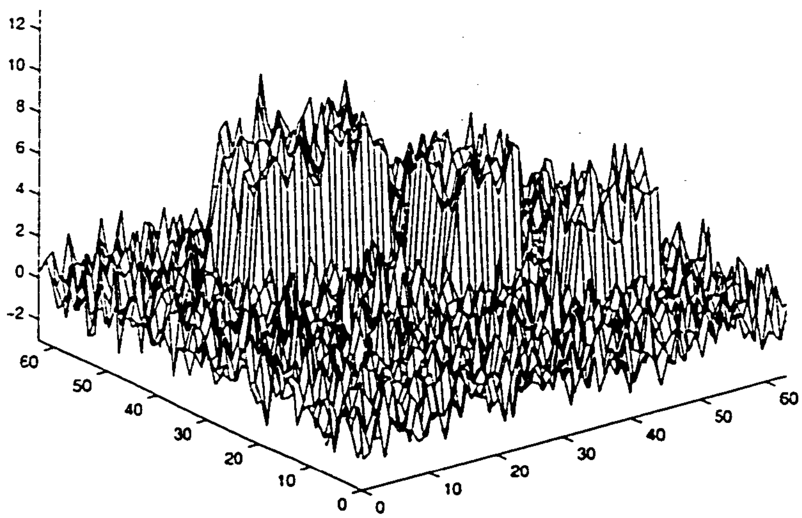

Figure 6.18: Signal denoised with the energy cost.

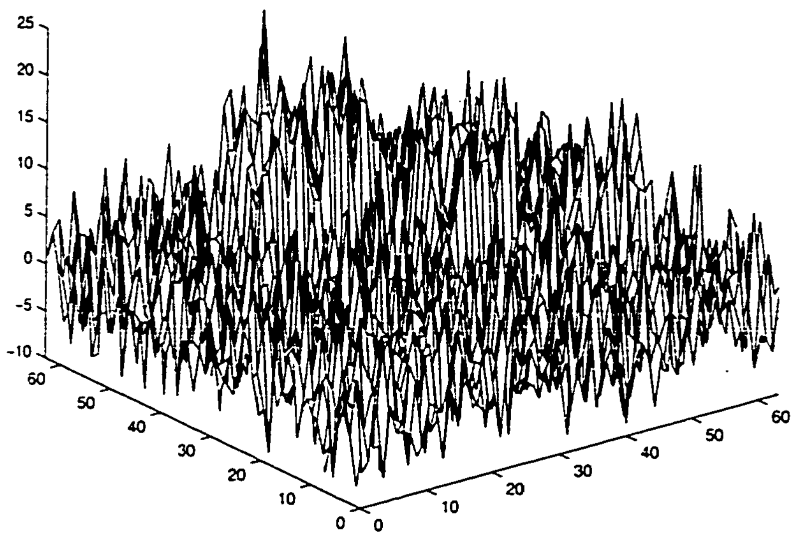

Figure 6.19: Noisy signal, SNR=-1.40 dB.

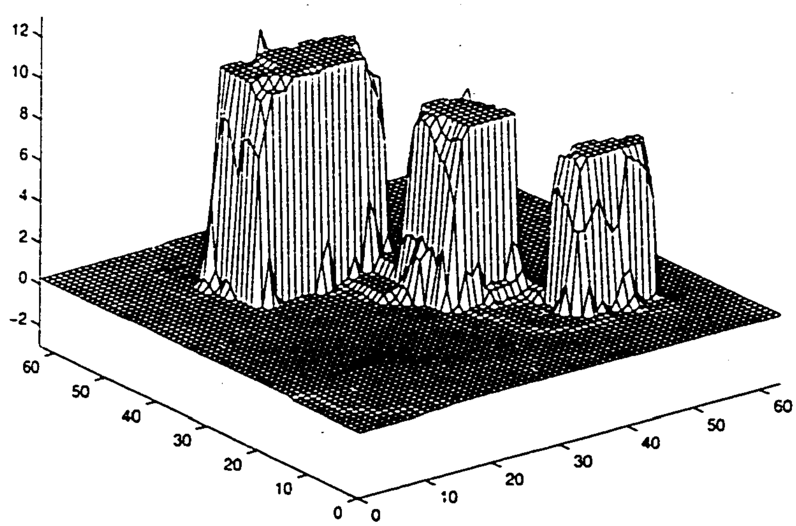

Figure 6.20: Signal denoised with the total variation cost.

