

# **Applied Mathematical Sciences**

Volume 147

*Editors*

S.S. Antman J.E. Marsden L. Sirovich

*Advisors*

J.K. Hale P. Holmes J. Keener  
J. Keller B.J. Matkowsky A. Mielke  
C.S. Peskin K.R.S. Sreenivasan

**Springer**

New York

Berlin

Heidelberg

Hong Kong

London

Milan

Paris

Tokyo

## Applied Mathematical Sciences

---

1. *John*: Partial Differential Equations, 4th ed.
2. *Sirovich*: Techniques of Asymptotic Analysis.
3. *Hale*: Theory of Functional Differential Equations, 2nd ed.
4. *Percus*: Combinatorial Methods.
5. *von Mises/Friedrichs*: Fluid Dynamics.
6. *Freiberger/Grenander*: A Short Course in Computational Probability and Statistics.
7. *Pipkin*: Lectures on Viscoelasticity Theory.
8. *Giacaglia*: Perturbation Methods in Non-linear Systems.
9. *Friedrichs*: Spectral Theory of Operators in Hilbert Space.
10. *Stroud*: Numerical Quadrature and Solution of Ordinary Differential Equations.
11. *Polovick*: Linear Multivariable Systems.
12. *Berkovitz*: Optimal Control Theory.
13. *Bluman/Cole*: Similarity Methods for Differential Equations.
14. *Yoshizawa*: Stability Theory and the Existence of Periodic Solution and Almost Periodic Solutions.
15. *Braun*: Differential Equations and Their Applications, 3rd ed.
16. *Lefschetz*: Applications of Algebraic Topology.
17. *Collatz/Wetterling*: Optimization Problems.
18. *Grenander*: Pattern Synthesis: Lectures in Pattern Theory, Vol. I.
19. *Marsden/McCracken*: Hopf Bifurcation and Its Applications.
20. *Driver*: Ordinary and Delay Differential Equations.
21. *Courant/Friedrichs*: Supersonic Flow and Shock Waves.
22. *Rouche/Habets/Laloy*: Stability Theory by Liapunov's Direct Method.
23. *Lamperti*: Stochastic Processes: A Survey of the Mathematical Theory.
24. *Grenander*: Pattern Analysis: Lectures in Pattern Theory, Vol. II.
25. *Davies*: Integral Transforms and Their Applications, 2nd ed.
26. *Kushner/Clark*: Stochastic Approximation Methods for Constrained and Unconstrained Systems.
27. *de Boor*: A Practical Guide to Splines: Revised Edition.
28. *Keilson*: Markov Chain Models—Rarity and Exponentiality.
29. *de Veubeke*: A Course in Elasticity.
30. *Sniatycki*: Geometric Quantization and Quantum Mechanics.
31. *Reid*: Sturmian Theory for Ordinary Differential Equations.
32. *Meis/Markowitz*: Numerical Solution of Partial Differential Equations.
33. *Grenander*: Regular Structures: Lectures in Pattern Theory, Vol. III.
34. *Kevorkian/Cole*: Perturbation Methods in Applied Mathematics.
35. *Carr*: Applications of Centre Manifold Theory.
36. *Bengtsson/Ghil/Källén*: Dynamic Meteorology: Data Assimilation Methods.
37. *Saperstone*: Semidynamical Systems in Infinite Dimensional Spaces.
38. *Lichtenberg/Lieberman*: Regular and Chaotic Dynamics, 2nd ed.
39. *Piccinini/Stampacchia/Vidossich*: Ordinary Differential Equations in  $\mathbb{R}^n$ .
40. *Naylor/Sell*: Linear Operator Theory in Engineering and Science.
41. *Sparrow*: The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors.
42. *Guckenheimer/Holmes*: Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields.
43. *Ockendon/Taylor*: Inviscid Fluid Flows.
44. *Pazy*: Semigroups of Linear Operators and Applications to Partial Differential Equations.
45. *Glashoff/Gustafson*: Linear Operations and Approximation: An Introduction to the Theoretical Analysis and Numerical Treatment of Semi-Infinite Programs.
46. *Wilcox*: Scattering Theory for Diffraction Gratings.
47. *Hale/Magalhães/Oliva*: Dynamics in Infinite Dimensions, 2nd ed.
48. *Murray*: Asymptotic Analysis.
49. *Ladyzhenskaya*: The Boundary-Value Problems of Mathematical Physics.
50. *Wilcox*: Sound Propagation in Stratified Fluids.
51. *Golubitsky/Schaeffer*: Bifurcation and Groups in Bifurcation Theory, Vol. I.
52. *Chipot*: Variational Inequalities and Flow in Porous Media.
53. *Majda*: Compressible Fluid Flow and Systems of Conservation Laws in Several Space Variables.
54. *Wasow*: Linear Turning Point Theory.
55. *Yosida*: Operational Calculus: A Theory of Hyperfunctions.
56. *Chang/Hwang*: Nonlinear Singular Perturbation Phenomena: Theory and Applications.
57. *Reinhardt*: Analysis of Approximation Methods for Differential and Integral Equations.
58. *Dwyer/Hussain/Voigt* (eds.): Theoretical Approaches to Turbulence.
59. *Sanders/Verhulst*: Averaging Methods in Nonlinear Dynamical Systems.

(continued following index)

Gilles Aubert Pierre Kornprobst

# Mathematical Problems in Image Processing

Partial Differential Equations and  
the Calculus of Variations

Foreword by Olivier Faugeras

With 93 Figures



Gilles Aubert  
Laboratoire J.-A. Dieudonne  
University of Nice-Sophia Antipolis  
U.M.R. no. 6621 du C.N.R.S.  
Parc Valrose  
06108 Nice Cedex 02  
France  
[gaubert@math.unice.fr](mailto:gaubert@math.unice.fr)

Pierre Kornprobst  
INRIA  
2004 Route des Lucioles  
06902 Sophia Antipolis  
France  
[Pierre.Kornprobst@sophia.inria.fr](mailto:Pierre.Kornprobst@sophia.inria.fr)

*Editors*

Stanislaw S. Antman  
Department of Mathematics  
and  
Institute for Physical Science  
and Technology  
University of Maryland  
College Park, MD 20742-4015  
USA

Jerrold E. Marsden  
Control and Dynamical  
Systems, 107-81  
California Institute of  
Technology  
Pasadena, CA 91125  
USA

Lawrence Sirovich  
Division of Applied  
Mathematics  
Brown University  
Providence, RI 02912  
USA

---

Mathematics Subject Classification (2000): 35J, 35L, 35Q, 49J, 49N

---

Library of Congress Cataloging-in-Publication Data

Aubert, Gilles.

Mathematical problems in image processing : partial differential equations and the

calculus of variations / Gilles Aubert, Pierre Kornprobst.

p. cm. — (Applied mathematical sciences ; 147)

Includes bibliographical references and index.

ISBN 0-387-95326-4 (alk. paper)

I. Image processing—Mathematics. I. Kornprobst, Pierre. II. Title. III. Applied  
mathematical sciences (Springer-Verlag New York, Inc.) ; v. 147.

QA1.A647

621.36'70151—dc21

2001041115

Printed on acid-free paper.

© 2002 Springer Verlag New York, LLC

All rights reserved. This work may not be translated or copied in whole or in part without the written  
permission of the publisher (Springer-Verlag New York, LLC, 175 Fifth Avenue, New York, NY  
10010, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in  
connection with any form of information storage and retrieval, electronic adaptation, computer  
software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.  
The use of general descriptive names, trade names, trademarks, etc., in this publication, even if  
the former are not especially identified, is not to be taken as a sign that such names, as understood by  
the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

Printed in the United States of America. (MVG)

9 8 7 6 5 4 3 2

SPIN 10987650

Springer Verlag is a part of *Springer Science+Business Media*

[springeronline.com](http://springeronline.com)

*To Jean-Michel Morel, whose ideas  
have deeply influenced the mathematical  
vision of image processing.*

# Foreword

Image processing, image analysis, computer vision, robot vision and machine vision are terms that refer to some aspects of the process of computing with images. This process has been made possible by the advent of computers powerful enough to cope with the large dimensionality of image data and the complexity of the algorithms that operate on them.

In brief these terms differ according to what kind of information is used and output by the process. In image processing the information is mostly the intensity values at the pixels, and the output is itself an image; in image analysis, the intensity values are enriched with some computed parameters, e.g. texture or optical flow, and by labels indicating such things as a region number or the presence of an edge; the output is usually some symbolic description of the content of the image, for example the objects present in the scene. Computer, robot and machine vision very often use three-dimensional information such as depth, three-dimensional velocity and perform some sort of abstract reasoning (as opposed to purely numerical processing) followed by decision making and action.

According to this rough classification this book deals with image processing and some image analysis.

These disciplines have a long history that can be traced back at least to the early sixties. For more than two decades, the field was occupied mostly by computer scientists and electrical engineers and did not attract much interest from mathematicians. Its rather low level of mathematical sophistication reflected the kind of mathematical training that computer scientists and electrical engineers were exposed to and, unfortunately, still are: it is roughly limited to a subset of 19th century mathematics. This is one reason. Another reason stems from the fact that simple heuristic methods, e.g. histogram equalisation, can produce apparently startling results; but these ad hoc approaches suffer from significant limitations, the main one being that there is no precise characterisation of why and when they work or don't work. The idea of the proof of correctness of an algorithm under a well-defined set of hypotheses has long been almost unheard of in image processing and analysis despite the strong connection with computer science.

It is clear that things have been changing at a regular pace for some time now. These changes are in my view due to two facts: first, the level of mathematical sophistication of researchers in computer vision has been steadily growing in the last twenty five years or so and second, the number

of professional mathematicians who develop an interest in this field of application has been regularly increasing thanks maybe to the examples set out by two Fields medallists, David Mumford and Pierre-Louis Lions. As a result of these facts the field of computer vision is going through a crucial mutation analog to the one that turned alchemy into modern chemistry.

If we now wonder what are the relevant mathematics to image processing and analysis we come up with a surprisingly long list: differential and Riemannian geometry, geometric algebra, functional analysis (calculus of variations and partial differential equations), probability theory (probabilistic inference, Bayesian probability theory), statistics (performance bounds, sampling algorithms), singularity theory (generic properties of solutions to partial differential equations) are all being successfully applied to image processing. It should be apparent that it is in fact the whole set of 20th century mathematics which is relevant to image processing and computer vision.

In what sense are those mathematics relevant? as I said earlier, many of the original algorithms were heuristic in nature, no proof was in general given of their correctness and no attempt was made at defining the hypotheses under which they would work or not. Mathematics clearly contribute to change this state of affairs by posing the problems in somewhat more abstract terms with the benefit of a clarification of the underlying concepts, e.g. what are the relevant functional spaces, the possibility of proving the existence and uniqueness of solutions to these problems under a set of well-defined hypotheses and the correctness of algorithms for computing these solutions. A further benefit of the increase of mathematical sophistication in machine vision may come out of the fact that the mathematical methods developed to analyse images with computers may be important for building a formal theory of biological vision: this was the hope of the late David Marr and should be considered as another challenge to mathematicians, computer vision scientists, psycho-physicists and neurophysiologists.

Conversely image processing and computer vision bring to mathematics a host of very challenging new problems and fascinating applications, they contribute to grounding them in the real world like physics does.

This book is a brilliant “tour de force” that shows the interest of using some of the most recent techniques of functional analysis and the theory of partial differential equations to study several fundamental questions in image processing such as how to restore a degraded image and how to segment it into meaningful regions. The reader will find early in the book a summary of the mathematical prerequisites as well as pointers to some specialised textbooks. These prerequisites are quite broad, ranging from direct methods in the calculus of variations (relaxation, Gamma-convergence) to the theory of viscosity solutions for Hamilton Jacobi equations and include the space of functions of bounded variations. Lebesgue theory of integration as well as Sobolev spaces are assumed to be part of the reader’s culture but pointers to some relevant textbooks are also provided.

The book can be read by professional mathematicians (and it is I think its prime target) as an example of the application of different parts of modern functional analysis to some attractive problems in image processing. These persons will find in the book most of the proofs of the main theorems (or pointers to these in the literature) and get a clear idea of the mathematical difficulty of these apparently simple problems. The proofs are well detailed, very clearly written and as a result, easy to follow. Moreover, since most theorems can also be turned into algorithms and computer programs, their conclusions are illustrated with spectacular results of processing performed on real images. Furthermore since the authors provide examples of several open mathematical questions my hope is that this book will attract more mathematicians to their study.

It can also be read by the mathematically inclined computer vision researcher. I do not want to convey the idea that I underestimate the amount of work necessary for such a person in order to grasp all the details of all the proofs but I think that it is possible as a first reading to get a general idea of the methods and the main results. Hopefully this person will then want to learn in more detail the relevant mathematics and this can be done by alternating reading the textbooks that are cited and studying the proofs in the book. My hope is that this will convince more image processing scientists that these mathematics must become part of the tools they use.

This book, written by two mathematicians with a strong interest in images, is a wonderful contribution to the mutation I was alluding to above, the transformation of image processing and analysis as well as computer, robot and machine vision into formalised fields, based on sets of competing scientific theories within which predictions can be performed and methods (algorithms) can be compared and evaluated. This is hopefully a step in the direction of understanding what it means to see.

Olivier Faugeras

# Preface

It is surprising when we realize just how much we are surrounded by images. Images allow us not only to perform complex tasks on a daily basis, but also to communicate, transmit information, represent and understand the world around us. Just think, for instance about digital television, medical imagery, video-surveillance, etc. The tremendous development in information technology accounts for most of this. We are now able to handle more and more data. Many day to day tasks are now fully or partially accomplished with the help of computers. Whenever images are involved we are entering the domains of computer vision and image processing. The requirements for this are reliability and speed. Efficient algorithms have to be proposed to process these digital data. It is also important to rely on a well-established theory to justify the well-founded nature of the methodology.

Amongst the numerous approaches which have been suggested, we focus on Partial Differential Equations (PDEs), and Variational Approaches in this book. Traditionally applied in physics, these methods have been successfully and widely transferred in Computer Vision over the last decade. One of the main interests in using PDEs is that the theory behind the concept is well-established. Of course, PDEs are written in a continuous setting referring to analog images, and once the existence and the uniqueness have been proven, we need to discretize them in order to find a numerical solution. It is our conviction that reasoning within a continuous framework makes the understanding of physical realities easier and stimulates the intuition necessary to propose new models. We hope that this book will illustrate this idea effectively.

The message we wish to put over is that the intuition which leads to certain formulations and the underlying theoretical study are often complementary. Developing a theoretical justification of a problem is not simply “art for art sake”. In particular, a deep understanding of the theoretical difficulties may lead to the development of suitable numerical schemes or different models.

This book is concerned with the mathematical study of certain image processing problems. Thus we target two audiences:

- The first is the mathematical community and is achieved by showing the contribution of mathematics to this domain by studying classical and challenging problems which come from Computer Vision. It is

also the occasion to highlight some difficult and unsolved theoretical questions.

- The second is the Computer Vision community: this is done by presenting a clear, self-contained and global overview of the mathematics involved for the problems of image restoration, image segmentation, sequence analysis and image classification.

We hope that this work will serve as a useful source of reference and inspiration for fellow researchers in Applied Mathematics and Computer Vision, as well as being a basis for advanced courses within these fields.

This book is divided into six main parts. Chapter 1 introduces the subject and gives a *detailed plan of the book*. In chapter 2, most of the mathematical notions used therein are recalled in an educative fashion and illustrated in detail. In Chapters 3 and 4 we examine how PDEs and variational methods can be successfully applied in the restoration and segmentation of one image. Chapter 5 is more applied and some challenging computer vision problems are described, such as sequence analysis or classification. As the final goal of any approach is to compute some numerical solution, we propose an introduction to the method of finite differences in the Appendix.

We would like to express our deep gratitude to the following people for their various contributions:

- The ARIANA group from INRIA Sophia Antipolis and in particular Laure Blanc-Féraud and Christophe Samson for providing the results regarding the classification problem.
- The ROBOTVIS group from INRIA Sophia Antipolis and especially Olivier Faugeras and Bertrand Thirion for their subsequent valuable comments.
- Agnés Desolneux, François Helt, Ron Kimmel, Etienne Mémin, Nikos Paragios, Luminita Vese and Joachim Weickert for their contribution to the writing of certain parts and for providing us with some experimental results.

# Contents

<i>Foreword</i> . . . . .	iii
<i>Preface</i> . . . . .	vii
<b>1 Introduction</b> . . . . .	
1.1 The image society . . . . .	1
1.2 What is a digital image? . . . . .	3
1.3 About Partial Differential Equations (PDEs) . . . . .	5
1.4 Detailed plan . . . . .	5
<b>Guide to relevant mathematical proofs</b> . . . . .	23
<b>Notations and symbols</b> . . . . .	25
<b>2 Mathematical preliminaries</b> . . . . .	
How to read this chapter? . . . . .	31
2.1 The direct method in the calculus of variations . . . . .	32
2.1.1 Topologies on Banach spaces . . . . .	32
2.1.2 Convexity and lower semi-continuity . . . . .	34
2.1.3 Relaxation . . . . .	39
2.1.4 About $\Gamma$ -Convergence . . . . .	42
2.2 The space of functions of bounded variation . . . . .	44
2.2.1 Basic definitions on measures . . . . .	44
2.2.2 Definition of $BV(\Omega)$ . . . . .	46
2.2.3 Properties of $BV(\Omega)$ . . . . .	48
2.2.4 Convex functions of measures . . . . .	51

2.3	Viscosity solutions in PDEs . . . . .	52
2.3.1	Around the eikonal equation . . . . .	52
2.3.2	Definition of viscosity solutions . . . . .	54
2.3.3	About the existence . . . . .	55
2.3.4	About the uniqueness . . . . .	56
2.4	Elements of differential geometry: the curvature . . . . .	58
2.4.1	Parametrized curves . . . . .	59
2.4.2	Curves as isolevel of a function $u$ . . . . .	60
2.4.3	Images as surfaces . . . . .	60
2.5	Other classical results used in this book . . . . .	61
2.5.1	Inequalities . . . . .	61
2.5.2	Calculus facts . . . . .	63
2.5.3	About convolution and smoothing . . . . .	64
2.5.4	Uniform convergence . . . . .	65
2.5.5	Dominated convergence theorem . . . . .	65
2.5.6	Well-posed problem . . . . .	65
<b>3</b>	<b>Image Restoration</b> . . . . .	<b>67</b>
	How to read this chapter? . . . . .	67
3.1	Image degradation . . . . .	68
3.2	The energy method . . . . .	70
3.2.1	An inverse problem . . . . .	70
3.2.2	Regularization of the problem . . . . .	71
3.2.3	Existence and uniqueness of a solution for the minimization problem . . . . .	74
3.2.4	Toward the numerical approximation . . . . .	77
3.2.5	Some invariances and the role of $\lambda$ . . . . .	85
3.2.6	Some remarks in the nonconvex case. . . . .	88
3.3	PDE-based methods . . . . .	92
3.3.1	Smoothing PDEs . . . . .	93
	The heat equation . . . . .	93
	Nonlinear diffusion . . . . .	96
	The Alvarez-Guichard-Lions-Morel scale space theory . . . . .	105
	Weickert's approach . . . . .	112
	Surface based approaches . . . . .	115
3.3.2	Smoothing-Enhancing PDEs . . . . .	120
	The Perona and Malik model [209] . . . . .	120
	Regularization of the Perona and Malik model: Catté et al [59] . . . . .	122
3.3.3	Enhancing PDEs . . . . .	127
	The Osher and Rudin's shock-filters [199] . . . . .	127
	A case study: construction of a solution by the method of characteristics . . . . .	128
	Comments on the shock-filter equation . . . . .	133

<b>4 The Segmentation Problem</b>	<b>137</b>
How to read this chapter? . . . . .	137
4.1 Definition and objectives . . . . .	138
4.2 The Mumford and Shah functional . . . . .	141
4.2.1 A minimization problem . . . . .	141
4.2.2 The mathematical framework for the existence of a solution . . . . .	141
4.2.3 Regularity of the edge set . . . . .	150
4.2.4 Approximations of the Mumford and Shah functional . . . . .	154
4.2.5 Experimental results . . . . .	159
4.3 Geodesic active contours and the level sets method . . . . .	161
4.3.1 The Kass-Witkin-Terzopoulos model [142] . . . . .	161
4.3.2 The Caselles-Kimmel-Sapiro geodesic active contours model [58] . . . . .	164
4.3.3 The level sets method . . . . .	170
4.3.4 Experimental results . . . . .	182
4.3.5 About some recent advances . . . . .	184
Global stopping criterion . . . . .	184
Toward more general shape representation . . . . .	186
<b>5 Other Challenging Applications</b>	<b>189</b>
How to read this chapter? . . . . .	189
5.1 Sequence analysis . . . . .	190
5.1.1 Introduction . . . . .	190
5.1.2 The optical flow: an apparent motion . . . . .	192
The Optical Flow Constraint (OFC) . . . . .	193
Solving the aperture problem . . . . .	194
Overview of a discontinuity preserving variational approach . . . . .	198
Alternatives of the OFC . . . . .	201
5.1.3 Sequence segmentation . . . . .	203
Introduction . . . . .	203
A variational formulation (the time-continuous case) .	204
Mathematical study of the time sampled energy .	207
Experiments . . . . .	210
5.1.4 Sequence restoration . . . . .	212
5.2 Image classification . . . . .	218
5.2.1 Introduction . . . . .	218
5.2.2 A level sets approach for image classification [221]	219
5.2.3 A variational model for image classification and restoration [222] . . . . .	224
<b>A Introduction to Finite Difference</b>	<b>237</b>
How to read this chapter? . . . . .	237

A.1	Definitions and theoretical considerations illustrated by the 1-D parabolic heat equation . . . . .	238
A.1.1	Getting started . . . . .	238
A.1.2	Convergence . . . . .	241
A.1.3	The Lax Theorem . . . . .	243
A.1.4	Consistency . . . . .	243
A.1.5	Stability . . . . .	245
A.2	Hyperbolic equations . . . . .	250
A.3	Difference schemes in image analysis . . . . .	259
A.3.1	Getting started . . . . .	259
A.3.2	Image restoration by energy minimization . . . . .	263
A.3.3	Image enhancement by the Osher and Rudin's shock-filters . . . . .	266
A.3.4	Curves evolution with the level sets method . . . . .	267
	Mean curvature motion . . . . .	268
	Constant speed evolution . . . . .	270
	The pure advection equation . . . . .	271
	Image segmentation by the geodesic active contour model . . . . .	271
<b>References</b>		<b>273</b>
<b>Index</b>		<b>291</b>

# 1

## Introduction

### 1.1 The image society

Our society is often designated as being an “information society”. It could also very well be defined as an “image society”. This is not only because image is a powerful and widely used media of communication, but also because it is an easy, compact and widespread way in which to represent the physical world. When we think about it, it is very surprising indeed (and maybe frightening) to realize just how much images are omni-present in our lives, and how much we rely upon them: just have a glance at the patchwork presented in Figure 1.1.

Advances made in acquisition devices are part of the origin of such a phenomenon. A huge amount of digital information is available. The second origin is naturally the increase in capacity of computers that enables us to process more and more data. This has brought about a new discipline known as Computer Vision.

For example, medical imagery made a substantial use of images from the earliest days. Many devices exist which are based on ultrasounds, Xrays and scanners, etc. Images produced by these can then be processed to improve their quality, enhance some features or efficiently combine different pieces of information (fusion).

Another important field that concerns us directly is remote sensing. This designates applications where we need to analyze, measure or interpret scenes at a distance. In addition to defense and video surveillance applications and road traffic analysis, the observation of the earths resources is

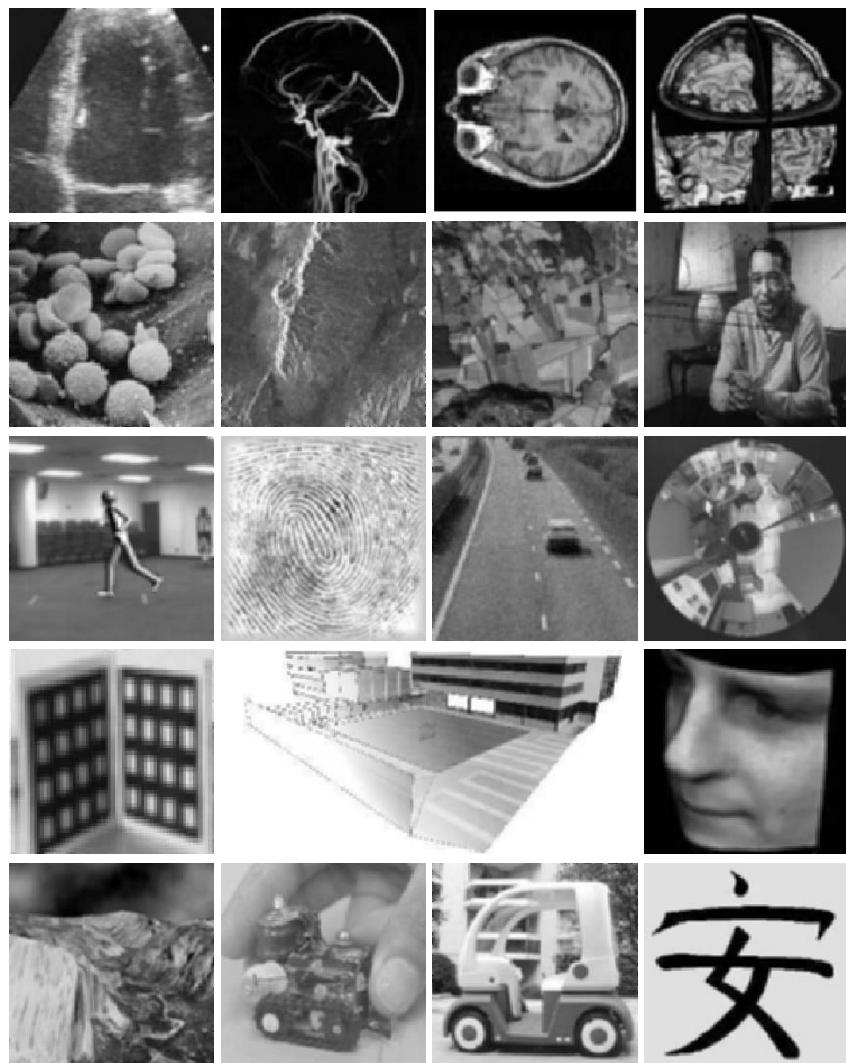


Figure 1.1. Illustration of some applications or systems that use image analysis. One may find in this patchwork examples from medicine, biology, satellite images, old movie restoration, forensic and video-surveillance, 3-D reconstructions and virtual reality, robotics or character recognition. Other applications include data compression, industrial quality control, fluids motion analysis, art (for instance for virtual databases or manuscript analysis), games, special effects, etc

another important field. Image processing provides tools to track and quantify changes in forests, water supplies, urbanization and pollution, etc. It is also widely used for weather forecasting to analyze huge amounts of data.

Video processing is clearly becoming an important area of investigation. This is because in many applications we need to process not only still images but also sequences of images. Motion analysis and segmentation are two important cues necessary for analyzing a sequence. This is necessary for instance for the forecasting of the weather in order to estimate and predict the movement of clouds. It is also determinant for more complex task such as compression. A clear understanding of the sequence in terms of background and foreground, with motion information enables us to describe a sequence with less given information. New challenges and subsequent problems are arising as video and cinema become digital: storage, special effects, video processing like the restoration of old movies, etc.

Beyond these general themes, we could also mention many different applications where image processing is involved. These include “World Wide Web”, character recognition, 3D reconstruction of scenes or objects from images, quality control, robotics, fingerprint analysis, virtual art databases, etc.

Without necessarily knowing it, we are consumers of image processing on a daily basis.

## 1.2 What is a digital image?

A digital image (also called discrete image) comes from a continuous world. It is obtained from an analog image by sampling and quantization. This process depends on the acquisition device and depends for instance on CCD's for digital cameras. Basically, the idea is to superimpose a regular grid on an analog image and to assign a digital number to each square of the grid, for example the average brightness in that square. Each square is called pixel, for picture element, and its value is the gray-level or brightness (see Figure 1.2).

Depending on the kind of image, the number of bits used to represent the pixel value may vary. Common pixel representations are unsigned bytes (0 to 255) and floating point. To describe a pixel, one may also need several channels (or bands). For instance, in the case of a color image, three channels are necessary, typically red, green, blue. In this book, we will only consider gray-scale images, with one channel.

The last important characteristic of an image is its size (or resolution). It is the number of rows and columns in the image. Just to give an idea, typical digital cameras now give images of size 320x240 and can reach 3060x2036 for professional ones. For digital cinema, we consider images of size 720x576 (standard video format), 1920x1440 (high definition) or higher, for medical

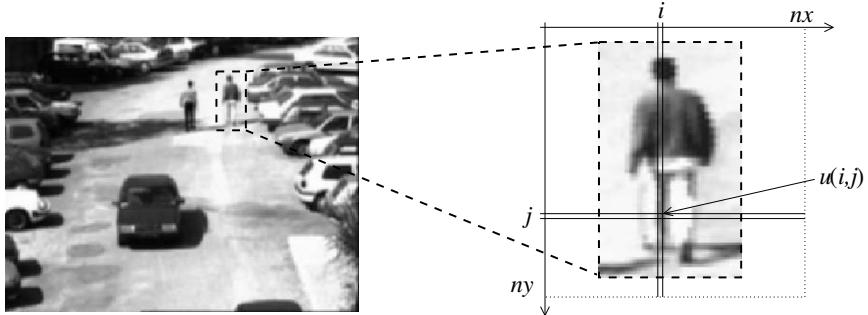


Figure 1.2. A digital image is nothing but a two dimensional array of pixels with assigned brightness values

imagery functional MRI images are about 128x128. In a way, the higher the resolution, the closer the digital image is from the physical world.

As in the real world, an image is composed of a wide variety of structures, and this is even more complex because of the digitalization and the limited number of gray-levels to represent it. To give an idea, we show in Figure 1.3 an image and some close-ups on different parts. This shows the effects of low resolution (some areas would need more pixels to be represented) and low contrasts, different kind of “textures”, progressive or sharp contours and fine objects. This gives an idea of the complexity to find an approach which permits to cope with the different problems or structures at the same time.

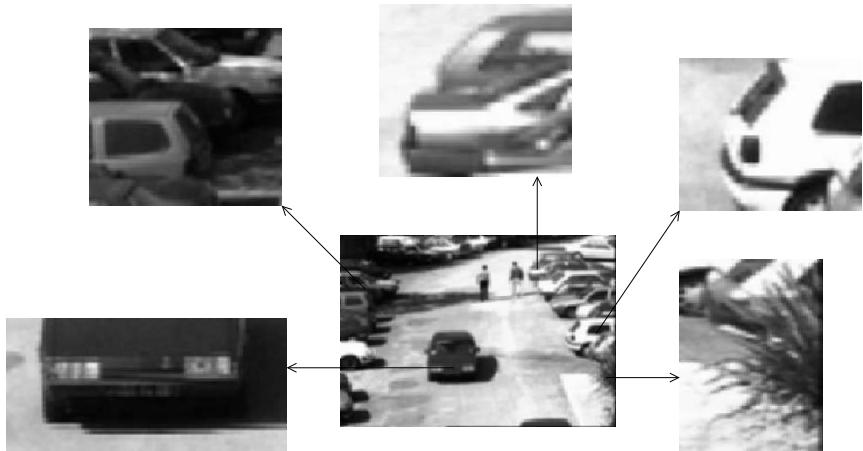


Figure 1.3. Digital image example. ↗ the close-ups show examples of low resolution, low contrasts, graduated shadings, sharp transitions and fine elements.

## 1.3 About Partial Differential Equations (PDEs)

Many approaches have been developed to process these digital images and it is difficult to say which one is more natural than the other. Image processing has a long history. Maybe the oldest methods come from 1-D signal processing techniques. They rely on the filter's theory (linear or not), on spectral analysis or on some basic concepts of probability or statistics. For an overview, we refer the interested reader to the book by Jain [140].

Today, more sophisticated tools have been developed. Three main directions emerge: stochastic modelization, wavelets and Partial Differential Equation (PDE) approaches. Stochastic modelization is widely based on the Markov Random Field theory (see [158, 118, 116]). It deals directly with digital images. The wavelet theory is inherited from signal processing and relies on decomposition techniques (see the monograph by S. Mallat[167] and [63, 95, 96, 98, 99, 100]) We do not consider here these approaches and focus instead on PDE based methods. They have been intensively developed in image analysis since the nineties.

PDEs, which are one of the most important parts of mathematical analysis, are closely related to the physical world. Every scientist comes across the wave equation or the heat equation, and the names of Euler, Poisson, Laplace, etc. are quite familiar. If PDEs originally come from physics and mechanics, one may encounter them more and more in other fields such as biology and finance and now in image analysis. One of the main interests of using PDEs is that the theory is well-established. Of course, PDEs are written in a continuous setting referring to analog images and once the existence and the uniqueness have been proven, we need to discretize them in order to find a numerical solution. It is our conviction that reasoning in a continuous framework makes the understanding of physical realities easier and provides the intuition to propose new models. We hope that this book will illustrate this idea well.

## 1.4 Detailed plan

This book is divided in five Chapters and an Appendix.

### **Chapter 2: Mathematical preliminaries**

---

We cover in this chapter most of the mathematics used in this book. We tried to make it as complete and educative as possible. Examples are given to emphasize the importance of certain notions, or to simply illustrate them. This chapter should be read carefully. It is divided in six sections.

**Section 2.1 The direct method in the calculus of variations.** Many

problems in computer vision can be formulated as minimization problems. Among the first concerns, we are interested in the existence of a solution. The presentation of the methodology to prove existence reveals two major notions: coercivity and lower semi-continuity (noted l.s.c., see Definition 2.1.3). Counter examples are proposed where these conditions are missing. Now, if the functional is not l.s.c., the idea is to look at the greatest l.s.c. functional equal or less than the initial one. It is called the relaxed functional (Definition 2.1.6). Some interesting properties link the relaxed and the initial problem (Theorem 2.1.6). This notion is quite technical but will be very useful in many occasions. Finally, we recall the notion of  $\Gamma$ -convergence introduced by De Giorgi [119, 85] (Section 2.1.4, Definition 2.1.7). This permits to define a convergence for a sequence of functionals. For instance, this can be used to approximate a problem for which characterizing the solution is not easy (as a typical application, we refer to Section 3.2.4). The main properties are stated in Theorem 2.1.7. The link between  $\Gamma$ -convergence and pointwise convergence is clarified in Theorem 2.1.8.

**Section 2.2 The space of functions of bounded variation.** It is the functional space that is commonly used in image analysis. The main reason is that, as opposed to classical Sobolev spaces [1], functions can be discontinuous across hypersurfaces. In terms of images, this means that images are discontinuous across edges. This is possible because derivatives are in fact measures. After recalling some basic definitions on measures in Section 2.2.1, the space  $BV(\Omega)$  is defined in Section 2.2.2. This space has interesting properties that we recall in Section 2.2.3: semi-continuity, a notion of trace, compactness and decomposability. The latter is maybe the more specific: it says that the derivative of a function  $u \in BV(\Omega)$  can be decomposed as:

$$Du = \nabla u \, dx + (u^+ - u^-) n_u \mathcal{H}_{|S_u}^{N-1} + C_u,$$

that is the sum of an absolutely continuous part with respect to the Lebesgue measure, a jump part and a Cantor measure. We finally generalize this decomposition to give a sense to convex functions of measures (Section 2.2.4).

**Section 2.3 Viscosity solutions in PDEs.** In many situations, we will encounter equations that do not come from variational principles, like in image restoration or segmentation by level sets. So we need to find a suitable framework to study these equations. To illustrate some of the problems, we consider in Section 2.3.1 the case of the 1-D eikonal equation:

$$\begin{cases} |u'(x)| = 1 & \text{in } (0, 1) \\ u(0) = u(1) = 0. \end{cases}$$

Existence, uniqueness and compatibility conditions are considered and reveal some difficulties. The framework of viscosity solution is then presented

(Section 2.3.2, Definition 2.3.1). Introduced in the eighties for first order PDEs by Crandall and P.L. Lions [82, 80, 160], the theory of viscosity solutions has shown a very successful development and has been extended for second order equations [81]. It is a notion of weak solutions that are continuous. To prove existence (Section 2.3.3), two approaches are classically used: the vanishing viscosity method (giving its name to the theory), and the Perron's method. As for uniqueness (Section 2.3.4), it is usually based on the Crandall-Ishii's lemma (Lemma 2.3.2). Since it is very technical, we try to illustrate on an easy example its role.

**Section 2.4 Elements of differential geometry: the curvature.** Image analysis has obviously some connections with differential geometry. For instance, curvature is a very important cue, and its definition depends on the “object” that we are considering. We define here the notion of curvature depending on we are interested in parametrized curves, images, or 3-D surfaces. We refer to [161] for more details.

**Section 2.5 Other classical and important result used in this book.** We finally mention some important results used in this book in different proofs. The motivation is to help the reader to find most of the results in this monograph, being as selfcontained as possible.

### Chapter 3: Image restoration

---

Many applications are based on images and then rely on their quality. Unfortunately, those images are not always of good quality for various reasons (defects in the sensors, natural noise, interferences, transmission, etc). Some “noise” is introduced, and it is important to consider automatic and efficient approaches to remove it. It is historically one of the oldest concerns and is still a necessary pre-processing for many applications. There exist many ways to tackle image restoration. Amongst others, we may quote the (linear) filtering approach [140], the stochastic modelization [114, 46, 73, 90, 116], and the variational/PDE based approaches. We focus here on the latter.

**Section 3.1 Image degradation.** The notion of noise is quite vague and it is often very hard to modelize its structure or its origins. Different models of degradation can be proposed and we briefly review some of them. In all the sequel, we will assume that the noise follows a Gaussian distribution.

**Section 3.2 The energy method.** Restoring an image can be seen as a minimization problem. We denote by  $u$  the original image and  $u_0$  the observed image, assuming the following model of degradation:

$$u_0 = R u + \eta$$

where  $\eta$  stands for an additive Gaussian noise and where  $R$  is a linear operator representing the blur (usually a convolution). We show in Section 3.2.2 that the problem of recovering  $u$  from  $u_0$  can be approximated by:

$$\inf_{u \in \{u \in L^2(\Omega); \nabla u \in L^1(\Omega)^2\}} E(u) = \frac{1}{2} \int_{\Omega} |u_0 - R u|^2 dx + \lambda \int_{\Omega} \phi(|\nabla u|) dx$$

where the function  $\phi$  is convex and of linear growth at infinity, allowing the preservation of discontinuities. The choice of this function  $\phi$  determines the smoothness of the resulting function  $u$ .

The mathematical study of this problem, considered in Section 3.2.3 reveals some major difficulties. The space  $V = \{u \in L^2(\Omega); \nabla u \in L^1(\Omega)^2\}$  is not reflexive and we cannot deduce any information from bounded minimizing sequences. So we relax the problem by studying it in the larger space  $BV(\Omega)$  and the correct formulation of the problem is (Theorem 3.2.1):

$$\begin{aligned} \inf_{u \in BV(\Omega)} \bar{E}(u) = & \frac{1}{2} \int_{\Omega} |u_0 - R u|^2 dx + \\ & + \lambda \int_{\Omega} \phi(|\nabla u|) dx + \lambda c \int_{S_u} (u^+ - u^-) d\mathcal{H}^1 + \lambda c \int_{\Omega - S_u} |Cu|. \end{aligned}$$

We demonstrate the existence and the uniqueness of a minimizer for  $\bar{E}(u)$  in the Theorem 3.2.2.

The Section 3.2.4 is about the numerical approximation of the solution. Although some characterization of the solution is possible in the distributional sense, it remains difficult to handle numerically. To circumvent the problem, we consider sequences of close functionals, using  $\Gamma$ -convergence. This is quite standard and shows clearly the importance of the notion of  $\Gamma$ -convergence.

The Section 3.2.5 investigates the role of the parameter  $\lambda$  with respect to the solution. This introduces the notion of scale and we can wonder about the meaning of an “optimal”  $\lambda$ . This question is usually handled in a stochastic framework and we mention some related work.

Until now, we have always considered the convex case. Surprisingly, we can observe very good numerical results using nonconvex functions  $\phi$ , for example with:

$$\phi(s) = \frac{s^2}{1+s^2}.$$

In the computer vision community, very little concerns is made on the interpretation of the result. It is a very hard problem indeed. For instance, we prove in Proposition 3.2.3 that the minimization problem:

$$\inf_{u \in W^{1,2}(\Omega)} E(u) = \int_{\Omega} |u - u_0^2| dx + \lambda \int_{\Omega} \frac{|\nabla u|^2}{1 + |\nabla u|^2} dx$$

has no minimizer in  $W^{1,2}(\Omega)$  and that the infimum is zero (if  $u_0$  is not a constant). It is still an open question to understand what the numerical solution is and we conclude by discussing some recent related work.

**Section 3.3 PDE-based methods.** A restored image can be seen as a version of the initial image at a special scale. More precisely, we consider here models where an image  $u$  is embedded in a evolution process. We denote it by  $u(t, \cdot)$ . At time  $t = 0$ ,  $u(0, \cdot) = u_0(\cdot)$  is the original image. It is then transformed through a process that can be written:

$$\frac{\partial u}{\partial t}(t, x) + F(x, u(t, x), \nabla u(t, x), \nabla^2 u(t, x)) = 0 \quad \text{in } \Omega \\ (\text{A second order differential operator})$$

which is a partial differential equation. In other words, the evolution of  $u$  may depend on its derivatives at different orders. This is a very generic form and we show in this section some possibilities for  $F$  to restore an image. We distinguished three categories.

- Smoothing PDEs (Section 3.3.1). Maybe the most famous PDE in image restoration is the heat equation:

$$\frac{\partial u}{\partial t}(t, x) - \Delta u(t, x) = 0. \quad (1.1)$$

We recall some of its main properties: the equivalence with a Gaussian convolution, a low-pass filter, and some invariances of the operator  $T_t$  defined by  $(T_t u_0)(x) = u(t, x)$ , where  $u(t, x)$  is the unique solution of (1.1). Because its oversmoothing property (edges get smeared), it is necessary to introduce some nonlinearity. We then consider the model:

$$\frac{\partial u}{\partial t} = \operatorname{div}(c(|\nabla u|^2) \nabla u), \quad (1.2)$$

where the function  $c$  is fixed here so that the equation remains parabolic. If we compare to the heat equation, this does not seem to introduce major changes, at least on a theoretical point of view. In order to preserve the discontinuities, we show that we would like to have  $c(s) \approx \frac{1}{\sqrt{s}}$  as  $s \rightarrow +\infty$ . Unfortunately, because of this degenerated behaviour, it is no longer possible to apply general results for parabolic equations. A well-adapted framework to study this equation is the nonlinear semi-group theory. The idea is to show that the divergence operator in (1.2) is maximal monotone. A convenient way to prove it is to identify the divergence operator with the subdifferential of a convex lower semi-continuous functional  $\bar{J}(u)$  given in (3.48). We can then establish in Theorem 3.3.1 the existence and the uniqueness of a solution. The characterization of the solution:

$$“u(t) \in \operatorname{Dom}(\partial \bar{J}) \quad \dots / \dots \quad -\frac{du}{dt} \in \partial \bar{J}(u(t))”.$$

is not very explicit and we try in Proposition 3.3.4 to better explain it.

In 1992 Alvarez-Guichard-Lions-Morel introduced the notion of scale-space via PDEs [4, 5]. Given some axioms and invariance properties for an “image oriented” operator  $T_t$ , the idea is to try to identify this operator. This is a major contribution and a very original work. More precisely, under previous assumptions, it can be established (Theorem 3.3.2) that  $u(t, x) = (T_t u_0)(x)$  is the unique viscosity solution of:

$$\frac{\partial u}{\partial t} = F(\nabla u, \nabla^2 u).$$

In other words, if  $T_t$  verify some natural assumptions, then it can be solved through a PDE depending only on the first and second derivatives of  $u$ . With more assumptions,  $F$  can be fully determined (Theorems 3.3.3 and 3.3.4).

We then present briefly the Weickert’s approach. We refer the interested reader to [249] for more details. Roughly speaking, it is a tensor based version of the equation (1.2) where the scalar coefficient  $c$  which controls the diffusion is replaced by a function of the diffusion *tensor*:

$$\nabla u \nabla u^t = \begin{pmatrix} u_x^2 & u_x u_y \\ u_x u_y & u_y^2 \end{pmatrix}.$$

Some Gaussian convolutions are introduced to take into account the scales. This approach permits to better take into account the directions of diffusion and is theoretically justified [249].

Last but not least, we mention some contributions by Fallah, Kimmel et al [233, 232]. The specificity of these approaches is to consider an image as a surface. The differential operators are then based on the metric properties of the surface. Instead of recalling each model separately, we try to present them in a common framework.

- Smoothing-enhancing PDEs (Section 3.3.2). We consider equations that can behave locally as inverse heat equations. This is motivated by the original choices of  $c$  by Perona and Malik [209] in the equation (1.2):

$$c(s) = \frac{1}{1 + \frac{s}{k}} \quad \text{or} \quad c(s) = e^{-\frac{s}{k}} \quad (1.3)$$

where  $k$  is a constant. If we denote by  $T$  and  $N$  the tangent and normal directions to the isophotes, the divergence term in (1.2) can be rewritten as:

$$\operatorname{div}(c(|\nabla u(t, x)|^2) \nabla u(t, x)) = c(|\nabla u(t, x)|^2) u_{TT} + b(|\nabla u(t, x)|^2) u_{NN}$$

where the function  $b$  depends on  $c$ . With choices of  $c$  like (1.3), it can be verified that the function  $b$  may become negative. Classical argu-

ments as in the previous section can no longer be applied. It is in fact the notion of “solution” itself that has to be defined differently. From Kichenassamy [143], a solution should consist of regions with “low” gradients separated by points of discontinuity where the gradient is infinite (Theorem 3.3.6). Therefore, the notion of solution must be understood in a measure sense, and it is still an open problem. Another possibility that we detail here is to introduce some regularization. Catté et al [59] proposed to solve the following partial differential equation instead of (1.2):

$$\frac{\partial u}{\partial t}(t, x) = \operatorname{div}(c(|(\nabla G_\sigma * u)(t, x)|^2) \nabla u(t, x)).$$

Interestingly, we show that this equation is now well-posed and we prove in Theorem 3.73 the existence and uniqueness of a solution in the distributional sense. The existence is shown using a Schauder’s fixed-point theorem.

Some challenging questions are still open and mentioned at the end of this section.

- Enhancing PDEs (Section 3.3.3). In this last section, we examine the case of deblurring (or enhancement). It is essentially devoted to the shock filter model proposed by Osher and Rudin [199] in the 1-D case:

$$u_t(t, x) = -|u_x(t, x)| \operatorname{sign}(u_{xx}(t, x)). \quad (1.4)$$

To better understand the action of this equation, we consider the following simpler case:

$$u_t(t, x) = -|u_x(t, x)| \operatorname{sign}((u_0)_{xx}(t, x))$$

with  $u_0(x) = \cos(x)$ . Using the method of characteristics, we construct explicitly a solution. Although this cannot be done in the general case (1.4), this presents interesting calculus. The understanding of (1.4) on a theoretical point of view remains an open question up to our knowledge.



## Chapter 4: The Segmentation Problem

---

In this chapter, we examine the segmentation problem which is, with image restoration, one of the most important question in image analysis. Though restoration and segmentation are not totally disconnected, image segmentation has its own objectives (see Section 4.1) and its own methodology. By segmentation, we mean that we are interested to know the constituant parts of an image rather than to improve its quality (which is a restoration task). The aim of Chapter 4 is to present two variational and PDE approaches for the segmentation problem. In the first approach, the idea is to consider

that segmenting an image consists, from an initial image  $u_0$ , in constructing a new image  $u$  closed to  $u_0$  and made of distinct homogeneous regions, the boundaries between these regions being sharp and piecewise regular. This is achieved by minimizing the Mumford-Shah functional (Section 4.2). The second approach (which can be seen as a dual point of view of the former) aims at detecting the contours of the objects lying in the image  $u_0$  (or a smooth version of it). These contours are modelized by closed curves to be identified. This edge detection method called geodesic active contours is presented in Section 4.3.

**Section 4.1: Definition and objectives.** Based on some real examples, we suggest what could image segmentation possibly mean. There are two different ways to consider it. The first is to have a simplified version of the original image, compounded of homogeneous regions separated by sharp edges. The second is to extract some “significant” contours. Still, in each case, the important features are edges and so we briefly recall some of the earliest edge detectors that appeared in literature.

**Section 4.2: The Mumford Shah functional.** In Section 4.2.1, we first present the formulation introduced by Mumford and Shah in 1985 [186]. It is an energy based method. For a given  $u_0(x)$  (the initial image), we search for a function  $u : \Omega \rightarrow R$  and a set  $K \subset \Omega$  (the discontinuity set) minimizing:

$$E(u, K) = \int_{\Omega} (u - u_0)^2 dx + \alpha \int_{\Omega - K} |\nabla u|^2 dx + \beta \oint_K d\sigma.$$

The first term measures the fidelity to the data, the second imposes that  $u$  is smooth in the region  $\Omega - K$  and the third term that the discontinuity set has minimal length and therefore is as smooth as possible. This type of functional forms part of a wider class of problems called *“free discontinuity problems”*. As we can imagine, there is an important literature related to the Mumford-Shah problem. We refer to the book of Morel and Solimini [185] for a complete review. Our aim is to present to the reader a clear overview of the main results.

Section 4.2.2 is concerned with the existence of a solution. As we can foresee, it is not an easy task and the first question is: what is the good functional framework? Following Ambrosio [9, 10, 11], we show why the problem cannot be directly solved and we demonstrate the necessity to define an equivalent formulation involving only one unknown  $u$ . This is achieved by considering the functional:

$$G(u) = \int_{\Omega} (u - u_0)^2 dx + \alpha \int_{\Omega} |\nabla u|^2 dx + \beta \mathcal{H}^{N-1}(S_u).$$

defined for  $u \in SBV(\Omega)$  and where  $S_u$  stands for the discontinuity set of  $u$ , i.e. the complement, up to a set of  $\mathcal{H}^{N-1}$  measure zero, of the set of Lebesgue points of  $u$ . It is proven that the problem  $\inf \{G(u), u \in SBV(\Omega) \cap L^\infty(\Omega)\}$  admits at least a solution  $u$  and that the couple  $(u, K)$  with  $K = \Omega \cap \overline{S_u}$  is a minimizer of the initial Mumford-Shah functional.

Once we get the existence of a minimizer, we would like to compute it. The natural way to do that is to search for optimality conditions. We establish these conditions (Theorem 4.2.3) assuming the regularity hypotheses:

- (C<sub>1</sub>)  $K$  is made of a finite number of  $C^{1,1}$ -curves, meeting  $\partial\Omega$  and meeting each other only at their endpoints.
- (C<sub>2</sub>)  $u$  is  $C^1$  on each connected component of  $\Omega - K$ .

Section 4.2.3 investigates the regularity hypotheses (C<sub>1</sub>), (C<sub>2</sub>). In their seminal work [186], Mumford and Shah conjectured that the functional  $F(u, K)$  admits a minimizer satisfying (C<sub>1</sub>), (C<sub>2</sub>) and then they proved that for such a minimizer, the only vertices of the curves  $\gamma_i$  forming  $K$  are:

- (i) Points  $P$  on  $\partial\Omega$  where one  $\gamma_i$  meets  $\partial\Omega$  perpendicularly.
- (ii) Triple points  $P$  where three  $\gamma_i$  meet with angles  $\frac{2\pi}{3}$ .
- (iii) Crack-tip where a  $\gamma_i$  ends and meets nothing.

We first show qualitative properties (i), (ii) and (iii) when  $K$  verifies (C<sub>1</sub>) and (C<sub>2</sub>). Then, we state the important result from Bonnet [41, 43] where the hypotheses (C<sub>1</sub>) and (C<sub>2</sub>) are removed and replaced by a connectedness assumption.

The approximation of the Mumford-Shah functional is examined in Section 4.2.4. The lack of differentiability for a suitable norm does not allow to use directly Euler equations. Therefore, it is natural to search for an approximation of  $F(u, K)$  by a sequence  $F_\varepsilon$  of regular functionals defined on Sobolev spaces. There exist many ways to approximate  $F(u, K)$  and we list some of them. We focus our attention on the Ambrosio-Tortorelli approximation [13] which is one of the most used in image analysis. In their approach, the set  $K$  (or  $S_u$ ) is replaced by an auxiliary variable  $v$  (a function) that approximates the characteristic function  $(1 - \chi_K)$ . The sequence of functionals  $F_\varepsilon$  is defined by:

$$\begin{aligned} F_\varepsilon(u, v) = & \int_{\Omega} (u - u_0)^2 dx + \int_{\Omega} (v^2 + h(\varepsilon)) |\nabla u|^2 dx + \\ & + \int_{\Omega} \left( \varepsilon |\nabla v|^2 + \frac{1}{4\varepsilon} (v - 1)^2 \right) dx \end{aligned}$$

where  $h(\varepsilon)$  is a sequence of constants such that  $\lim_{\varepsilon \rightarrow 0} h(\varepsilon) = 0$ .  $F_\varepsilon(u, v)$  are elliptic functionals which  $\Gamma$ -converge to the Mumford and Shah functional.

We close Section 4.2 by saying some words about the numerical computation and giving some experimental results (Section 4.2.5). We point out the monograph of Braides [48] to anyone interested in the approximation of free discontinuity problems.

**Section 4.3: Geodesic active contours and the level sets method.** As suggested in Section 4.1 another approach for segmentation is to detect the “significant” contours, for example the contours of the different objects in a scene rather than to construct a partition made of homogeneous regions. This is the active contour model. The idea of this approach is that contours are characterized by sharp variations of the image intensity, and so by infinite gradients. The principle consists in matching deformable curves to the contours objects by means of energy minimization.

In Section 4.3.1, we begin by describing the Kass-Witkin-Terzopoulos model [142] which is, to the best of our knowledge, one of the first in this direction. In this model, the energy to be minimized is defined on a set  $C$  of closed, piecewise regular, parametric curves (called snakes) and is given by:

$$J_1(c) = \int_a^b |c'(q)|^2 dq + \beta \int_a^b |c''(q)|^2 dq + \lambda \int_a^b g^2(|\nabla I(c(q))|) dq,$$

$c \in C$ , where  $g$  is a decreasing monotonic function vanishing at infinity. Thanks to the third term, minimizing curves are attracting by the edges of the objects forming the image. We point out some drawbacks of this model (in particular, the non intrinsic nature of  $J_1$  which depends upon the parametrization) that lead Caselles, Kimmel and Sapiro to propose their geodesic active contour model [58] (Section 4.3.2). They observed that in the energy  $J_1$ , we can choose  $\beta = 0$ , and that, in a sense to be precised (Definition 4.3.1), the minimization of  $J_1(c)$  (with  $\beta = 0$ ) is equivalent to the minimization of:

$$J_2(c) = 2\sqrt{\lambda} \int_a^b g(|\nabla I(c(q))|) |c'(q)| dq.$$

Now,  $J_2$  is invariant under change of parametrization and can be seen as a weighted Euclidian length. Starting from an initial curve  $c_0(q)$ , we detect the image contours by evolving a family of curves  $c(t, q)$  according to the decreasing gradient flow associated to  $J_2$ :

$$\frac{\partial c}{\partial t} = (\kappa g - \langle \nabla g, N \rangle) N. \quad (1.5)$$

where  $N$  is the unit outward normal to  $c(t, q)$  and  $\kappa$  its curvature. Thanks to the definition of  $g$ , the evolution of  $\{c(t, q)\}_{t \geq 0}$  is stopped when edges are detected. However, due to its parametric formulation, this model has still

some drawbacks. For example, a change of topology (corresponding to the detection of several objects) during the curve evolution is not allowed. To circumvent this type of problem, Osher and Sethian proposed in a pioneering work [200] the so-called level set method that we examine in Section 4.3.3. The idea is simple: a curve in  $R^2$  can be seen as the zero-level line of a function  $u$  defined from  $R^2$  to  $R$ . We reformulate the geodesic active contour model in terms of the function  $u$  and we deduce that equation (1.5) can be rewritten as:

$$\frac{\partial u}{\partial t} = \kappa g |\nabla u(t, x)| - \langle \nabla g, \nabla u \rangle. \quad (1.6)$$

Then, we list the major advantages of this formulation and we show that equations of the form (1.6) are well-posed in the viscosity sense. We develop in details the mathematical analysis of (1.6) (existence, uniqueness, maximum principles).

Section 4.3.4 is concerned with the experimental results obtained with (1.6). The main difficulty in the numerical approximation of (1.6) comes from the presence of hyperbolic terms. It is suggested why the discretization has to be carried out carefully, which leads to the introduction of entropic conservatives schemes. The interested reader will find more details in the Appendix (Sections A.2 and A.3.4).

Finally, we point out in Section 4.3.5 how some limitations from the snakes or level-set formulations may be overcome. Two recent contributions are presented. The first one is concerned with the stopping criterium. Classically based on the intensity gradient (corresponding to sharp edges), this criterium may not be suitable for some kind of images with soft contours or “perceptual” contours (imagine an image with dots not evenly distributed). The idea proposed by Chan and Vese [66] is to consider what is inside the regions instead of only focusing on the boundaries. The second one is related to the curves representation using level sets. Although it is very convenient numerically, it can only deal with closed non-intersecting hypersurfaces. We present some recent developments [219, 124] who describe a more flexible representation that should be more investigated in the future.



## Chapter 5: Other Challenging Applications

---

The scope of this last chapter is more applied. Two main problems are analyzed: sequence analysis and classification (a supervised segmentation). Although some theoretical results are given and proved, the goal of this chapter is mainly to show how the previous material may be used for more “applied” and complex problems.

### Section 5.1: Sequence analysis

As the terms “sequence analysis” may appear quite wide and unprecise, we begin in Section 5.1.1 by giving some generalities about sequences. The aim is to make the reader realize that one uses all its knowledge of the environment or scenario to analyze a sequence. Without this, it becomes a very difficult problem.

One of the first difficulty is the motion estimation (Section 5.1.2). Beyond the fact that motions may be of different types and amplitudes, it can only be recovered thanks to intensity variations. This explains why it is only possible to recover an apparent motion (*a priori* different from the projection of the 3D real motion field). This section is essentially a review of existing variational approaches for optical flow. We first present the so-called optical flow constraint, which is a scalar equation corresponding to the conservation of intensity along trajectories. It links the intensity of the sequence  $u(t, x) : R^+ \times \Omega \rightarrow R$  (the data) to the instantaneous motion field at time  $t = t_0$ ,  $\sigma(x) : \Omega \rightarrow R^2$  (the unknown), by:

$$\sigma(x) \cdot \nabla u(t, x_0) + u_t(t, x_0) = 0.$$

As it is not sufficient to solve the problem (one equation for two unknowns: the two components of the flow field), many solutions have been proposed and we recall some of them. We focus on regularization approaches and mention some of the numerous research based on this method. We then present in more details a discontinuity preserving approach by Aubert, Deriche and Kornprobst [92, 17, 18] where some theoretical results have been established. The problem is to minimize:

$$E(\sigma) = \underbrace{\int_{\Omega} |\sigma \cdot Du + u_t|}_{A(\sigma)} + \underbrace{\alpha^s \sum_{j=1}^2 \int_{\Omega} \phi(D\sigma_j)}_{S(\sigma)} + \underbrace{\alpha^h \int_{\Omega} c(Du) |\sigma|^2}_{H(\sigma)} dx$$

where  $\alpha^s, \alpha^h$  are positive constants. From now on, unless specified otherwise, all the derivative are written in a formal setting (i.e.  $Du$  is the gradient of  $u$  in the distributional sense). Since we look for discontinuous optical flows, the suitable theoretical background to study this problem will be the space of bounded variation  $\mathbf{BV}(\Omega)$  (see Section 2.2). The energy is compounded of three terms:

- $A(\sigma)$  is the “ $L^1$ ”-norm of the optical flow constraint. In fact, as it is formal here, it has to be interpreted as a measure.
- $S(\sigma)$  is the smoothing term, chosen as for image restoration (see Section 3.2), in order to keep the discontinuities of the flow.
- $H(\sigma)$  is related to homogeneous regions. The idea is that if there is no texture, that is to say no gradient, there is no way to estimate correctly the flow field. Then one may force it to be zero.

We conclude this section by discussing the well-foundness of the optical flow constraint in some situations, and try to indicate some future paths for this still challenging problem.

The second problem that we consider in Section 5.1.3 is the problem of sequence segmentation. Here, segmentation means finding the different objects in a scene, and it is naturally in relation with velocity estimation. Motion based segmentation perform well on some sequences (low noise, good sampling,...) but will fail otherwise. Another idea is to use the background redundancy to segment the scene into layers (foreground/background). It is shown on a synthetic example how this idea may be considered and what are the limits of the naive approach. In particular, in the case of noisy sequences, obtaining a reference image is as difficult as segmenting the sequence. So it would be certainly more efficient to look for the reference image and the segmentation at the same time. This can be formalized in terms of PDEs as done by Kornprobst, Deriche and Aubert [154], where a coupled approach is proposed. Let  $N(t, x)$  denotes the given noisy sequence, for which the background is assumed to be static. We look simultaneously for:

- The restored background  $B(x)$ .
- The sequence  $C(t, x)$  which indicates the moving regions. Typically,  $C(t, x) = 0$  if the pixel  $x$  belongs to a moving object at time  $t$ , and 1 otherwise.

The minimization problem is:

$$\inf_{B, C} \underbrace{\iint_V C^2(B - N)^2 + \alpha_c(C - 1)^2 dxdt}_{A(\sigma)} + \underbrace{\alpha_b^r \int_{\Omega} \phi_1(DB) + \alpha_c^r \iint_V \phi_2(DC)}_{S(\sigma)}$$

where  $\alpha_c$ ,  $\alpha_b^r$ ,  $\alpha_c^r$  are positive constants. The energy is compounded of two kinds of terms:

- $A(\sigma)$  realizes the coupling between the two variables  $B$  and  $C$ . The second term forces the function  $C$  to be equal to 1, which corresponds to the background. However if the current image  $N$  is too different from the background (meaning that an object is present), then the first term will be too high, which forces  $C$  to be 0.
- $S(\sigma)$  is the smoothing term. As usual, the functions  $\phi_i$  are chosen in order to preserve the discontinuities (see Section 3.2.2).

Although this functional is globally nonconvex and degenerated, existence and uniqueness can be proven on the space of bounded variation, with a condition on the coefficient  $\alpha_c$ . The performance of the algorithm is illustrated on several real sequences.

Finally, we consider in Section 5.1.4 the problem of sequence restoration and more precisely movie restoration. Little research has been carried out on this topic, and especially using PDEs. The aim of this section is essentially to make the reader realize the difficulties of this task. It is mainly due to the wide variety of ways a film can be damaged (chemical, mechanical, handling, etc). Faced with this variety of defects, we propose to classify them, by taking into account the degree of human interaction which is necessary to detect and/or correct them. Interestingly, this degree of interaction/automation can also be found in the different systems that are proposed nowadays. Something important which comes out of this discussion is that no system is able to deal with every kind of defects. Finally, by trying to define what should be a “perfect” restoration, it turns out that it has to be perceptually correct: it will be sufficient for a defect to no longer be perceived even if it has not been completely removed. More than any other domain in computer vision, movie restoration should benefit from advances and studies of human perception (see [141] for instance). Some recent contributions using the PDE framework conclude this part.

**Section 5.2 Image classification**, which consists in assigning a label to each pixel of an image, has rarely been introduced in a variational formulation (continuous models), mainly because the notion of classes has a discrete nature. Many classification models have been developed with structural notions as region growing methods for example [207], or by stochastic approach (discrete models) with the use of Markov Random Field theory [33, 258]. It is considered that stochastic methods are robust but nonetheless time consuming. The goal of this section is to show how PDEs and variational techniques can also solve some image classification problems. Two variational models are presented. For these two models, it is assumed that each pixel is only characterized by its intensity level, that each class  $C_i$  has a Gaussian distribution of intensity  $N(\mu_i, \sigma_i)$  and that the number of classes  $K$  and the parameters  $(\mu_i, \sigma_i)$  are known (supervised classification) (see Section 5.2.1).

In Section 5.2.2, we present a level-set formulation [221]. Here, classification is seen as a partitioning problem. If  $\Omega$  is the image domain and  $u_0$  is the data, we search for a partition of  $\Omega$  defined by:

$$\Omega = \bigcup_{i=1}^K (\Omega_i \cup \Gamma_i) \quad \text{and} \quad \Omega_i \cap \Omega_j = \emptyset, \quad i \neq j$$

where  $\Gamma_i = \partial\Omega_i \cap \Omega$  is the intersection of the boundary of  $\Omega_i$  with  $\Omega$  and  $\Gamma_{ij} = \Gamma_{ji} = \Gamma_i \cap \Gamma_j$ ,  $i \neq j$ , the interface between  $\Omega_i$  and  $\Omega_j$ . Moreover, to be admissible, the partitioning process has to satisfy some constraints:

- Taking into account the Gaussian distribution property of the classes.
- Ensuring the regularity of each interface.

This can be achieved by identifying each  $\Omega_i$  to an upper level set of a signed function  $\phi_i(\cdot)$ , i.e.  $\phi_i(x) > 0$  if  $x \in \Omega_i$ ,  $\phi_i(x) = 0$  if  $x \in \partial\Omega_i$  and  $\phi_i(x) < 0$  otherwise.

These three requirements (partitioning, Gaussian distribution of classes, regularity of interfaces) can be attained through the minimization of the following global energy defined on  $\Omega$ :

$$\begin{aligned} F(\phi_1, \dots, \phi_K) = & \int_{\Omega} \left( \sum_{i=1}^K H(\phi_i(x)) - 1 \right)^2 dx + \\ & + \sum_{i=1}^K e_i \int_{\Omega} H(\phi_i(x)) \frac{(u_0(x) - \mu_i)^2}{\sigma_i^2} dx + \sum_{i=1}^K \gamma_i \int_{\phi_i=0} ds \end{aligned}$$

where  $H(s)$  is the Heaviside function:  $H(s) = 1$  if  $s > 0$  and  $H(s) = 0$  if  $s < 0$ . Unfortunately, the functional  $F$  has some drawbacks (non differentiability, presence of a boundary term,...). Thus, instead of  $F$ , we propose to minimize an approximated  $F_\alpha$  close to  $F$  (as  $\alpha \rightarrow 0$ ) and we show on several experimental results the soundness of the model.

Section 5.2.3 presents another model of classification coupled with a restoration process [222]. It is based on the theory of phase transitions developed in mechanics [3, 53]. The model still relies on the minimization of an energy whose generic form is:

$$J_\varepsilon(u) = \int_{\Omega} (u(x) - u_0(x))^2 dx + \lambda^2 \varepsilon \int_{\Omega} \varphi(|\nabla u(x)|) dx + \frac{\eta^2}{\varepsilon} \int_{\Omega} W(u(x), \mu, \sigma) dx.$$

The first two integrals in  $J_\varepsilon$  are the usual terms in restoration (we can choose for instance  $\phi(s) = s^2$ , or  $\phi(s) = \sqrt{1+s^2}$ ). The third integral is a classification term. The role of the function  $W(u)$  is to attract the values of  $u$  towards the labels of the classes. Since the chosen labels are the mean  $\mu_i$  of the classes, good candidates for  $W$  are those which satisfy  $W(\mu_i) = 0$ ,  $i = 1,..K$ ,  $W(u) \geq 0$  for all  $u$ . Such functions are known as multiple wells potentials. The parameter  $\varepsilon > 0$  is to be destined to tend to zero and its contribution is major in the restoration-classification process. Roughly speaking, as  $\varepsilon$  decreases, during the first steps of the algorithm, the weight of  $\frac{1}{\varepsilon} \int_{\Omega} W(u(x), \mu, \sigma) dx$  is negligible with respect to the two others and

only the restoration process runs. As  $\varepsilon$  becomes smaller, the diffusion is progressively stopped while the classification procedure has become active. This phenomenon is illustrated on experimental results.

Concerning the mathematical aspect of the approach, we show that the model is well-posed when the regularization function is  $\phi(s) = s^2$ . We study the asymptotical behaviour (as  $\varepsilon \rightarrow 0$ ) of  $J_\varepsilon(u)$  as well as the one of a sequence of minimizers  $u_\varepsilon$  of  $J_\varepsilon$ . The proofs rely on the  $\Gamma$ -convergence theory and are borrowed from previous works related to the theory of phase

transitions [179, 234, 235]. We conclude this section by giving some illustrative examples on synthetic and real images. 

## Appendix: Introduction to finite difference schemes in image analysis

The scope of this monograph is to explain how PDEs can help to modelize certain image processing problems and to justify the models. Until now, we have always considered a continuous setting ( $x \in \Omega \subset R^2$ ) referring to analog images. However, working in this field requires also to test the models in the digital (discrete) world. So we need to know how to discretize the equations. Although several kind of approaches can be considered (like finite elements or spectral methods), the success of finite differences in image analysis is due to the structure of digital images for which we can associate a natural regular fixed grid. The aim of this chapter is to give the basis and the main notions of finite differences methods. It is also to write the discretization of some complex operators presented in this book.

**Section A.1: Definitions and theoretical considerations illustrated by the 1-D heat equation.** This section gives the main notions and definitions for discrete schemes (convergence, consistency and stability). Every notion is illustrated by developing explicit calculus for the case of the 1-D heat equation.

**Section A.2: Hyperbolic equations.** This section is concerned with hyperbolic partial differential equations which are the more difficult equations to discretize properly. Just to make the reader aware of it, we first consider the linear 1-D advection equation:

$$\frac{\partial u}{\partial t} + a \frac{\partial u}{\partial x} = 0$$

where  $a$  is a constant. Interestingly, we show that if we do not choose a suitable discretization for the spatial derivative, then the scheme will be unconditionally unstable. The right way is to choose an upwind scheme that is a scheme which takes into account the direction of the propagation.

We then consider the nonlinear Burgers equation:

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = 0$$

and show by using the method of characteristics that a shock appear even if we start from a continuous initial data, and we show that this also brings some difficulties on a numerical point of view. The notions of monotone and conservative schemes are then necessary to capture the correct entropy solution.

**Section A.3: Difference schemes in image analysis.** After introduc-

ing the main notations we present the discretization of certain equations encountered in this book. Here the aim is to offer the reader the possibility to re-implement these equations. The problems considered are:

- Restoration by energy minimization (from Section 3.2): we detail the discretization of a divergence term which can also be found for the Perona and Malik equation.
- Enhancement by Osher and Rudin's shock filters (from Section 3.3.3): the main interest is to use a flux limiter called minmod.
- Curves evolution with the level sets method. Having in mind the geodesic active contours equation from Section 4.3.3, we decompose the problem by studying separately the different terms. We start with the classical mean curvature motion and we show on a simple example that a re-initialization is in general necessary to keep a distance function. The second example is about motions with a constant speed. We mention the possibility for such evolutions (and more generally with monotone speed) to use a fast marching approach. The third equation is the pure advection equation. Finally, we come back to the segmentation problem with geodesic active contours. Some examples illustrate the equations.

We hope that this Appendix will give the reader effective ideas to discretize and implement PDEs studied throughout this monograph and in their own research.



# Guide to main mathematical concepts and their application

This book is mainly organized by image processing problems and not by classes of mathematical concepts. The aim of this guide is to highlight the different concepts used and especially where they are applied.

## *Direct method in the calculus of variations (Section 2.1.2)*

This terminology is used when the problem is to minimize an integral functional, for example:

$$\inf_{\Omega} \left\{ F(u) = \int_{\Omega} f(x, u(x), \nabla u(x)) \, dx, \quad u \in V \right\}. \quad (\mathcal{F})$$

The classical (or direct) method consists in defining a minimizing sequence  $u_n \in V$ , bounding uniformly  $u_n$  in  $V$  and extracting a subsequence converging to an element  $u \in V$  (compactness property, Section 2.1.1) and proving that  $u$  is a minimizer (lower semi-continuity property, Section 2.1.2). This technique has been applied in two cases:

- Image restoration (Section 3.2.3, Theorem 3.2.2)
- Sequence segmentation (Section 5.1.3)

## *Relaxation (Section 2.1.3)*

When for a minimization problem  $(\mathcal{F})$  the direct method does not apply (because the energy is not lower semi-continuous (l.s.c.), or the space is

not reflexive for example) it is then classical to associate to  $(\mathcal{F})$  another problem called  $(R\mathcal{F})$  (relaxed problem), that is another functional  $R\mathcal{F}$  (relaxed functional).  $(R\mathcal{F})$  is related to  $(\mathcal{F})$  thanks to the two following properties. The first is that  $(R\mathcal{F})$  is well-posed i.e.  $(R\mathcal{F})$  does have solutions and  $\min\{R\mathcal{F}\} = \inf\{\mathcal{F}\}$ . The second is that we can extract from minimizing sequences of  $(\mathcal{F})$  subsequences converging to a solution of  $(R\mathcal{F})$ . We have used this concept for:

- Image restoration for which the initial formulation was mathematically ill-posed (Section 3.2.3, Theorem 3.2.1 and Section 3.3.1, Proposition 3.46).

### $\Gamma$ -convergence (Section 2.1.4)

The  $\Gamma$ -convergence is a notion of convergence for functionals. It is particularly well adapted to deal with free discontinuity problems. Roughly speaking, if the sequence of functionals  $F_h$   $\Gamma$ -converge to another functional  $F$ , and if  $u_h$  is a sequence of minimizers of  $F_h$  and  $u$  a minimizer of  $F$  then (up to sequence):  $\lim_{h \rightarrow 0} F_h(u_h) = F(u)$  and  $\lim_{h \rightarrow 0} u_h = u$ . This notion is illustrated in the two cases:

- Approximation of the Mumford and Shah segmentation functional (Section 4.2.4, Theorem 4.2.8).
- Image classification (Section 5.2.1).

### Viscosity solutions (Section 2.3)

The theory of viscosity solutions aims at proving the existence and the uniqueness of a solution for the fully nonlinear PDEs of the form:

$$\frac{\partial u}{\partial t} + F(x, u(x), \nabla u(x), \nabla^2 u(x)) = 0$$

This is a very weak notion because solutions are expected to be only continuous. We have used this theory for:

- The Alvarez-Guichard-Lions-Morel scale space theory (Section 3.3.1, Theorem 3.3.2).
- Geodesic active contours and the level sets methods (Section 4.3.3, Theorem 4.3.2).

# Notations and symbols

## About functionals

For  $\Omega$  an open subset of  $R^N$  we define the following real valued functions spaces:

$\mathcal{B}(\Omega)$	Borel subset of $\Omega$ .
$S^{N-1}$	Unit sphere in $R^N$ .
$dx$	Lebesgue measure in $R^N$ .
$\mathcal{H}^{N-1}$	Hausdorff measure of dimension $N - 1$ .
$BV(\Omega)$	Space of bounded variation.
$BV - w^*$	The weak* topology of $BV(\Omega)$ .
$C_0^p(\Omega)$	Space of real valued functions, $p$ continuously differentiable with compact support.
$C_0^\infty(\Omega)$	Space of real valued functions, infinitely continuously differentiable with compact support.
$C^{0,\gamma}(\Omega)$	For $0 < \gamma \leq 1$ : space of continuous functions $f$ on $\Omega$ such that $ f(x) - f(y)  \leq C x - y ^\gamma$ , for some constant $C$ , $x, y \in \Omega$ . It is called the space of Hölder continuous functions with exponent $\gamma$ .
$C^{k,\gamma}(\Omega)$	Space of $k$ -times continuously differentiable functions whose $k^{\text{th}}$ partial derivatives belong to $C^{0,\gamma}(\Omega)$ .

$(\mathcal{C}_0^\infty(\Omega))'$	Dual of $\mathcal{C}_0^\infty(\Omega)$ , i.e. the space of distributions on $\Omega$ .
$L^p(\Omega)$	Space of Lebesgue measurable functions $f$ such that $\int_{\Omega}  f ^p dx < \infty.$
$L^\infty(\Omega)$	Space of Lebesgue measurable functions $f$ such that there exists a constant $c$ with $ f(x)  \leq c$ , a.e. $x \in \Omega$ .
$\mathcal{M}(\Omega)$	Space of Radon measures.
$SBV(\Omega)$	Space of Special functions of bounded variation.
$LSC(\Omega)$	Space of lower semi-continuous functions on $\Omega$ .
$USC(\Omega)$	Space of upper semi-continuous functions on $\Omega$ .
$W^{1,p}(\Omega)$	With $1 \leq p \leq \infty$ : Sobolev space of functions $f \in L^p(\Omega)$ such that all derivative up to order $p$ belong to $L^p(\Omega)$ . $W^{1,\infty}(\Omega)$ identifies with the space of locally Lipschitz functions.
$W_0^{1,p}(\Omega)$	$\{u \in W^{1,p}(\Omega) : u _{\partial\Omega} = 0\}$ .

*Vector valued spaces will be denoted in bold font*  
 $\mathbf{BV}(\Omega)$ ,  $\mathbf{C}_0^p(\Omega)$ ,  $\mathbf{L}^p(\Omega)$ ,  $\mathbf{M}(\Omega)$ ,  $\mathbf{W}^{1,p}(\Omega)$ ,  $\mathbf{SBV}(\Omega)$

For  $X$  a Banach space with a norm  $|\cdot|_X$  and  $v : (0, T) \rightarrow X$ :

$C^m(0, T; X)$	With $m \geq 0$ , $0 < T < \infty$ : space of functions from $[0, T]$ to $X$ $m$ -times continuously differentiable. It is a Banach space with the norm
	$ v _{C^m(0, T; X)} = \max_{0 \leq l \leq m} \left( \sup_{0 \leq t \leq T} \left  \frac{\partial^l v}{\partial t^l}(t) \right _X \right).$
$L^p(0, T; X)$	With $1 \leq p < \infty$ : space of functions $v \rightarrow v(t)$ measurable on $(0, T)$ for the measure $dt$ (i.e. the scalar functions $t \rightarrow  v _X$ are $dt$ -measurable). It is a Banach space with the norm
	$ v _{L^p(0, T; X)} = \left( \int_0^T  v _X^p dt \right)^{1/p} < +\infty.$
$L^\infty(0, T; X)$	Space of functions $v$ such that $ v _{L^\infty(0, T; X)} = \inf_c \{ v _X \leq c, \text{ a.e. } t\}.$

For a functional  $F : X \rightarrow ]-\infty, +\infty]$  where  $X$  is a Banach space:

$\operatorname{Argmin} F$	$\{u \in X : F(u) = \inf_X F\}$ .
$R_\tau(F)$ , $RF$ , $\bar{F}$	Relaxed functional of $F$ (for the $\tau$ -topology).

l.s.c. (sequentially)

Lower semi-continuous:  $F$  is called l.s.c.  
 if and only if for every sequence  $(u^n)$  converging to  $u$ , we have:  $\lim_{n \rightarrow \infty} F(u^n) \geq F(u)$ .

u.s.c. (sequentially)

Upper semi-continuous:  $F$  is called u.s.c.  
 if and only if for every sequence  $(u^n)$  converging to  $u$ , we have:  $\lim_{n \rightarrow \infty} F(u^n) \leq F(u)$ .

## About Measures

For  $\mu$  et  $\nu$  two Radon measures

 $|\mu|$ 

Total variation of the measure  $\mu$ . If  $\mu$  is vector valued, we also denote  $|\mu| = |\mu_1| + \dots + |\mu_N|$ .

 $\nu \ll \mu$ 

$\nu$  is absolutely continuous with respect to  $\mu$  if and only if:  $\mu(A) = 0 \Rightarrow \nu(A) = 0$  for all  $A \in R^N$ .

 $\nu \perp \mu$ 

$\nu$  is singular with respect to  $\mu$  if and only if there exists a Borel  $B \subset R^N$  such that:

$$\mu(R^N - B) = \nu(B) = 0.$$

## About functions

For a function  $f : \Omega \subset R^N \rightarrow R$  and a sequence of functions  $(f^n)_{n \in N}$  belonging to a Banach space  $X$ :

 $f^n \xrightarrow{X} f$ 

The sequence  $(f^n)$  converges strongly to  $f$  in  $X$ .

 $f^n \xrightarrow{X} f$ 

The sequence  $(f^n)$  converges weakly to  $f$  in  $X$ .

 $f^n \xrightarrow[X]{*} f$ 

The sequence  $(f^n)$  converges to  $f$  for the weak\* topology of  $X$ .

 $\overline{\lim}_{n \rightarrow +\infty} f^n$ 

$$\overline{\lim}_{n \rightarrow +\infty} f^n(x) = \inf_k \sup \{f^k(x), f^{k+1}(x), \dots\}.$$

 $\underline{\lim}_{n \rightarrow +\infty} f^n$ 

$$\begin{aligned} \underline{\lim}_{n \rightarrow +\infty} f^n(x) &= -\overline{\lim}_{n \rightarrow +\infty} f^n(x) \\ &= \sup_k \inf \{f^k(x), f^{k+1}(x), \dots\}. \end{aligned}$$

 $|f|_X$ 

Norm of  $f$  in  $X$ .

 $\text{spt}(f)$ 

For a measurable function  $f : \Omega \subset R^N \rightarrow R$ , let  $(w_i)_{i \in I}$  be the family of all open subsets such that  $w_i \subset \Omega$  and for each  $i \in I$ ,  $f = 0$  a.e. on  $w_i$ , then  $\text{spt}$  (the support of  $f$ ) is defined by  $\text{spt} f = \Omega - \bigcup_i w_i$ .

 $Df$ 

Distributional gradient of  $f$ .

 $D^2 f$ 

Hessian matrix of  $f$  (in the distributional sense).

 $\nabla f$ 

Gradient of  $f$  in the classical sense. It corresponds to the absolutely continuous part of  $Df$  with respect to the Lebesgue measure  $dx$ .