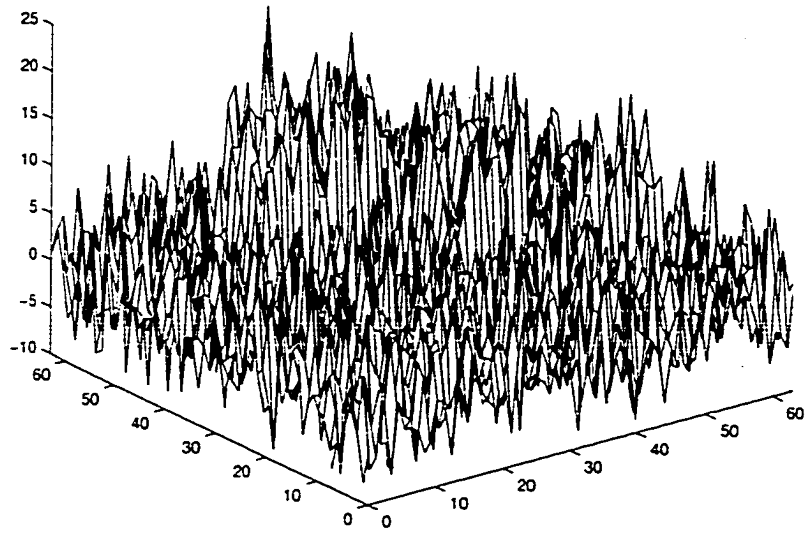

Figure 6.21: Noisy signal, SNR=-1.40 dB.

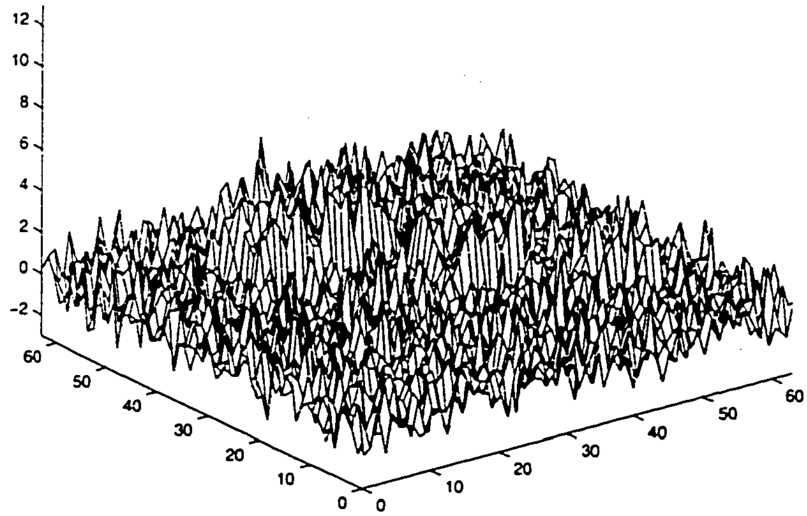

Figure 6.22: Signal denoised with the energy cost.

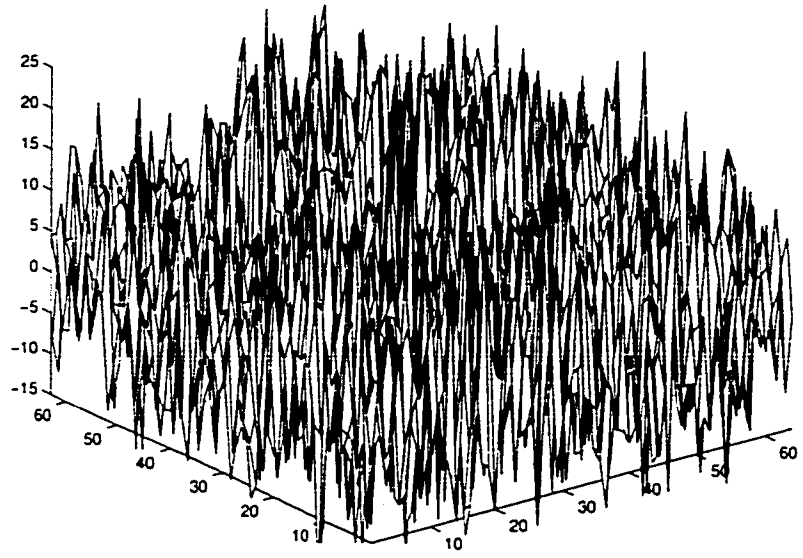

Figure 6.23: Noisy signal, SNR=-7.42 dB.