$\operatorname{div}(f)$	Divergence operator: $\operatorname{div}(f) = \sum_{i=1}^N \frac{\partial f}{\partial x_i}.$
$\nabla^2 f$	Hessian matrix of $f$ in the classical sense: $(\nabla^2 f)_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}.$
$\Delta f$	Laplacian operator: $\Delta f = \sum_{i=1}^N \frac{\partial^2 f}{\partial x_i^2}.$
$\int_{\Omega} f \, dx$	Mean value of $f$ over $\Omega$ : $\int_{\Omega} f \, dx = \frac{1}{ \Omega } \int_{\Omega} f \, dx.$
$\overset{\bullet}{\mathcal{P}}_{\Omega}^{\pm}$	Precise representation of $f$ . “superjets”.

For a function  $\phi : R^N \rightarrow R$

$\phi^*(.)$	The Fenchel-Legendre conjugate.
$\phi^{\infty}(.)$	The recession function defined by $\phi^{\infty}(z) = \lim_{s \rightarrow \infty} \phi(sz)/s.$

### Misc notations

$A \xhookrightarrow{\text{strong}} B$	$A$ is relatively compact in $B$ .
$A \xhookrightarrow{\text{weak}} B$	$A$ is weakly relatively compact in $B$ .
$O^*$	The adjoint operator of $O$ .
$  \cdot  $	Euclidian norm in $R^N$ .
$G_{\sigma}$	The Gaussian kernel defined by: $G_{\sigma}(x) = \frac{1}{2\pi\sigma^2} \exp\left(\frac{- x ^2}{2\sigma^2}\right).$
$B(x, r) \subset R^N$	Ball of center $x$ and radius $r$ .
$\mathcal{S}(N)$	Subset of $N \times N$ symmetric matrices .
$SNR(I_1/I_2)$	<i>Signal to Noise Ratio:</i> used to estimate the quality of an image $I_2$ with respect to a reference image $I_1$ . It is defined by: $SNR(I_1/I_2) = 10 \log_{10} \left[ \frac{\sigma^2(I_2)}{\sigma^2(I_1 - I_2)} \right]$ where $\sigma$ is the variance.
$\alpha \vee s \wedge \beta$	Truncature function equal to $\alpha$ if $s \leq \alpha$ , $\beta$ if $s \geq \beta$ , $s$ otherwise.
$\operatorname{sign}(s)$	The sign function equal to 1 if $s > 0$ , 0 if $s = 0$ and -1 if $s < 0$ .
$\chi_R$	The characteristic function of $R$ : $\chi_R(x) = \begin{cases} 1 & \text{if } x \in R \\ 0 & \text{otherwise.} \end{cases}$
$\operatorname{Per}_{\Omega}(R)$	Perimeter of $R$ in $\Omega$ defined as the total variation of $\chi_R$ .

*Symbols for readers convenience*

Indicates general references, books, reviews or other parts of the book where to find complementary informations.



Summary of an *important idea*.



Symbol marking the *end* of a proof, an example or a remark.



This symbols indicates *unsolved or challenging unsolved problems* that need to be further investigated.

# 2

## Mathematical preliminaries

### How to read this chapter?

This chapter introduces the mathematics used in this book. It covers some of the main notions in the calculus of variations and the theory of viscosity solutions. It is introductory and we tried to make it as self contained as possible. A careful reading is advised to better understand the coming analysis. The prerequisite for this chapter is a good course in advanced calculus.

- Section 2.1 introduces the basic tools concerning optimization problems in Banach spaces. We answer the following questions: what are the good hypotheses ensuring the existence and the uniqueness of a minimizer? What can be said when some assumptions are missing? We also introduce the notion of relaxed problem when the original one is ill-posed (Section 2.1.3). Section 2.1.4 concerns the notion of  $\Gamma$ -convergence which is a notion of convergence for functionals. The  $\Gamma$ -convergence theory is particularly useful to approximate free discontinuity or ill-posed problems. This notion will be used several times in all the book.
- Section 2.2 presents the space  $BV(\Omega)$  of functions of bounded variation. This space appears to be the suitable functional space in image analysis since it contains functions that can be discontinuous across curves (i.e. across edges). Note that many theoretical results given in

this book will be established in a *BV*-framework: see Sections 3.2, 4.2 and 5.1.

- Section 2.3 introduces the concept of viscosity solutions for nonlinear second order PDEs. This notion is interesting since it allows to prove the existence of a solution in a very weak sense and also its uniqueness which is the most difficult part. An example of application will be detailed further in Section 4.3.
- Some basic element of differential geometry used throughout this book are recalled in Section 2.4.
- We conclude in Section 2.5 by giving various results of interest (inequalities and theorems) to be as much as possible self contained. This section may be consulted just when necessary.

For a systematic study of the classical mathematics presented therein, the following references may be useful:

-  Functional analysis [51, 109, 217, 218, 255], partial differential equations [105, 117], Sobolev spaces [1], integration [216], differential geometry [161].

## 2.1 The direct method in the calculus of variations

### 2.1.1 Topologies on Banach spaces

Let us introduce some definitions. Let  $(X, \|\cdot\|)$  a real Banach space<sup>1</sup>, we denote by  $X'$  the topological dual space of  $X$ :

$$X' = \left\{ l : X \rightarrow \mathbb{R} \text{ linear such that } \|l\|_{X'} = \sup_{x \neq 0} \frac{|l(x)|}{\|x\|_X} < \infty \right\}.$$

Classically  $X$  can be endowed with two topologies (we only work with sequences):

#### Definition 2.1.1 (topologies on $X$ )

- (i) *The strong topology, noted  $x_n \xrightarrow[X]{} x$  is defined by  $|x_n - x|_X \rightarrow 0$  ( $n \rightarrow \infty$ ).*
- (ii) *The weak topology, noted  $x_n \xrightarrow[X]{} x$  is defined by  $l(x_n) \rightarrow l(x)$  ( $n \rightarrow \infty$ ) for every  $l \in X'$ .*

---

<sup>1</sup>A Banach space is a complete, normed linear space. Complete means that any Cauchy sequence is convergent.

The strong convergence implies the weak convergence, but the reverse is false in general.

**Example** Let us show it on a counter example. We consider the sequence  $f_h(x) = \sin(2\pi xh)$ ,  $x \in (0, 1)$  when  $h \rightarrow 0$ . We are going to establish that  $f_h(x) \xrightarrow{L^2(\Omega)} 0$  and that  $f_h(x) \xrightarrow{L^2(\Omega)} 0$  is not true.

To show the weak convergence, we have for all  $\varphi \in \mathcal{C}_0^\infty(0, 1)$  and by using the mean value Theorem:

$$\int_0^1 \sin(2\pi xh)\varphi(x) dx = \frac{1}{h} \int_0^h \sin(2\pi y)\varphi\left(\frac{y}{h}\right) dy = \sin(2\pi\xi)\varphi\left(\frac{\xi}{h}\right)$$

with  $\xi \in [0, h]$ . So, it is clear that  $\langle f_h, \varphi \rangle \rightarrow 0$  when  $h \rightarrow 0$  for all  $\varphi \in \mathcal{C}_0^\infty(0, 1)$ . By density, this result can be generalized for all  $\varphi \in L^2(0, 1)$ .

Now, to prove that there is no strong convergence, we can remark that:

$$\int_0^1 \sin^2(2\pi xh) dx = \frac{1}{2} \int_0^1 (1 - \cos(4\pi xh)) dx = \frac{1}{2} + \frac{1}{2} \frac{\sin(4\pi h)}{4\pi h} \rightarrow 1$$

when  $h \rightarrow 0$ . ■

The dual  $X'$  can also be endowed with the strong and the weak topology:

**Definition 2.1.2 (topologies on  $X'$ )**

- (i) *The strong topology, noted  $l_n \xrightarrow{X'} l$ , is defined by*  
 $|l_n - l|_{X'} \rightarrow 0$    *or equivalently*    $\sup_{x \neq 0} \frac{|l_n(x) - l(x)|}{|x|_X} \rightarrow 0$  ( $n \rightarrow \infty$ ).
- (ii) *The weak topology, noted  $l_n \xrightarrow{X'} l$ , is defined by*  
 $z(l_n) \rightarrow z(l)$  ( $n \rightarrow \infty$ ) *for every*  $z \in (X')'$ , *the bidual space of  $X$ .*

*In some cases it is more convenient to equip  $X'$  with a third topology:*

- (iii) *The weak\* topology, noted  $l_n \xrightarrow{X'}^* l$ , is defined by*  
 $l_n(x) \rightarrow l(x)$  ( $n \rightarrow \infty$ ) *for every*  $x \in X$ .

The interest of the weak\* topology will be clear later (see Theorem 2.1.1). We recall that the space  $X$  is called *reflexive* if  $(X')' = X$  and that  $X$  is *separable* if it contains a countable dense subset.

**Examples** Let  $\Omega$  be an open subset of  $R^N$ .

- $X = L^p(\Omega)$  is reflexive for  $1 \leq p < \infty$  and separable for  $1 \leq p < \infty$ . The dual space of  $L^p(\Omega)$  is  $L^{p'}(\Omega)$  for  $1 \leq p < \infty$  with  $\frac{1}{p} + \frac{1}{p'} = 1$ .
- $X = L^1(\Omega)$  is non-reflexive and  $X' = L^\infty(\Omega)$ . ■

The main properties associated to these different topologies are summarized in the following theorem.

**Theorem 2.1.1 (weak sequential compactness)**

- (i) Let  $X$  be a reflexive Banach space, let  $K > 0$  and  $x_n \in X$  a sequence such that  $|x_n|_X \leq K$ , then there exist  $x \in X$  and a subsequence  $x_{n_j}$  of  $x_n$  such that  $x_{n_j} \xrightarrow[X]{} x$  ( $n \rightarrow \infty$ ).
- (ii) Let  $X$  be separable Banach space, let  $K > 0$  and  $l_n \in X'$  such that  $|l_n|_{X'} \leq K$ , then there exist  $l \in X'$  and a subsequence  $l_{n_j}$  of  $l_n$  such that  $l_{n_j} \xrightarrow[X']{} l$  ( $n \rightarrow \infty$ ).

The interest of the weak\* topology is that it allows to get compactness results even if  $X$  is not reflexive. Notice that nothing can be said about the strong convergence of the sequences.

### 2.1.2 Convexity and lower semi-continuity

Let  $X$  be a Banach space,  $F : X \rightarrow \mathbb{R}$ , and consider the minimization problem

$$\inf_{x \in X} F(x).$$

Let us first consider the problem of the existence of a solution. Proving it is usually achieved by the following steps, which is the direct method of the calculus of variations:

- (A) One constructs a *minimizing sequence*  $x_n \in X$ , i.e. a sequence verifying  $\lim_{n \rightarrow \infty} F(x_n) = \inf_{x \in X} F(x)$ .
- (B) If  $F$  is *coercive* ( $\lim_{|x| \rightarrow \infty} F(x) = +\infty$ ), one can obtain a uniform bound  $|x_n|_X \leq C$ . If  $X$  is reflexive, thanks to Theorem 2.1.1, one deduces the existence of  $x_0 \in X$  and of a subsequence  $x_{n_j}$  such that  $x_{n_j} \xrightarrow[X]{} x_0$ .
- (C) To prove that  $x_0$  is minimum point of  $F$  it suffices to have the inequality  $\lim_{x_{n_j} \rightarrow x_0} F(x_{n_j}) \geq F(x_0)$  which obviously implies that  $F(x_0) = \min_{x \in X} F(x)$ .

This latter property which appears here naturally, is called the weak *lower semi-continuity*. More precisely we have the definition:

**Definition 2.1.3 (lower semi-continuity)**

$F$  is called lower semi-continuous (l.s.c.) for the weak topology if and only

if for all sequence  $x_n \rightarrow x_0$  we have:

$$\lim_{x_n \rightarrow x_0} F(x_n) \geq F(x_0). \quad (2.1)$$

The same definition can be given with a strong topology.

- In the direct method, the notion of weak l.s.c. emerges very naturally.

Unfortunately, it is in general difficult to prove the weak l.s.c.. A sufficient condition that implies weak l.s.c. is the convexity:

**Definition 2.1.4 (convexity)**  $F$  is convex on  $X$  if and only if:

$$F(\lambda x + (1 - \lambda)y) \leq \lambda F(x) + (1 - \lambda)F(y)$$

for all  $x, y \in X$  and  $\lambda \in [0, 1]$ .

**Theorem 2.1.2 (l.s.c. strong and weak)** Let  $F : X \rightarrow R$  is convex. Then  $F$  is weakly l.s.c. if and only if  $F$  is strongly l.s.c..

This theorem is useful since in most cases the strong l.s.c. is not very hard to prove.

If  $F$  is an integral functional, we can even say more about the link between convexity and l.s.c.. Let  $\Omega \subset R^N$  be a bounded open set,  $f : \Omega \times R \times R^N \rightarrow R$  be a continuous function satisfying:

$$0 \leq f(x, u, \xi) \leq a(x, |u|, |\xi|) \quad (2.2)$$

where  $a$  is increasing with respect to  $|u|$  and  $|\xi|$ , and integrable in  $x$ . Let  $W^{1,p}(\Omega)$  be the Sobolev space:

$$W^{1,p}(\Omega) = \{u \in L^p(\Omega), Du \in L^p(\Omega)\},$$

where  $Du$  is the distributional gradient of  $u$  (Notice that in this case  $Du$  is a function and we can also denote it  $\nabla u$ ). For  $u \in W^{1,p}(\Omega)$  we consider the functional:

$$F(u) = \int_{\Omega} f(x, u(x), Du(x)) dx.$$

**Theorem 2.1.3 (l.s.c. and convexity)**  $F(u)$  is (sequentially) weakly l.s.c. on  $W^{1,p}(\Omega)$ ,  $1 \leq p < \infty$  (weakly\* l.s.c. if  $p = \infty$ ) if and only if  $f$  is convex in  $\xi$ .

## Remarks

- We emphasize that convexity is a sufficient condition for existence. There exist nonconvex problems admitting a solution [19].

- The previous theorem has been established assuming that  $u : \Omega \subset R^N \rightarrow R^m$  with  $m = 1$ . When dealing with gray scale images, we are in this situation since  $N = 2$ ,  $m = 1$ . This result is also true in the case  $N = 1$  and  $m > 1$ . Both cases can be referred to as scalar (either  $N$  or  $m$  is equal to 1). However, this theorem is no longer true in the vectorial case, when  $N > 1$  and  $m > 1$ . A weaker definition of the convexity has to be introduced (the quasi-convexity), and similar results can be obtained. We refer the interested reader to [83] for more details. ■

Therefore, in the scalar case, that is  $u(x) \in R$ , the natural condition to impose on the integrand  $f(x, u, \xi)$  to obtain the existence of a minimizer for  $F$ , is the convexity in  $\xi$ . More precisely, we have:

**Theorem 2.1.4** *Let  $\Omega \subset R^N$  bounded and  $f : \Omega \times R \times R^N \rightarrow R$  continuous verifying*

- (i)  $f(x, u, \xi) \geq a(x) + b |u|^p + c |\xi|^p$  for every  $(x, u, \xi)$  and for some  $a \in L^1(\Omega)$ ,  $b > 0$ ,  $c > 0$  and  $p > 1$ .
- (ii)  $\xi \rightarrow f(x, u, \xi)$  is convex for every  $(x, u)$ .
- (iii) There exists  $u_0 \in W^{1,p}(\Omega)$  such that  $F(u_0) < \infty$ ,  
then the problem  $\inf \left\{ F(u) = \int_{\Omega} f(x, u(x), \nabla u(x)) dx, u \in W^{1,p}(\Omega) \right\}$   
admits a solution. Moreover, if  $(u, \xi) \rightarrow f(x, u, \xi)$  is strictly convex for every  $x$ , then the solution is unique.

In Theorem 2.1.4, the coercivity condition (i) implies the boundness of the minimizing sequences. The condition (ii) permits to pass to the limit on these sequences. The condition (iii) ensures that the problem has a meaning. This can be summarized as follow:

☞ Convexity is used to get the l.s.c. while coercivity is related to the compactness.

Before going further, let us illustrate on three examples the importance of coercivity, reflexivity and convexity.

**Examples** Let  $\Omega = ]0, 1]$ . We propose below some classical examples where either coercivity, reflexivity or convexity are no longer true:

- (A) Weierstrass ( $N = m = 1$ )

Let  $f$  defined by  $f(x, u, \xi) = x\xi^2$  and let us denote:

$$m = \inf \left\{ \int_0^1 x (u'(x))^2 dx \text{ with } u(0) = 1 \text{ and } u(1) = 0 \right\}.$$

Then, we can show that this problem does not have any solution. The function  $f$  is convex, but *the  $W^{1,2}(\Omega)$ -coercivity with respect to  $u$  is not verified because the integrand  $f(x, \xi) = x\xi^2$  vanishes at  $x = 0$ .*

Let us first prove that  $m = 0$ . The idea is to propose the following minimizing sequence:

$$u_n(x) = \begin{cases} 1 & \text{if } x \in \left(0, \frac{1}{n}\right) \\ -\frac{\log(x)}{\log(n)} & \text{if } x \in \left(\frac{1}{n}, 1\right) \end{cases}$$

It is then easy to verify that  $u_n \in W^{1,\infty}(0, 1)$ , and that

$$F(u_n) \equiv \int_0^1 x (u'_n(x))^2 dx = \frac{1}{\log(n)} \rightarrow 0.$$

So we have  $m = 0$ . If there exists  $\bar{u}$  a minimum, then we should have  $F(\bar{u}) = 0$ , that is  $\bar{u}' = 0$  almost everywhere (a.e.) in  $(0, 1)$ , which is clearly incompatible with the boundary conditions.

(B) Minimal surfaces ( $p = 1$ ).

Let  $f$  defined by  $f(x, u, \xi) = \sqrt{u^2 + \xi^2}$ . Then the associated functional  $F$  is convex and coercive on the *non reflexive Banach space  $W^{1,1}(\Omega)$*  ( $F(u) \geq \frac{1}{2} |u|_{W^{1,1}(\Omega)}$ ). This example shows the importance of reflexivity.

Let us denote:

$$m = \inf \left\{ \int_0^1 \sqrt{u^2 + u'^2} dx \mid u(0) = 0 \text{ and } u(1) = 1 \right\}$$

Let us prove that  $m = 1$ . First, we can remark that:

$$F(u) \equiv \int_0^1 \sqrt{u^2 + u'^2} dx \geq \int_0^1 |u'| dx \geq \int_0^1 u' dx = 1.$$

So we have  $m \geq 1$ . Then, if we consider the sequence

$$u_n(x) = \begin{cases} 0 & \text{if } x \in \left(0, 1 - \frac{1}{n}\right) \\ 1 + n(x - 1) & \text{if } x \in \left(1 - \frac{1}{n}, 1\right) \end{cases}.$$

we can verify that:  $F(u_n) \rightarrow 1$  ( $n \rightarrow \infty$ ). So  $m = 1$ . If we assume the existence of a solution  $\bar{u}$ , then we should have:

$$1 = F(\bar{u}) = \int_0^1 \sqrt{\bar{u}^2 + \bar{u}'^2} dx \geq \int_0^1 |\bar{u}'| dx \geq \int_0^1 \bar{u}' dx = 1 \Rightarrow \bar{u} = 0,$$

which obviously does not satisfy the boundary conditions. As a conclusion, there is no solution to this problem.

(C) Bolza.

Let  $f$  defined by  $f(x, u, \xi) = u^2 + (\xi^2 - 1)^2$ . The Bolza problem is:

$$\inf \left\{ F(u) = \int_0^1 ((1 - u'^2)^2 + u^2) dx ; u \in W^{1,4}(0, 1) \right. \\ \left. u(0) = u(1) = 0 \right\}. \quad (2.3)$$

The functional is clearly *nonconvex*. It is easy to see that  $\inf_u F = 0$ . Indeed, for  $n$  an integer and  $0 \leq k \leq n - 1$ , if we choose

$$u_n(x) = \begin{cases} x - \frac{k}{n} & \text{if } x \in \left(\frac{2k}{2n}, \frac{2k+1}{2n}\right) \\ -x + \frac{k+1}{n} & \text{if } x \in \left(\frac{2k+1}{2n}, \frac{2k+2}{2n}\right) \end{cases}$$

then  $u_n \in W^{1,\infty}(0, 1)$  and:

$$\begin{aligned} 0 \leq u_n(x) \leq \frac{1}{2n} \text{ for every } x \in (0, 1) \\ |u'_n(x)| = 1 \text{ a.e. in } (0, 1) \\ u_n(0) = u_n(1) = 0 \end{aligned}.$$

Therefore  $0 \leq \inf_u F(u) \leq F(u_n) \leq \frac{1}{4n^2}$ . Letting  $n \rightarrow \infty$ , we get  $\inf_u F(u) = 0$ . However there exists no function  $u \in W^{1,4}(0, 1)$ , such that  $u(0) = u(1) = 0$ , and verifying  $F(u) = 0$ . So the problem (2.3) does not have a solution in  $W_0^{1,4}(0, 1)$ .

■

Once we have the existence of a minimum, the natural second step is to write the optimality conditions. We need for that the definition of the Gâteaux-derivative.

**Definition 2.1.5 (Gâteaux derivative)** Let  $X$  be a Banach space and  $F : X \rightarrow R$ . We call directional derivative of  $F$  at  $u$  in the direction  $v$  the limit if it exists:  $F'(u; v) = \lim_{\lambda \rightarrow 0^+} \frac{F(u + \lambda v) - F(u)}{\lambda}$ . Moreover if there exists  $\tilde{u} \in X'$  such that  $F'(u; v) = \tilde{u}(v)$ ,  $\forall v \in X$ , we say that  $F$  is Gâteaux-differentiable at  $u$  and we note  $F'(u) = \tilde{u}$ .

If  $F$  is Gâteaux-differentiable and if the problem  $\inf_{v \in X} F(v)$  has a solution  $u$ , then we have

$$F'(u) = 0.$$

Conversely if  $F$  is convex, then a solution  $u$  of  $F'(u) = 0$  is a solution of the minimization problem. The equation  $F'(u) = 0$  is called an *Euler-Lagrange equation* (also referred to as Euler-Lagrange equation). Let us write

it explicitly for a functional  $F$  defined by:

$$F(u) = \int_{\Omega} f(x, u(x), \nabla u(x)) dx$$

where  $f$  is of class  $C^1$  with respect to  $(u, \xi)$ , satisfies conditions (i) and (iii) of Theorem 2.1.4, and the growth condition for the derivatives:

$$\begin{cases} \left| \frac{\partial f}{\partial u}(x, u, \xi) \right| \leq a' (1 + |u|^{p-1} + |\xi|^p) \\ |\nabla_{|\xi|} f(x, u, \xi)| \leq a'' (1 + |u|^p + |\xi|^{p-1}) \end{cases} \quad \text{a.e. } x, \forall (u, \xi) \quad (2.4)$$

for some constants  $a', a'' > 0$ . Then we can prove that, for  $u \in W^{1,p}(\Omega)$ :

$$F'(u) = \frac{\partial f}{\partial u}(x, u, \nabla u) - \sum_{i=1}^{i=N} \frac{\partial}{\partial x_i} \left( \frac{\partial f}{\partial \xi_i}(x, u, \nabla u) \right). \quad (2.5)$$

We let the proof to the reader as an exercise.

**Remark** Notice that the growth conditions (2.4) permit to apply the Lebesgue dominated convergence theorem to prove (2.5). ■

### 2.1.3 Relaxation

In this section, we examine the case when the functional  $F$  is not weakly l.s.c.. As we will see on a counterexample there is no hope to obtain, in general, the existence of a minimum for  $F$ . We could however associate to  $F$  another functional  $RF$  whose minima should be weak-cluster points of minimizing sequences of  $F$ . This idea is important and will be used several times in this book.

As an illustration, let us consider again the Bolza problem (2.3). In this example, the integrand  $f(u, \xi) = (1 - \xi^2)^2 + u^2$  is nonconvex in  $\xi$  and the functional  $F$  is not weakly l.s.c. on  $W^{1,4}(0, 1)$ . In such a situation, it is classical to define the lower semi-continuous envelope (or relaxed function)  $R_\tau F$  of  $F$ .

■ Relaxation [45, 27, 44, 52, 84, 101, 83].

Let  $X$  a Banach space and  $F : X \rightarrow \bar{R}$ . In the sequel, we equally denote by  $\tau$  the strong or the weak topology of  $X$ .

**Definition 2.1.6 (relaxed functional)** *The  $\tau$ -lower semi-continuous envelope (also called relaxed functional)  $R_\tau F$  of  $F$  is defined for every  $x \in X$  by:  $R_\tau F(x) = \sup \{G(x), G \in \Gamma\}$  where  $\Gamma$  is the set of all  $\tau$ -lower semi-continuous functions on  $X$  such that  $G(y) \leq F(y)$  for every  $y \in X$ .*

To compute  $R_\tau F$ , the following characterization is useful:

**Theorem 2.1.5 (characterization of the relaxed functional)**

$R_\tau F$  is characterized by the following properties:

- (i) For every sequence  $x_n$   $\tau$ -converging to  $x$  in  $X$  then  

$$R_\tau F(x) \leq \varliminf_{n \rightarrow \infty} F(x_n).$$
- (ii) For every  $x$  in  $X$  there exists a sequence  $x_n$   $\tau$ -converging to  $x$  in  $X$  such that  $R_\tau F(x) \geq \varlimsup_{n \rightarrow \infty} F(x_n).$

We consider now the relation between the original problem  $\inf \{F(x), x \in X\}$  and the relaxed problem  $\inf \{RF_\tau(x), x \in X\}$ .

**Theorem 2.1.6 (main properties)** Let  $X$  be a reflexive Banach space and let  $\tau$  be the weak topology. Assume that  $F : X \rightarrow \bar{R}$  is coercive, then the following properties hold:

- (i)  $R_\tau F$  is coercive and  $\tau$ -lower semi-continuous.
- (ii)  $R_\tau F$  has a minimum point in  $X$ .
- (iii)  $\min_{x \in X} R_\tau F(x) = \inf_{x \in X} F(x).$
- (iv) Every cluster-point of a minimizing sequence for  $F$  is a minimum point for  $R_\tau F$ .
- (v) Every minimum point for  $R_\tau F$  is the limit of a minimizing sequence for  $F$ .

In summary, starting with a minimization problem which has no solution, we can define a relaxed problem whose connections with the original one are clearly stated in Theorem 2.1.6. For integral functionals, one possibility to compute the relaxed functional is to use the polar and bipolar functions. Let  $f : R^N \rightarrow R$ , we define the *polar* of  $f$  the function  $f^* : R^N \rightarrow R$  as:

$$f^*(\eta) = \sup_{\xi \in R^N} \{\eta \cdot \xi - f(\xi)\}$$

and the bipolar of  $f$  :

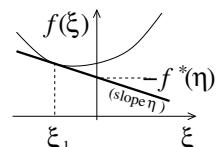
$$f^{**}(\xi) = \sup_{\eta \in R^N} \{\eta \cdot \xi - f^*(\eta)\}.$$

Since  $f^*$  and  $f^{**}$  are the supremum of affine functions it turns out that  $f^*$  and  $f^{**}$  are convex. In fact from convex analysis results  $f^{**}$  is the convex envelope of  $f$ , i.e. the greatest convex function less than  $f$ .

Notice that the function  $f^*$  has an interesting geometric interpretation. The definition of  $f^*$  implies:

$$f(\xi) \geq \eta \cdot \xi - f^*(\eta) \quad \text{for all } \xi \in R^N,$$

that is, the affine function  $h(\xi) = \eta \cdot \xi - f^*(\eta)$  is everywhere below the graph of  $f$ . If this supremum



is reached, for instance at  $\xi = \xi_1$ ,  $-f^*(\eta)$  is the intersection of this function with the vertical axis.

### Examples

- If  $f(\xi) = \frac{1}{p}|\xi|^p$ ,  $1 < p < \infty$ , then  $f^*(\eta) = \frac{1}{q}|\eta|^q$ , with  $\frac{1}{p} + \frac{1}{q} = 1$ .
- If  $f(\xi) = |\xi|$ , then  $f^*(\eta) = \begin{cases} 0 & \text{if } |\eta| \leq 1 \\ +\infty & \text{otherwise} \end{cases}$ .
- If  $f(\xi) = e^\xi$ ,  $\xi \in R$ , then  $f^*(\xi) = \begin{cases} \eta \log(\eta) - \eta & \text{if } \eta > 0 \\ 0 & \text{if } \eta = 0 \\ +\infty & \text{if } \eta < 0. \end{cases}$  ■

Let us mention that the notion of polarity also exists in infinite dimensional spaces. So, let us consider the functional

$$F(u) = \int_{\Omega} f(x, \nabla u(x)) dx$$

where  $\Omega$  is a bounded open subset of  $R^N$  and  $f : \Omega \times R^N \rightarrow R$  is a continuous function such that for every  $\xi \in R^N$  and a.e.  $x$ :

$$a |\xi|^p \leq f(x, \xi) \leq b(|\xi|^p + 1), \quad 1 < p < \infty$$

for some constant  $a, b > 0$ . If  $\tau$  is the weak topology of  $W^{1,p}(\Omega)$  then

$$R_\tau F(u) = \int_{\Omega} f^{**}(x, \nabla u(x)) dx,$$

where the polar functions are always computed with respect to the gradient variable. To illustrate this, we compute in the example below the relaxed functional for the Bolza problem.

**Example** For the previous Bolza problem:

$$\inf \left\{ F(u) = \int_0^1 ((1 - u'^2)^2 + u^2) dx ; u \in W^{1,4}(0, 1), u(0) = u(1) = 0 \right\}$$

the lack of weak l.s.c., as mentioned above, is due to the presence of the nonconvex function  $\xi \rightarrow (1 - \xi^2)^2$ . The second term  $\int_0^1 u^2 dx$  is weakly continuous according to the compact embedding  $W^{1,4}(0, 1) \subset L^2(0, 1)$ . Then, the relaxed functional of  $F$  is

$$RF(u) = \int_0^1 (((1 - u'^2)_+)^2 + u^2) dx$$

where  $t_+ = t$  if  $t \geq 0$  and  $t_+ = 0$  otherwise.

The problem  $\inf\{RF(u), u \in W^{1,4}(0,1), u(0) = u(1) = 0\}$  admits a unique solution  $u_0$  and the set  $E = \{x \in (0,1); |u'_0(x)| < 1\}$  has a positive Lebesgue measure (otherwise if  $|E| = 0$  then  $|u'_0(x)| \geq 1$  a.e.  $x$  and  $F$  would have a minimizer which is wrong).  $\blacksquare$

### 2.1.4 About $\Gamma$ -Convergence

We recall in this section the main results concerning  $\Gamma$ -convergence. Introduced by De Giorgi [119, 85], the aim is to give a meaning to the convergence of a sequence of functionals.

#### $\Gamma$ -convergence [119, 14, 84]

This notion will be particularly useful for approximating nonconvex problems in numerous applications throughout this book.

Let  $X$  be a separable Banach space endowed with a topology  $\tau$  and let  $F_h : X \rightarrow \overline{\mathbb{R}}$  be a sequence of functionals.

**Definition 2.1.7 ( $\Gamma$ -convergence)** We say that  $F_h$   $\Gamma$ -converges to  $F$  ( $F = \Gamma - \lim F_h$ ) for the topology  $\tau$ , if and only if:

- (i) For every  $x$  in  $X$  and for every sequence  $x_n$   $\tau$ -converging to  $x$  in  $X$  then

$$F(x) \leq \underline{\lim}_{h \rightarrow \infty} F_h(x_h).$$

- (ii) For every  $x$  in  $X$ , there exists a sequence  $x_n$   $\tau$ -converging to  $x$  in  $X$  such that

$$F(x) \geq \overline{\lim}_{h \rightarrow \infty} F_h(x_h).$$

**Definition 2.1.8 (equicoercivity)** We say that the sequence of functionals is equicoercive if and only if for every  $t \geq 0$  there exists  $K_t$  a compact subset of  $X$  such that:  $\{x \in X; F_h(x) \leq t\} \subset K_t$  for all  $h$ .

Main properties of  $\Gamma$ -convergence are summarized in the following theorem.

**Theorem 2.1.7 (main properties of the  $\Gamma$ -limit)** Let  $X$  be a separable Banach space and let  $F_h$  be a sequence of equicoercive functionals from  $X$  into  $\overline{\mathbb{R}}$  then:

- (i) The  $\Gamma$ -limit of  $F_h$ , if it exists, is unique and l.s.c..
- (ii) There exists a subsequence  $F_{h_k}$  and  $F$  such that  $F = \Gamma - \lim F_{h_k}$ .
- (iii) If  $F = \Gamma - \lim F_h$  then  $F + G = \Gamma - \lim_{h \rightarrow \infty} (F_h + G)$  for all  $G: X \rightarrow \mathbb{R}$  continuous.

- (iv) Set  $F = \Gamma\text{-}\lim F_h$  and let us suppose that  $F(x)$  admits a unique minimum point  $x_0$  on  $X$  and let  $x_h \in X$  such that  

$$\left|F_h(x_h) - \inf_{x \in X} F_h(x)\right| \leq \varepsilon^h \text{ where } \varepsilon^h \rightarrow 0 \text{ (} h \rightarrow \infty \text{)} \text{ then } x_h \text{ converges to } x_0 \text{ in } X \text{ and } \lim_{h \rightarrow \infty} F_h(x_h) = F(x_0).$$

In general, there is no relation between  $\Gamma$ -convergence and pointwise convergence. However there exist some connections.

**Theorem 2.1.8 ( $\Gamma$ -convergence and pointwise convergence)**

- (i) If  $F_h$  converges to  $F$  uniformly then  $F_h$   $\Gamma$ -converges to  $F$ .
- (ii) If  $F_h$  is a decreasing sequence converging to  $F$  pointwise then  $F_h$   $\Gamma$ -converges to  $RF$  the lower semi-continuous envelope of  $F$ .

Now we illustrate these ideas by giving an example [84].

**Example** Let  $\Omega$  be an open of  $R^N$  and let  $a_h$  be a sequence of functions satisfying for all  $h$ :  $0 < c_1 \leq a_h(x) \leq c_2$  a.e.  $x \in \Omega$  for some constants  $c_1$  and  $c_2$ . Up to subsequences there exist from Theorem 2.1.1  $a, b \in L^\infty(\Omega)$  such that  $a_h \xrightarrow[L^\infty(\Omega)]{*} a$  and  $\frac{1}{a_h} \xrightarrow[L^\infty(\Omega)]{*} b$ . Then let us consider the sequence of functionals  $F_h : L^2(\Omega) \rightarrow R$  defined by

$$F_h(u) = \int_{\Omega} a_h u^2 dx.$$

We claim that  $F_h$   $\Gamma$ -converges to  $F(u) = \int_{\Omega} \frac{u^2}{b} dx$  for the weak topology of  $L^2(\Omega)$ . According to the definition, we have to prove:

- (i) For every  $u \in L^2(\Omega)$  and for every  $u_h \xrightarrow[L^2(\Omega)]{} u$  then  

$$\varliminf_{h \rightarrow \infty} F_h(u_h) \geq F(u).$$
- (ii) For every  $u \in L^2(\Omega)$  there exists a sequence  $u_h \in L^2(\Omega)$ ,  $u_h \xrightarrow[L^2(\Omega)]{} u$  such that  $\overline{\lim}_{h \rightarrow \infty} F_h(u_h) \leq F(u)$ .

We begin to examine (ii).

Let  $u \in L^2(\Omega)$  and let us define  $u_h = \frac{u}{b a_h}$ . Since  $\frac{1}{a_h} \xrightarrow[L^\infty(\Omega)]{*} b$  it is clear that  $u_h \xrightarrow[L^2(\Omega)]{} u$  and  $\overline{\lim}_{h \rightarrow \infty} F_h(u_h) = \overline{\lim}_{h \rightarrow \infty} \int_{\Omega} \frac{u^2}{b^2 a_h} dx = \int_{\Omega} \frac{u^2}{b} dx$ , so  $\overline{\lim}_{h \rightarrow \infty} F_h(u_h) \leq F(u)$  (in fact the equality holds).

For proving (i), let  $u \in L^2(\Omega)$  and  $v_h$  any sequence such that  $v_h \xrightarrow{L^2(\Omega)} u$ .

Let  $u_h = \frac{u}{b a_h}$ , then from the inequality  $a_h (v_h - u_h)^2 \geq 0$ , we deduce

$$a_h v_h^2 \geq a_h u_h^2 + 2 a_h u_h (v_h - u_h) = -a_h u_h^2 + 2 \frac{u}{b} v_h$$

which yields

$$F_h(v_h) \geq -F_h(u_h) + 2 \int_{\Omega} \frac{u}{b} v_h \, dx$$

therefore

$$\varliminf_{h \rightarrow \infty} F_h(v_h) \geq -\overline{\lim}_{h \rightarrow \infty} F_h(u_h) + 2 \int_{\Omega} \frac{u^2}{b} \, dx \geq F(u),$$

which concludes the proof. ■

**Remark** As an exercise, the reader can see for himself that the  $\Gamma$ -limit of  $F_h$  for the strong topology of  $L^2(\Omega)$  is  $G(u) = \int_{\Omega} a u^2 \, dx!$  ■

## 2.2 The space of functions of bounded variation

In most computer vision problems the discontinuities in the images are significant and important features. So we need to be able to represent discontinuous functions. Unfortunately, classical Sobolev spaces do not allow to take into account such phenomenon since the gradient of a Sobolev function is a function. When  $u$  is discontinuous the gradient of  $u$  has to be understood as a measure and the space  $BV(\Omega)$  of functions of *bounded variation* [106, 120, 89, 123] is well-adapted for this purpose. In this section we recall some basic definitions about measures, then  $BV(\Omega)$  is defined and its main properties are examined. We end this section by introducing the notion of convex functions of measures.

### 2.2.1 Basic definitions on measures

**Definition 2.2.1 (algebra)** Let  $X$  be a non empty set and let  $\mathfrak{I}$  be a collection of subsets of  $X$ .

- (i) We say that  $\mathfrak{I}$  is an algebra if  $\emptyset \in \mathfrak{I}$  and  $E_1 \cup E_2 \in \mathfrak{I}$ ,  $X - E_1 \in \mathfrak{I}$  whenever  $E_1, E_2 \in \mathfrak{I}$ .

- (ii) We say that an algebra  $\mathfrak{I}$  is a  $\sigma$ -algebra if for any sequences  $(E_h) \subset \mathfrak{I}$  its union  $\bigcup_h E_h$  belongs to  $\mathfrak{I}$ .  $\sigma$ -algebras are closed under countable intersections.
- (iii) If  $(X, \tau)$  is a topological space we denote by  $B(X)$  the  $\sigma$ -algebra generated by the open subsets of  $X$  (the smallest  $\sigma$ -algebra containing the open subsets).

We can now give the definition of a positive measure.

**Definition 2.2.2 (positive measure)** Let  $\mu : \mathfrak{I} \rightarrow [0, +\infty]$  we say that  $\mu$  is a positive measure if  $\mu(\emptyset) = 0$  and  $\mu$  is  $\sigma$ -additive on  $\mathfrak{I}$  i.e. for any sequences  $(E_h)$  of pairwise disjoint elements of  $\mathfrak{I}$ :

$$\mu\left(\bigcup_{h=0}^{\infty} E_h\right) = \sum_{h=0}^{\infty} \mu(E_h).$$

We say that  $\mu$  is bounded if  $\mu(X) < \infty$ .

We will also use the notion of vector-valued measures.

**Definition 2.2.3 (vector-valued measure)** Let  $(X, \mathfrak{I})$  be a measure space,  $m \in N$ ,  $m \geq 1$ . We say that  $\mu : \mathfrak{I} \rightarrow R^m$  is a measure if  $\mu(\emptyset) = 0$  and for any sequences  $(E_h)$  of pairwise disjoint elements of  $\mathfrak{I}$ ,  $\mu\left(\bigcup_{h=0}^{\infty} E_h\right) = \sum_{h=0}^{\infty} \mu(E_h)$ . If  $m = 1$  we say that  $\mu$  is a real or a signed measure and if  $m > 1$  we say that  $\mu$  is a vector-valued measure. We will denote by  $M$  the space of vector-valued Radon measures<sup>2</sup>. If  $\mu$  is a measure, we define its total variation  $|\mu|$  for every  $E \in \mathfrak{I}$  as follows:

$$|\mu|(E) = \sup \left\{ \sum_{h=0}^{\infty} \left| \mu(E_h) ; E_h \in \mathfrak{I} \text{ pairwise disjoint}, E = \bigcup_{h=0}^{\infty} E_h \right| \right\}.$$

$|\mu|$  is bounded measure.

**Definition 2.2.4 ( $\mu$ -negligible)** Let  $\mu$  be a positive measure, we say that  $A \subset X$  is  $\mu$ -negligible if there exists  $E \in \mathfrak{I}$  such that  $A \subset E$  and  $\mu(E) = 0$ . A property  $P(x)$  depending on the point  $x \in X$  holds  $\mu$ -a.e. in  $X$  if the set where  $P$  fails is a  $\mu$ -negligible set.

We end this subsection by recalling what is perhaps the most important tool in integration theory, the Lebesgue decomposition. Let us first introduce some definitions.

**Definition 2.2.5 (Radon Nikodym derivative)** Let  $\mu$  be a bounded positive measure, and  $\nu$  be a measure. Let  $B(x, r)$  be the ball of center

---

<sup>2</sup>A Radon measure on  $R^N$  is a measure which is finite in each compact set  $K \subset R^N$ .

$x$  and radius  $r$ , we note

$$\Delta(x, r) = \begin{cases} \frac{\nu(B(x, r))}{\mu(B(x, r))} & \text{if } \mu(B(x, r)) > 0 \\ +\infty & \text{if } \mu(B(x, r)) = 0 \end{cases}$$

and  $\overline{D_\mu}\nu(x) = \lim_{r \rightarrow 0} \Delta(x, r)$ ,  $\underline{D_\mu}\nu(x) = \lim_{r \rightarrow 0} \Delta(x, r)$ .

If  $\overline{D_\mu}\nu(x) = \underline{D_\mu}\nu(x) < \infty$ , we say that  $\nu$  is differentiable with respect to  $\mu$  in  $x$  and we note  $\frac{d\nu}{d\mu}(x) = \overline{D_\mu}\nu(x) = \underline{D_\mu}\nu(x)$  the Radon Nikodym derivative of  $\nu$  with respect to  $\mu$ .

**Definition 2.2.6 (absolutely continuous, mutually singular)** Let  $\mu$  be a positive measure, and  $\nu$  be a measure, we say that  $\nu$  is absolutely continuous with respect to  $\mu$  and we write  $\nu \ll \mu$  if  $\mu(E) = 0 \Rightarrow \nu(E) = 0$ . We say that  $\mu$  and  $\nu$  are mutually singular and we write  $\mu \perp \nu$  if there exists a set  $E$  such that  $\mu(R^N - E) = \nu(E) = 0$ .

The theorem is the following:

**Theorem 2.2.1 (Lebesgue decomposition)** Let  $\mu$  be a positive bounded measure on  $(X, \mathfrak{F})$  and  $\nu$  a vector-valued measure on  $(X, \mathfrak{F})$  then there exists a unique pair of measures  $\nu_{ac}$  and  $\nu_s$  such that:

$$\nu = \nu_{ac} + \nu_s, \quad \nu_{ac} \ll \mu, \quad \nu_s \perp \mu$$

Moreover  $\frac{d\nu}{d\mu} = \frac{d\nu_{ac}}{d\mu}, \frac{d\nu_s}{d\mu} = 0$  and  $\nu(A) = \int_A \frac{d\nu}{d\mu} d\mu + \nu_s(A)$  for all  $A \in \mathfrak{F}$ , where  $\nu_{ac}$  and  $\nu_s$  are the absolute part and the singular part of  $\nu$ .

## 2.2.2 Definition of $BV(\Omega)$

Let  $\Omega$  be a bounded open subset of  $R^N$  and let  $u$  be a function in  $L^1(\Omega)$ . We note

$$\int_{\Omega} |Du| = \sup \left\{ \int_{\Omega} u \operatorname{div} \varphi dx; \varphi = (\varphi_1, \varphi_2, \dots, \varphi_N) \in C_0^1(\Omega)^N, |\varphi|_{L^\infty(\Omega)} \leq 1 \right\}$$

where  $\operatorname{div} \varphi = \sum_{i=1}^N \frac{\partial \varphi_i}{\partial x_i}(x)$ ,  $dx$  is the Lebesgue measure, and  $C_0^1(\Omega)^N$  is the space of continuously differentiable functions with compact support in  $\Omega$ . The inequality  $|\varphi|_{L^\infty(\Omega)} \leq 1$  means that all the components of the vector-valued function  $\varphi$  have a  $L^\infty(\Omega)$ -norm less than one.

## Examples

- If  $u \in C^1(\Omega)$ , then  $\int_{\Omega} u \operatorname{div} \varphi \, dx = - \int_{\Omega} \nabla u \cdot \varphi \, dx$  and  

$$\int_{\Omega} |Du| = \int_{\Omega} |\nabla u(x)| \, dx.$$
- Let  $u$  be defined in  $(-1, +1)$  by  $u(x) = -1$  if  $-1 \leq x < 0$  and  $u(x) = +1$  if  $0 < x \leq 1$ , then  $\int_{-1}^{+1} u \varphi' \, dx = -2\varphi(0)$  and  $\int_{-1}^{+1} |Du| = 2$ .

We can remark that  $Du$  the distributional derivative of  $u$  is equal to  $2 \delta_0$  where  $\delta_0$  is the Dirac measure in 0. In fact, we have the decomposition  $Du = 0 \, dx + 2\delta_0$ . ■

**Definition 2.2.7 ( $BV(\Omega)$ )** We define  $BV(\Omega)$  the space of functions of bounded variation as:

$$BV(\Omega) = \left\{ u \in L^1(\Omega); \int_{\Omega} |Du| < \infty \right\}.$$

We are going to show that if  $u \in BV(\Omega)$ , then  $Du$  (the distributional gradient of  $u$ ) can be identified to a Radon vector-valued measure. Let  $u \in BV(\Omega)$  and  $L: C_0^1(\Omega)^N \rightarrow R$  be the functional defined by:

$$L(\varphi) = \int_{\Omega} u \operatorname{div} \varphi \, dx.$$

$L$  is linear and since  $u \in BV(\Omega)$  we have:

$$\sup \{L(\varphi); \varphi \in C_0^1(\Omega)^N, |\varphi|_{L^\infty(\Omega)} \leq 1\} = c < \infty$$

where  $c$  is a constant depending only upon  $\Omega$  and  $u$ . So for all  $\varphi \in C_0^1(\Omega)^N$ :

$$|L(\varphi)| \leq c |\varphi|_{L^\infty(\Omega)}. \quad (2.6)$$

Now let  $K \subset \Omega$  be a compact set and  $\varphi \in C_0(\Omega)^N$ ,  $\operatorname{supp} \varphi \subset K$ , we can always find a sequence  $\varphi_k \in C_0^1(\Omega)^N$  such that:

$$\begin{aligned} \varphi_k &\rightarrow \varphi \text{ uniformly, } k \rightarrow \infty \\ |\varphi_k|_{L^\infty(\Omega)} &\leq |\varphi|_{L^\infty(\Omega)}, \forall k. \end{aligned}$$

Let  $\bar{L}(\varphi) = \lim_{k \rightarrow \infty} L(\varphi_k)$ . From (2.6) this limit exists and is independent of the choice of the sequence  $\varphi_k$ . Therefore  $L$  uniquely extends to a linear continuous functional:

$$\bar{L}: C_0(\Omega)^N \rightarrow R.$$

From the Riesz representation Theorem [216], there exists a Radon measure  $\mu$  (a positive measure finite on compact sets of  $R^N$ ) and a  $\mu$ -measurable function  $\sigma$  such that:

$$|\sigma(x)| = 1 \quad \mu\text{-a.e. } x$$

$$\int_{\Omega} u \operatorname{div} \varphi \, dx = - \int_{\Omega} \sigma \cdot \varphi \, d\mu \quad \text{for all } \varphi \in C_0^1(\Omega)^N$$

which means that  $Du$  is a vector-valued Radon measure ( $Du = \sigma \, d\mu$ ).

An important example is the case when  $u = \chi_A$ , the characteristic function of a subset  $A$  of  $\mathbb{R}^N$ . Then:

$$\int_{\Omega} |Du| = \sup \left\{ \int_A \operatorname{div} \varphi \, dx; \varphi \in C_0^1(\Omega)^N, |\varphi|_{L^\infty(\Omega)} \leq 1 \right\}.$$

If this supremum is finite,  $A$  is called a set of finite perimeter in  $\Omega$  and we note:

$$\int_{\Omega} |Du| = \operatorname{Per}_{\Omega}(A).$$

If  $\partial A$  is smooth,  $\operatorname{Per}_{\Omega}(A)$  coincides with the classical length ( $N = 2$ ) or surface area ( $N = 3$ ).

### 2.2.3 Properties of $BV(\Omega)$

We summarize below the main properties of  $BV(\Omega)$  that we will use in the sequel. We assume that  $\Omega$  is bounded and has a lipschitz boundary.

(P<sub>1</sub>) *Lower-semicontinuity*

Let  $u_j \in BV(\Omega)$  and  $u_j \xrightarrow[L^1(\Omega)]{} u$ , then  $\int_{\Omega} |Du| \leq \lim_{j \rightarrow \infty} \int_{\Omega} |Du_j|$ .

(P<sub>2</sub>) *Trace*

The trace operator  $\operatorname{tr} : u \rightarrow u|_{\partial\Omega}$ , from  $BV(\Omega)$  to  $L^1(\partial\Omega, \mathcal{H}^{N-1})$ , is linear continuous for the strong topology of  $BV(\Omega)$ .  $\mathcal{H}^{N-1}$  denotes the  $N - 1$  dimensional measure (see Definition 2.2.8).

(P<sub>3</sub>) *A weak\* topology*

$BV(\Omega)$  is a Banach space endowed with the norm

$|u|_{BV(\Omega)} = |u|_{L^1(\Omega)} + \int_{\Omega} |Du|$ . We will not use this topology which

has no good compactness properties. Classically in  $BV(\Omega)$  one works with the  $BV - w^*$  topology defined as:

$$u_j \xrightarrow[BV-w^*]{*} u \Leftrightarrow u_j \xrightarrow[L^1(\Omega)]{} u \text{ and } Du_j \xrightarrow[M]{*} Du \quad (2.7)$$

where  $Du_j \xrightarrow[M]{*} Du$  means  $\int_{\Omega} \varphi Du_j \rightarrow \int_{\Omega} \varphi Du$  for all  $\varphi$  in  $C_0(\Omega)^N$ .

Equipped with this topology,  $BV(\Omega)$  has some interesting compactness properties.

(P<sub>4</sub>) *Compactness*

Every uniformly bounded sequences  $u_j$  in  $BV(\Omega)$  is relatively compact in  $L^p(\Omega)$  for  $1 \leq p < \frac{N}{N-1}$ ,  $N \geq 1$ . Moreover there exist a subsequence  $u_{j_k}$  and  $u$  in  $BV(\Omega)$  such that  $u_{j_k} \xrightarrow[\text{BV-w}^*]{*} u$ . We also recall that  $BV(\Omega)$  is continuously embedded in  $L^p(\Omega)$  with  $p = +\infty$  if  $N = 1$ , and  $p = \frac{N}{N-1}$  otherwise.

(P<sub>5</sub>) *Decomposability of  $BV(\Omega)$* 

We are going to show that  $Du$  can decomposed as the sum of a regular measure and a singular measure. Before doing that we need the definition of the Hausdorff measure.

**Definition 2.2.8 (Hausdorff measure)** *Let  $k \in [0, +\infty]$  and  $A \subset R^N$ . The  $k$ -dimensional Hausdorff measure of  $A$  is given by:*

$$\mathcal{H}^k(A) = \lim_{\delta \rightarrow 0} \mathcal{H}_\delta^k(A)$$

where for  $0 < \delta \leq \infty$ ,  $\mathcal{H}_\delta^k(A)$  is defined by

$$\mathcal{H}_\delta^k(A) = \frac{w_k}{2^k} \inf \left\{ \sum_{i \in I} |\text{diam}(A_i)|^k, \text{ diam}(A_i) \leq \delta, A \subset \bigcup_{i \in I} A_i \right\}$$

for finite or countable covers  $(A_i)_{i \in I}$ ;  $\text{diam}(A_i)$  denotes the diameter of the set  $A_i$  and  $w_k$  is a normalization factor equal to  $\pi^{\frac{k}{2}} \Gamma(1 + \frac{k}{2})$

where  $\Gamma(t) = \int_0^\infty s^{t-1} e^{-s} ds$  is the  $\Gamma$ -function ( $w_k$  coincides with the

Lebesgue measure of the unit ball of  $R^k$  if  $k \geq 1$  is an integer).

We define the Hausdorff dimension of  $A$  by:

$$\mathcal{H}\text{-dim}(A) = \inf\{k \geq 0; \mathcal{H}(A) = 0\}.$$

$\mathcal{H}^k$  is a measure in  $R^N$ ,  $\mathcal{H}^N$  coincides with the Lebesgue measure  $L^N$  and for  $1 \leq k \leq N$ , integer,  $\mathcal{H}^k(A)$  is the classical  $k$ -dimensional area of  $A$  if  $A$  is a  $C^1$   $k$ -dimensional manifold embedded in  $R^N$ . Moreover if  $k > k' \geq 0$  then  $\mathcal{H}^k(A) > 0 \Rightarrow \mathcal{H}^{k'}(A) = +\infty$ .

Let us go back to  $BV(\Omega)$ . If  $u$  belongs to  $BV(\Omega)$  and if in Theorem 2.2.1 we choose  $\mu = dx$ , the  $N$ -dimensional Lebesgue measure and  $\nu = Du$ , we get

$$Du = \nabla u \, dx + D_s u$$

where  $\nabla u(x) = \frac{d(Du)}{dx}(x) \in L^1(\Omega)$  and  $D_s u \perp dx$ .  $\nabla u$  is also called the approximate derivative of  $u$  (see [9]). In fact, we can say more for  $BV(\Omega)$  functions. In [9] Ambrosio showed that the singular part

$D_s u$  of  $Du$  can be decomposed into a "jump" part  $J_u$  and a "Cantor" part  $C_u$ . Before specifying what is exactly  $J_u$  we have to define the notion of approximate limit. Let  $B(x, r)$  be the ball of center  $x$  and radius  $r$  and let  $u \in BV(\Omega)$ , we define the approximate upper limit  $u^+(x)$  and the approximate lower limit  $u^-(x)$  by (see Figure 2.1)

$$u^+(x) = \inf \left\{ t \in [-\infty, +\infty] ; \lim_{r \rightarrow 0} \frac{dx(\{u > t\} \cap B(x, r))}{r^N} = 0 \right\}$$

$$u^-(x) = \sup \left\{ t \in [-\infty, +\infty] ; \lim_{r \rightarrow 0} \frac{dx(\{u < t\} \cap B(x, r))}{r^N} = 0 \right\}.$$

If  $u \in L^1(\Omega)$ , then:

$$\lim_{r \rightarrow 0} \frac{1}{|B(x, r)|} \int_{B(x, r)} |u(x) - u(y)| dy = 0 \quad \text{a.e. } x. \quad (2.8)$$

A point  $x$  for which (2.8) holds is called a Lebesgue point of  $u$  and we have:

$$u(x) = \lim_{r \rightarrow 0} \frac{1}{|B(x, r)|} \int_{B(x, r)} u(y) dy, \quad (2.9)$$

and  $u(x) = u^+(x) = u^-(x)$ . We denote by  $S_u$  the jump set, that is to say the complement, up to a set of  $\mathcal{H}^{N-1}$  zero measure, of the set of Lebesgue points:

$$S_u = \{x \in \Omega ; u^-(x) < u^+(x)\}.$$

$S_u$  is countably rectifiable and for  $\mathcal{H}^{N-1}$ - a.e.  $x \in \Omega$ , we can define a normal  $n_u(x)$ .

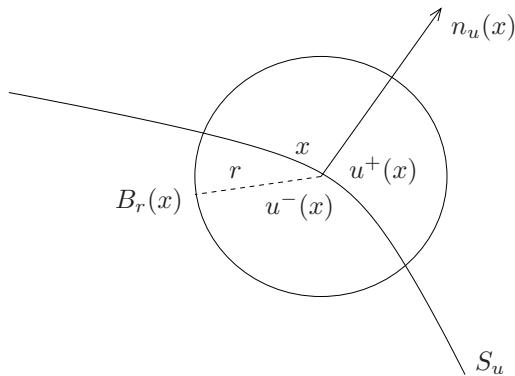


Figure 2.1. Definition of  $u^+$ ,  $u^-$  and the jump set  $S_u$

So, the result proved in [9] is:

$$Du = \nabla u dx + (u^+ - u^-) n_u \mathcal{H}_{|S_u}^{N-1} + C_u. \quad (2.10)$$

$J_u = (u^+ - u^-) n_u \mathcal{H}_{|S_u}^{N-1}$  is the jump part and  $C_u$  is the Cantor part of  $D_s u$ . We have  $C_u \perp dx$  and  $C_u$  is diffuse i.e.  $C_u\{x\} = 0$ . More generally we have  $C_u(B) = 0$  for all  $B$  such that  $\mathcal{H}^{N-1}(B) < \infty$ , that is to say the Hausdorff dimension of the support of  $C_u$  is strictly greater than  $N - 1$ . From (2.10) we can deduce the total variation of  $Du$ :

$$\begin{aligned} |Du|(\Omega) &= \int_{\Omega} |Du| = \\ &= \int_{\Omega} |\nabla u|(x) dx + \int_{S_u} |u^+ - u^-| d\mathcal{H}^{N-1} + \int_{\Omega - S_u} |C_u|. \end{aligned}$$

**Remark** About differentiation. An important fact is that a summable function is “approximately continuous” at almost every point. This means that if  $u$  is simply in  $L^1(\Omega)$ , the right-hand side of (2.9) exists  $dx$  a.e.. However, if  $u$  is also in  $BV(\Omega)$ , we can say more. Let us define:

$$\dot{u}(x) = \frac{u^+(x) + u^-(x)}{2} \quad \mathcal{H}^{N-1} \text{ a.e. on } S_u. \quad (2.11)$$

Then, it can be shown [247, 106] that  $\dot{u}$  is well-defined  $\mathcal{H}^{N-1}$  a.e. on  $S_u$ . An interesting property of  $\dot{u}$  is that we have the following approximation result :

$$\dot{u}(x) = \lim_{\varepsilon \rightarrow 0} \eta_{\varepsilon} \star u(x) \quad \mathcal{H}^{N-1} \text{ a.e.} \quad (2.12)$$

where  $(\eta_{\varepsilon})$  are the usual mollifiers (see Section 2.5.3). The function  $\dot{u}$  is called the *precise representation* of  $u$  since it permits in some way to define  $u$ ,  $\mathcal{H}^{N-1}$  a.e.. Remark that  $\dot{u}$  and  $u$  are in fact the same elements in  $BV(\Omega)$  (they belong to the same equivalence class of  $dx$  a.e. equal functions) therefore their distributional derivatives are the same. ■

#### 2.2.4 Convex functions of measures

We would like to give a sense to the formal writing:

$$\int_{\Omega} \Psi(Du)$$

when  $u$  is a  $BV(\Omega)$  function i.e. when  $Du$  is a measure. According to the above discussion we are for the moment only able to define the total variation of  $Du$  i.e. when  $\Psi(\xi) = |\xi|$ . Let us extend this to more general  $\Psi$ .

■ Convex functions of measures [123, 89].

Let  $\phi : R \rightarrow R^+$  be convex, even, non decreasing on  $R^+$  with linear growth at infinity and let  $\phi^\infty$  be the recession function of  $\phi$  defined by

$\phi^\infty(z) = \lim_{s \rightarrow \infty} \frac{\phi(sz)}{s}$ . Then for  $u \in BV(\Omega)$  and if  $\Psi(\xi) = \phi(|\xi|)$ , we set

$$\int_{\Omega} \Psi(Du) = \int_{\Omega} \phi(|\nabla u(x)|) dx + \phi^\infty(1) \int_{S_u} |u^+ - u^-| d\mathcal{H}^{N-1} + \phi^\infty(1) \int_{\Omega - S_u} |C_u|.$$

Of course, if  $\Psi(\xi) = |\xi|$ , this definition coincides with the total variation of  $Du$ . The main consequence of this definition is that:

$$u \rightarrow \int_{\Omega} \Psi(Du) \text{ is l.s.c. for the } BV - w^* \text{ topology.}$$

We will use this notion in image restoration (Chapter 3).

## 2.3 Viscosity solutions in PDEs

### 2.3.1 Around the eikonal equation

Until now, we have recalled some mathematical tools necessary to tackle problems in computer vision from a variational point of view. But in many situations, e.g. nonlinear filtering in restoration, equations we have to solve do not come from variational principles. They are PDEs that are not Euler-Lagrange equations of functionals so we need different tools.

To convince the reader of the difficulties in studying nonlinear PDEs, let us consider this very simple 1-D example:

$$\begin{cases} |u'(x)| = 1 & \text{in } (0, 1) \\ u(0) = u(1) = 0. \end{cases} \quad (2.13)$$

which is the eikonal equation. Several questions appear unclear:

- The existence. Clearly (2.13) cannot admit a  $C^1$  solution since in this case from Rolle's Theorem we would deduce the existence of  $x_0 \in ]0, 1[$  such that  $u'(x_0) = 0$  which is in contradiction with  $|u'(x_0)| = 1$ . So the gradient of  $u$  has to "break" and a theory involving non-regular solutions has to be developed.
- The uniqueness. For example the function  $u^+(x) = \frac{1}{2} - \left| \frac{1}{2} - x \right|$  is a solution of (2.13) for almost every  $x \in (0, 1)$ . But  $u^-(x) = -u^+(x)$  is also a solution. In fact there is an infinite number of solutions of the form:

$$\begin{aligned} u_n(0) &= u_n(1) = 0 \\ u'_n(x) &= 1 \quad \text{if } x \in \left[ \frac{2k}{2^n}, \frac{2k+1}{2^n} \right] \quad k = 0, \dots, 2^n - 1. \\ u'_n(x) &= -1 \quad \text{if } x \in \left[ \frac{2k+1}{2^n}, \frac{2k+2}{2^n} \right] \end{aligned}$$

Some examples are given in Figure 2.2. Moreover  $0 \leq u_n(x) \leq \frac{1}{2^n}$  so  $u_n(x) \rightarrow 0$  uniformly ( $n \rightarrow \infty$ ). However, 0 is not a solution of (2.13). Note that  $u^+(x)$  is the greatest solution of (2.13).

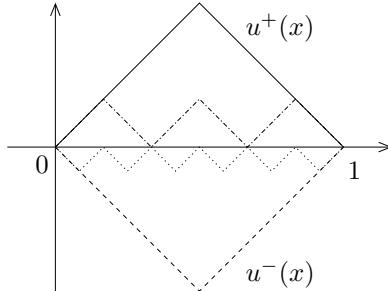


Figure 2.2. Examples of admissible solutions of the 1-D eikonal equation

- The compatibility conditions. Let us consider the  $N$ -dimensional eikonal equation.

$$\begin{cases} |\nabla u(x)| = f(x) & \text{in } \Omega, \text{ a bounded open set of } R^N \\ u|_{\partial\Omega} = u_0(x). \end{cases} \quad (2.14)$$

Let  $x, y \in \bar{\Omega}$  and  $\xi(t) : [0, T] \rightarrow R^N$  be a lipschitz path such that  $\xi(0) = x$ ,  $\xi(T) = y$  and  $|\xi'(t)| \leq 1$  a.e.  $t \in (0, 1)$ . We formally have

$$u(y) - u(x) = \int_0^T \nabla u(\xi(s)) \cdot \xi'(s) \, ds$$

thus

$$|u(y) - u(x)| \leq \int_0^T f(\xi(s)) \, ds$$

from which we deduce

$$|u(y) - u(x)| \leq L(x, y)$$

where

$$L(x, y) = \inf_{\xi, T} \left\{ \int_0^T f(\xi(s)) \, ds; \xi(0) = x; \xi(T) = y; |\xi'(t)| \leq 1; \xi(t) \in \bar{\Omega} \right\}.$$

Writing this necessary condition for  $x, y$  belonging to the boundary  $\partial\Omega$  of  $\Omega$  we get

$$|u_0(y) - u_0(x)| \leq L(x, y) \quad (2.15)$$

which is a compatibility condition on the data.

## Remarks

- It can be shown [160] that in fact (2.15) is also a sufficient existence condition.
- If  $f \equiv 1$  and if  $\Omega$  is convex, (2.15) writes as  $|u_0(y) - u_0(x)| \leq |y - x|$ . ■

### 2.3.2 Definition of viscosity solutions

As it is suggested in the previous example, there is a need to define a suitable and well-defined framework. Initiated in the eighties for first order PDEs by Crandall and PL Lions [82, 80, 160], the theory of viscosity solutions has shown a very successful development and has been extended for second order equations [81].

#### Theory of viscosity solutions [81, 24, 74]

Generally speaking the theory of viscosity solutions deals with equations called *Hamilton Jacobi equations* of the form:

$$\frac{\partial u}{\partial t}(t, x) + H(t, x, \nabla u(x), \nabla^2 u(x)) = 0 \quad t \geq 0, x \in \Omega \quad (2.16)$$

with boundary and initial conditions.  $H : ]0, T] \times \Omega \times R \times R^N \times S^N \rightarrow \mathfrak{I}$  is called an *hamiltonian*,  $S^N$  is the set of  $N \times N$  symmetric matrices and  $\nabla^2 u$  stands for the hessian matrix of  $u$ .  $H$  will be always supposed degenerate elliptic:

$$H(t, x, u, p, S) \geq H(t, x, u, p, S') \text{ if } S \leq S' \quad (2.17)$$

( $S^N$  is ordered with the natural order:  $S \leq S' \Leftrightarrow \xi^t(S' - S)\xi \geq 0$  for all  $\xi \in R^N$ ). The theory aims to define generalized solutions of (2.16) and particularly solutions that are only continuous. The following theorem, stated for the stationary case, is useful to understand the definition of viscosity solutions.

**Theorem 2.3.1** *Let  $H : \Omega \times R \times R^N \times S^N \rightarrow \mathfrak{I}$  be continuous and degenerate elliptic and let  $u \in C^2(\Omega)$  then  $u$  is solution of  $H(x, u(x), \nabla u(x), \nabla^2 u(x)) = 0$  in  $\Omega$  if and only if*

(i)  $\forall \phi \in C^2(\Omega)$ ,  $\forall x_0 \in \Omega$  local maximum of  $(u - \phi)(x)$  then

$$H(x_0, u(x_0), \nabla \phi(x_0), \nabla^2 \phi(x_0)) \leq 0.$$

(ii)  $\forall \phi \in C^2(\Omega)$ ,  $\forall x_0 \in \Omega$  local minimum of  $(u - \phi)(x)$  then

$$H(x_0, u(x_0), \nabla \phi(x_0), \nabla^2 \phi(x_0)) \geq 0.$$

**Proof** Let  $\phi \in C^2(\Omega)$ , and  $x_0 \in \Omega$  be a local maximum of  $(u - \phi)(x)$ , then from classical arguments:

$$\begin{aligned}\nabla u(x_0) &= \nabla \phi(x_0) \\ \nabla^2 u(x_0) &\leq \nabla^2 \phi(x_0).\end{aligned}$$

Hence from (2.17):

$$\begin{aligned}H(x_0, u(x_0), \nabla \phi(x_0), \nabla^2 \phi(x_0)) &= H(x_0, u(x_0), \nabla u(x_0), \nabla^2 \phi(x_0)) \leq \\ &\leq H(x_0, u(x_0), \nabla u(x_0), \nabla^2 u(x_0)) = 0.\end{aligned}$$

The inequality (ii) is proven by similar arguments.

Reciprocally if (i) and (ii) are true, by choosing  $\phi = u$  and since each point  $x \in \Omega$  is both local maximum and local minimum of  $(u - \phi)(x) = 0$ , we get

$$\begin{aligned}H(x_0, u(x_0), \nabla u(x_0), \nabla^2 u(x_0)) &\leq 0 \\ \text{and } H(x_0, u(x_0), \nabla u(x_0), \nabla^2 u(x_0)) &\geq 0 \\ \text{i.e. } H(x_0, u(x_0), \nabla u(x_0), \nabla^2 u(x_0)) &= 0.\end{aligned}$$

■

► In the former equivalence (Theorem 2.3.1),  $u$  only needs to be continuous. The derivatives are in fact evaluated on the test functions  $\phi$ .

This observation leads to the following definition:

**Definition 2.3.1 (viscosity subsolution, supersolution, solution)**  
Let  $H : ]0, T] \times \Omega \times \mathbb{R} \times \mathbb{R}^N \times \mathbb{S}^N \rightarrow \mathfrak{I}$  be continuous, satisfying (2.17) and let  $u \in C([0, T] \times \Omega)$  then

- (i)  $u$  is a viscosity subsolution of (2.16) if and only if  $\forall \phi \in C^1([0, T] \times \Omega)$ ,  $\forall (t_0, x_0)$  local maximum of  $(u - \phi)(t, x)$  then  
 $H(t_0, x_0, u(t_0, x_0), \nabla \phi(t_0, x_0), \nabla^2 \phi(t_0, x_0)) \leq 0$ .
- (ii)  $u$  is a viscosity supersolution of (2.16) if and only if  $\forall \phi \in C^1([0, T] \times \Omega)$ ,  $\forall (t_0, x_0)$  local minimum of  $(u - \phi)(t, x)$  then  
 $H(t_0, x_0, u(t_0, x_0), \nabla \phi(t_0, x_0), \nabla^2 \phi(t_0, x_0)) \geq 0$ .
- (iii)  $u$  is a viscosity solution of (2.16) if  $u$  is both a viscosity subsolution and a viscosity supersolution.

### 2.3.3 About the existence

We can wonder why the word "viscosity" is used in the former definition. This terminology comes from historical reasons. Initial work on Hamilton Jacobi equations were first concerned with first order PDEs like:

$$H(x, u(x), \nabla u(x)) = 0. \quad (2.18)$$

The way to solve (2.18) was to introduce in (2.18) an additional regularizing (viscosity) term:

$$-\varepsilon \Delta u(x) + H(x, u(x), \nabla u(x)) = 0. \quad (2.19)$$

To prove existence, one can follow this two steps:

- (i) Under suitable hypotheses, we show that (2.19) admits a unique regular solution  $u_\varepsilon$  that we uniformly bound.
- (ii) The second step is to study the behaviour of  $u_\varepsilon$  as  $\varepsilon \rightarrow 0$  and to pass to the limit in (2.19).

This method is well known in mechanics as the *vanishing viscosity method*. We will use it in Chapter 4 (for active contours models).

In this approach, the main concern is in fact the behaviour of the solution as  $\varepsilon \rightarrow 0$ . We have the following stability result:

**Lemma 2.3.1 (stability [24])** *Let  $H_\varepsilon(x, u, p, M)$  be a sequence of continuous functions on  $\Omega \times R \times R^N \times S^N$  ( $\Omega \subset R_N$ ) satisfying the ellipticity condition:*

$$H_\varepsilon(x, u, p, M_1) \leq H_\varepsilon(x, u, p, M_2) \quad \text{if } M_1 - M_2 \geq 0$$

*and let  $u_\varepsilon \in C(\Omega)$  be a viscosity solution of  $H_\varepsilon(x, u_\varepsilon, \nabla u_\varepsilon, \nabla^2 u_\varepsilon) = 0$  in  $\Omega$ . If  $u_\varepsilon \rightarrow u$  in  $C(\Omega)$  and if  $H_\varepsilon \rightarrow H$  in  $C(\Omega \times R \times R^N \times S^N)$  then  $u$  is a viscosity solution of  $H(x, u, \nabla u, \nabla^2 u) = 0$  in  $\Omega$ .*

Another way to get a viscosity solution is to use the *Perron's method*. Roughly speaking the method runs as follows. Under appropriate assumptions:

- (i) One proves that the set  $S$  of subsolutions is not void.
- (ii) Let  $w(x) = \sup \{v(x); v \in S\}$ . By stability  $w(x)$  is a subsolution. It remains to show that  $w(x)$  is a supersolution.

### 2.3.4 About the uniqueness

Theorems concerning uniqueness are perhaps the strong point of this theory. But, as expected, it is also the most difficult part. For second order Hamilton Jacobi equations uniqueness results are build on a powerful technical lemma due to Crandall-Ishii. Before stating it, we need the definitions of super and sub-jet of  $u$ .

**Definition 2.3.2 (super-jets, sub-jets)** *Let  $u: \Omega_t = ]0, T] \times \Omega \rightarrow R$  then we call super-jet of  $u(s, z)$  the set  $P_\Omega^+ u(s, z)$  defined by :  $(a, p, X) \in R \times R^N \times S^N$  lies in  $P_\Omega^+ u(s, z)$  if  $(s, z) \in \Omega_t$  and*

$$u(t, x) \leq u(s, z) + a(t-s) + p \cdot (x-z) + \frac{1}{2} (x-z)^t X (x-z) + o(|t-s| + |x-z|^2)$$

as  $\Omega_t \supset (t, x) \rightarrow (s, z)$ . Similarly the sub-jet of  $u(s, z)$  is defined by:

$$P_{\Omega}^- u(s, z) = -P_{\Omega}^+(-u)(s, z).$$

We can show [24, 81] that  $u$  is a subsolution (a supersolution) of (2.15) if and only:

$$a + H(t, x, u(t, x), p, X) \leq 0 \quad (\geq 0) \quad (2.20)$$

for all  $(t, x) \in \Omega_t$  and  $(a, p, X) \in P_{\Omega_t}^+ u(t, x)$  ( $\in P_{\Omega_t}^- u(t, x)$ ). The interest of this equivalent definition of sub (super) solution is that using of test functions  $\phi$  is not required anymore.

We can now state the Crandall-Ishii's Lemma.

**Lemma 2.3.2 (Crandall-Ishii's Lemma [79, 81])** *Let  $\Omega_i$  be locally compact subsets of  $R^{N_i}$  and let  $u_i: ]0, T] \times \Omega_i \rightarrow R$  be upper semi-continuous functions,  $i = 1 \dots k$ . Let  $\varphi$  be defined on an open neighbourhood of  $(0, T) \times \Omega_1 \times \dots \times \Omega_k$  and such that  $(t, x_1, \dots, x_k) \rightarrow \varphi(t, x_1, \dots, x_k)$  is once continuously differentiable and twice continuously differentiable in  $(x_1, \dots, x_k) \in \Omega_1 \times \dots \times \Omega_k$ . Suppose that  $\bar{t} \in (0, T)$ ,  $\bar{x}_i \in \Omega_i$ ,  $i = 1 \dots k$  and*

*$w(t, x_1, \dots, x_k) \equiv u(t, x_1) + \dots + u(t, x_k) - \varphi(t, x_1, \dots, x_k) \leq w(\bar{t}, \bar{x}_1, \dots, \bar{x}_k)$  for  $0 < t < T$  and  $x_i \in \Omega_i$ . Assume moreover that there is  $r > 0$  such that for  $M > 0$  there is a constant  $C$  such that for  $(b_i, q_i, X_i) \in P_{\Omega_i}^+ u_i(t, x_i)$ ,  $|x_i - \bar{x}_i| + |t - \bar{t}| \leq r$  and  $|u_i(t, x_i)| + |q_i| + |X_i| \leq M$  we have*

$$b_i \leq C, \quad i = 1 \dots K \quad (2.21)$$

*then for each  $\varepsilon > 0$  there exists  $X_i \in S^{N_i}$  such that*

$$(b_i, \nabla_{x_i} \varphi(\bar{t}, \bar{x}_1, \dots, \bar{x}_k), X_i) \in P_{\Omega_i}^+ u(\bar{t}, \bar{x}_i) \quad \text{for } i = 1 \dots k \quad (2.22)$$

$$-\left(\frac{1}{\varepsilon} + |A|\right)I \leq \begin{pmatrix} X_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & X_k \end{pmatrix} \leq A + \varepsilon A^2 \quad (2.23)$$

$$b_1 + b_2 + \dots + b_k = \frac{\partial \varphi}{\partial t}(\bar{t}, \bar{x}_1, \dots, \bar{x}_k) \quad (2.24)$$

where  $A = (\nabla_{x_i}^2 \varphi)(\bar{t}, \bar{x}_1, \dots, \bar{x}_k)$ .

This Lemma is technical and it is rather hard to understand its role to prove uniqueness. Let us describe it in a caricatural example to show where it is important.

**Example** Let  $\Omega$  be bounded, we assume that  $u$  and  $v$  are two continuous functions respectively sub and super-solution of

$$H(x, u, \nabla u, \nabla^2 u) = 0 \quad x \in \Omega, \quad (2.25)$$

and  $u \leq v$  on  $\partial\Omega$ . We want to show that  $u \leq v$  on  $\overline{\Omega}$ . This is a “maximum principle like result” and it gives uniqueness as soon as (2.25) is associated with Dirichlet conditions.

Let us first assume that  $u$  and  $v$  are in fact classical sub and super solutions of (2.25), that is to say twice differentiable. If the function  $w = u - v$  admits a local maximum  $\hat{x} \in \Omega$ , then we have:

$$\nabla u(\hat{x}) = \nabla v(\hat{x}) \quad (2.26)$$

$$\nabla^2 u(\hat{x}) \leq \nabla^2 v(\hat{x}) \quad (2.27)$$

and so:

$$H(\hat{x}, u(\hat{x}), \nabla u(\hat{x}), \nabla^2 u(\hat{x})) \leq 0 \leq H(\hat{x}, v(\hat{x}), \boxed{\nabla v(\hat{x})}, \boxed{\nabla^2 v(\hat{x})}) \quad (2.28)$$

$$\leq H(\hat{x}, v(\hat{x}), \boxed{\nabla u(\hat{x})}, \boxed{\nabla^2 u(\hat{x})}). \quad (2.29)$$

Remark that we have choosen as test functions  $u$  and  $v$  which was possible here because we assumed  $u$  and  $v \in C^2(\Omega)$ . Now, if we assume<sup>3</sup> that  $H(x, r, p, X)$  is strictly increasing with respect to the variable  $r$ , then  $u - v$  is negative at  $\hat{x} \in \Omega$  and then  $u \leq v$  on  $\overline{\Omega}$  (since we have assumed that  $u \leq v$  on  $\partial\Omega$ ).

In the generic case, one cannot choose anymore  $u$  and  $v$  as test functions and write (2.26)-(2.27). Let us highlight what needs to be adapted from the simple proof to the general case:

- Choosing  $u$  and  $v$  as test functions, and looking at the maximum point of  $u(x) - v(x)$  allowed us to get the equality of the derivative at the maximum point  $\hat{x}$  (2.26). This is used in (2.28)-(2.29) (see terms  $\boxed{\phantom{0}}$ ). In the general case, the classical idea is to duplicate the variables.
- In the same way, we used in (2.28)-(2.29) the comparison between the second derivatives (2.27) (see terms  $\boxed{\phantom{0}}$ ). This comparison will be given by the Crandall-Ishii’s Lemma, more precisely by (2.23).

See the detailed the proof of Theorem 4.3.2 in Section 4.3.3. ■

## 2.4 Elements of differential geometry: the curvature

In this section we recall some basic definitions and properties on differential geometry and focus on the notion of curvature. The curvature has different meaning depending on the “object” that we are considering. Typically, the

---

<sup>3</sup>Again, we recall that this example is just for educative purposes. In practice, such a strong assumption will not be necessary.

“objects” found in image analysis will be parametrized curves (for snakes for instance), images that can be represented by their isovalues, or images seen as surfaces where the height is the gray-scale intensity (see Figure 2.3). We precise the notion of curvatures in these three cases. Naturally, there are many other notions of curvature, especially for surfaces, and we refer to [161] for more details.