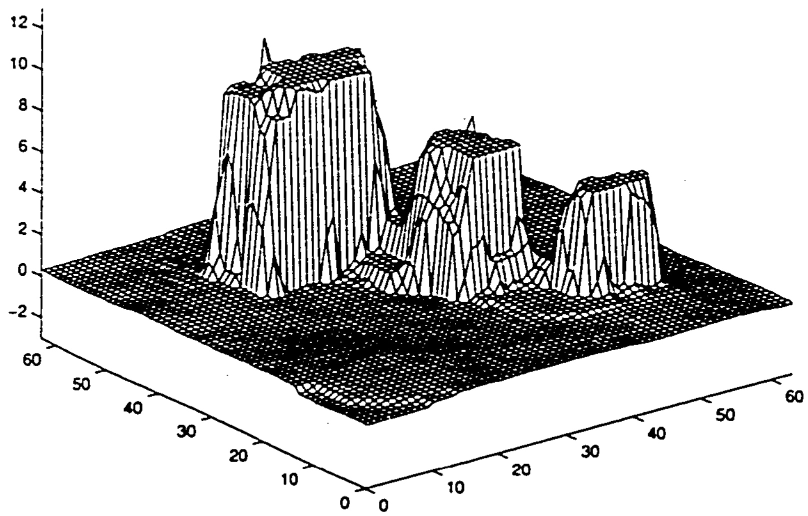

Figure 6.24: Signal denoised with the total variation cost.

Chapter 7

Conclusion

7.1 Summary

Many signal restoration problems can be formulated in the form of the constrained convex optimization problem (1.2). In recent years, there has been a growing interest in the use of nondifferentiable optimality criteria in signal denoising and restoration, primarily because of their ability to preserve and restore edges.

In this dissertation, a general subgradient level method was proposed to solve (1.2) with nondifferentiable costs. Applications to several constrained minimum total variation signal restoration and denoising problems were demonstrated. Unlike the methods presented in [9, 10, 29, 47, 48] the proposed method allows multiple convex constraints and general nondifferentiable convex costs. Moreover, the methods of [9, 10, 29, 47, 48] are for a specific type of cost and allow only one quadratic constraint.

Let us note that in the proposed subgradient level method the number of iterations required to detect infeasibility in (4.44) increases significantly as the iterates get closer and closer to an optimal solution, a behavior which is typical of level set methods [23]. It is therefore important to set the tolerance parameter ϵ at a realistic value, as too small a value will increase the number of iterations with no practical improvement on the solution. Since the asymptotic behavior of Algorithm 4.4 depends on many factors, i.e., ϵ, β, ξ, and $\underline{\alpha}_0$, as well as κ and γ for infeasibility detection at Step 7, we chose not to report run times in Chapters 5 and 6.

7.2 Directions for Future Research

For a nondifferentiable convex function $f\colon \mathsf{E} \to \mathbb{R}$, as seen in Section 1.3.3.4, an arbitrary subgradient may provide a poor description of the local behavior of f. Consequently the choice of a subgradient will affect the efficiency of a subgradient projection method. A standard way to remedy this problem is to use bundle methods to enrich the local information on f [18, 21].