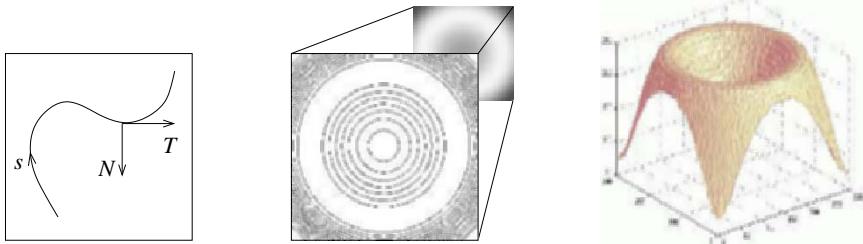


Figure 2.3. Different “objects” encountered in image analysis: parametrized curves, curves as isolevels of an image, image as a surface

#### 2.4.1 Parametrized curves

Let  $x(p) = (x_1(p), x_2(p))$  be a regular planar oriented curve in  $R^2$ ,  $0 \leq p \leq 1$ . We note:

$T(p) = x'(p) = (x'_1(p), x'_2(p))$  the tangent vector at  $x(p)$ ,

$N(p) = (-x'_2(p), x'_1(p))$  the normal vector at  $x(p)$ ,

$$s(p) = \int_0^p \sqrt{(x'_1(r))^2 + (x'_2(r))^2} dr \text{ the curvilinear abscissa (or arc length).}$$

If  $x(p)$  is regular we can parametrize it by  $s$  and then  $T(s) = \frac{dx}{ds}(s)$  is such that  $|T(s)| = 1$ . The curvature tensor is defined by  $\frac{dT}{ds}(s) = \frac{d^2x}{ds^2}(s)$ . One can show that the curvature tensor is collinear to  $\frac{N(s)}{|N(s)|}$  i.e.  $\frac{dT}{ds}(s) = \kappa(s) \frac{N(s)}{|N(s)|}$  where  $\kappa(s)$  is the curvature and  $\frac{1}{|\kappa(s)|}$  is the radius of curvature. For any parametrization we have :

$$\kappa(p) = \frac{x'_1(p) x''_2(p) - x'_2(p) x''_1(p)}{((x'_1(p))^2 + x'_2(p))^{\frac{3}{2}}} \quad (2.30)$$

and

$$\frac{1}{|x'(p)|} \frac{\partial}{\partial p} \left( \frac{x'(p)}{|x'(p)|} \right) = \kappa(p) \frac{N(p)}{|N(p)|}.$$

### 2.4.2 Curves as isolevel of a function $u$

Let us now consider the case where  $x(s)$ , parametrized by its curvilinear abscissa, is the  $k$ -level of a function  $u : R^2 \rightarrow R$  that is to say

$$x(s) = \{(x_1(s), x_2(s)) ; u((x_1(s), x_2(s))) = k\}.$$

By differentiating with respect to  $s$  the equality  $u((x_1(s), x_2(s))) = k$  we get:

$$x'_1(s) u_{x_1} + x'_2(s) u_{x_2} = 0, \quad (2.31)$$

where  $u_{x_i}$  stands for  $\frac{\partial u}{\partial x_i}(x_1(s), x_2(s))$ . Therefore the vectors  $(x'_1(s), x'_2(s))$  and  $(-u_{x_2}, u_{x_1})$  are collinear. For some  $\lambda$  we have

$$\begin{cases} x'_1(s) = -\lambda u_{x_2} \\ x'_2(s) = \lambda u_{x_1}, \end{cases} \quad (2.32)$$

so the vectors  $(u_{x_1}, u_{x_2})$  and  $(-u_{x_2}, u_{x_1})$  are respectively normal and tangent to the curve  $x(s)$ . If we differentiate again (2.31) with respect to  $s$ , we obtain:

$$(x'_1(s))^2 u_{x_1^2} + (x'_2(s))^2 u_{x_2^2} + 2 x'_1(s) x'_2(s) u_{x_1 x_2} + x''_1(s) u_{x_1} + x''_2(s) u_{x_2} = 0.$$

hence with (2.32):

$$\lambda^2 ((u_{x_1})^2 u_{x_2^2} + (u_{x_2})^2 u_{x_1^2} - 2 u_{x_1} u_{x_2} u_{x_1 x_2}) + \frac{1}{\lambda} (x''_1(s) x'_2(s) - x''_2(s) x'_1(s)) = 0.$$

But since  $|x'(s)| = 1$ , we get from (2.32)  $\lambda^2 = \frac{1}{|\nabla u|^2}$ . Therefore with (2.30) we deduce the expression of the curvature (of course, we suppose  $|\nabla u(x)| \neq 0$ ):

$$\kappa = \frac{(u_{x_1})^2 u_{x_2^2} + (u_{x_2})^2 u_{x_1^2} - 2 u_{x_1} u_{x_2} u_{x_1 x_2}}{((u_{x_1})^2 + (u_{x_2})^2)^{\frac{3}{2}}} \quad (2.33)$$

and we leave it as an exercise to the reader to verify that:

$$\kappa = \operatorname{div} \left( \frac{\nabla u}{|\nabla u|} \right). \quad (2.34)$$

### 2.4.3 Images as surfaces

Let us now examine quickly the case of 3-D surfaces. Denote by  $D$  an open set in  $R^2$ ,  $S : D \rightarrow R^3$ ,  $(u, v) \mapsto S(u, v)$  a regular parametrized surface. We assume that the vectors  $S_u$  and  $S_v$  are non-collinear for every  $(u, v) \in D$ , so they form a basis of the tangent plane.  $N(u, v) = \frac{S_u \wedge S_v}{|S_u \wedge S_v|}$  is the unit

normal vector to  $S(u, v)$ . Some classical notations are:

$$\left. \begin{array}{l} E(u, v) = |S_u|^2 \\ F(u, v) = S_u \cdot S_v \\ G(u, v) = |S_v|^2 \end{array} \right\} \begin{array}{l} \text{coefficients of the first} \\ \text{quadratic fundamental form} \\ \text{of } S(u, v) \end{array}$$

$$\left. \begin{array}{l} L(u, v) = S_{uu} \cdot N = -S_u \cdot N_u \\ M(u, v) = S_{uv} \cdot N = -\frac{1}{2}(S_u \cdot N_v + S_v \cdot N_u) \\ P(u, v) = S_{vv} \cdot N = S_v \cdot N_v \end{array} \right\} \begin{array}{l} \text{coefficients of the second} \\ \text{quadratic fundamental form} \\ \text{of } S(u, v) \end{array}$$

Using these notations, we have  $|S_u \wedge S_v| = \sqrt{EG - F^2}$ , the surface element is  $ds = \sqrt{EG - F^2} dudv$  and the mean curvature  $H$  can be rewritten as:

$$H = \frac{EP + GL - 2FM}{2(EG - F^2)}.$$

 Differential geometry [161]

## 2.5 Other classical results used in this book

We now summarize some classical theorems, propositions or inequalities. The purpose is not to give a complete list of important results but only to help the reader to find the one that are used in this book. For a complete and very clear overview, we refer the interested reader to the appendices of the book by Evans [105] from which we have selected the relevant results.

 Functional analysis [105]

### 2.5.1 Inequalities

**Cauchy's inequality with  $\varepsilon$ :**  $ab \leq \varepsilon a^2 + \frac{b^2}{4\varepsilon}$  ( $a, b > 0, \varepsilon > 0$ ).

**Cauchy-Schwarz inequality:**  $|x \cdot y| \leq |x||y|$  ( $x, y \in R^N$ ).

**Gronwall's inequality (differential form).**

- (i) Let  $\eta(\cdot)$  be a nonnegative, absolutely continuous function on  $[0, T]$ , which satisfies for a.e.  $t$  the differential inequality:

$$\eta'(t) \leq \phi(t)\eta(t) + \Psi(t),$$

where  $\phi(t)$  and  $\Psi(t)$  are nonnegative, integrable functions on  $[0, T]$ . Then:

$$\eta(t) \leq e^{\int_0^t \phi(s)ds} \left[ \eta(0) + \int_0^t \Psi(s)ds \right] \quad \text{for all } 0 \leq t \leq T.$$

- (ii) In particular, if  $\eta' \leq \phi\eta$  on  $[0, T]$  and  $\eta(0) = 0$ , then  
 $\eta \equiv 0$  on  $[0, T]$ .

**Gronwall's inequality (integral form).**

- (i) Let  $\xi(t)$  be a nonnegative, integrable function on  $[0, T]$  which satisfies for a.e.  $t$  the integral inequality:

$$\xi(t) \leq C_1 \int_0^t \xi(s) ds + C_2$$

for some constants  $C_1, C_2 \geq 0$ , then:

$$\xi(t) \leq C_2 (1 + C_1 t e^{C_1 t}) \quad \text{for a.e. } 0 \leq t \leq T.$$

- (ii) In particular, if  $\xi(t) \leq C_1 \int_0^t \xi(s) ds$  for a.e.  $0 \leq t \leq T$ , then:

$$\xi(t) = 0 \text{ a.e.}$$

**Hölder's inequality:** assume  $1 \leq p, q \leq \infty$ ,  $\frac{1}{p} + \frac{1}{q} = 1$ . Then, if  $u \in L^p(\Omega)$ ,  $v \in L^q(\Omega)$ , we have:

$$\int_{\Omega} |uv| dx \leq \|u\|_{L^p(\Omega)} \|v\|_{L^q(\Omega)}.$$

**Jensen's inequality:** assume that  $f : R \rightarrow R$  is convex, and  $\Omega \subset R^N$  is open, bounded ( $|\Omega| = 1$ ). Let  $u : \Omega \rightarrow R$  be integrable. Then:

$$f \left( \int_{\Omega} u dx \right) \leq \left( \int_{\Omega} f(u) dx \right),$$

where  $\int_{\Omega} u dx$  denotes the mean value of  $u$  over  $\Omega$ .

**Minkowski's inequality:** assume  $1 \leq p \leq \infty$  and  $u, v \in L^p(\Omega)$ . Then:

$$|u + v|_{L^p(\Omega)} \leq \|u\|_{L^p(\Omega)} + \|v\|_{L^p(\Omega)}.$$

**Poincaré inequality:** let  $\Omega$  be an open bounded set of  $R^N$  and  $u \in W_0^{1,p}(\Omega) = \{u / u \in W^{1,p}(\Omega); u|_{\partial\Omega} = 0\}$ ,  $1 \leq p < n$ , then:

$$\|u\|_{L^q(\Omega)} \leq c_1 \|\nabla u\|_{L^p(\Omega)}, \quad q \in \left[ 1, \frac{np}{n-p} \right]$$

for some constant  $c_1$  depending only on  $p, n, q$  and  $\Omega$ .

**Poincaré-Wirtinger inequality:** let  $\Omega$  be open, bounded and connected

with a  $C^1$  boundary, then for all  $u \in W^{1,p}(\Omega)$ ,  $1 \leq p \leq +\infty$ ,

$$\left| u - \int_{\Omega} u \, dx \right|_{L^p(\Omega)} \leq c_2 |\nabla u|_{L^p(\Omega)}$$

for some constant  $c_2$  depending only on  $p$ ,  $n$  and  $\Omega$ . Remark that the same inequality holds for functions of bounded variation, where  $|\nabla u|_{L^p(\Omega)}$  is replaced by the total variation  $|Du|(\Omega)$ .

**Young's inequality:** let  $1 < p, q < \infty$ ,  $\frac{1}{p} + \frac{1}{q} = 1$ , then:

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q} \quad (a, b > 0).$$

**Young's inequality with  $\varepsilon$ :** let  $1 < p, q < \infty$ ,  $\frac{1}{p} + \frac{1}{q} = 1$ , then:

$$ab \leq \varepsilon a^p + C(\varepsilon) b^q \quad \text{for } C(\varepsilon) = (\varepsilon p)^{-q/p} q^{-1}.$$

### 2.5.2 Calculus facts

$\Omega$  is supposed to be a bounded, open subset of  $R^N$ , and  $\partial\Omega$  is supposed to be  $C^1$ .

**Theorem 2.5.1 (Gauss-Green Theorem)** Suppose  $u \in C^1(\bar{\Omega})$ . Then

$$\int_{\Omega} u_{x_i} \, dx = \int_{\partial\Omega} u \nu^i \, ds \quad (i = 1, \dots, N).$$

where  $\nu$  is the outward unit normal of  $\partial\Omega$ .

**Theorem 2.5.2 (Integration by part formula)** Let  $u, v \in C^1(\bar{\Omega})$ . Then

$$\int_{\Omega} u_{x_i} v \, dx = - \int_{\Omega} u v_{x_i} \, dx + \int_{\partial\Omega} u v \nu^i \, ds \quad (i = 1, \dots, N).$$

**Theorem 2.5.3 (Green's formulas)** Let  $u, v \in C^2(\Omega)$ . Then:

$$(i) \int_{\Omega} \Delta u \, dx = \int_{\partial\Omega} \frac{\partial u}{\partial \nu} \, ds.$$

$$(ii) \int_{\Omega} \nabla v \cdot \nabla u \, dx = - \int_{\Omega} u \Delta v \, dx + \int_{\partial\Omega} \frac{\partial v}{\partial \nu} u \, ds.$$

$$(iii) \int_{\Omega} u \Delta v - v \Delta u \, dx = \int_{\partial\Omega} u \frac{\partial v}{\partial \nu} - v \frac{\partial u}{\partial \nu} \, ds.$$

**Theorem 2.5.4 (Coarea formula)** Let  $u : R^N \rightarrow R$  be Lipschitz continuous and assume that for a.e.  $r \in R$ , the level set

$$\{x \in R^N \mid u(x) = r\}$$

is a smooth,  $(n - 1)$ -dimensional hypersurface in  $R^N$ . Suppose also  $f : R^N \rightarrow R$  is continuous and integrable. Then:

$$\int_{R^N} f |\nabla u| dx = \int_{-\infty}^{+\infty} \left( \int_{\{u=r\}} f ds \right) dr.$$

### 2.5.3 About convolution and smoothing

For  $\Omega \subset R^N$ , we denote  $\Omega_\varepsilon = \{x \in \Omega \mid \text{dist}(x, \partial\Omega) > \varepsilon\}$ .

**Definition 2.5.1 (mollifier)**

(i) Define  $\eta \in \mathcal{C}_c^\infty(R^N)$  by:

$$\eta(x) = \begin{cases} C \exp\left(\frac{1}{|x|^2 - 1}\right) & \text{if } |x| < 1 \\ 0 & \text{if } |x| \geq 1, \end{cases}$$

the constant  $C > 0$  is selected so that  $\int_{R^N} \eta dx = 1$ .  $\eta$  is called a standard mollifier.

(ii) For each  $\varepsilon > 0$ , set:

$$\eta_\varepsilon(x) = \frac{1}{\varepsilon^N} \eta\left(\frac{x}{\varepsilon}\right).$$

The functions  $\eta_\varepsilon$  are  $\mathcal{C}_c^\infty(R^N)$  and satisfy:

$$\int_{R^N} \eta_\varepsilon dx = 1, \quad \text{spt}(\eta_\varepsilon) \subset B(0, \varepsilon).$$

**Definition 2.5.2 (mollification)** If  $f : \Omega \rightarrow R$  is locally integrable, define its mollification by:

$$f^\varepsilon = \eta_\varepsilon * f \quad \text{in } \Omega_\varepsilon$$

that is:

$$f^\varepsilon(x) = \int_{\Omega} \eta_\varepsilon(x - y) f(y) dy = \int_{B(0, \varepsilon)} \eta_\varepsilon(y) f(x - y) dy \quad \text{for } x \in \Omega_\varepsilon.$$

**Theorem 2.5.5 (properties of mollifiers)**

(i)  $f^\varepsilon \in \mathcal{C}_c^\infty(\Omega_\varepsilon)$ .

(ii)  $f^\varepsilon \rightarrow f$  a.e. as  $\varepsilon \rightarrow 0$ .

- (iii) If  $f \in \mathcal{C}(\Omega)$ , then  $f^\varepsilon \rightarrow f$  uniformly on compact subsets of  $\Omega$ .
- (iv) If  $1 \leq p < \infty$  and  $f \in L_{\text{loc}}^p(\Omega)$ , then  $f^\varepsilon \rightarrow f$  in  $L_{\text{loc}}^p(\Omega)$ .

### 2.5.4 Uniform convergence

**Theorem 2.5.6 (Arzelà-Ascoli compactness criterion)** Suppose that  $\{f_k\}_{k=1}^\infty$  is a sequence of real-valued functions defined on  $R^N$  such that:

$$|f_k(x)| \leq M \quad (k = 1, \dots, x \in R^N)$$

for some constant  $M$ , and the  $\{f_k\}_{k=1}^\infty$  are uniformly equicontinuous. Then there exist a subsequence  $\{f_{k_j}\}_{j=1}^\infty \subseteq \{f_k\}_{k=1}^\infty$  and a continuous function  $f$ , such that:

$$f_{k_j} \rightarrow f \text{ uniformly on compact subsets of } R^N.$$

We recall that saying that  $\{f_k\}_{k=1}^\infty$  are uniformly equicontinuous means that for each  $\varepsilon > 0$ , there exists  $\delta > 0$ , such that  $|x - y| < \delta$  implies  $|f_k(x) - f_k(y)| < \varepsilon$ , for  $x, y \in R^N$ ,  $k = 1, \dots$

### 2.5.5 Dominated convergence theorem

**Theorem 2.5.7 (dominated convergence theorem)** Assume the functions  $\{f_k\}_{k=1}^\infty$  are Lebesgue integrable and:

$$f_k \rightarrow f \text{ a.e.}$$

Suppose also:

$$|f_k| \leq g \text{ a.e.,}$$

for some integrable function  $g$ . Then:

$$\int_{R^N} f_k \, dx \rightarrow \int_{R^N} f \, dx.$$

This theorem which is fundamental in the Lebesgue theory of integration will be very often used in this book.

### 2.5.6 Well-posed problem

Finally, recall the classical definition concerning the well-posedness of a minimization problem or a PDE.

**Definition 2.5.3 (well-posed)** When a minimization problem or a PDE admit a unique solution which depends continuously on the data, we say that the minimization problem or the PDE are well-posed in the sense of Hadamard.

If one of the following conditions: existence, uniqueness or continuity fails, we say that the minimization problem or the PDE are ill-posed.

# 3

## Image Restoration

### How to read this chapter?

Image restoration is historically one of the oldest concerns and is still a necessary pre-processing for many applications. So we start with a precise study of this problem which will give to the reader a broad overview of the variational and PDE based approaches as applied to image analysis.

- We first give in Section 3.1 some precisions about what we mean by degradation or noise. This is actually a difficult question and we focus on a simple model with additive noise and convolution by a linear operator for the blur.
- Section 3.2 presents the restoration through the minimization of a functional involving two terms: a fidelity term to the data (based on the model of noise) plus a regularization term. We discuss in Section 3.2.2 some qualitative properties we would like for the restored image. This leads to a certain functional that we study in details in Section 3.2.3 (existence and uniqueness of a solution). This subsection is rather mathematical and shows an example of relaxation in the BV-framework. Section 3.2.4 concerns the the numerical computation of the solution found previously. We develop an algorithm called half-quadratic minimization and present some experimental results. We finally mention in Section 3.2.5 some scale invariance properties and conclude in Section 3.2.6 by considering the nonconvex case.

- Section 3.3 is a survey of some PDEs based models proposed in literature over the last decade, for restoration and enhancement. Three types are distinguished:
  - Smoothing or parabolic PDEs (Section 3.3.1) mainly used in pure restoration.
  - Smoothing-enhancing or parabolic-hyperbolic PDEs (Section 3.3.2) concerning restoration-enhancement processes.
  - Hyperbolic PDEs (Section 3.3.3) for enhancing blurred images, focusing on shock-filters.

For each case we develop in details how to get the model and its mathematical justification (when it is possible and instructive). The mathematical background involves the theory of maximal operators, the notion of viscosity solutions (Section 3.3.1) and fixed point techniques (Section 3.3.2). As for Section 3.3.3 no rigorous results are available. We only mention a conjecture by Osher and Rudin regarding the existence of a weak solution for shock-filters.

### 3.1 Image degradation

It is well-known that during formation, transmission or recording processes, images are deteriorated. Classically this degradation is the result of two phenomena. The first one is deterministic and is related to the mode of image acquisition, to possible defects of the imaging system (blur created by a wrong lens adjustment, by motion...) or other phenomena such as atmospheric turbulences. The second phenomenon is a random one and corresponds to the *noise* coming from any signal transmission. When it is possible, it is important to choose a degradation model, as close as possible to the reality. Each model is usually characterized by a probabilistic distribution. In many cases, a Gaussian distribution is assumed. However some applications require more specific ones like Gamma distribution for radar images (speckle noise), the Poisson distribution for tomography, etc. We show in Figure (3.2) few examples of possible degradations.

Our aim in this chapter is to explain the methods which allow to remove or diminish the effects of these degradations. To fix the terminology, we will designate this processing as *restoration*.

Unfortunately, it is usually impossible for a given real image to identify the kind of noise involved. If no model of degradation is available (for instance, we may know the defects of the satellite used for the acquisition), some assumptions have to be made. A commonly used model is the following. Let  $u : \Omega \subset R^2 \rightarrow R$  an original image describing a real scene, and let  $u_0$  be the observed image of the same scene (i.e. a degradation of  $u$ ). We



Figure 3.1. “Borel building” image (building from INRIA Sophia-Antipolis)

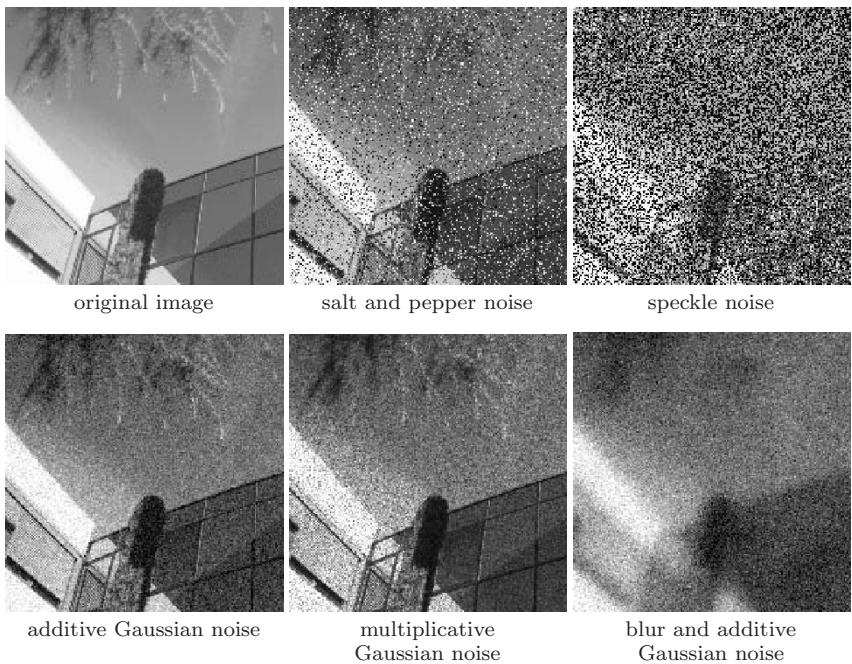


Figure 3.2. Examples of degradations on the top left-hand side corner of the “Borel building” image

assume that:

$$u_0 = R u + \eta, \quad (3.1)$$

where  $\eta$  stands for a white additive Gaussian noise and where  $R$  is a linear operator representing the blur (usually a convolution). Given  $u_0$ , the problem is then to reconstruct  $u$  knowing (3.1). As we will see, the problem is *ill-posed* and we are only able to carry out an approximation of  $u$ .

## 3.2 The energy method

### 3.2.1 An inverse problem

Let  $u$  be the original image describing a real scene (the unknown), and let  $u_0$  be the observed image (the data). Let us assume valid the model of degradation (3.1). Recovering  $u$  from  $u_0$  knowing (3.1) is a typical example of inverse problem.

#### Inverse problems [147]

This is not an easy task since we know little things about the noise  $\eta$ . We only know some statistics as its mean, its variance... Of course, since  $\eta$  is a random variable, a natural way to interpret equation (3.1) is to use probabilities (see for instance [46, 73, 90, 116]). It is not our goal to develop such a theory. Let us only mention that by supposing that  $\eta$  is a white Gaussian noise, and according to the maximum likelihood principle, we can find an approximation of  $u$  by solving the least square problem:

$$\inf_u \int_{\Omega} |u_0 - Ru|^2 \, dx \quad (3.2)$$

where  $\Omega$  is the domain of the image. To fix ideas, let us imagine for a moment that  $u_0$  and  $u$  are discrete variables in  $R^M$ , that  $R$  is a  $M \times M$  matrix and that  $|\cdot|$  stands for the Euclidian norm. If a minimum  $u$  of (3.2) exists then it necessarily verifies the following equation:

$$R^* u_0 - R^* R u = 0 \quad (3.3)$$

where  $R^*$  is the adjoint of  $R$ . Solving (3.3) is in general an ill-posed problem.  $R^* R$  is not always one to one and even if  $R^* R$  were one to one, its eigenvalues may be small causing numerical instabilities. Therefore the idea is to regularize the problem (3.2) that is considering a connected problem which admits a unique solution.

 From now on, we suppose that  $u_0 \in L^\infty(\Omega)$  ( $\Omega \subset R^2$ , bounded) and that  $R$  is a linear operator of  $L^2(\Omega)$ . We do not use any probabilistic argument.

The noise  $\eta$  is regarded as a perturbation causing spurious oscillations in the image. One of the goals of the restoration is to remove these oscillations while preserving salient features as edges.

### 3.2.2 Regularization of the problem

A classical way to overcome ill-posed minimization problems is to add a regularization term to the energy. This idea was introduced in 1977 by Tikhonov and Arsenin [239]. The authors proposed to consider the minimization problem:

$$F(u) = \int_{\Omega} |u_0 - Ru|^2 dx + \lambda \int_{\Omega} |\nabla u|^2 dx. \quad (3.4)$$

The first term in  $F(u)$  measures the fidelity to the data. The second one is a smoothing term. In other words we search for  $u$  fitting the best to the data so that its gradient is low (so that noise will be removed). The parameter  $\lambda$  is a positive weighting constant.

To study this problem, the functional space for which both terms are well-defined is:

$$W^{1,2}(\Omega) = \{u \in L^2(\Omega); \nabla u \in L^2(\Omega)^2\}.$$

Under suitable assumptions on  $R$  (we will come back later on these assumptions) the problem  $\inf \{F(u), u \in W^{1,2}(\Omega)\}$  admits a unique solution characterized by the Euler-Lagrange equation

$$R^*Ru - R^*u_0 - \lambda \Delta u = 0 \quad (3.5)$$

with the Neumann boundary condition

$$\frac{\partial u}{\partial N} = 0 \text{ on } \partial\Omega \text{ (}N\text{ is the outward normal exterior to } \partial\Omega\text{)}.$$

Is the solution  $u$  of (3.5) a good candidate to our original restoration problem? The answer is no since it is well known that the Laplacian operator has very strong isotropic smoothing properties and does not preserve edges (see Figure 3.3). Equivalently, this over smoothing can be explained by looking at the energy (3.4). The  $L^p$  norm with  $p = 2$  of the gradient allows to remove the noise but unfortunately penalizes too much the gradients corresponding to edges. One should then decrease  $p$  in order to keep as much as possible the edges. One of the first work in this direction were done by Rudin, Osher and Fatemi [215, 214] who proposed to use the  $L^1$  norm of the gradient of  $u$  in (3.5), also called the total variation, instead of the  $L^2$  norm.



Figure 3.3. Restoration of the noisy “Borel building” image (additive Gaussian noise) by minimizing (3.4); edges are lost (in this case  $R = \text{Id}$ )

In order to study more precisely the influence of the smoothing term, let us consider the following energy [20, 245]:

$$E(u) = \frac{1}{2} \int_{\Omega} |u_0 - Ru|^2 dx + \lambda \int_{\Omega} \phi(|\nabla u|) dx. \quad (3.6)$$

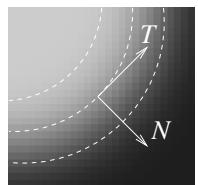
- We need to find the properties on  $\phi$  so that the solution of the minimization problem is close to a piecewise constant image, that is formed by homogeneous regions separated by sharp edges.

Let us suppose that  $E(u)$  has a minimum point  $u$ . Then it formally verifies the Euler-Lagrange equation:

$$R^* R u - \lambda \operatorname{div} \left( \frac{\phi'(|\nabla u|)}{|\nabla u|} \nabla u \right) = R^* u_0. \quad (3.7)$$

Equation (3.7) can be written in an expanded form by developing formally the divergence term.

We are going to show that it can be decomposed using the local image structures, that is the tangent and normal directions to the isophote lines (lines along which the intensity keeps constant). More precisely, for each point  $x$  where  $|\nabla u(x)| \neq 0$  we can define the vectors  $N(x) = \frac{\nabla u(x)}{|\nabla u(x)|}$  and  $T(x)$ ,  $|T(x)| = 1$ ,  $T(x)$  orthogonal to  $N(x)$ . With the usual notations  $u_{x_1}, u_{x_2}, u_{x_1 x_1 \dots}$  for the first and second partial derivatives of  $u$ , we can



rewrite (3.7) as

$$R^*Ru - \lambda \left( \frac{\phi'(|\nabla u|)}{|\nabla u|} u_{TT} - \phi''(|\nabla u|) u_{NN} \right) = R^*u_0 \quad (3.8)$$

where we denote by  $u_{TT}$  and  $u_{NN}$  the second derivatives of  $u$  in the  $T$ -direction and  $N$ -direction respectively:

$$\begin{aligned} u_{TT} &= {}^t T \nabla^2 u T = \frac{1}{|\nabla u|^2} (u_{x_1}^2 u_{x_2 x_2} + u_{x_2}^2 u_{x_1 x_1} - 2u_{x_1} u_{x_2} u_{x_1 x_2}) \\ u_{NN} &= {}^t N \nabla^2 u N = \frac{1}{|\nabla u|^2} (u_{x_1}^2 u_{x_1 x_1} + u_{x_2}^2 u_{x_2 x_2} + 2u_{x_1} u_{x_2} u_{x_1 x_2}). \end{aligned}$$

In fact, decomposing the divergence term as a weighted sum of the two directional derivatives along  $T$  and  $N$  can be done for most classical diffusion operators [153]. This allows to see clearly the action of the operators in directions  $T$  and  $N$ .

In our case, this is also useful to determine how the function  $\phi$  should be chosen:

- At locations where the variations of the intensity are weak (low gradients), we would like to encourage smoothing, the same way in any directions. Assuming that the function  $\phi$  is regular, this isotropic smoothing condition may be achieved by imposing:

$$\phi'(0) = 0, \lim_{s \rightarrow 0^+} \frac{\phi'(s)}{s} = \lim_{s \rightarrow 0^+} \phi''(s) = \phi''(0) > 0. \quad (3.9)$$

Therefore at points where  $|\nabla u|$  is small, (3.8) becomes:

$$R^*Ru - \lambda \phi''(0)(u_{TT} + u_{NN}) = R^*u_0$$

i.e. since  $u_{TT} + u_{NN} = \Delta u$ :

$$R^*Ru - \lambda \phi''(0)\Delta u = R^*u_0. \quad (3.10)$$

So, at these points,  $u$  locally satisfies (3.10), which is a uniformly elliptic equation having strong regularizing properties in all directions.

- In a neighborhood of an edge  $C$ , the image presents a strong gradient. If we want to preserve this edge, it is preferable to diffuse along  $C$  (in the  $T$ -direction) and not across it. To do this, it is sufficient in (3.7) to annihilate, for strong gradients, the coefficient of  $u_{NN}$  and to assume that the coefficient of  $u_{TT}$  does not vanish:

$$\lim_{s \rightarrow +\infty} \phi''(s) = 0, \lim_{s \rightarrow +\infty} \frac{\phi'(s)}{s} = \beta > 0. \quad (3.11)$$

Unfortunately, these two conditions are not compatible. One must find a compromise. For example,  $\phi''(s)$  and  $\frac{\phi'(s)}{s}$  both converge to

zero as  $s \rightarrow +\infty$ , but with different speed:

$$\lim_{s \rightarrow +\infty} \phi''(s) = \lim_{s \rightarrow +\infty} \frac{\phi'(s)}{s} = 0 \quad \text{and} \quad \lim_{s \rightarrow +\infty} \frac{\phi''(s)}{\frac{\phi'(s)}{s}} = 0. \quad (3.12)$$

Notice that many functions  $\phi$  verifying the conditions (3.9)-(3.12) can be found. For example the function

$$\phi(s) = \sqrt{1+s^2}, \quad (3.13)$$

which is usually called the *hypersurface minimal function*.

**Remark** The assumptions (3.9) and (3.12) on  $\phi$  are qualitative. They have been imposed in order to describe the regularization conditions. Naturally, they are not sufficient to ensure that the model is mathematically well-posed. Other hypotheses such as convexity, linear growth are necessary. This is developed in the coming section. ■

### 3.2.3 Existence and uniqueness of a solution for the minimization problem

This section is devoted to the mathematical study of:

$$\inf \left\{ E(u) = \frac{1}{2} \int_{\Omega} |u_0 - Ru|^2 dx + \lambda \int_{\Omega} \phi(|\nabla u|) dx \right\}. \quad (3.14)$$

In order to use the direct method of the calculus of variations, we have to assume some minimal hypotheses on  $\phi$ :

$$\begin{aligned} \phi &\text{ is a strictly convex, nondecreasing function from } R^+ \\ &\text{ to } R^+, \text{ with } \phi(0) = 0 \text{ (without a loss of generality)} \end{aligned} \quad (3.15)$$

$$\lim_{s \rightarrow +\infty} \phi(s) = +\infty. \quad (3.16)$$

This latter growth condition must not be too strong because it must not penalize strong gradients, i.e. the formation of edges (see what happened with  $\phi(s) = s^2$ ). Hence we assume that  $\phi$  grows at most linearly:

$$\begin{aligned} &\text{There exist two constants } c > 0 \text{ and } b \geq 0 \text{ such that} \\ &cs - b \leq \phi(s) \leq cs + b \quad \forall s \geq 0. \end{aligned} \quad (3.17)$$

**Remark** To recover and preserve edges in an image, it would certainly be preferable to impose a growth condition of the type:  $\lim_{s \rightarrow +\infty} \phi(s) = \beta > 0$ . In this case the contribution of the term  $\phi(|\nabla u|)$  in  $E(u)$  would not penalize the formation of strong gradients since "it would cost nothing". Unfortunately, as we also want that  $\phi$  has a quadratic behavior near zero, then necessarily  $\phi$  should have a nonconvex shape, which is an undesirable property (see for instance the Bolza problem discussed in Sections 2.1.2 and

2.1.3). ■

According to (3.17), the natural space on which we would be able to seek a solution, is the space

$$V = \{u \in L^2(\Omega); \nabla u \in L^1(\Omega)^2\}.$$

Unfortunately, this space is not reflexive. In particular, we cannot say anything about minimizing sequences that are bounded in  $V$ . However, an interesting remark is that sequences bounded in  $V$  are also bounded in  $BV(\Omega)$ . Therefore, they are compact for the  $BV - w^*$  topology. Still, the energy  $E$  is not lower semi-continuous for this topology... In this case, it is classical to compute the relaxed energy.

**Theorem 3.2.1** *The relaxed functional of (3.14) for the  $BV - w^*$  topology is defined by:*

$$\begin{aligned} \overline{E}(u) = & \frac{1}{2} \int_{\Omega} |u_0 - Ru|^2 dx + \\ & + \lambda \int_{\Omega} \phi(|\nabla u|) dx + \lambda c \int_{S_u} (u^+ - u^-) d\mathcal{H}^1 + \lambda c \int_{\Omega - S_u} |Cu| \end{aligned} \quad (3.18)$$

where  $c = \lim_{s \rightarrow +\infty} \frac{\phi(s)}{s}$ .

**Proof** Let us define:

$$e(u) = \begin{cases} \frac{1}{2} \int_{\Omega} |u_0 - Ru|^2 dx + \lambda \int_{\Omega} \phi(|\nabla u|) dx & \text{if } u \in V \\ +\infty & \text{if } u \in BV(\Omega) - V. \end{cases}$$

We note that  $e(u) = E(u)$  if  $u \in V$ . Since  $e(u)$  is not l.s.c. for the  $BV - w^*$  topology we need to compute its l.s.c. envelope (for the  $BV - w^*$  topology), i.e. the greatest l.s.c. functional  $\bar{e}(u)$  less than or equal to  $e(u)$ . Since  $\overline{E}(u)$  is l.s.c. (see Section 2.2.3), we have  $\bar{e}(u) \geq \overline{E}(u)$ . Thus, we have to show that  $\bar{e}(u) \leq \overline{E}(u)$ .

Thanks to [89], for each  $u \in BV(\Omega)$  there exists a sequence  $u_n \in C^\infty(\Omega) \cap V$  such that  $u_n \xrightarrow[BV-w^*]{} u$  and  $\overline{E}(u) = \lim e(u_n)$ . Therefore:

$$\overline{E}(u) = \lim e(u_n) \geq \inf_{\substack{u_n \in BV(\Omega) \\ u_n \xrightarrow[BV-w^*]{} u}} \{ \underline{\lim} e(u_n) \} = \bar{e}(u)$$

which concludes the proof. ■

In the sequel, we also assume that:

$$R : L^2(\Omega) \rightarrow L^2(\Omega) \text{ is a linear continuous operator, and } R.1 \neq 0 \quad (3.19)$$

The second assumption of (3.19) means that  $R$  does not annihilate the constants which guarantees the  $BV$ -coercivity of  $\bar{E}(u)$ .

**Theorem 3.2.2** *Under assumptions (3.15)-(3.17) and (3.19), the minimization problem:*

$$\inf_{u \in BV(\Omega)} \bar{E}(u) \quad (3.20)$$

where  $\bar{E}$  is defined by (3.18), admits a unique solution  $u \in BV(\Omega)$ .

**Proof** The proof follows [245, 246].

Step 1: Existence

Let  $u_n$  a minimizing sequence for (3.20). Thanks to (3.16), we have:

$$\left\{ \begin{array}{l} |Du_n|(\Omega) = \int_{\Omega} |\nabla u_n| dx + \int_{S_{u_n}} |u_n^+ - u_n^-| d\mathcal{H}^1 + \int_{\Omega - S_{u_n}} |Cu_n| \leq M \\ \int_{\Omega} |Ru_n - u_0|^2 dx \leq M \end{array} \right.$$

where  $M$  denotes an universal strictly positive constant which may differ from line to line. The first above inequality says that the total variation of  $Du_n$  is uniformly bounded. It remains to prove that  $|u_n|_{L^1(\Omega)}$  is bounded.

Let  $w_n = \frac{1}{|\Omega|} \int_{\Omega} u_n dx$  and  $v_n = u_n - w_n$ . Then  $\int_{\Omega} v_n dx = 0$  and  $Dv_n = Du_n$ .

Hence  $|Dv_n| \leq M$ . Using the generalized Poincaré-Wirtinger inequality , we get:

$$|v_n|_{L^2(\Omega)} \leq K |Dv_n|(\Omega) \leq M \text{ where } K \text{ is a constant} \quad (3.21)$$

Now, from the inequality  $\int_{\Omega} |Ru_n - u_0|^2 dx \leq M$ , we deduce:

$$\begin{aligned} |Rw_n|_{L^2(\Omega)} & \left[ |Rw_n|_{L^2(\Omega)} - 2 \left( |R| |v_n|_{L^2(\Omega)} + |u_0|_{L^2(\Omega)} \right) \right] \leq \\ & \leq |Rw_n|_{L^2(\Omega)} \left[ |Rw_n|_{L^2(\Omega)} - 2 |Ru_n - u_0|_{L^2(\Omega)} \right] \leq \\ & \leq \left[ |Ru_n - u_0|_{L^2(\Omega)} - |Rw_n|_{L^2(\Omega)} \right] \leq \\ & \leq |Rv_n + Rw_n - u_0|_{L^2(\Omega)}^2 = \\ & = |Ru_n - u_0|_{L^2(\Omega)}^2 \leq M \end{aligned}$$

where  $|R|$  denotes the norm of the operator  $R$ . Let  $x_n = |Rw_n|_{L^2(\Omega)}$  and  $a_n = |R| |v_n|_{L^2(\Omega)} + |u_0|_{L^2}$ , then the above inequality writes as:

$$x_n(x_n - 2a_n) \leq M$$

with  $0 \leq a_n \leq |R|M + |u_0|_{L^2(\Omega)} \leq M$ . Therefore, we obtain:

$$0 \leq x_n \leq a_n + \sqrt{a_n^2 + M} \leq M$$

i.e.

$$|Ru_n|_{L^2(\Omega)} = \left| \frac{1}{|\Omega|} \int_{\Omega} u_n dx \right| |R.1|_{L^2(\Omega)} \leq M$$

and thanks to (3.19), we obtain that  $\left| \int_{\Omega} u_n dx \right|$  is uniformly bounded.

Applying again the Poincaré-Wirtinger inequality, it results from (3.21):

$$|u_n|_{L^2(\Omega)} = \left| v_n + \frac{1}{|\Omega|} \int_{\Omega} u_n dx \right|_{L^2(\Omega)} \leq |v_n|_{L^2(\Omega)} + \left| \frac{1}{|\Omega|} \int_{\Omega} u_n dx \right| \leq M.$$

Hence  $u_n$  is bounded in  $L^2(\Omega)$  and in  $L^1(\Omega)$  ( $\Omega$  is bounded). Since  $|Du_n|(\Omega)$  is also bounded, we get that  $u_n$  is bounded in  $BV(\Omega)$ .

Thus, up to a subsequence, there exists  $u$  in  $BV(\Omega)$  so that  $u_n \xrightarrow[BV-w^*]{} u$  and  $Ru_n \xrightarrow[L^2(\Omega)]{} Ru$ .

Finally, from the weak semi-continuity property of the convex function of measures and the weak semi-continuity of the  $L^2$ -norm, we get

$$\begin{aligned} \int_{\Omega} |Ru - u_0|^2 dx &\leq \liminf_{n \rightarrow +\infty} \int_{\Omega} |Ru_n - u_0|^2 dx \\ \int_{\Omega} \phi(Du) &\leq \liminf_{n \rightarrow +\infty} \int_{\Omega} \phi(Du_n) \end{aligned}$$

that is to say:

$$\bar{E}(u) \leq \liminf_{n \rightarrow +\infty} \bar{E}(u_n) = \inf_{v \in BV(\Omega)} \bar{E}(v)$$

i.e.  $u$  is minimum point of  $\bar{E}(u)$ .

### Step 2: Uniqueness

Let  $u$  and  $v$  be two minima of  $\bar{E}(u)$ . From the strict convexity of  $\phi$  we easily get that  $Du = Dv$ , which implies that  $u = v + c$ . But as the function  $u \rightarrow \int_{\Omega} |Ru - u_0|^2 dx$  is also strictly convex, we deduce that  $Ru = Rv$ , therefore  $R.c = 0$  and from (3.19) we conclude that  $c = 0$  and  $u = v$ . ■

#### 3.2.4 Toward the numerical approximation

The next question is to characterize the solution of the problem (3.20) in order to get a numerical approximation. If we try to write directly the

Euler-Lagrange equations, the difficulty is to define variations on  $BV(\Omega)$  because of the presence of measures. Some interesting results can still be obtained. For instance, it can be proved that the solution of (3.20) verify:

$$(i) \quad R^*Ru - R^*u_0 - \lambda \operatorname{div} \left( \frac{\phi'(|\nabla u|)}{|\nabla u|} \nabla u \right) = 0 \text{ in } L^2(\Omega).$$

$$(ii) \quad \frac{\phi'(|\nabla u|)}{|\nabla u|} \frac{\partial u}{\partial N} = 0 \text{ on } \partial\Omega.$$

(iii) For all  $w \in BV(\Omega)$  with  $Dw = \nabla w dx + D_s w$  and  $D_s w = \rho D_s u + \mu'$ :

$$\begin{aligned} \int_{\Omega} (Ru - u_0)Rw \, dx + \int_{\Omega} \frac{\phi'(|\nabla u|)}{|\nabla u|} \nabla u \cdot \nabla w \, dx + \int_{\Omega} \rho |D_s u| + \int_{\Omega} |\mu'| \geq \\ \geq - \int_{\Omega} \operatorname{div} \left( \frac{\phi'(|\nabla u|)}{|\nabla u|} \nabla u \right) w \, dx \end{aligned}$$

where  $\nabla u$  is the approximate derivative of  $u$ . This is fully detailed in Proposition 3.3.4 (Section 3.3.1) which involves similar operators. So, it is mathematically uncorrect to use only (i) and (ii) to find a numerical solution, which is usually done. However the condition (iii) remains difficult to handle...

To circumvent this difficulty, we can consider a “close” energy for which the Euler-Lagrange equation will be easier to implement. This can be done through  $\Gamma$ -convergence. The general idea is the following (see also Figure 3.4):

- We construct a sequence of energy  $E_\varepsilon$  so that for each  $\varepsilon > 0$ , the associated minimization problem admits a unique minimum  $u_\varepsilon$  in the Sobolev space  $W^{1,2}(\Omega)$ . Then, we prove, via the  $\Gamma$ -convergence theory, that  $u_\varepsilon$  converges in  $L^1(\Omega)$ -strong to the minimum of  $\overline{E}(u)$ .
- Then, for  $\varepsilon > 0$  fixed, we propose a suitable numerical scheme called the *half-quadratic algorithm* for which we give a convergence result. The idea is to introduce an additional variable, also called *dual variable*, so that the extended functional  $J_\varepsilon(u, b)$  has the same minimum value as  $E_\varepsilon(u)$ . More precisely, we show that the minimizing sequence of  $J_\varepsilon$ ,  $(u_n, b_n)$  is convergent and that  $u_n$  converges for the  $L^1(\Omega)$ -strong topology to  $u_\varepsilon$ .

This methodology has been originally proposed with the total variation ( $\phi(s) = s$ ) by Chambolle and Lions [64]. It has been then extended to convex functions of measures and applied for several problems (see for instance Sections 5.1.2 and 5.1.3). It is general in the sense that it is a way to approximate the smoothing term which is usually the difficult point. Let us present in detail this two-steps approach.

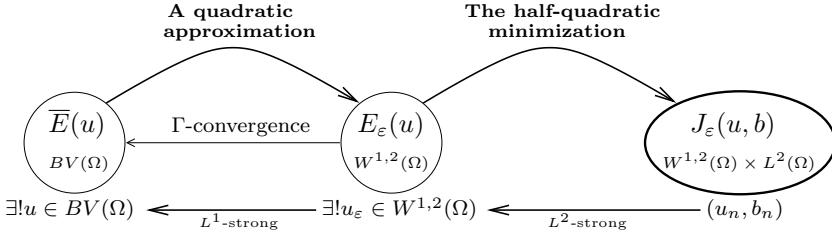


Figure 3.4. Overview of the approach

### A quadratic approximation

For a function  $\phi$  verifying (3.15)-(3.17), let us define  $\phi_\varepsilon$  by

$$\phi_\varepsilon(s) = \begin{cases} \frac{\phi'(\varepsilon)}{2\varepsilon} s^2 + \phi(\varepsilon) - \frac{\varepsilon\phi'(\varepsilon)}{2} & \text{if } 0 \leq s \leq \varepsilon \\ \phi(s) & \text{if } \varepsilon \leq s \leq 1/\varepsilon \\ \frac{\varepsilon\phi'(1/\varepsilon)}{2} s^2 + \phi(1/\varepsilon) - \frac{\phi'(1/\varepsilon)}{2\varepsilon} & \text{if } s \geq \varepsilon. \end{cases}$$

We have  $\forall \varepsilon, \phi_\varepsilon \geq 0$  and  $\forall s, \lim_{\varepsilon \rightarrow 0} \phi_\varepsilon(s) = \phi(s)$ . Now, let us define the functional  $E_\varepsilon$  by

$$E_\varepsilon(u) = \begin{cases} \frac{1}{2} \int_{\Omega} |Ru - u_0|^2 dx + \lambda \int_{\Omega} \phi_\varepsilon(|\nabla u|) dx & \text{if } u \in W^{1,2}(\Omega) \\ +\infty & \text{otherwise.} \end{cases} \quad (3.22)$$

From now on, we suppose that  $\phi$  and  $R$  satisfy hypotheses (3.15)-(3.17) and (3.19). The existence and the uniqueness of  $u_\varepsilon$  are quite obvious and derive from classical arguments.

**Proposition 3.2.1** *For each  $\varepsilon > 0$ , the functional  $E_\varepsilon$  has a unique minimum  $u_\varepsilon$  in  $W^{1,2}(\Omega)$ .*

Now, to show that  $u_\varepsilon$  converges in  $L^1(\Omega)$ -strong to the unique minimum of  $\bar{E}(u)$ , we are going to use the notion of  $\Gamma$ -convergence. In particular we use the two following results that we recall for the convenience of the reader (see Chapter 2, Theorems 2.1.7 and 2.1.8):

Let  $X$  be a topological space, endowed with a  $\tau$ -topology and let  $F_h, F : X \rightarrow \bar{R}$ , then

**Theorem 2.1.7** .../... Let us assume that  $(F_h)$  is equi-coercive and  $\Gamma$ -converges to  $F$ . Let us suppose that  $F$  has a unique minimum  $x_0$  in  $X$ . If  $(x_h)$  is a sequence in  $X$  such that  $x_h$  is a minimum for  $F_h$ , then  $(x_h)$  converges to  $x_0$  in  $X$  and  $(F_h(x_h))$  converges to  $F(x_0)$ .

**Theorem 2.1.8** .../... If  $(F_h)$  is a decreasing sequence converging to  $F$  pointwise, then  $(F_h)$   $\Gamma$ -converges to the lower semi-continuous envelop of  $F$  in  $X$ , denoted  $R_F$ .

Then, we have the following result:

**Theorem 3.2.3** *The sequence  $u_\varepsilon$  from Proposition 3.2.1 converges in  $L^1(\Omega)$ -strong to the unique minimum  $u$  of  $\bar{E}(u)$  and  $E_\varepsilon(u_\varepsilon)$  converges to  $\bar{E}(u)$ .*

**Proof** In our case  $X = L^1(\Omega)$ . Let us denote by  $\tilde{E}(u) : BV(\Omega) \rightarrow \bar{R}$  the functional defined by

$$\tilde{E}(u) = \begin{cases} \bar{E}(u) & \text{if } u \in W^{1,2}(\Omega) \\ +\infty & \text{otherwise.} \end{cases}$$

By construction, we observe that  $E_\varepsilon(u)$  is a decreasing sequence converging pointwise to  $\tilde{E}(u)$ . Therefore  $E_\varepsilon$   $\Gamma$ -converges to the lower semi-continuous envelope  $R_\tau \tilde{E}$  of  $\tilde{E}$ . In order to apply the Theorem 2.1.7, we need to check that  $\bar{E} = R_\tau \tilde{E}$ .

Step 1:  $\bar{E}$  is l.s.c. in  $L^1(\Omega)$  with respect to the  $L^1(\Omega)$ -strong topology.

Indeed, let  $u_h \in L^1(\Omega)$  such that  $u_h \xrightarrow[L^1(\Omega)]{} u$  as  $h \rightarrow +\infty$  and

$\lim_{h \rightarrow +\infty} \bar{E}(u_h) < \infty$ . Then, as  $\bar{E}(u_h)$  is bounded, we deduce that  $u_h$  is uniformly bounded in  $BV(\Omega)$ . Thus, up to a subsequence,  $u_h \xrightarrow[BV - w^*]{} u$  and

$\lim_{h \rightarrow +\infty} \bar{E}(u_h) \geq \bar{E}(u)$ , i.e.  $\bar{E}$  is l.s.c. with respect to the  $L^1(\Omega)$  topology.

Step 2: Let us show that  $\bar{E} = R_\tau \tilde{E}$ .

From Step 1, it suffices to prove that for  $u$  in  $BV(\Omega)$  there exists a sequence  $u_h \in W^{1,2}(\Omega)$  such that  $u_h \xrightarrow[L^1(\Omega)]{} u$  and  $\bar{E}(u) = \lim_{h \rightarrow +\infty} \tilde{E}(u_h)$ . Such a sequence can be constructed by using classical approximation arguments [106, 89].

Finally, applying the  $\Gamma$ -convergence result from Theorem 2.1.7, we conclude that  $u_\varepsilon$ , the unique minimum of  $E_\varepsilon$ , converges in  $L^1(\Omega)$ -strong to the unique minimum  $u$  of  $\bar{E}$ . ■

### The half-quadratic minimization

Now, it remains to compute  $u_\varepsilon$  numerically. To do this, we can use the Euler-Lagrange equation verified by  $u_\varepsilon$ :

$$R^* R u_\varepsilon - \operatorname{div} \left( \frac{\phi'(|\nabla u_\varepsilon|)}{|\nabla u_\varepsilon|} \nabla u_\varepsilon \right) = R^* u_0. \quad (3.23)$$

Equation (3.23) is a highly nonlinear equation. To overcome this difficulty we propose the half-quadratic algorithm based on the following "duality" result [72, 115].

**Proposition 3.2.2** *Let  $\phi : [0, +\infty[ \rightarrow [0, +\infty[$  be such that  $\phi(\sqrt{s})$  is concave on  $]0, +\infty[$  and  $\phi(s)$  is non-decreasing. Let  $L$  and  $M$  be defined by*

$L = \lim_{s \rightarrow +\infty} \frac{\phi'(s)}{2s}$  and  $M = \lim_{s \rightarrow 0} \frac{\phi'(s)}{2s}$ . Then there exists a convex and decreasing function  $\psi : ]L, M] \rightarrow [\beta_1, \beta_2]$  so that:

$$\phi(s) = \inf_{L \leq b \leq M} (bs^2 + \psi(b)) \quad (3.24)$$

where  $\beta_1 = \lim_{s \rightarrow 0^+} \phi(s)$  and  $\beta_2 = \lim_{s \rightarrow +\infty} \left( \phi(s) - \frac{s\phi'(s)}{2} \right)$ . Moreover, for every  $s \geq 0$ , the value  $b$  for which the minimum is reached is given by  $b = \frac{\phi'(s)}{2s}$ . The additional variable  $b$  is usually called the dual variable.

**Proof** Let  $\theta(s) = -\phi(\sqrt{s})$ . By construction  $\theta(s)$  is convex. Thus,  $\theta(s)$  identifies with its convex envelope, i.e.

$$\theta(s) = \theta^{**}(s) = \sup_{s^*} (ss^* - \theta^*(s^*))$$

where  $\theta^*(s^*)$  is the polar function of  $\theta(s)$  defined as:

$$\theta^*(s^*) = \sup_s (ss^* - \theta(s)).$$

Therefore:

$$\phi(\sqrt{s}) = \inf_{s^*} (-ss^* + \theta^*(s^*)).$$

Let  $b = -s^*$  and  $s = \sqrt{s}$ , then  $\phi$  writes as:

$$\phi(s) = \inf_b (bs^2 + \theta^*(-b)) \quad (3.25)$$

that gives the first part of the theorem with  $\psi(b) = \theta^*(-b)$ . Now,

$$\theta^*(-b) = \sup_s (-sb - \theta(s))$$

and, since the application  $s \rightarrow -sb - \theta(s)$  is concave, the supremum is given by the zero of its derivative:

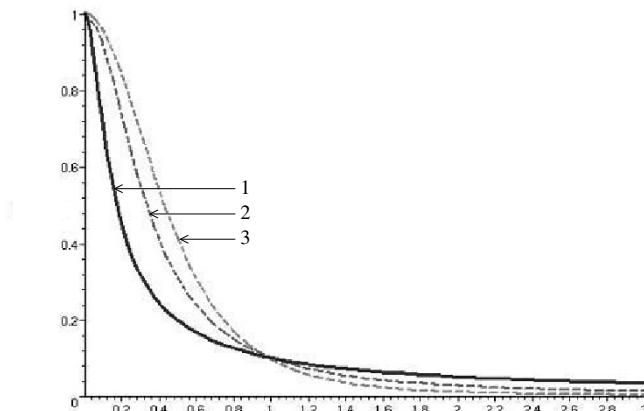
$$-b - \theta'(s) = 0$$

that is to say  $b = \frac{\phi'(\sqrt{s})}{2s}$ . As the application  $s \rightarrow \frac{\phi'(s)}{2s}$  is non-increasing, it is easy to see that the infimum in (3.24) is achieved for  $b \in [L, M]$ . The expressions of  $\beta_1$  and  $\beta_2$  follow immediately. ■

**Remark** It is interesting to observe that Proposition 3.2.2 applies for convex and nonconvex functions  $\phi$ . But, in all cases the function  $\psi$  appearing in (3.24) is always convex. We list in table 3.1 three examples of functions  $\phi$  and their corresponding functions  $\psi$ .

We mention that the first one is often called hypersurface minimal function. The last column presents scaled versions of  $\frac{\phi'(s)}{s}$  so that they are close to 0.1 for  $s = 1$  (see also Figure 3.5). This permits to better compare them from a numerical point of view. ■

	$\phi(s)$	convex?	$\psi(b)$	$\frac{\phi'(s)}{2s}$	Scaled $\frac{\phi'(s)}{2s}$
1	$2\sqrt{1+s^2} - 2$	Yes	$b + \frac{1}{b}$	$\frac{1}{\sqrt{1+s^2}}$	$\frac{1}{\sqrt{1+(10s)^2}}$
2	$\log(1+s^2)$	No	$b - \log(b) - 1$	$\frac{1}{1+s^2}$	$\frac{1}{1+(3s)^2}$
3	$\frac{s^2}{1+s^2}$	No	$b - 2\sqrt{b} + 1$	$\frac{1}{1+s^{2^2}}$	$\frac{1}{1+(3s/2)^{2^2}}$

Table 3.1. Examples of  $\phi$  functionsFigure 3.5. Functions  $\frac{\phi'(s)}{2s}$  with different choices of  $\phi$  (see table 3.1)

Now, let us look at how we may apply Proposition 3.2.2 for solving the problem:

$$\inf_u \left\{ E_\varepsilon(u) = \frac{1}{2} \int_{\Omega} |Ru - u_0|^2 dx + \lambda \int_{\Omega} \phi_\varepsilon(|\nabla u|) dx, \quad u \in W^{1,2}(\Omega) \right\}.$$

Let us assume that  $\phi_\varepsilon$  fulfills the hypotheses of Proposition 3.2.2. Then there exists  $L_\varepsilon$ ,  $M_\varepsilon$  and  $\psi_\varepsilon$  so that:

$$E_\varepsilon(u) = \frac{1}{2} \int_{\Omega} |Ru - u_0|^2 dx + \lambda \int_{\Omega} \inf_{L_\varepsilon \leq b \leq M_\varepsilon} (b |\nabla u|^2 + \psi_\varepsilon(b)) dx.$$

Supposing that we can invert the infimum with respect to  $b$  and the integral (this can be justified), we have:

$$\begin{aligned} \inf_u E_\varepsilon(u) &= \inf_u \inf_b \left[ \frac{1}{2} \int_{\Omega} |Ru - u_0|^2 dx + \lambda \int_{\Omega} (b |\nabla u|^2 + \psi_\varepsilon(b)) dx \right] \\ &= \inf_b \inf_u \left[ \frac{1}{2} \int_{\Omega} |Ru - u_0|^2 dx + \lambda \int_{\Omega} (b |\nabla u|^2 + \psi_\varepsilon(b)) dx \right]. \end{aligned}$$

If we introduce the functional:

$$J_\varepsilon(u, b) = \frac{1}{2} \int_{\Omega} |Ru - u_0|^2 dx + \lambda \int_{\Omega} (b |\nabla u|^2 + \psi_\varepsilon(b)) dx$$

then:

- $J_\varepsilon$  is convex in  $u$  and for each  $u$  fixed in  $W^{1,2}(\Omega)$  it is convex in  $b$ .

Of course,  $J_\varepsilon(u, b)$  is not convex in the pair  $(u, b)$ . So, this leads to the alternate semi-quadratic algorithm described in Table 3.2. To illustrate this, we display in Figure 3.6 the obtained result on the “Borel building” image. The sequence of functions  $b^n(x)$  can be seen as an indicator of



Figure 3.6. Result with the half-quadratic minimization (compare with Figure 3.3: noise is removed while discontinuities are kept)

contours. If  $\phi$  verifies the edge-preserving hypotheses:  $\lim_{s \rightarrow +\infty} \frac{\phi(s)}{2s} = 0$  and  $\lim_{s \rightarrow 0^+} \frac{\phi(s)}{2s} = 1$ , then

- If  $b^n(x) = 0$  then  $x$  belongs to a contour.

---

For  $(u^0, b^0)$  given

---

- $u_\varepsilon^{n+1} = \underset{u}{\operatorname{argmin}} J_\varepsilon(u, b^n)$ . Since the problem is convex (and quadratic), this equivalent to solve:

$$\begin{cases} R^*Ru - \operatorname{div}(b^n \nabla u) = 0 & \text{in } \Omega \\ b^n \frac{\partial u}{\partial N} = 0 & \text{on } \partial\Omega. \end{cases} \quad (3.26)$$

Once discretized, the linear system can be solved with a Gauss-Seidel for example.

- $b^{n+1} = \underset{b}{\operatorname{argmin}} J_\varepsilon(u_\varepsilon^{n+1}, b)$ . According to Proposition 3.2.2, the minimum in  $b$  is reached for

$$b^{n+1} = \frac{\phi'(|\nabla u_\varepsilon^{n+1}|)}{2|\nabla u_\varepsilon^{n+1}|}. \quad (3.27)$$

- Go back to first step until convergence.

The limit  $(u_\varepsilon^\infty, b^\infty)$  is the solution (see Theorem 3.2.4)

Table 3.2. Presentation of the half-quadratic algorithm also called ARTUR (see [70, 72]). On a numerical point of view, the only difficulty is the discretization of the term  $\operatorname{div}(b \nabla u)$  in (3.26) where  $b$  is given and defined by (3.27). Several possibilities can be considered. This is detailed in the Section A.3.2 of the Appendix.

- If  $b^n(x) = 1$  then  $x$  belongs to an homogeneous region.

So the property of this iterative algorithm is to detect and take into account progressively the discontinuities of the image. This is illustrated in Figure 3.7 where the initial condition was such that  $(u^0, b^0) \equiv (0, 1)$ .

Now, as far as the convergence of this algorithm is concerned, we have:

**Theorem 3.2.4** *If  $\phi_\varepsilon$  and  $R$  satisfy (3.15)-(3.17) and (3.19), and if  $\phi_\varepsilon$  fulfills the assumptions of Proposition 3.2.2, then the sequence  $(u^n, b^n)$  is convergent in  $L^2(\Omega)$ -strong  $\times L^\infty(\Omega)$ -weak. Moreover  $u^n$  converges strongly in  $L^2(\Omega)$  (and weakly in  $W^{1,2}(\Omega)$ ) to the unique solution  $u_\varepsilon$  of  $E_\varepsilon$ .*

We do not reproduce here the proof of Theorem 3.2.4. It is rather long and technical. For more details, we refer the reader to [64, 17].

## Remarks

- The only assumption ensuring that the semi-quadratic algorithm works is  $\phi(\sqrt{s})$  concave. That is to say, we may even apply this algorithm to nonconvex functions  $\phi$ . Of course, we will be able to give a convergence result only for convex functions.

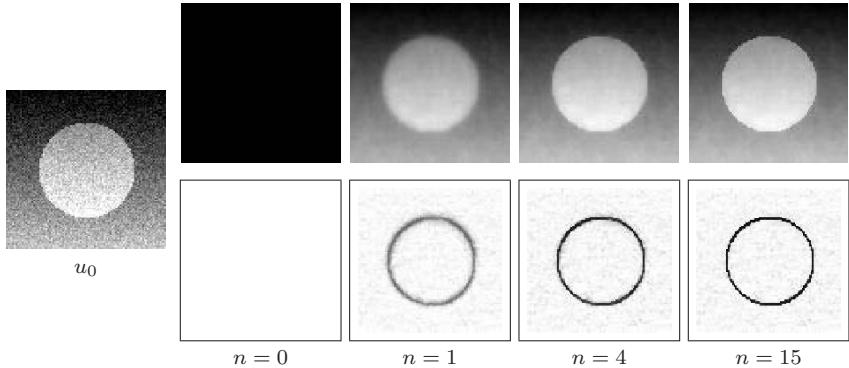


Figure 3.7. Illustration of the behaviour of half-quadratic algorithm on a synthetic image  $u_0$ . On the right-hand side, four iterations of the algorithm are displayed ( $n = 0, 1, 4, 15$ ). First row is  $u^n$  while the second one is  $b^n$ . Initialization is  $u^0 \equiv 0$  and  $b^0 \equiv 1$ . This shows the interpretation of  $b$  as an edge detector which becomes more precise as time evolves.

- The half-quadratic approach can also be realized with another duality result based on the Legendre transform. This has been studied in [16, 71].

■

### 3.2.5 Some invariances and the role of $\lambda$

In this section we set some elementary properties to highlight the invariance and the "scale" nature of the parameter  $\lambda$  in front of the regularization term in  $E(u)$ :

$$E(u) = \frac{1}{2} \int_{\Omega} |u - u_0|^2 dx + \lambda \int_{\Omega} \phi(|\nabla u|) dx.$$

Here we assume that  $u_0 \in L^2(\Omega)$  and that  $\phi$  verify hypotheses (3.15)-(3.17). Let  $u(x, \lambda)$  be the unique minimizer of  $E(u)$ . To simplify, we suppose that

$$u(., \lambda) \in W^{1,1}(\Omega) \cap L^2(\Omega).$$

Let us define the operator  $T_\lambda : L^2(\Omega) \rightarrow L^2(\Omega)$  by  $T_\lambda u_0 = u(\lambda)$  where  $u(\lambda) = u(x, \lambda)$ . By definition, we have for all  $v \in W^{1,1}(\Omega) \cap L^2(\Omega)$

$$\begin{aligned} \frac{1}{2} \int_{\Omega} |u(x, \lambda) - u_0(x)|^2 dx + \lambda \int_{\Omega} \phi(|\nabla u(x, \lambda)|) dx &\leq \\ &\leq \frac{1}{2} \int_{\Omega} |v(x) - u_0(x)|^2 dx + \lambda \int_{\Omega} \phi(|\nabla v(x)|) dx. \end{aligned} \quad (3.28)$$

Moreover,  $u(x, \lambda)$  necessarily verifies the Euler-Lagrange equation:

$$\begin{cases} u(x, \lambda) - u_0(x) = \lambda \operatorname{div} \left( \frac{\phi'(|\nabla u(x, \lambda)|)}{|\nabla u(x, \lambda)|} \nabla u(x, \lambda) \right) \text{ in } \Omega \\ \frac{\phi'(|\nabla u(x, \lambda)|)}{|\nabla u(x, \lambda)|} \frac{\partial u}{\partial N}(x, \lambda) = 0 \text{ on } \partial\Omega. \end{cases} \quad (3.29)$$

We begin by carrying out some invariance properties which can be easily proved. These invariances with respect to some image transformation  $Q$  express the fact that  $T_\lambda$  and  $Q$  can commute.

(A1) *Gray level invariance:*

$$T_\lambda 0 = 0 \text{ and } T_\lambda(u_0 + c) = T_\lambda u_0 + c, \text{ for every constant } c.$$

(A2) *Translation invariance:*

$$\text{Define the translation } \tau_h \text{ by } (\tau_h)(f)(x) = f(x + h), \\ \text{then } T_\lambda(\tau_h u_0) = \tau_h(T_\lambda u_0).$$

(A3) *Isometry invariance:*

Let us denote  $(Rf)(x) = f(Rx)$  for any isometry of  $R^2$  and function  $f$  from  $R^2$  to  $R$ . Then  $T_\lambda(Ru_0) = R(T_\lambda u_0)$ . Of course, this invariance is true because the regularization function depends on the norm of  $\nabla u$ .

Now, let us examine some properties of the correspondence:

$$\lambda \rightarrow T_\lambda u_0(x) = u(x, \lambda).$$

Before proving them, it is interesting to do some quantitative tests on the “Borel building” image. From these experiments, we can observe that

- The SNR reaches a maximum rapidly and decreases fastly (see Figure 3.8).
- $|u(., \lambda)|_{L^2(\Omega)}$  seems to be constant.
- $\bar{u}(., \lambda)$ , the mean of  $u$ , is constant.
- $\int_{\Omega} |u(x, \lambda) - \bar{u}_0| dx$  tends to zero, which means that  $u$  converges in the  $L^1(\Omega)$ -strong topology to the average of the initial data.

Let us see if we can prove some of theses empiric properties.

**Property 1** *The  $L^2$  norm of  $u(., \lambda)$  is bounded by a constant independant of  $\lambda$ .*

**Proof** By letting  $v = 0$  in (3.28) and remembering that  $\phi(0) = 0$ , we get:

$$\int_{\Omega} |u(x, \lambda) - u_0|^2 dx \leq \int_{\Omega} |u_0|^2 dx$$

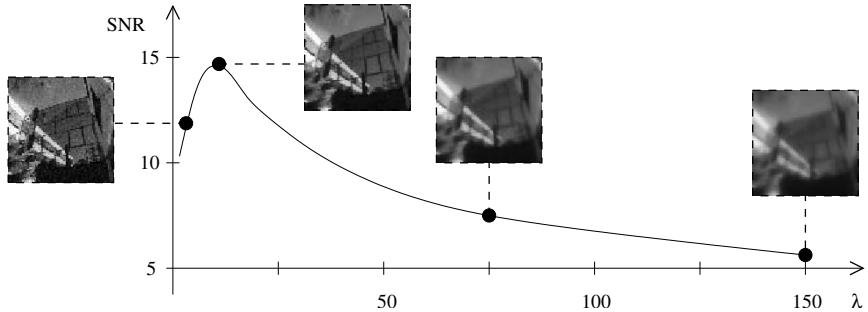


Figure 3.8. SNR as a function of  $\lambda$ . After being optimal, high values of  $\lambda$  smooth too much the image

from which, we deduce:

$$\int_{\Omega} |u(x, \lambda)|^2 dx \leq 2 \int_{\Omega} |u_0|^2 dx$$

i.e. the  $L^2$ -norm of  $u$  is bounded by a constant independent of  $\lambda$ . ■

**Property 2** For every  $\lambda$ , we have  $\int_{\Omega} u(x, \lambda) dx = \int_{\Omega} u_0(x) dx$ .

**Proof** From (3.29), we get

$$\int_{\Omega} u(x, \lambda) dx = \int_{\Omega} u_0(x) dx + \lambda \int_{\Omega} \operatorname{div} \left( \frac{\phi'(|\nabla u(x, \lambda)|)}{|\nabla u(x, \lambda)|} \nabla u(x, \lambda) \right) dx.$$

But thanks to the Green formula and the boundary condition in (3.29):

$$\int_{\Omega} \operatorname{div} \left( \frac{\phi'(|\nabla u(x, \lambda)|)}{|\nabla u(x, \lambda)|} \nabla u(x, \lambda) \right) dx = \int_{\partial\Omega} \frac{\phi'(|\nabla u(x, \lambda)|)}{|\nabla u(x, \lambda)|} \frac{\partial u}{\partial N} d\Gamma = 0$$

which concludes the proof. ■

**Property 3**  $u(\cdot, \lambda)$  converges in  $L^1(\Omega)$ -strong to the average of the initial data.

**Proof** Again from (3.28), letting  $v = 0$ , we obtain:

$$0 \leq \lambda \int_{\Omega} \phi(|\nabla u(x, \lambda)|) dx \leq \frac{1}{2} \int_{\Omega} u_0^2(x) dx.$$

Therefore:

$$0 \leq \lim_{\lambda \rightarrow +\infty} \int_{\Omega} \phi(|\nabla u(x, \lambda)|) dx \leq \lim_{\lambda \rightarrow +\infty} \frac{1}{2\lambda} \int_{\Omega} u_0^2(x) dx = 0$$

i.e.  $\lim_{\lambda \rightarrow +\infty} \int_{\Omega} \phi(|\nabla u(x, \lambda)|) dx = 0$ . Since  $\phi$  is strictly convex with linear growth, we easily deduce from the above equality that:

$$\lim_{\lambda \rightarrow +\infty} |\nabla u(., \lambda)|_{L^1(\Omega)} = 0. \quad (3.30)$$

On the other hand, thanks to the Poincaré-Wirtinger inequality:

$$\int_{\Omega} |u(x, \lambda) - \bar{u}_0| dx \leq cte |\nabla u(x, \lambda)|_{L^1(\Omega)} \quad \text{with } \bar{u}_0 = \frac{1}{\Omega} \int_{\Omega} u_0(x, \lambda) dx.$$

Thus, with (3.30),  $\lim_{\lambda \rightarrow +\infty} \int_{\Omega} |u(x, \lambda) - \bar{u}_0| dx = 0$ , i.e.  $u(., \lambda)$  converges in  $L^1(\Omega)$ -strong to the average of the initial data. ■

This latter property shows that  $\lambda$  can be interpreted as a *scale parameter*. Starting from the initial image  $u_0$  we construct a family of images  $\{u(x, \lambda)\}_{\lambda > 0}$  of gradually simplified (smoothed) versions of it. This scale notion, also called *scale space theory*, plays a central role in the image analysis and has been investigated by many authors [28, 185, 4, 148, 251]. As we will see in the Section 3.3, "scale space ideas" are strongly present in the PDE theory.

### 3.2.6 Some remarks in the nonconvex case.

As noticed at the beginning of Section 3.2.3 a "good" edge-preserving behaviour for  $\phi$  would be such that:

$$\phi(s) \approx cs^2 \quad \text{as } s \rightarrow 0^+ \quad (3.31)$$

$$\lim_{s \rightarrow +\infty} \phi(s) \approx \gamma > 0. \quad (3.32)$$

Unfortunately, conditions (3.31) and (3.32) imply that  $\phi$  is nonconvex. Of course, there is no longer existence of a solution to the minimization problem and one cannot prove any convergence result. Nevertheless, potentials satisfying (3.31)-(3.32) like:

$$\phi(s) = \frac{s^2}{1+s^2}$$

seem to provide better (sharper) results than convex potentials with linear growth... This is illustrated in Figures 3.9 and 3.10. So what can be said for this case?

First of all we have just mentioned that using a nonconvex potential yields to an ill-posed problem. This may be not straightforward indeed. Let us

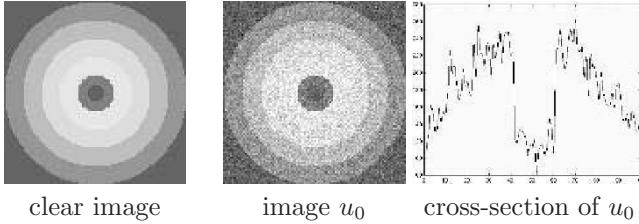


Figure 3.9. Clear image (without noise), original noisy image (obtained by adding a Gaussian noise of variance 20) and a cross-section of it passing through the center of the image.

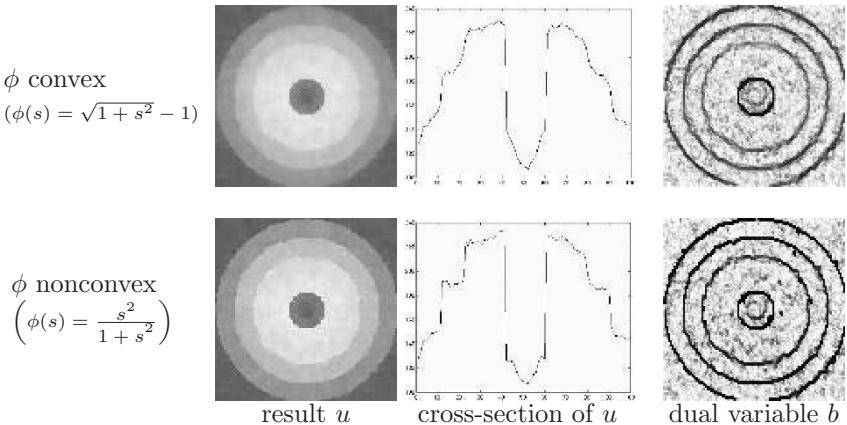


Figure 3.10. Results obtained using a convex and a nonconvex  $\phi$  function. We display the result  $u$ , a cross-section passing through the center of the image and the dual variable  $b$ . One can observe that nonconvex  $\phi$  functions allow to reconstruct sharper images.

prove it for the following energy:

$$E(u) = \int_{\Omega} |u - u_0|^2 dx + \lambda \int_{\Omega} \frac{|\nabla u|^2}{1 + |\nabla u|^2} dx.$$

The function  $\phi(s) = \frac{s^2}{1+s^2}$  satisfies (3.31) and (3.32) with  $\gamma = 1$ . We are going to show that  $E(u)$  has no minimizer. We assume that  $u_0 \in L^\infty(\Omega)$ .

**Proposition 3.2.3** *If  $u_0(x)$  is not a constant, the functional  $E(u)$  has no minimizer in  $W^{1,2}(\Omega)$  and  $\inf \{E(u); u \in W^{1,2}(\Omega)\} = 0$ .*

**Proof** For clarity's sake, we prove the proposition in the one-dimensional case:  $\Omega = [a, b]$ . The same proof goes for  $N \geq 2$ . We follow the proof given by Chipot et al [75].

By density, we always may find a sequence of step functions  $\overline{u_n}$  such that:

$$|\overline{u_n}| \leq |u_0|_{L^\infty}, \quad \lim_{n \rightarrow \infty} |\overline{u_n} - u_0|_{L^2(\Omega)} = 0.$$

In fact, we can find a partition  $a = x_0 < x_1 < \dots < x_n = b$  such that  $\overline{u_n}$  is constant on each interval  $(x_{i-1}, x_i)$ ,  $h_n = \max_i(x_{i-1} - x_i) < 1$  with  $\lim_{n \rightarrow \infty} h_n = 0$ . Let us denote  $\sigma_i = x_i - x_{i-1}$ . Next, we define a sequence of continuous functions  $u_n$  by:

$$u_n(x) = \begin{cases} \overline{u_n}(x) & \text{if } x \in [x_{i-1}, x_i - \sigma_i^2] \\ \frac{(\overline{u_{n,i+1}} - \overline{u_{n,i}})}{\sigma_i^2} + \overline{u_{n,i}} & \text{if } x \in [x_i - \sigma_i^2, x_i] \end{cases}$$

where  $\overline{u_{n,i}} = \overline{u_n}(x)_{[x_{i-1}, x_i]}$ . It is easy to check that:

$$|\overline{u_n} - u_n|_{L^2(\Omega)} \leq 2 |u_0|_{L^\infty}^2 \sum_{i=1}^n \sigma_i^2 \leq 2 |u_0|_{L^\infty}^2 (b-a) h_n.$$

Therefore  $\lim_{n \rightarrow \infty} |\overline{u_n} - u_n|_{L^2(\Omega)} = 0$ . Since:

$$\lim_{n \rightarrow \infty} |u_n - u_0|_{L^2(\Omega)} \leq \lim_{n \rightarrow \infty} |u_n - \overline{u_n}|_{L^2(\Omega)} + \lim_{n \rightarrow \infty} |\overline{u_n} - u_0|_{L^2(\Omega)}$$

we also deduce that  $\lim_{n \rightarrow \infty} |u_n - u_0| = 0$ . Moreover:

$$\begin{aligned} \int_a^b \frac{|u'_n|^2(x)}{1 + |u'_n|^2(x)} dx &= \sum_i \int_{x_i - \sigma_i^2}^{x_i} \frac{(\overline{u_{n,i+1}} - \overline{u_{n,i}})^2}{\sigma_i^2 + (\overline{u_{n,i+1}} - \overline{u_{n,i}})^2} dx \leq \\ &\leq \sum_i \sigma_i^2 \leq h_n \sum_i \sigma_i = h_n(b-a) \end{aligned}$$

thus:

$$\lim_{n \rightarrow \infty} \int_a^b \frac{|u'_n|^2(x)}{1 + |u'_n|^2(x)} dx = 0$$

and finally:

$$0 \leq \inf_{u \in W^{1,2}(\Omega)} E(u) \leq \lim_{n \rightarrow \infty} E(u_n) = 0$$

i.e.

$$\inf_{u \in W^{1,2}(\Omega)} E(u) = 0.$$

Now, if there exists a minimizer  $u \in W^{1,2}(\Omega)$ , necessarily  $E(u) = 0$ , which implies:

$$\begin{cases} \int_a^b |u - u_0|^2 dx = 0 \Leftrightarrow u = u_0 \text{ a.e.} \\ \int_a^b \frac{u'^2}{1 + u'^2} dx = 0 \Leftrightarrow u' = 0 \text{ a.e.} \end{cases}$$

The first equality is only possible if  $u_0 \in W^{1,2}(\Omega)$  (since  $u \in W^{1,2}(\Omega)$ ) and in this case the second equality implies  $u'_0 = 0$ , which is only possible if  $u_0$  is a constant. Therefore excluding this trivial case,  $E(u)$  has no minimizer in  $W^{1,2}(\Omega)$ . ■

☞ *By density arguments, there is no hope to obtain the existence of a minimizer for  $E(u)$  in any reasonable space.*

Then what can we do? A possibility is to regularize the functional  $E(u)$  either by adding constraints, or by adding a supplementary term. This latter idea has been investigated by Chipot et al [75]. They introduced the following energy:

$$E_\varepsilon(u) = \int_{\Omega} |u - u_0|^2 dx + \lambda \int_{\Omega} \frac{|\nabla u|^2}{1 + |\nabla u|^2} dx + \varepsilon \int_{\Omega} |\nabla u|^2 dx.$$

$E_\varepsilon(u)$  is convex for  $\varepsilon \geq \frac{\lambda}{4}$  and nonconvex for  $\varepsilon < \frac{\lambda}{4}$ . The former case is not interesting since it is too regularizing and not edge-preserving. Though  $E_\varepsilon(u)$  is nonconvex for  $\varepsilon < \frac{\lambda}{4}$ , it has a quadratic growth at infinity. This fact allows us to use convexification tools that permit to obtain the existence of a minimizer for  $E_\varepsilon(u)$  in the one-dimensional case [75].

⌘ *For dimensions greater than one, the problem is quite open. The behaviour of the minimizing sequences is also a challenging problem which is closely related to the Perona-Malik anisotropic diffusion, as we shall see further.*

Another attempt would be to work directly with the discrete version of  $E(u)$ . For example, we mention a recent paper by Rosati [212] who studies the asymptotic behaviour of a discrete model related to  $E(u)$ . In his paper, Rosati chooses  $\phi(s) = \frac{s^2}{1 + \mu s^2}$  and with the condition that  $\mu$  is proportional to the mesh-size  $h$ , he proves that the discrete model  $\Gamma$ -converges to a modified Mumford-Shah functional as  $h$  tends to 0. This is surely a wise approach since if we make such efforts to introduce nonconvex potentials, it is because they give very good numerical results in restoration problems. Discrete problems have generally solutions even in the nonconvex case. It

should be very interesting to investigate better the relationships between discrete and continuous models.

We also mention that choosing  $\phi(s) = \log(1 + s^2)$  leads also very good results and is often used in experiments. This is again another situation since  $\phi$  is nonconvex but with sublinear growth at infinity. This would be also very interesting to understand the problem from a theoretical point of view.

### 3.3 PDE-based methods

In the previous section, we have considered a class of approach which consists in setting the best energy according to our needs. The equations that were to be solved numerically were the Euler-Lagrange equations associated to the minimization problems. Another possibility is to work directly on the equations, without thinking of any energy. This is the aim of this section to present some classical PDE-based methods for restoration, trying to follow the chronological order in which they appeared in literature. These models can be formally written in the general form:

$$\begin{cases} \frac{\partial u}{\partial t}(t, x) + F(x, u(t, x), \nabla u(t, x), \nabla^2 u(t, x)) = 0 & \text{in } \Omega \\ & \text{(a second order differential operator)} \\ \frac{\partial u}{\partial N}(t, x) = 0 & \text{on } \partial\Omega \quad \text{(Neumann boundary condition)} \\ u(0, x) = u_0(x) & \text{(initial condition)} \end{cases} \quad (3.33)$$

where  $u(t, x)$  is the restored version of the initial degraded image  $u_0(x)$ . As usual  $\nabla u$  and  $\nabla^2 u$  stand respectively for the gradient and the Hessian matrix of  $u$  with respect to the space variable  $x$ . Let us comment.

One of the main difference with the equations encountered up to now is the presence of the parameter  $t$ . Starting from the initial image  $u_0(x)$  and by running (3.33) we construct a family of functions (i.e. images)  $\{u(t, x)\}_{t>0}$  representing successive versions of  $u_0(x)$ . As  $t$  increases we expect that  $u(t, x)$  changes into a more and more simplified image, or in other words structures for large  $t$  constitute simplifications of corresponding structures at small  $t$ . Moreover no new structure must be created. For these reasons  $t$  is called a *scale variable*.