Let $x \in \mathsf{E}$. For every $\epsilon > 0$ there exists a neighborhood $\mathcal{V}$ of x such that $\partial_\epsilon f(x) \supset \bigcup_{y \in \mathcal{V}} \partial f(y)$. Thus $\partial_\epsilon f(x)$ provides better information of the local behavior of f at or near a nondifferentiable point and a good direction of descent can be found in $\partial_\epsilon f(x)$. For a given f, an analytic form of $\partial_\epsilon f(x)$ is hardly available. Instead of selecting $t \in \partial_\epsilon f(x)$ one picks a vector $t \in \mathcal{P}$, where $\mathcal{P}$ is an inner polytope approximation to $\partial_\epsilon f(x)$. The resulting problem usually requires excessive storage for a large size problem. In [21, 22, 23, 24] methods involving projections onto an affine approximation to f derived from the aggregation of several subgradients rather than from an arbitrary subgradient are used to improve the efficiency. Such methods require less storage and are attractive for large scale problems. Although it seems that various bundle methods can be adopted directly by our method, implementation issues need to be further investigated.

New York, January 28, 2000

Bibliography

[1] R. Acar and C. R. Vogel, "Analysis of bounded variation penalty methods for ill-posed problems," *Inverse Problems,* vol. 10, pp. 1217-1229, 1994.

[2] E. Allen, R. Helgason, J. Kennington, and B. Shetty, "A generalization of Polyak's convergence result for subgradient optimization," *Mathematical Programming,* vol. 37, pp. 309-317, 1987.

[3] S. Alliney, "A property of the minimum vectors of a regularizing functional defined by means of the absolute norm," *IEEE Transactions on Signal Processing,* vol. 45, pp. 913-917, April 1997.

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

[5] H. C. Andrews and B. R. Hunt, *Digital Image Restoration.* Englewood, NJ: Prentice Hall, 1977.

[6] G. Archer and D. M. Titterington, "On some Bayesian/regularization methods for image restoration," *IEEE Transactions on Image Processing,* vol. 4, pp. 989-995, July 1995.

[7] H. H. Bauschke and J. M. Borwein "On projection algorithms for solving convex feasibility problems," *SIAM Review,* vol. 38, pp. 367-426, September 1996.

[8] Y. Censor, "Cyclic subgradient projections," *Mathematical Programming,* vol. 24, pp. 233-235, 1982.

[9] A. Chambolle and P. Lions, "Image recovery via total variation minimization and related problems," *Numerische Mathematik,* vol. 76, pp. 167-188, 1997.

[10] T. F. Chan and P. Mulet, "On the convergence of the lagged diffusivity fixed point method in total variation image restoration," *SIAM Journal on Numerical Analysis,* vol. 36, no. 2, pp. 354-367, 1999.

[11] P. L. Combettes, "Signal recovery by best feasible approximation," *IEEE Transactions on Image Processing,* vol. 2, pp. 269-271, April 1993.

[12] P. L. Combettes, "The convex feasibility problem in image recovery," in *Advances in Imaging and Electron Physics,* vol. 95, pp. 155-270. New York: Academic, 1996.

[13] P. L. Combettes, "Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections," *IEEE Transactions on Image Processing,* vol. 6, pp. 493-506, April 1997.

[14] P. L. Combettes, "Hilbertian convex feasibility problem: Convergence of projection methods," *Applied Mathematics and Optimization,* vol. 35, pp. 311-330, May 1997.

[15] P. L. Combettes and T. J. Chaussalet, "Combining statistical information in set theoretic estimation," *IEEE Signal Processing Letters,* vol. 3, pp. 61-62, March 1996.

[16] E. Giusti, *Minimal Surfaces and Functions of Bounded Variation.* Boston: Birkhäuser, 1984.

[17] J.-L. Goffin and K. C. Kiwiel, "Convergence of a simple subgradient level method," *Mathematical Programming,* vol. 85, pp. 207-211, 1999.

[18] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms.* New York: Springer-Verlag, 1993.

[19] B. R. Hunt, "The application of constrained least squares estimation to image reconstruction," *IEEE Transactions on Computers,* vol. 22, pp. 805-812, September 1973.

[20] S. Kim, H. Ahn, and S. Cho, "Variable target value subgradient method," *Mathematical Programming,* vol. 49, pp. 359-369, 1991.

[21] K. C. Kiwiel, "New variants of bundle methods," *Mathematical Programming,* vol. 69, pp. 111-147, 1995.

[22] K. C. Kiwiel, "Block-iterative surrogate projection methods for convex feasibility problems," *Linear Algebra Application,* vol. 215, pp. 225-260, 1995.