As we will see further, the choice of  $F$  in (3.33) is determining since we would like to attain two goals that may seem *a priori* contradictory. The first is that  $u(t, x)$  should represent a smooth version of  $u_0(x)$  where the noise has been removed. The second is to be able to preserve some features such as edges, corners, T-junctions, which may be viewed as singularities.

Finally, a natural question is how to classify PDE-based models. The answer is inevitably subjective. Maybe, the simplest way is to choose the classical PDEs classification, namely forward parabolic PDEs, backward parabolic PDEs and hyperbolic PDEs corresponding respectively to

smoothing, enhancing and smoothing-enhancing processes. Let us follow this classification.

### 3.3.1 Smoothing PDEs

The heat equation

The oldest and most investigated equation in image processing is probably the parabolic linear heat equation [28, 4, 148]:

$$\begin{cases} \frac{\partial u}{\partial t}(t, x) - \Delta u(t, x) = 0 & x \in R^2, t \geq 0 \\ u(0, x) = u_0(x). \end{cases} \quad (3.34)$$

Notice that we have here  $x \in R^2$ . In fact, we consider that  $u_0(x)$  is primarily defined on the square  $[0, 1]^2$ . By symmetry we extend it to  $C = [-1, +1]^2$  and then in the whole  $R^2$  by periodicity (see Figure 3.11). This way of extending  $u_0(x)$  is classical in image processing. The motivation will become clearer in the sequel. If  $u_0(x)$  extended in this way satisfies in addition

$\int_C |u_0(x)| dx < +\infty$ , we will say that  $u_0 \in L^1_{\#}(C)$ .

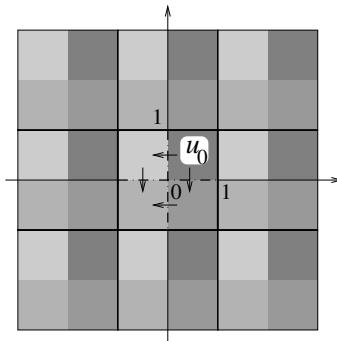


Figure 3.11. Extension of  $u_0$  primarily defined on  $[0, 1]^2$  to  $R^2$  by symmetry and periodicity

The motivation to bring out such an equation came from the following remark: solving (3.34) is equivalent to carry out a Gaussian linear filtering which was widely used in signal processing. More precisely, let  $u_0$  be in  $L^1_{\#}(C)$ , then the explicit solution of (3.34) is given by:

$$u(t, x) = \int_{R^2} G_{\sqrt{2t}}(x - y) u_0(y) dy = (G_{\sqrt{2t}} * u_0)(x) \quad (3.35)$$

where  $G_\sigma(x)$  denotes the two-dimensional Gaussian kernel:

$$G_\sigma(x) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|x|^2}{2\sigma^2}\right). \quad (3.36)$$

The convolution by a positive kernel is the basic operation in linear image filtering. It corresponds to a low-pass filtering (see Figure 3.12). This formula gives the correspondence between the time  $t$  and the scale parameter  $\sigma$  of the Gaussian kernel.



Figure 3.12. Examples of the test image at different scales

The action of the Gaussian kernel can also be interpreted in the frequency domain. Let us define the Fourier transform:

$$F[f](w) = \int_{R^2} f(x) \exp(-i w \cdot x) dx$$

where  $w \in R^2$ . It is well-known that:

$$F[G_\sigma * f](w) = F[G_\sigma](w)F[f](w)$$

and since:

$$F[G_\sigma](w) = \exp\left(-\frac{|w|^2}{2/\sigma^2}\right)$$

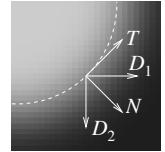
then:

$$F[G_\sigma * f](w) = \exp\left(-\frac{|w|^2}{2/\sigma^2}\right) F[f](w)$$

i.e. the convolution by a Gaussian is a low-pass filter that inhibits high frequencies (oscillations in the space domain).

**Remark** As we can observe in Figure 3.12, the smoothing is isotropic: it does not depend on the image and it is the same in all directions. In particular, edges are not preserved.

In fact, if we introduce two arbitrary orthonormal directions  $D_1$  and  $D_2$ , we have  $\Delta u = u_{D_1 D_1} + u_{D_2 D_2}$ . If we rewrite this equality with the directions  $D_1 = N = \frac{\nabla u}{|\nabla u|}$  and  $D_2 = T$  with  $T \cdot N = 0$ ,  $|T| = 1$ , then  $\Delta u = u_{NN} + u_{TT}$ . The isotropy means that the diffusion is equivalent in the two directions.



As it will be shown in the sequel, most of the diffusion operators can be decomposed as a weighted sum of  $u_{NN}$  and  $u_{TT}$  [153]. ■

To set the properties satisfied by  $u(t, x)$ , we can either use the fact that  $u$  is convolution product or we can deduce these properties from the general theory of uniformly parabolic equation. Choosing the latter approach, we summarize below some of the main properties of  $u(t, x)$  (see [130] for the proofs).

**Proposition 3.3.1** *Let  $u_0$  be in  $L^1_\#(C)$  and define  $u(t, x)$  by (3.35). Then  $u(t, x)$  satisfies for all  $t > 0$  and  $x \in R^2$ , the heat equation with initial value  $u_0$ :*

$$\frac{\partial u}{\partial t}(t, x) = \Delta u(t, x) \quad \text{and} \quad \lim_{t \rightarrow 0} \int_C |u(t, x) - u_0(x)| dx = 0$$

$$u(., t) \in L^1_\#(C) \quad \text{and} \quad u \in C^\infty(R^2 \times (0, T)) \quad \text{for all } T > 0.$$

Moreover, if  $t_1$  is any positive real number, there exists a constant  $c(t_1)$  so that for  $t \in [t_1, +\infty[$

$$\sup_{x \in R^2} |u(t, x)| \leq c(t_1) \|u_0\|_{L^1_\#(C)}. \quad (3.37)$$

If  $u_0 \in L^\infty_\#(C)$  then we have a maximum principle

$$\inf_{x \in R^2} u_0(x) \leq u(t, x) \leq \sup_{x \in R^2} u_0(x). \quad (3.38)$$

$u(t, x)$  given by (3.35) is the unique solution of the heat equation satisfying conditions (3.37) and (3.38).

**Remark** The uniqueness of  $u$  in Proposition 3.3.1 has been obtained in the class of periodic functions. If we drop this assumption we have to replace it by another one. It is well-known that the heat equation  $\frac{\partial u}{\partial t} - \Delta u = 0$  in  $R^2 \times (0, T)$ ,  $u(0, x) = 0$  has infinitely many solutions. Each of the solutions besides  $u \equiv 0$  grows very rapidly as  $|x| \rightarrow +\infty$ . To get an uniqueness result in this case, it sufficient to impose that  $u$  satisfy the growth estimate:

$$|u(t, x)| \leq A e^{a|x|^2},$$

for some constants  $A$  and  $a > 0$ . We refer to [105] for more details. ■

Now, let  $T_t$ ,  $t > 0$ , be the family of scale-operators from  $L^1_\#(C)$  into

$L^1_{\#}(C)$  defined by  $(T_t u_0)(x) = u(t, x)$ , where  $u(t, x)$  is the unique solution of (3.34) given by (3.35).  $T_t$  is a family of linear operators. We list below some invariance properties of  $T_t$  which can be easily proved by remarking that  $G_\sigma(x) > 0$ ,  $\int_{R^2} G_\sigma(x) dx = 1$ , and by using the periodicity of  $u_0$  and classical theorems of integration theory such as the Fubini's Theorem:

(A1) *Gray level shift invariance:*

$$T_t(0) = 0 \text{ and } T_t(u_0 + c) = T_t u_0 + c, \text{ for any constant } c.$$

(A2) *Translation invariance:*

$$T_t(\tau_h u_0) = \tau_h(T_t u_0), \text{ where } \tau_h \text{ is the translation } \tau_h(f)(x) = f(x+h).$$

(A3) *Scale invariance:*

$$T_t(H_\lambda u_0) = H_\lambda(T_{t'} u_0) \text{ with } t' = t \lambda^2, \text{ where } (H_\lambda f)(x) = f(\lambda x).$$

(A4) *Isometry invariance:*

$$T_t(R u_0) = R(T_t u_0), \text{ for any orthogonal transformation } R \text{ of } R^2 \text{ where } (Rf)(x) = f(Rx).$$

(A5) *Conservation of average value:*

$$T_t(M u_0) = M(T_t u_0), \text{ where } Mf = \int_C f(x) dx.$$

(A6) *Semi-group property:*

$$T_{t+s} u_0 = T_t(T_s u_0).$$

(A7) *Comparison principle:*

$$\text{If } u_0 \leq v_0 \text{ then } (T_t u_0) \leq (T_t v_0).$$

These invariance properties are quite natural from an image analysis point of view. For example the gray level shift invariance means that the analysis must be independent of the range of the brightness of the initial image. The other geometric properties traduce the invariance of image analysis under the respective positions of percipients and perceptum.

Are these properties sufficient to ensure correct qualitative properties for  $T_t u$ ? The answer is no. Though the heat equation has been (and is) successfully applied in image processing, it has really some drawbacks: it is too smoothing. In fact, whatever the regularity of the initial data,  $u(t, x)$  is  $C^\infty$  in  $x$ ,  $\forall t > 0$ : edges are lost. We sometimes say that the heat equation has infinite speed of propagation. Of course, this instantaneous regularity is not a desirable property since in particular edges can be lost or severely blurred.

### Nonlinear diffusion

We are going to describe models that are generalizations of the heat equation. What we would like to do is to find models (if possible, well-posed models) removing the noise while preserving the edges at best. By now,

the domain image will be a bounded open set  $\Omega$  of  $R^2$ . Let us consider the following equation initially proposed by Perona and Malik [209]:

$$\begin{cases} \frac{\partial u}{\partial t} = \operatorname{div}(c(|\nabla u|^2) \nabla u) & \text{in } \Omega \times (0, T) \\ \frac{\partial u}{\partial N} = 0 & \text{on } \partial\Omega \times (0, T) \\ u(0, x) = u_0(x) & \text{in } \Omega \end{cases} \quad (3.39)$$

with  $c(s) : [0, +\infty[ \rightarrow ]0, +\infty[$ . Before going further, we can remark that if we choose  $c \equiv 1$ , then we recover the heat equation. Now, imagine that  $c(s)$  is a decreasing function satisfying  $c(0) = 1$  and  $\lim_{s \rightarrow +\infty} c(s) = 0$ . With this choice:

- Inside the regions where the magnitude of the gradient of  $u$  is weak, equation (3.39) acts like the heat equation resulting in an isotropic smoothing.
- Near the region boundaries where the magnitude of the gradient is large, the regularization is “stopped” and the edges are preserved.

Indeed, we can be more precise if we interpret this divergence operator using the directions  $T, N$  associated to the image (as for (3.7)-(3.8)). By developing formally the divergence operator, we get (with the usual notations  $u_x, u_{xx}, \dots$ )

$$\begin{aligned} \operatorname{div}(c(|\nabla u|^2) \nabla u) &= \\ &= 2(u_x^2 u_{xx} + u_y^2 u_{yy} + 2u_x u_y u_{xy}) c'(|\nabla u|^2) + c(|\nabla u|^2) (u_{xx} + u_{yy}). \end{aligned}$$

If we define  $b(s) = c(s) + 2sc'(s)$ , then (3.39) reads as

$$\frac{\partial u}{\partial t}(t, x) = c(|\nabla u|^2) u_{TT} + b(|\nabla u|^2) u_{NN}. \quad (3.40)$$

Therefore (3.40) may be interpreted as a sum of a diffusion in the  $T$ -direction plus a diffusion in the  $N$ -direction, the functions  $c$  and  $b$  acting as weighting coefficients. Of course, since  $N$  is normal to the edges it would be preferable to smooth more in the tangential direction  $T$  than in the normal direction  $N$ . Thus, we impose  $\lim_{s \rightarrow +\infty} \frac{b(s)}{c(s)} = 0$ , or equivalently, according to the definition of  $b$ :

$$\lim_{s \rightarrow +\infty} \frac{s c'(s)}{c(s)} = -\frac{1}{2}. \quad (3.41)$$

If we restrict ourselves to functions  $c(s) > 0$  with power growth then the above limit implies that  $c(s) \approx \frac{1}{\sqrt{s}}$  as  $s \rightarrow +\infty$ . The question now is to know whether (3.39) is well-posed or not.

Firstly, we have to examine the parabolicity of equation (3.39). To do this, we observe that (3.39) can be written as:

$$\frac{\partial u}{\partial t} = a_{11}(|\nabla u|^2) u_{xx} + 2 a_{12}(|\nabla u|^2) u_{xy} + a_{22}(|\nabla u|^2) u_{yy} \quad (3.42)$$

with

$$\begin{aligned} a_{11}(|\nabla u|^2) &= 2 u_x^2 c'(|\nabla u|^2) + c(|\nabla u|^2) \\ a_{12}(|\nabla u|^2) &= 2 u_x u_y c'(|\nabla u|^2) \\ a_{22}(|\nabla u|^2) &= 2 u_y^2 c'(|\nabla u|^2) + c(|\nabla u|^2) \end{aligned}$$

and (3.42) is parabolic if and only if  $\sum_{i=1,2} a_{ij}(|\nabla u|^2) \xi_i \xi_j \geq 0$ ,  $\forall \xi \in R^2$ .

An easy algebraic calculation shows that this condition reduces to the only inequality:

$$b(s) > 0.$$

To summarize, the assumptions imposed on  $c(s)$  are:

$$\begin{cases} c : [0, +\infty[ \rightarrow ]0, +\infty[ \text{ decreasing} \\ c(0) = 1, \quad c(s) \approx \frac{1}{\sqrt{s}} \quad \text{as } s \rightarrow +\infty \\ b(s) = c(s) + 2s c'(s) > 0. \end{cases} \quad (3.43)$$

A canonical example of function  $c(s)$  verifying (3.43) is  $c(s) = \frac{1}{\sqrt{1+s}}$ .

 *With the assumption the (3.43) the nonlinear diffusion model (3.39) acts as a forward parabolic equation smoothing homogeneous regions while preserving edges.*

**Remark** If we release the condition  $b(s) > 0$ , by supposing for example that for some  $s_0$ ,  $b(s) > 0$  for  $s \leq s_0$ , and  $b(s) < 0$  for  $s > s_0$  then (3.39) changes into a backward parabolic equation for  $|\nabla u|^2 > s_0$ , or equivalently into a smoothing-enhancing model. We will come back latter on this model (see Sectionsec:Rest:smoothing-enhancing). ■

What can be said concerning the existence of a solution for (3.39)? Unfortunately, with the assumption (3.43) we cannot directly apply general results for parabolic equations. The difficulty comes from the highly degenerated behaviour of (3.39) due to the vanishing condition  $c(s) \approx \frac{1}{\sqrt{s}}$  as  $s$  tends to infinity. As a matter of fact, one can find some classical results for equations of the form:

$$\frac{\partial u}{\partial t} - \operatorname{div} a(t, x, u, \nabla u) = 0$$

where the function  $a$  satisfies the structure conditions:

$$a(t, x, u, \nabla u) \cdot \nabla u \geq \alpha_0 |\nabla u|^p - \beta_0(t, x) \quad (3.44)$$

$$|a(t, x, u, \nabla u)| \leq \alpha_1 |\nabla u|^{p-1} + \beta_1(t, x) \quad (3.45)$$

a.e.  $(t, x)$  with  $p > 1$ ;  $(\alpha_i)_{i=0,1}$  are given constants and  $(\beta_i)_{i=0,1}$  are given non-negative functions satisfying some integrability conditions [97]. Here we have  $p = 1 \dots$

• In fact the difficulties to get an existence result for (3.39) have to be compared to the ones encountered in Section 3.1.2 for variational problems. We saw that a linear growth assumption on the potential required enlarging the problem and working on the space  $BV(\Omega)$  of functions of bounded variation.

A well-adapted framework to solve (3.39) with assumptions (3.43) is the *nonlinear semi-group theory* and the notion of *maximal operator*. We only recall some basic definitions referring the reader to [50, 60] for the complete theory. Let  $(H, < . >)$  be an Hilbert space and  $A : H \rightarrow \mathcal{P}(H)$  an operator where  $\mathcal{P}(H)$  is the set of subsets of  $H$ . The domain of  $A$  is the set  $D(A) = \{x \in H ; Ax \neq \emptyset\}$  and the range of  $A$  is  $R(A) = \bigcup_{x \in H} Ax$ . If for any  $x \in H$  the set  $Ax$  contains more than one element, we say that  $A$  is multivalued. The graph of  $A$  is the set  $G(A) = \{(x, y) \in D(A) \times H ; y \in Ax\}$ .

**Definition 3.3.1 (monotone operator)**  $A : H \rightarrow \mathcal{P}(H)$  is said monotone if and only if:

$$\forall x_1, x_2 \in D(A) : \langle Ax_1 - Ax_2, x_1 - x_2 \rangle_{H \times H} \geq 0$$

or  $\forall y_1 \in Ax_1, \forall y_2 \in Ax_2 : \langle y_1 - y_2, x_1 - x_2 \rangle_{H \times H} \geq 0$  if  $A$  is multivalued.

The set  $\mathcal{A}$  of monotone operators can be ordered by a graph inclusion. We will say that  $A_1 \leq A_2$  if and only if  $G(A_1) \subset G(A_2)$ , which is equivalent to  $A_1(x) \subset A_2(x) \forall x \in H$ . It can be shown that every totally ordered subset of  $\mathcal{A}$  has an upper bound. Then, thanks to the well-known Zorn Lemma [216],  $\mathcal{A}$  contains at least one maximal element, which is called a maximal monotone operator. Accordingly, a monotone operator  $A : H \rightarrow \mathcal{P}(H)$  is maximal monotone if and only if  $G(A) \subset G(B)$  implies  $A = B$ , where  $B : H \rightarrow \mathcal{P}(H)$  is an arbitrary monotone operator.

In practice, to show that a monotone operator is maximal, it is easier to use the following characterizing property:

**Proposition 3.3.2** Let  $A : H \rightarrow \mathcal{P}(H)$ ,  $A$  monotone, then  $A$  is maximal monotone if and only if:

- (i) The operator  $(A+I)$  is surjective, i.e.  $R(A+I) = H$  ( $I$  is the identity operator).

or

(ii)  $\forall \lambda > 0$   $(I + \lambda A)^{-1}$  is a contraction on the whole  $H$ .

**Example** Let  $\varphi : H \rightarrow ]-\infty, +\infty]$  convex, proper ( $\varphi \neq +\infty$ ). Then for any  $x \in H$ , the subdifferential of  $\varphi$  at  $x$  defined as:

$$\partial\varphi(x) = \{y \in H ; \forall \xi \in H, \varphi(\xi) \geq \varphi(x) + \langle y, \xi - x \rangle_{H \times H}\}$$

is monotone. It can be proved [50] that if  $\varphi$  is a lower semi-continuous proper convex function then  $\partial\varphi(x)$  is maximal monotone. ■

The main interest of this notion, in our context, is that it permits to solve certain nonlinear evolution PDEs:

**Proposition 3.3.3 [50]** *Let  $A : H \rightarrow \mathcal{P}(H)$ ,  $A$  maximal monotone, and  $u_0 \in D(A)$ , then there exists a unique function  $u(t) : [0, +\infty[ \rightarrow H$  such that:*

$$\begin{cases} 0 \in \frac{du}{dt} + Au(t) \\ u(0) = u_0. \end{cases} \quad (3.46)$$

So, if we want to solve equations like (3.46) we only have to check, according to Proposition 3.3.2, that  $(A + I)$  is surjective. Thus, the study of an evolution equation reduces to the study of a stationary one which represents a big advantage. Let us apply these results to (3.39). Our aim is to show that the divergence operator in (3.39):

$$Au = -\operatorname{div}(c(|\nabla u|^2) \nabla u)$$

is maximal monotone. As suggested before, a classical and convenient way is to identify  $A$  with the subdifferential of a convex l.s.c. functional. Let  $\Phi(t)$  be the function defined by:

$$\Phi(s) = \int_0^s \tau c(\tau^2) d\tau + 1.$$

We have  $\Phi(0) = 1$ ,  $\Phi'(s) = s c(s^2)$  and if  $c(s)$  satisfies (3.43) then  $\Phi(s)$  is strictly convex. Let us set:

$$J(u) = \begin{cases} \int_{\Omega} \Phi(|\nabla u(x)|) dx & \text{if } u \in W^{1,1}(\Omega) \\ +\infty & \text{if } u \in L^2(\Omega) - W^{1,1}(\Omega). \end{cases} \quad (3.47)$$

We can easily verify that  $Au = \operatorname{div} \left( \frac{\Phi'(|\nabla u|)}{|\nabla u|} \nabla u \right)$  identifies with the subdifferential of  $J$  at  $u$ ,  $\partial J(u)$ , with:

$$\begin{aligned} \operatorname{Dom}(A) = & \left\{ u \in W^{1,1}(\Omega), \operatorname{div} \left( \frac{\Phi'(|\nabla u|)}{|\nabla u|} \nabla u \right) \in L^2(\Omega), \right. \\ & \left. \frac{\Phi'(|\nabla u|)}{|\nabla u|} \frac{\partial u}{\partial N} = 0 \text{ on } \partial\Omega \right\}. \end{aligned}$$

Unfortunately  $J(u)$  is not lower semi-continuous on  $L^2(\Omega)$  and then  $A$  is not maximal monotone. To overcome this difficulty, like in the variational case, we introduce the relaxed functional:

$$\bar{J}(u) = \begin{cases} \int_{\Omega} \Phi(|\nabla u(x)|) dx + \alpha |D_s u| & \text{if } u \in BV(\Omega) \\ +\infty & \text{if } u \in L^2(\Omega) - BV(\Omega) \end{cases} \quad (3.48)$$

where  $\nabla u dx + D_s u$  is the Lebesgue decomposition of the measure  $Du$  and  $\alpha = \lim_{s \rightarrow +\infty} \frac{\Phi(s)}{s}$ .  $\bar{J}(u)$  is convex and l.s.c. on  $L^2(\Omega)$ . Then, we associate to  $\bar{J}(u)$  the evolution problem on  $L^2(\Omega)$ :

$$\begin{cases} 0 \in \frac{du}{dt} + \partial \bar{J}(u) & \text{on } ]0, +\infty[ \\ u(0, x) = u_0(x). \end{cases} \quad (3.49)$$

We can check that  $\partial \bar{J}$  is maximal monotone and from general results concerning evolution equations governed by maximal monotone operator, it is proved in [245] the following theorem:

**Theorem 3.3.1** [245] *Let  $\Omega$  be an open, bounded and connected subset of  $R^2$ , with Lipschitz boundary  $\Gamma = \partial\Omega$ . Let  $u_0 \in \operatorname{Dom}(\partial \bar{J}) \cap L^\infty(\Omega)$ , then there exists a unique function  $u(t) : [0, +\infty[ \rightarrow L^2(\Omega)$  such that*

$$u(t) \in \operatorname{Dom}(\partial \bar{J}), \forall t > 0, \frac{du}{dt} \in L^\infty((0, +\infty); L^2(\Omega)) \quad (3.50)$$

$$-\frac{du}{dt} \in \partial \bar{J}(u(t)), \text{ a.e. } t > 0, u(0) = u_0. \quad (3.51)$$

If  $\hat{u}$  is a solution with  $\hat{u}_0$  instead of  $u_0$ , then  $(3.52)$

$$|u(t) - \hat{u}(t)|_{L^2(\Omega)} \leq |u_0 - \hat{u}_0|_{L^2(\Omega)} \text{ for all } t \geq 0.$$

## Remarks

- For each  $t > 0$ , the map  $u_0 \rightarrow u(t)$  is a contraction of  $\operatorname{Dom}(\partial \bar{J})$  in  $\overline{\operatorname{Dom}(\partial \bar{J})}$ . We denote by  $S(t)$  its unique extension by continuity to  $\overline{\operatorname{Dom}(\partial \bar{J})} = \operatorname{Dom} \bar{J} = BV(\Omega)$ . If  $u_0$  is in  $BV(\Omega)$ , then  $u(t) = S(t)u_0$  is called the generalized solution of (3.39).

- We observe that if  $u(t, x)$  is a solution of (3.39) then a.e.  $t$ :

$$\int_{\Omega} u(t, x) \, dx = \int_{\Omega} u_0(x) \, dx.$$

To prove it, we differentiate  $I(t) = \int_{\Omega} u(t, x) \, dx$  and apply the Green's formula:

$$\begin{aligned} I'(t) &= \int_{\Omega} \frac{du}{dt}(t, x) \, dx = \int_{\Omega} \operatorname{div} \left( c(|\nabla u|^2(t, x)) \nabla u(t, x) \right) \, dx = \\ &= \int_{\Gamma} c(|\nabla u|^2(t, x)) \frac{du}{dN}(t, x) \, ds = 0. \end{aligned}$$

Thus  $I(t) = I(0) = \int_{\Omega} u_0(x) \, dx$ . ■

Although this theorem ensures the existence of the solution, it remains difficult to understand. In particular, if we look at results (3.50) or (3.51), one would like to know more about  $\partial\bar{J}$ . Let us characterize  $\partial\bar{J}(u)$  for  $u \in BV(\Omega)$ . Let us assume that  $\partial\bar{J}(u) \neq \emptyset$  and  $\xi \in L^2(\Omega)$  is such that  $\xi \in \partial\bar{J}(u)$ . By definition we have:

$$\bar{J}(u + sw) \geq \bar{J}(u) + s < \xi, w >_{L^2(\Omega) \times L^2(\Omega)} \quad \forall s, \forall w \in L^2(\Omega). \quad (3.53)$$

From (3.53), we can deduce some conditions by choosing successively functions  $w \in \mathcal{C}_0^\infty(\Omega)$ ,  $\mathcal{C}_0^\infty(\bar{\Omega})$  and  $BV(\Omega)$ . So we have the following result,

**Proposition 3.3.4** *If  $\partial\bar{J}(u) \neq \emptyset$ , then:*

(i)  $\partial\bar{J}(u)$  has only one element given by

$$\xi = \operatorname{div} \left( \frac{\Phi'(|\nabla u|)}{|\nabla u|} \nabla u \right) \in L^2(\Omega).$$

(ii)  $\frac{\Phi'(|\nabla u|)}{|\nabla u|} \frac{\partial u}{\partial N} = 0$  on  $\partial\Omega$ .

(iii) For all  $w \in BV(\Omega)$  with  $Dw = \nabla w dx + D_s w$  and  $D_s w = \rho D_s u + \mu'$ :

$$\int_{\Omega} \frac{\Phi'(|\nabla u|)}{|\nabla u|} \nabla u \cdot \nabla w \, dx + \int_{\Omega} \rho |D_s u| + \int_{\Omega} |\mu'| \geq - \int_{\Omega} \operatorname{div} \left( \frac{\Phi'(|\nabla u|)}{|\nabla u|} \nabla u \right) w \, dx.$$

**Proof** Starting from the definition (3.53), we look for necessary conditions on  $\xi$ . Several choices of  $w$  are made:

Step 1:  $w \in \mathcal{C}_0^\infty(\Omega)$

We have  $Du + sDw = (\nabla u + s\nabla w)dx + D_s u$  and (3.53) rewrites:

$$\begin{aligned} \frac{\bar{J}(u + sw) - \bar{J}(u)}{s} &= \int_{\Omega} \frac{\Phi(|\nabla u + s\nabla w|) - \Phi(|\nabla u|)}{s} dx \geq \\ &\geq \langle \xi, w \rangle_{L^2(\Omega) \times L^2(\Omega)} \quad \forall s, \forall w \in \mathcal{C}_0^\infty(\Omega). \end{aligned}$$

When  $s \rightarrow 0^+$ , we have after integrating by part:

$$-\int_{\Omega} \operatorname{div} \left( \frac{\Phi'(|\nabla u|)}{|\nabla u|} \nabla u \right) w dx \geq \langle \xi, w \rangle_{L^2(\Omega) \times L^2(\Omega)} \quad \forall w \in \mathcal{C}_0^\infty(\Omega).$$

By changing  $w$  into  $-w$ , we obtain:

$$\xi = -\operatorname{div} \left( \frac{\Phi'(|\nabla u|)}{|\nabla u|} \nabla u \right) \quad \text{in the distributional sense.} \quad (3.54)$$

Since  $\xi \in L^2(\Omega)$ , this equality is also true in  $L^2(\Omega)$ .

**Step 2:**  $w \in C^\infty(\bar{\Omega})$

We also have  $Du + sDw = (\nabla u + s\nabla w)dx + D_s u$ , and by (3.53):

$$\int_{\Omega} \frac{\Phi'(|\nabla u|)}{|\nabla u|} \nabla u \cdot \nabla w dx \geq \langle \xi, w \rangle_{L^2(\Omega) \times L^2(\Omega)} \quad \forall w \in C^\infty(\bar{\Omega}).$$

After integrating by part we have now:

$$\begin{aligned} -\int_{\Omega} \operatorname{div} \left( \frac{\Phi'(|\nabla u|)}{|\nabla u|} \nabla u \right) w dx + \int_{\partial\Omega} \frac{\Phi'(|\nabla u|)}{|\nabla u|} \nabla u \cdot n w ds &\geq \\ &\geq \langle \xi, w \rangle_{L^2(\Omega) \times L^2(\Omega)} \quad \forall w \in C^\infty(\bar{\Omega}). \end{aligned}$$

Notice that the term on  $\partial\Omega$  is well-defined since  $\sigma = \frac{\Phi'(|\nabla u|)}{|\nabla u|} \nabla u \in L^2(\Omega)^2$  and  $\operatorname{div}(\sigma) \in L^2(\Omega)$  thanks to (3.54) (see [159]). Thanks to (3.54), we easily deduce from the previous inequality the condition (ii).

**Step 3:**  $w \in BV(\Omega)$

Let  $D_s w = \rho D_s u + \mu'$  be the Lebesgue decomposition of the measure  $D_s w$  with respect to  $D_s u$ . So, we have

$$Du + sDw = (\nabla u + s\nabla w)dx + D_s u(1 + s\rho) + s\mu'.$$

Thus:

$$\begin{aligned} \frac{\bar{J}(u + sw) - \bar{J}(u)}{s} &= \\ &= \int_{\Omega} \frac{\Phi(|\nabla u + s\nabla w|) - \Phi(|\nabla u|)}{s} dx + \int_{\Omega} \frac{|1 + s\rho| - 1}{s} |D_s u| + \int_{\Omega} |\mu'|, \end{aligned}$$

and by letting  $s \rightarrow 0$ , with (3.53), we obtain the inequality (iii). ■

### Remarks

- It is useless to search for more characterization choosing  $w \in L^2(\Omega)$  (and not in  $BV(\Omega)$ ) since in this case  $\bar{J}(u + sw) = +\infty$ .
- In fact, since  $\Phi$  is convex, it is easy to show that if  $u \in BV(\Omega)$  verifies (i), (ii) and (iii), then  $\partial\bar{J}(u) \neq \emptyset$ . Therefore, (i), (ii) and (iii) are necessary and sufficient conditions ensuring that  $\partial\bar{J}(u) \neq \emptyset$ .
- The existence and uniqueness theorem from Brezis tells that  $-\frac{du}{dt} \in \partial\bar{J}(u(t))$ . We just showed that this means  $\frac{du}{dt} = \frac{\Phi'(|\nabla u|)}{|\nabla u|} \nabla u$  in  $L^2(\Omega)$ . ■

Let us now interpret the last inequality (iii) in a particular case which is representative in image analysis. Before that, we need the following lemma:

**Lemma 3.3.1** *Let  $\Omega$  be regular,  $\sigma \in C(\Omega)^M$  bounded,  $\operatorname{div}\sigma \in L^2(\Omega)$  and  $w \in BV(\Omega) \cap L^2(\Omega)$ . Then we have the following Stokes formula:*

$$\int_{\Omega} \operatorname{div} \sigma w \, dx = \int_{\partial\Omega} \sigma \cdot n w \, ds - \int_{\Omega} \sigma \cdot Dw$$

$$\text{where } \int_{\Omega} \sigma \cdot Dw = \int_{\Omega} \sigma \cdot \nabla w \, dx + \int_{\Omega} \sigma \cdot D_s w.$$

**Proof** By regularization, we can find a sequence  $w_n \in \mathcal{C}^\infty(\overline{\Omega})$  such that:

$$\begin{aligned} w_n &\xrightarrow[L^2(\Omega)]{} w \\ |Dw_n| &\longrightarrow |Dw|. \end{aligned}$$

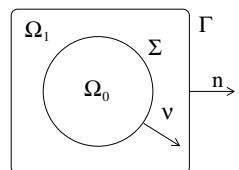
Since  $w_n$  is regular, we have

$$\int_{\Omega} \operatorname{div} \sigma w \, dx = \int_{\partial\Omega} \sigma \cdot n w_n \, ds - \int_{\Omega} \sigma \cdot \nabla w_n \, dx.$$

Moreover the convergence for the strong topology of  $BV(\Omega)$  induces the  $BV - w^*$  convergence and the convergence of the trace operator. The result is obtained as  $n$  tend to infinity. ■

As announced previsouly, let us explain the equality (iii) from Proposition 3.3.4 where we suppose that  $u$  admits a discontinuity along a single curve  $\Sigma$  and  $\Omega = \Omega_0 \cup \Sigma \cup \Omega_1$ . So  $u \in C^1(\Omega_0 \cup \Omega_1)$  and we have:

$$D_s u = [u]\nu \, d\mathcal{H}^1|_{\Sigma} = (u_1 - u_0)\nu \, d\mathcal{H}^1|_{\Sigma}.$$



We can still decompose  $D_s w$  in the following way:  $D_s w = \rho[u] d\mathcal{H}^1|_\Sigma + \mu'$ . Now, the idea is to look at the term:

$$A = \int_{\Omega} \operatorname{div} \left( \frac{\Phi'(|\nabla u|)}{|\nabla u|} \nabla u \right) w \, dx \equiv \int_{\Omega} \sigma w \, dx \quad \text{with } \sigma = \frac{\Phi'(|\nabla u|)}{|\nabla u|} \nabla u.$$

By decomposing  $A$  on  $\Omega_0$  and  $\Omega_1$  and by applying Lemma 3.3.1, we have:

$$\begin{aligned} A &= - \int_{\Omega_0} \sigma \cdot \nabla w \, dx - \int_{\Omega_0} \rho \sigma \cdot D_s u - \int_{\Omega_0} \sigma \mu' + \int_{\Sigma} \sigma \cdot \nu w \, ds \\ &\quad - \int_{\Omega_1} \sigma \cdot \nabla w \, dx - \int_{\Omega_1} \rho \sigma \cdot D_s u - \int_{\Omega_1} \sigma \mu' - \int_{\Sigma} \sigma \cdot \nu w \, ds + \int_{\partial\Omega} \underbrace{\sigma \cdot n}_{=0} w \, ds \\ &= - \int_{\Omega} \sigma \cdot \nabla w \, dx - \int_{\Omega} \rho \sigma \cdot D_s u - \int_{\Omega} \sigma \mu'. \end{aligned}$$

By replacing this expression in (iii) with  $D_s u = [u]\nu d\mathcal{H}^1|_\Sigma$ , we obtain after some computation:

$$\int_{\Sigma} \rho \left\{ |[u]| - \frac{\Phi'(|\nabla u|)}{|\nabla u|} \nabla u \cdot \nu[u] \right\} \, ds + \int_{\Omega} \left\{ |\mu'| - \frac{\Phi'(|\nabla u|)}{|\nabla u|} \nabla u \mu' \right\} \, dx \geq 0.$$

The second integral is always positive (if  $|\Phi'| \leq 1$ ) but  $\rho$  is arbitrary. So we have:

$$|[u]| - \frac{\Phi'(|\nabla u|)}{|\nabla u|} \nabla u \cdot \nu[u] = 0 \quad \text{a.e. on } \Sigma,$$

that we can rewrite (if  $[u] \neq 0$ ):

$$\operatorname{sign}[u] = \frac{\Phi'(|\nabla u|)}{|\nabla u|} \nabla u \cdot \nu \quad \text{a.e. on } \Sigma. \quad (3.55)$$

This condition is an interpretation of (iii). Naturally, this could be generalized to the case where  $u$  admits discontinuities on a finite or countable set of curves. It would be interesting to see if this condition could be used numerically.

### The Alvarez-Guichard-Lions-Morel scale space theory

In this section we examine the remarkable work of Alvarez et al [4]. In this paper the connection between scale space analysis and PDEs is rigorously established. Starting from a very natural filtering axiomatic (based on desired image properties) they prove that the resulting filtered image must necessarily be the viscosity solution of a PDE. In addition, they completely describe these PDEs. Most of their results have been collected in a very recent monograph by F. Guichard and J.M. Morel [130]. It is not our intention to set out all the details and refinements of their axiomatic. Our

purpose is to review some aspects of this very nice theory. We follow the presentation given in the original Alvarez et al paper [4, 5].

As said previously, we define a multi-scale analysis as a family of operators  $\{T_t\}_{t \geq 0}$  which, applied to the original image  $u_0(x)$ , yield to a sequence of images  $u(t, x) = (T_t u_0)(x)$ . We are going to list below a series of axioms to be verified by  $\{T_t\}_{t \geq 0}$ . These formal properties are very natural from an image analysis point of view. For simplicity, we suppose for all  $t \geq 0$ ,  $T_t : C_b^\infty(R^2) \rightarrow C_b(R^2)$  where  $C_b^\infty(R^2)$  (respectively  $C_b(R^2)$ ) is the space of bounded functions having derivatives at any order (respectively bounded continuous functions). This is not restrictive since, as usual in mathematical analysis, once properties are proved for regular functions we can extend them to non regular functions by density arguments.

List of axioms and invariance properties ( $X$  denotes the space  $C_b^\infty(R^2)$ )

(A1) *Recursivity:*

$$T_0(u) = u, \quad T_s \circ T_t(u) = T_{s+t}(u) \text{ for all } s, t \geq 0 \text{ and all } u \in X.$$

(A2) *Regularity:*

$$|T_t(u + h v) - (T_t(u) + h v)|_{L^\infty} \leq c h t \text{ for all } h \text{ and } t \text{ in } [0, 1] \text{ and all } u, v \in X.$$

(A3) *Locality:*

$$(T_t(u) - T_t(v))(x) = o(t), \quad t \rightarrow 0^+ \text{ for all } u \text{ and } v \in X \text{ such that } \nabla^\alpha u(x) = \nabla^\alpha v(x) \text{ for all } |\alpha| \geq 0 \text{ and all } x \text{ (\nabla}^\alpha u \text{ stands for the derivative of order } \alpha).$$

(A4) *Comparison principle:*

$$T_t(u) \leq T_t(v) \text{ on } R^2, \text{ for all } t \geq 0 \text{ and } u, v \in X \text{ such that } u \leq v \text{ on } R^2.$$

(I1) *Gray level shift invariance:*

$$T_t(0) = 0, \quad T_t(u + c) = T_t(u) + c \text{ for all } u \text{ in } X \text{ and all constant } c.$$

(I2) *Translation invariance:*

$$T_t(\tau_h \cdot u) = \tau_h \cdot (T_t u) \text{ for all } h \text{ in } R^2, \quad t \geq 0, \text{ where } (\tau_h \cdot u)(x) = u(x + h).$$

We emphasize that these axioms and invariance properties are quite natural from an image analysis point of view. A1 means that a coarser analysis of the original image can be deduced from a finer one without any dependence upon the original picture. A2 states a continuity assumption of  $T_t$ . A3 means that  $(T_t u)(x)$  is determined by the behaviour of  $u$  near  $x$ . A4 expresses the idea that if an image  $v$  is brighter than another image  $u$ , this ordering is preserved across scale. Finally, I1 and I2 state respectively that no *a priori* assumption is made on the range of the brightness and that all points are equivalent.

We are now in position to give the main result. We denote by  $S^2$  the space of all  $2 \times 2$  symmetric matrices endowed with its natural ordering.

**Theorem 3.3.2 [5]** *Under assumptions A1, A2, A3, A4, I1 and I2:*

- (i) *There exists a continuous function  $F : R^2 \times S^2 \rightarrow R$  satisfying  $F(p, A) \geq F(p, B)$  for all  $p \in R^2$ ,  $A$  and  $B$  in  $S^2$  with  $A \geq B$  such that:*

$$\delta_t(u) = \frac{T_t(u) - u}{t} \rightarrow F(\nabla u, \nabla^2 u), \quad t \rightarrow 0^+$$

*uniformly for  $x \in R^2$ , uniformly for  $u \in X$ .*

- (ii) *If  $u_0 \in C_b(R^2)$ , then  $u(t, x) = (T_t u_0)(x)$  is the unique viscosity solution of:*

$$\begin{cases} \frac{\partial u}{\partial t} = F(\nabla u, \nabla^2 u) \\ u(0, x) = u_0(x) \end{cases}$$

*and  $u(t, x)$  is bounded, uniformly continuous on  $R^2$ .*

**Proof** We only give the main steps, omitting sometimes technical details and referring for the original proofs to [4, 5].