[23] K. C. Kiwiel, "The efficiency of subgradient projection methods for convex optimization, Parts I and II," *SIAM Journal on Control and Optimization,* vol. 34, pp. 660-676 and pp. 677-697, 1996.

[24] K. C. Kiwiel and B. Lopuch, "Surrogate projection methods for finding fixed points of firmly nonexpansive mappings," *SIAM Journal on Optimization,* vol. 7, pp. 1084-1102, 1997.

[25] D. P. Kolba and T. W. Parks, "Optimal estimation for band-limited signals including time domain considerations," *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. 31, pp. 113-122, February 1983.

[26] R. M. Leahy and C. E. Goutis, "An optimal technique for constraint-based image restoration and reconstruction," *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. 34, pp. 1629-1642, December 1986.

[27] J. Luo and P. L. Combettes, "An adaptive level set method for nondifferentiable constrained image recovery," Technical report, Spring 2000.

[28] C. Lemaréchal, "Nondifferentiable optimization," in *Handbooks in Operations Research and Management Science,* vol. 1, pp. 529-572. New York: North-Holland, 1989.

[29] Y. Li and F. Santosa, "A computational algorithm for minimizing total variation in image reconstruction," *IEEE Transactions on Image Processing,* vol. 5, pp. 987-995, 1996.

[30] D. G. Luenberger, *Optimization by Vector Space Methods.* New York: John Wiley, 1969.

[31] V. A. Morozov, *Methods for Solving Incorrectly Posed Problems.* New York: Springer-Verlag, 1984.

[32] D. Noll, "Reconstruction with noisy data: An approach via eigenvalue optimization," *SIAM Journal on Optimization,* vol. 8, pp. 82-104, February 1998.

[33] E. Polak, "On the mathematical foundations of nondifferentiable optimization in engineering design," *SIAM Review,* vol. 29, pp. 21-89, 1987.

[34] B. T. Polyak, "Minimization of unsmooth functionals," *USSR Computational Mathematics and Mathematical Physics,* vol. 9, pp. 14-29, 1969.

[35] B. T. Polyak, *Introduction to Optimization.* New York: Optimization Software, Inc., Publications Division, 1987.

[36] L. C. Potter and K. S. Arun, "Energy concentration in band-limited extrapolation," *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. 37, pp. 1027-1041, July 1989.

[37] R. T. Rockafellar, *Convex Analysis.* Princeton, NJ: Princeton University Press, 1970.

[38] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D,* vol. 60, pp. 259-268, 1992.

[39] A. Sabharwal and L. C. Potter, "Convexly constrained linear inverse problems: Iterative least-squares and regularization," *IEEE transactions on Signal Processing,* vol. 46, pp. 2345-2352, September 1998.

[40] C. Sánchez-Avila, "An adaptive regularized method for deconvolution of signals with edges by convex projections," *IEEE Transactions on Signal Processing,* vol. 42, pp. 1849-1851, July 1994.

[41] M. I. Sezan and H. J. Trussell "Prototype image constraints for set-theoretic image restoration," *IEEE Transactions on Signal Processing,* vol. 39, pp. 2275-2285, October 1991.

[42] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions.* New York: Springer-Verlag, 1985.

[43] H. Stark (Editor), *Image Recovery: Theory and Application.* San Diego, California: Academic Press, 1987.

[44] A. M. Tekalp and H. J. Trussell, "Comparative study of some statistical and set-theoretic methods for image restoration," *CVGIP: Graphical Models and Image Processing,* vol. 53, pp. 108-120, March 1991.

[45] H. J. Trussell, "*A priori* knowledge in algebraic reconstruction methods," in *Advances in Computer Vision and Image Processing* (T. S. Huang, Ed.), vol. 1, pp. 265-316. Greenwich, CT: JAI Press, 1984.

[46] H. J. Trussell and M. R. Civanlar, "The feasible solution in signal restoration," *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. 32, pp. 201-212, April 1984.

[47] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM Journal on Scientific Computing,* vol. 17, pp. 227-238, January 1996.

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

[49] D. C. Youla and H. Webb, "Image restoration by the method of convex projections: Part 1 – Theory," *IEEE Transactions on Medical Imaging,* vol. 1, pp. 81-94, October 1982.