**Step 1:** Existence of an infinitesimal generator

There exists an operator  $S : X \rightarrow C_b(R^2)$  such that

$$\delta_t(u) = \frac{T_t(u) - u}{t} \rightarrow S[u], \quad t \rightarrow 0^+$$

uniformly on  $R^2$  and for all  $u \in X$ . See [4, 5] for this very technical proof.

**Step 2:** A general lemma

Let  $X, Y, Z$  be three sets,  $A : X \rightarrow Y$  and  $G : X \rightarrow Z$ . Let us suppose that the equality  $G(x) = G(x')$  implies  $A(x) = A(x')$ , then there exists  $F : G(X) \subset Z \rightarrow Y$  such that for all  $x$  in  $X$ ,  $A(x) = F(G(x))$ , i.e.  $A$  is only a function of  $G$ .

**Step 3:** Let us show that if  $u$  and  $v$  satisfy

$$\begin{aligned} \{u(0) = v(0); \nabla u(0) = \nabla v(0) = p \in R^2; \nabla^2 u(0) = \nabla^2 v(0) = A \in S^2\}, \\ \text{then } \{S[u](0) = S[v](0)\}. \end{aligned}$$

Let  $z(x)$  a function in  $C_b^\infty(R^2)$  such that  $z(x) \geq 0$  and  $z(x) = |x|^2$  for  $x$  near 0. Set  $u^\varepsilon(x) = u(x) + \varepsilon z(x)$ . We claim that  $u^\varepsilon(x) \geq v(x)$  for  $|x|$  small enough. Indeed, thanks to the Taylor formula:

$$u^\varepsilon(x) = u(0) + x \cdot \nabla u(0) + \frac{1}{2} \nabla^2 u(0) x \cdot x + o(|x|^2) + \varepsilon z(x).$$

But, since  $u, v \in C_b^\infty(R^2)$  and  $z(x) = |x|^2$  in a neighborhood of zero, it is clear that there exists a constant  $c > 0$  so that  $o(|x|^2) + \varepsilon z(x) \geq 0$  for  $|x| \leq c\varepsilon$ . Therefore  $u^\varepsilon(x) \geq v(x)$  for  $|x| \leq c\varepsilon$ . Then, let us set  $w^\varepsilon(x) = w(x/\varepsilon)$  where  $w \in C_b^\infty(R^2)$ ,  $0 \leq w \leq 1$ ,  $w(x) = 1$  if  $|x| \leq c/2$  and  $w(x) = 0$  if  $|x| \geq c$ . Finally, let us define  $\bar{u}^\varepsilon(x) = w^\varepsilon(x) u^\varepsilon(x) + (1 - w^\varepsilon(x)) v(x)$ . The function  $\bar{u}^\varepsilon(x)$  verifies:

$$\begin{aligned}\nabla^\alpha \bar{u}^\varepsilon(0) &= \nabla^\alpha u^\varepsilon(0), \quad \forall \alpha \\ \bar{u}^\varepsilon(x) &\geq v(x) \quad \text{on } R^2.\end{aligned}$$

Thus, from A4, we get  $T_t(\bar{u}^\varepsilon) \geq T_t(v)$  on  $R^2$  and (since  $\bar{u}^\varepsilon(0) = u^\varepsilon(0) = u(0) = v(0) = 0$ ):

$$\frac{T_t(\bar{u}^\varepsilon)(0) - \bar{u}^\varepsilon(0)}{t} \geq \frac{T_t(v)(0) - v(0)}{t}.$$

If  $t \rightarrow 0^+$ , we get  $S[\bar{u}^\varepsilon](0) \geq S[v](0)$ . But from A3, we have:  $(T_t \bar{u}^\varepsilon)(0) - (T_t u^\varepsilon)(0) = o(t)$ , which implies  $S[u^\varepsilon](0) \geq S[v](0)$ . Now, letting  $\varepsilon \rightarrow 0$  and using the uniform convergence in Step 1, we deduce  $S[u](0) \geq S[v](0)$  and by symmetry (changing  $z(x)$  into  $-z(x)$ ) we finally obtain:

$$S[u](0) = S[v](0).$$

**Step 4:** Let us show the same property as in step 3 for any  $x_0 \in R^2$

Let  $u$  and  $v$  satisfying:

$$\{u(x_0) = v(x_0); \nabla u(x_0) = \nabla v(x_0) = p \in R^2; \nabla^2 u(x_0) = \nabla^2 v(x_0) = A \in S^2\}.$$

We have to prove  $S[u](x_0) = S[v](x_0)$ . Without a loss of generality, we can suppose that  $u(x_0) = v(x_0) = 0$ . Let us define:

$$\begin{aligned}u_{x_0}(x) &= (\tau_{x_0} \cdot u)(x) = u(x_0 + x) \\ v_{x_0}(x) &= (\tau_{x_0} \cdot v)(x) = v(x_0 + x).\end{aligned}$$

We have:  $u_{x_0}(0) = v_{x_0}(0) = 0$ ,  $\nabla u_{x_0}(0) = \nabla v_{x_0}(0)$ ,  $\nabla^2 u_{x_0}(0) = \nabla^2 v_{x_0}(0)$ .

Step 3 implies  $S[u_{x_0}](0) = S[v_{x_0}](0)$ . But, from I2:

$$\begin{aligned}S[u_{x_0}](0) &= S[\tau_{x_0} \cdot u](0) = \lim_{t \rightarrow 0^+} \frac{T_t(\tau_{x_0} \cdot u)(0) - (\tau_{x_0} \cdot u)(0)}{t} = \\ &= \lim_{\substack{\text{from I2} \\ t \rightarrow 0^+}} \frac{\tau_{x_0} \cdot (T_t u - u)(0)}{t} = (\tau_{x_0} \cdot S[u])(0) = S[u](x_0)\end{aligned}$$

and the same stands for  $S[v](x_0)$ . Therefore:

$$S[u](x_0) = S[v](x_0).$$

To summarize, we have for all  $x \in R^2$ , the following property:

if  $\{(u(x), \nabla u(x), \nabla^2 u(x)) = (v(x), \nabla v(x), \nabla^2 v(x))\}$  then

$S[u](x) = S[v](x)$ . So we apply the general lemma of Step 2 with:

$$\begin{aligned}X &= \{u(x); u \in C_b^\infty(R^2)\}, \\ Y &= \{S[u](x); u \in C_b^\infty(R^2)\}, \\ Z &= \{(u(x), \nabla u(x), \nabla^2 u(x)); u \in C_b^\infty(R^2)\}\end{aligned}$$

and  $G : u(x) \rightarrow (u(x), \nabla u(x), \nabla^2 u(x))$  ( $G : X \rightarrow Z$ ). We have:

$$G(u(x)) = G(v(x)) \text{ implies } S[u](x) = S[v](x).$$

Therefore, there exists a function  $F : G(X) \subset Z \rightarrow Y$  such that

$$S[u](x) = F(x, u(x), \nabla u(x), \nabla^2 u(x)).$$

**Step 5:**  $F$  does not depend on  $x$  and  $u$

This a direct consequence of I1 and I2. From I2, we have for all  $h$

$$\tau_h.S[u](x) = S[\tau_h.u](x)$$

i.e. for all  $h$ :

$$\begin{aligned} F(x+h, u(x+h), \nabla u(x+h), \nabla^2 u(x+h)) &= \\ &= F(x, u(x+h), \nabla u(x+h), \nabla^2 u(x+h)) \end{aligned}$$

thus  $F$  does not depend on  $x$ . Now, from I1, if  $c$  is any constant:

$$S[u+c] = \lim_{t \rightarrow 0^+} \frac{T_t(u+c) - (u+c)}{t} = \lim_{t \rightarrow 0^+} \frac{T_t(u) - u}{t} = S[u]$$

i.e. for  $c$ :

$$F(u(x) + c, \nabla u(x), \nabla^2 u(x)) = F(u(x), \nabla u(x), \nabla^2 u(x))$$

thus  $F$  does not depend on  $u$ .

**Step 6:**  $F$  is continuous and nondecreasing with respect to its second argument.

The continuity of  $F$  follows from Step 1. Let us show that for all  $p \in R^2$ ,  $A, B \in S^2$  such that  $A \geq B$ , then  $F(p, A) \geq F(p, B)$ . Let us define

$$u(x) = \left( p \cdot x + \frac{1}{2} A x \cdot x \right) w(x) \quad \text{and} \quad v(x) = \left( p \cdot x + \frac{1}{2} B x \cdot x \right) w(x)$$

where  $w(x)$  is the function defined in Step 3. We have:

$$u(0) = v(0) = 0, \quad \nabla u(0) = \nabla v(0) = p, \quad \nabla^2 u(0) = A, \quad \nabla^2 v(0) = B.$$

Moreover

$$A - B \geq 0 \Rightarrow u(x) \geq v(x) \Rightarrow (T_t u)(x) \geq (T_t v)(x) \quad (\text{from A4})$$

and the last inequality at  $x = 0$  implies:

$$S[u](0) = F(p, A) \geq S[v](0) = F(p, B).$$

**Step 7:**  $u(t, x) = (T_t u_0)(x)$  is the unique viscosity solution of:

$$\begin{cases} \frac{\partial u}{\partial t} = F(\nabla u, \nabla^2 u) \\ u(0, x) = u_0(x). \end{cases} \quad (3.56)$$

Let us first prove that  $u(t, x)$  is a sub-solution of (3.56). Let  $\phi(t, x)$  be a test function and  $(x_0, t_0)$  be a global maximum of  $(u - \phi)(t, x)$ . Without a loss of generality, we may suppose:

$$u(x_0, t_0) - \phi(x_0, t_0) = \max_{(t, x)} (u - \phi)(t, x) = 0$$

and that  $\phi(t, x)$  is of the form:  $\phi(t, x) = f(x) + g(t)$  with  $g(t_0) = 0$ . These two simplifications imply:

$$\begin{aligned} u(t, x) &\leq \phi(t, x) \quad \text{for all } (t, x) \\ u(x_0, t_0) &= \phi(x_0, t_0) = f(x_0). \end{aligned}$$

Now, let  $h$  in  $]0, t_0[$ . From the recursivity axiom A1, we get:

$$T_h(u(x_0, t_0 - h)) = (T_h \circ T_{t_0-h})(u_0(x_0)) = T_{t_0}(u_0(x_0)) = u(x_0, t_0) = f(x_0)$$

but since  $u \leq \phi$ :

$$\begin{aligned} f(x_0) &= u(x_0, t_0) = T_h(u(x_0, t_0 - h)) \leq \\ &\leq T_h(\phi(x_0, t_0 - h)) = T_h(f(x_0) + g(t_0 - h)) = T_h(f(x_0)) + g(t_0 - h). \end{aligned}$$

Thus, since  $g(t_0) = 0$

$$\frac{1}{h}(g(t_0) - g(t_0 - h)) + \frac{1}{h}(f(x_0) - T_h(f(x_0))) \leq 0$$

and if  $h \rightarrow 0^+$

$$g'(t_0) - F(\nabla f(x_0), \nabla^2 f(x_0)) \leq 0 \quad \text{i.e.} \quad \frac{\partial \phi}{\partial t}(x_0, t_0) - F(\nabla f(x_0), \nabla^2 f(x_0)) \leq 0$$

which means that  $u(t, x)$  is a sub-solution of (3.56). It can be proved using similar arguments that  $u(t, x)$  is a super-solution. So  $u(t, x)$  is a viscosity solution of (3.56).

This last step concludes the proof of Theorem 3.3.2. ■

If the multi-scale analysis satisfies additional invariance properties then the function  $F$  can be written in an explicit form. We state below two important cases.

**Theorem 3.3.3** [4, 5] *Let us suppose that  $T_t$  satisfy the assumptions of Theorem 3.3.2 and*

(I3) *Isometry invariance:*

$T_t(R.u)(x) = R.(T_t u)(x)$  for all orthogonal transformation  $R$  on  $R^2$ , where  $(R.u)(x) = u(Rx)$ .

We also assume that  $u \rightarrow T_t u$  is linear. Then  $u(t, x) = (T_t u_0)(x)$  is the solution of the heat equation  $\frac{\partial u}{\partial t} = c \Delta u$ ,  $u(0, x) = u_0(x)$ , where  $c$  is a positive constant.

**Theorem 3.3.4** [4, 5] *Let us suppose that  $T_t$  satisfy the assumptions of Theorem 3.3.2 and*

(I4) *Gray scale invariance:*

$$T_t(\varphi(u)) = \varphi(T_t(u)) \text{ for all nondecreasing real function } \varphi.$$

(I5) *Scale invariance:*

$\forall \lambda, t > 0$ , there exists  $t'(t, \lambda) > 0$  such that  $H_\lambda.(T_t u) = T_t(H_\lambda.u)$ , where  $(H_\lambda.u)(x) = u(\lambda x)$ . Moreover, we suppose that  $t'(t, \lambda)$  is differentiable with respect to  $\lambda$  at  $\lambda = 1$  and that the function  $g(t) = \frac{\partial t'}{\partial \lambda}(t, 1)$  is continuous and positive for  $t > 0$ .

(I6) *Projection invariance:*

For all  $A : R^2 \rightarrow R^2$  linear, for all  $t > 0$ , there exists  $t'(t, A) > 0$  such that  $A.(T_t u) = T_t(A.u)$ .

Then  $u(t, x) = (T_t u_0)(x)$  is the solution of:

$$\begin{cases} \frac{\partial u}{\partial t} = |\nabla u| (t \operatorname{curv} u)^{1/3} \\ u(0, x) = u_0(x) \end{cases}$$

$$\text{where } \operatorname{curv} u = \frac{u_{xx} u_y^2 + u_{yy} u_x^2 - 2 u_{xy} u_x u_y}{|\nabla u|^3}.$$

► These two last theorems are very interesting since they express that the Alvarez et al theory is a very natural extension of the linear theory (Theorem 3.3.3) but also because the multi-scale axiomatic leads to new nonlinear filters (Theorem 3.3.4).

## Remarks

- If in Theorem 3.3.4 we suppose that  $T_t$  satisfy I4, I5 and I3 instead of I6, we only get the PDE:

$$\frac{\partial u}{\partial t} = |\nabla u| \beta (t \operatorname{curv} u)$$

where  $\beta$  is a continuous non-decreasing function.

- The previous scale-space theory can be extended to the analysis of movies [5, 128, 129, 181, 182, 180].
- There are strong connections between PDEs described in this section and morphological operators (i.e. monotone, translation and contrast invariant operators). In fact, let  $\mathcal{F}$  be a set of functions containing continuous functions and characteristic functions of level sets of elements of  $\mathcal{F}$ , then it can be proven [174] that any morphological operator on  $\mathcal{F}$  is of the form

$$(Tu)(x) = \inf_{B \in \mathcal{B}} \sup_{y \in B} u(x + y)$$

where  $\mathcal{B}$  is a family of structuring elements. As a very interesting result, one can prove that if we adequately scale morphological op-

erators, then by iterating the resulting operators we retrieve all the equations given in this section. The interested reader can consult [55, 130] and the references therein. ■

Weickert's approach

In order to take into account local variations of the gradient orientation, we need to define a more general descriptor than the magnitude of the gradient only. Let us start with simple remarks.

As seen in previous sections, it is a natural idea to say that the preferred smoothing direction is the one that minimizes gray value fluctuations. Let  $d(\theta)$  be the vector  $(\cos \theta, \sin \theta)$ . An elementary calculation shows that the function  $F(\theta) = (d(\theta) \cdot \nabla u(x))^2$  is maximal if  $d$  is parallel to  $\nabla u$  and is minimal if  $d$  is orthogonal to  $\nabla u$ . We can also remark that maximizing (respectively minimizing)  $F(\theta)$  is equivalent to maximize (respectively minimize) the quadratic form  $d^t \nabla u \nabla u^t d$ . The matrix

$$\nabla u \nabla u^t = \begin{pmatrix} u_{x_1}^2 & u_{x_1} u_{x_2} \\ u_{x_1} u_{x_2} & u_{x_2}^2 \end{pmatrix} \quad (3.57)$$

is positive semidefinite, its eigenvalues are  $\lambda_1 = |\nabla u|^2$  and  $\lambda_2 = 0$  and there exists an orthonormal basis of eigenvectors  $v_1$  parallel to  $\nabla u$  and  $v_2$  orthogonal to  $\nabla u$ .

So, it would be tempting to define at  $x$  an orientation descriptor as a function of  $\nabla u \nabla u^t(x)$ . But, by proceeding like this, we do not take into account possible information contained in a neighborhood of  $x$ . To this end, the idea proposed by Weickert is to introduce smoothing kernels at different scales. We only sketch the main ideas since Weickert himself has written a monograph [249] based on his work.

To avoid false detections due to the noise,  $u(x)$  is first convolved with a Gaussian kernel  $k_\sigma$ :  $u_\sigma(x) = (k_\sigma * u)(x)$ . The local information is averaged by convolving componentwise  $\nabla u_\sigma \nabla u_\sigma^t$  with a Gaussian kernel  $k_\rho$ . The result is a symmetric, positive semidefinite matrix

$$J_\rho(\nabla u_\sigma) = k_\rho * \nabla u_\sigma \nabla u_\sigma^t. \quad (3.58)$$

The matrix  $J_\rho(\nabla u_\sigma)$  has orthonormal eigenvectors  $v_1, v_2$  with

$$v_1 \text{ parallel to } \begin{pmatrix} 2j_{12} \\ j_{22} - j_{11} + \sqrt{(j_{22} - j_{11})^2 + 4j_{12}^2} \end{pmatrix}$$

where  $j_{lk}$  are the elements of the matrix  $J_\rho(\nabla u_\sigma)$ . The corresponding eigenvalues are given by

$$\mu_1 = \frac{1}{2} \left[ j_{11} + j_{22} + \sqrt{(j_{11} - j_{22})^2 + 4j_{12}^2} \right]$$

and

$$\mu_2 = \frac{1}{2} \left[ j_{11} + j_{22} - \sqrt{(j_{11} - j_{22})^2 + 4j_{12}^2} \right].$$

They describe average contrast in the eigendirections within a neighborhood of size  $O(\rho)$ . The noise parameter  $\sigma$  makes the descriptor insensitive to details of scale smaller than  $O(\sigma)$ . The vector  $v_1$  indicates the orientation maximizing the gray value fluctuations while  $v_2$  gives the preferred local direction of smoothing. The eigenvalues  $\mu_1$  and  $\mu_2$  convey shape information. Isotropic structures are characterized by  $\mu_1 \cong \mu_2$ , line like structure by  $\mu_1 \gg \mu_2 \approx 0$ , corners by  $\mu_1 \geq \mu_2 \gg 0$ .

Now, the nonlinear diffusion process is governed by a parabolic equation that can be viewed as an extension of (3.39):

$$\begin{cases} \frac{\partial u}{\partial t} = \operatorname{div}(D(J_\rho(\nabla u_\rho))\nabla u) & \text{in } \Omega \times ]0, T[ \\ u(0, x) = u_0(x) & \text{on } \Omega \\ \langle D(J_\rho(\nabla u_\rho))\nabla u, N \rangle = 0 & \text{on } \partial\Omega \times ]0, T[ \end{cases} \quad (3.59)$$

where  $D$  is an operator to be precised next and  $N$  is the unit outward normal to  $\partial\Omega$ . Notice the boundary condition which is the natural condition associated to the divergence operator<sup>1</sup>. We have the following result:

**Theorem 3.3.5 [249]** *Let us assume that:*

- (i) *The diffusion tensor  $D = (d_{ij})$  belongs to  $C^\infty(S^2, S^2)$  where  $S^2$  denotes the set of symmetric matrices.*
- (ii) *Uniform positive definiteness: for all  $w \in L^2(\Omega, R^2)$  with  $|w(x)| \leq k$  on  $\bar{\Omega}$ , there exists a positive lower bound  $\nu(k)$  for the eigenvalues of  $D(J_\rho(w))$ .*

*Then for all  $u_0 \in L^\infty(\Omega)$  equation (3.59) has a unique solution  $u(t, x)$  satisfying*

$$\begin{aligned} u &\in C([0, T]; L^2(\Omega)) \cap L^2([0, T]; W^{1,2}(\Omega)) \\ \frac{\partial u}{\partial t} &\in L^2((0, T); W^{1,2}(\Omega)) \end{aligned}$$

*Moreover,  $u \in C^\infty(\bar{\Omega} \times ]0, T[)$ . This solution depends continuously on  $u_0$  with respect to the  $L^2$ -norm and it fulfils the extremum principle:*

$$\inf_{\Omega} u_0(x) \leq u(t, x) \leq \sup_{\Omega} u_0(x).$$

Related results have been proved for semi-discrete and fully discrete versions of the model. For the proofs as well as further properties (invariances, image simplification properties, behaviour as  $t$  tend to infinity, we refer to Weickert [249].

Let us now describe two possibilities of how to choose the diffusion tensor  $D(J_\rho)$ . Since the eigenvectors of  $D$  should reflect the local image structure,

---

<sup>1</sup>This can be compared with the boundary condition which was associated to the divergence operator in (3.29)

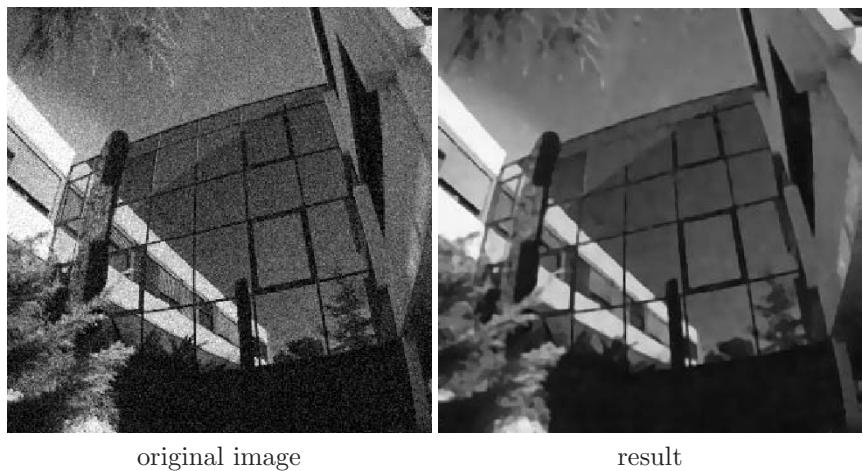


Figure 3.13. Example of Weickert's edge-enhancing (3.60) approach applied on the noisy "Borel building" image. It combines isotropic smoothing within flat regions with diffusion along edges. Diffusion across edges is reduced.

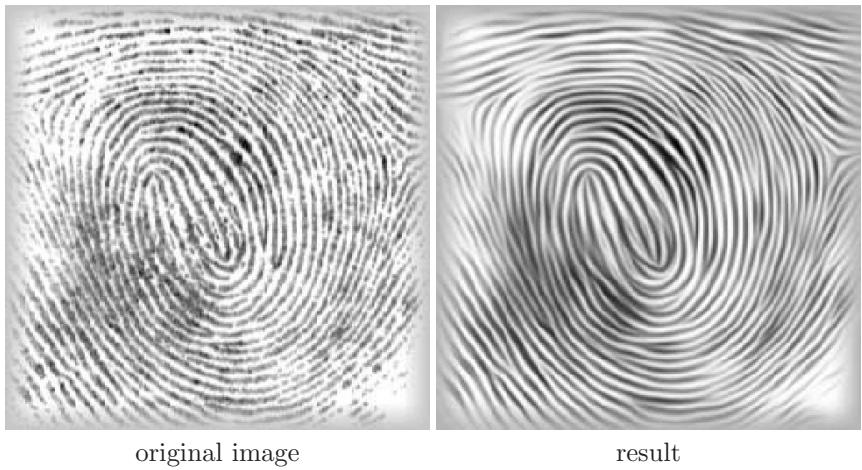


Figure 3.14. Example of Weickert's coherence-enhancing approach (3.61), from [250]. Interrupted lines are closed and the semantically important singularities are not destroyed. A typical application is for fingerprint enhancement where structure is especially important. The right-hand side image presents the result.

one should choose the same orthonormal basis of eigenvectors as one gets from  $J_\rho$ . The choice of the corresponding eigenvalues  $\lambda_1$  and  $\lambda_2$  of  $D$  depends on the desired goal:

- **Edge-enhancing anisotropic diffusion** [248]. If one wants to smooth preferably within each region and aims to preserve edges, then one should reduce the diffusivity  $\lambda_1$  perpendicular to edges all the more as the contrast  $\mu_1$  is large. This behaviour may be accomplished by the following choice:

$$\begin{aligned}\lambda_1 &= \begin{cases} 1 & \text{if } \mu_1 = 0, \\ 1 - \exp\left(\frac{-3.315}{\mu_1^4}\right) & \text{otherwise,} \end{cases} \\ \lambda_2 &= 1.\end{aligned}\quad (3.60)$$

Figure 3.13 illustrates such a process.

- **Coherence-enhancing anisotropic diffusion** [250]. If one wants to enhance flow-like structures and close interrupted lines, one should smooth preferably along the coherence direction  $v_2$  with a diffusivity  $\lambda_2$  which increases with respect to the coherence  $(\mu_1 - \mu_2)^2$ . This may be achieved by the following choice of the eigenvalues of  $D(J_\rho)$ :

$$\begin{aligned}\lambda_1 &= \alpha, \\ \lambda_2 &= \begin{cases} \alpha & \text{if } \mu_1 = \mu_2, \\ \alpha + (1-\alpha) \exp\left(\frac{-1}{(\mu_1 - \mu_2)^2}\right) & \text{otherwise,} \end{cases}\end{aligned}\quad (3.61)$$

where the small positive parameter  $\alpha \in (0, 1)$  keeps the diffusion tensor uniformly positive definite. Figure 3.14 shows the restoration properties of this diffusion filter as applied to a degraded fingerprint image.

### Surface based approaches

In [233, 232] Sochen, Kimmel, and Malladi introduced the concept of images as embedded maps and minimal surfaces, and applied it to processing movies, color images, texture, and volumetric medical images (see [145]). According to their geometrical framework for image processing, intensity images are considered as surfaces in the spatial-feature space. The image is thereby a two dimensional surface in three dimensional space (see Figure 3.15).

Let us briefly explain the main ideas for gray level images. As mentioned, an image is not considered as a function  $u(x)$  from a domain  $\Omega$  into  $R$ , but as an embedded surface  $\mathcal{M}$  in  $R^3$  defined by:

$$\begin{aligned}(\sigma_1, \sigma_2) \rightarrow X(\sigma_1, \sigma_2) &= (X_1(\sigma_1, \sigma_2), X_2(\sigma_1, \sigma_2), X_3(\sigma_1, \sigma_2)) \\ \Sigma \rightarrow \mathcal{M}\end{aligned}$$

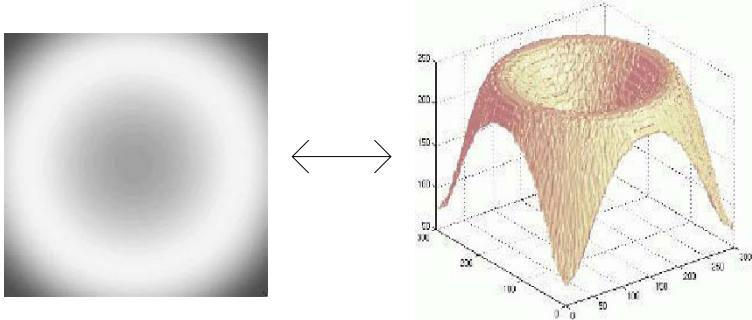


Figure 3.15. Interpretation of an image as a surface. The height is equal to the gray level value

where  $(\sigma_1, \sigma_2)$  denote the local coordinates of the surface<sup>2</sup>. The main point is that  $\Sigma$  and  $\mathcal{M}$  are viewed as Riemannian manifolds equipped with suitable (Riemannian) metrics. To better understand, let us consider the following example corresponding to a particular choice of  $\Sigma$  and  $X$ :

$$\begin{aligned}\Sigma &= \Omega, \text{ the image domain,} \\ \sigma_1 &= x_1, \quad \sigma_2 = x_2 \text{ the classical cartesian coordinates,} \\ X(x) &= (x, u(x)) \text{ where } u(x) \text{ is the gray level intensity.}\end{aligned}$$

The metric on  $\Sigma$  is the usual one,  $dx_1^2 + dx_2^2$ , and the induced metric on  $\mathcal{M}$  is  $ds^2 = dx_1^2 + dx_2^2 + du^2$ , that is, from an elementary calculus:

$$\begin{aligned}ds^2 &= dx_1^2 + dx_2^2 + (u_{x_1}dx_1 + u_{x_2}dx_2)^2 \\ &= (1 + u_{x_1}^2)dx_1^2 + 2u_{x_1}u_{x_2}dx_1dx_2 + (1 + u_{x_2}^2)dx_2^2.\end{aligned}$$

This can be rewritten:

$$ds^2 = (dx_1, dx_2) \begin{pmatrix} 1 + u_{x_1}^2 & u_{x_1}u_{x_2} \\ u_{x_1}u_{x_2} & 1 + u_{x_2}^2 \end{pmatrix} \begin{pmatrix} dx_1 \\ dx_2 \end{pmatrix}$$

i.e. the metric is given by the symmetric definite positive matrix:

$$G = \begin{pmatrix} 1 + u_{x_1}^2 & u_{x_1}u_{x_2} \\ u_{x_1}u_{x_2} & 1 + u_{x_2}^2 \end{pmatrix}.$$

How these concepts can be useful from an image analysis point of view? We know that most images are noisy or deteriorated. To obtain a restored approximation of a degraded image, we search for  $\mathcal{M}$  having a minimal area. By this way, singularities are smoothed. So, if  $g$  denotes the determinant of  $G$ , we have to minimize with respect to  $u$  the integral:

$$\mathcal{S}(\mathcal{M}) = \iint_{\Omega} \sqrt{g} dx dy = \iint_{\Omega} \sqrt{1 + u_{x_1}^2 + u_{x_2}^2} dx.$$

---

<sup>2</sup> $\Sigma$  is called the image manifold and  $\mathcal{M}$  the space feature manifold

If a minimizer  $u(x)$  exists, necessarily it verifies the Euler-Lagrange equation:

$$\frac{\partial}{\partial x_1} \left( \frac{u_{x_1}}{\sqrt{1+u_{x_1}^2+u_{x_2}^2}} \right) + \frac{\partial}{\partial x_2} \left( \frac{u_{x_2}}{\sqrt{1+u_{x_1}^2+u_{x_2}^2}} \right) = 0,$$

i.e.

$$\frac{u_{x_1 x_1}(1+u_{x_2}^2) + u_{x_2 x_2}(1+u_{x_1}^2) - 2u_{x_1}u_{x_2}u_{x_1 x_2}}{(1+u_{x_1}^2+u_{x_2}^2)^{3/2}} = 0 \quad (3.62)$$

which is equivalent to say that the mean curvature  $H$  of  $\mathcal{M}$  is zero. Surfaces of zero mean curvature are known as minimal surfaces.

For computing numerically a solution, we embed equation (3.62) into a dynamical scheme:

$$\frac{dX}{dt}(t) = F \quad (3.63)$$

where  $F$  is an arbitrary flow field defined on  $\mathcal{M}(t)$ . If  $X(t)$  is of the form:

$$X(t) = (x_1, x_2, u(t, x_1, x_2))^T,$$

then we have

$$\frac{dX}{dt}(t) = \left( 0, 0, \frac{\partial u}{\partial t}(t, x_1, x_2) \right)^T.$$

Therefore, the motion is necessarily in the  $z$  direction. If we choose  $F = (0, 0, \alpha H)$ , we obtain the scalar equation:

$$\frac{\partial u}{\partial t} = \alpha H = \alpha \frac{u_{x_1 x_1}(1+u_{x_2}^2) + u_{x_2 x_2}(1+u_{x_1}^2) - 2u_{x_1}u_{x_2}u_{x_1 x_2}}{(1+u_{x_1}^2+u_{x_2}^2)^{3/2}} \quad (3.64)$$

The coefficient  $\alpha \in R$  can be interpreted as a weighting parameter. If  $\alpha = \frac{1}{\sqrt{1+u_{x_1}^2+u_{x_2}^2}}$  then (3.64) rewrites as

$$\frac{\partial u}{\partial t} = \frac{u_{x_1 x_1}(1+u_{x_2}^2) + u_{x_2 x_2}(1+u_{x_1}^2) - 2u_{x_1}u_{x_2}u_{x_1 x_2}}{(1+u_{x_1}^2+u_{x_2}^2)^2}. \quad (3.65)$$

The right-hand side term in (3.65) is known as the Laplace Beltrami operator and the equation (3.65) can be viewed as the projection on the  $z$ -axis of the flow  $\frac{dX}{dt} = H N$  where  $N = \frac{1}{\sqrt{1+u_{x_1}^2+u_{x_2}^2}}(u_{x_2}, -u_{x_1}, 1)^T$  is the unit normal to  $\mathcal{M}(t)$  (see Figure 3.16).

The quantity  $\frac{1}{\sqrt{1+u_{x_1}^2+u_{x_2}^2}} = \frac{1}{\sqrt{g}}$  has in fact a remarkable interpretation. Let us consider the ratio:

$$r = \frac{A^{\text{domain}}}{A^{\text{surface}}}$$

where  $A^{\text{domain}}$  is the area of an infinitesimal surface in the image domain  $(x_1, x_2)$  and  $A^{\text{surface}}$  is the corresponding area on the surface  $\mathcal{M}$  (see also Figure 3.16).

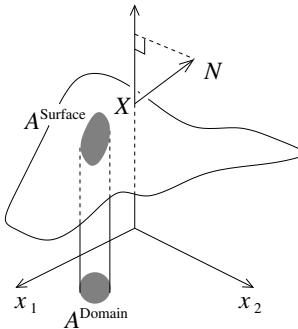


Figure 3.16. Representation of the normal  $N$ ,  $A^{\text{domain}}$  and  $A^{\text{surface}}$ .

This ratio can be interpreted as an indicator of the height variation on the surface.  $r$  is equal to 1 for flat surfaces and is close to 0 near edges. In fact  $r$  is related to the metric of the surface since:

$$r = \frac{dx_1 dx_2}{\sqrt{g} dx_1 dx_2} = \frac{1}{\sqrt{g}} = \frac{1}{\sqrt{1+u_{x_1}^2+u_{x_2}^2}}.$$

Hence, from a restoration point of view, it would be desirable to incorporate  $r$  in the model. For example, in (3.64), we can choose  $\alpha = r^\gamma$ , so the flow becomes:

$$\frac{\partial u}{\partial t} = r^{\gamma+3}(u_{x_1 x_1}(1+u_{x_2}^2) + u_{x_2 x_2}(1+u_{x_1}^2) - 2u_{x_1} u_{x_2} u_{x_1 x_2}). \quad (3.66)$$

By selecting different  $\gamma$  we recover some flows already proposed:

- For  $\gamma = -1$ , (3.66) is the mean curvature flow projected onto the normal [233, 232].
- For  $\gamma = 0$ , (3.66) is the flow proposed by [102].
- For  $\gamma = 1$ , (3.66) is the Laplace Beltrami flow. A result using this operator is shown in Figure 3.17.

Observing the result in Figure 3.17, we can remark that it is quite similar with the half-quadratic minimization and the total variation model de-

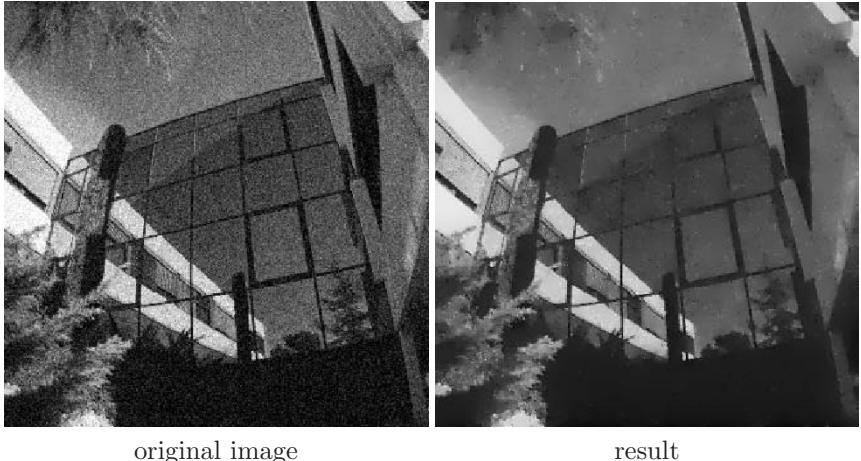


Figure 3.17. Example of the Laplace Beltrami equation on the “Borel building” image.

scribed in Section 3.2. In fact, as shown in [233, 232] the Laplace Beltrami equation has a direct relation with the total variation based regularization methods. If we choose on  $\Sigma$  the metric  $\varepsilon^2 dx_1^2 + \varepsilon^2 dx_2^2$  ( $\varepsilon > 0$ ) and on  $\mathcal{M}$  the metric  $ds^2 = \varepsilon^2 dx_1^2 + \varepsilon^2 dx_2^2 + du^2$ , then the associated symmetric definite positive  $G$  is given by

$$G = \begin{pmatrix} \varepsilon^2 + u_{x_1}^2 & u_{x_1} u_{x_2} \\ u_{x_1} u_{x_2} & \varepsilon^2 + u_{x_2}^2 \end{pmatrix}$$

and a minimal surface is obtained by minimizing the functional

$$\mathcal{S}_\varepsilon(\mathcal{M}) = \varepsilon \iint_{\Omega} \sqrt{\varepsilon^2 + |\nabla u|^2} \, dx$$

which is obviously a regularization of the total variation energy.

In fact, the real interest of this approach on a numerical point of view lies in the vectorial case. This Riemannian formalism can be developed in a wide variety of cases: textures, color, etc. For example, for color images, the space feature manifold  $\mathcal{M}$  is defined by:

$$X(\sigma_1, \sigma_2) = (X_1(\sigma_1, \sigma_2), X_2(\sigma_1, \sigma_2), X_r(\sigma_1, \sigma_2), X_g(\sigma_1, \sigma_2), X_b(\sigma_1, \sigma_2)).$$

$X$  is an embedded surface in  $R^5$ , and  $X_r$ ,  $X_g$ ,  $X_b$  are the three brightness components in the (Red, Green, Blue) system. We leave it as an exercise to the reader to write the associated metric and refer to [233, 232] for more details and experiments.

### 3.3.2 Smoothing-Enhancing PDEs

The Perona and Malik model [209]

Smoothing PDEs can be viewed as low-pass filters and act as forward diffusion processes. What about introducing locally high-pass filters? We show in this section how a suitable choice of the weighting coefficient  $c(\cdot)$  in the Perona and Malik model (3.39) permits to get this result<sup>3</sup>.

Let us first consider the 1-D case ( $x \in R$ ):

$$\begin{cases} \frac{\partial u}{\partial t}(t, x) = [c(u_x^2(t, x)) u_x(t, x)]_x \\ u(0, x) = u_0(x) \end{cases} \quad (3.67)$$

and give some formal definitions:

**Definition 3.3.2 (edge)** For a fixed time  $t$ , we say that  $\bar{x}$  is an edge of a function  $u(t, x)$  if  $u_x(\bar{x}) = \max_x u_x(t, x)$ . If  $u$  is smooth enough, we necessarily have at  $\bar{x}$ :  $u_{xx}(t, \bar{x}) = 0$  and  $u_{xxx}(t, \bar{x}) \leq 0$ . We will say that an edge  $\bar{x}$  is blurred by a PDE if in a neighborhood of  $\bar{x}$ ,  $u_x(t, x)$  decreases as  $t$  increases, or in other words if  $\frac{\partial}{\partial t}(u_x(t, x)) \leq 0$ .

We will say that an edge  $\bar{x}$  is enhanced by a PDE if in a neighbourhood of  $\bar{x}$ ,  $u_x(t, x)$  increases as  $t$  increases, i.e. if  $\frac{\partial}{\partial t}(u_x(t, x)) \geq 0$ .

Let us examine the relationship between the coefficient  $c(\cdot)$  in the PDE (3.67) and the blurring/enhancing of an edge. From (3.67), we have formally:

$$\frac{\partial}{\partial t}(u_x) = \left( \frac{\partial u}{\partial t} \right)_x = [c(u_x^2) u_x]_{xx} = u_{xxx} b(u_x^2) + 2 u_{xx}^2 b'(u_x^2)$$

where  $b(s) = 2sc'(s) + c(s)$ . If  $x$  is an edge at time  $t$ , then  $u_{xx}(t, x) = 0$  and  $u_{xxx}(t, x) \leq 0$ . Thus:

$$\text{sign} \left( \frac{\partial u}{\partial t}(t, x) \right)_x = \text{sign}(-b(u_x^2)(t, x)).$$

Therefore, we see that the blurring/enhancing process is governed by the sign of  $b(u_x^2)$ :

- If  $b(u_x^2) > 0$ , which means that (3.67) is a forward parabolic equation, the edge is blurred.
- If  $b(u_x^2) < 0$ , which means that (3.67) is a backward parabolic equation, the edge is sharpened.

---

<sup>3</sup>We recall that in Section 3.3.1, it was assumed  $b(s) = c(s) + 2sc'(s) > 0$

Now, let us go back to the general 2-D Perona and Malik model [209]:

$$\begin{cases} \frac{\partial u}{\partial t}(t, x) = \operatorname{div}(c(|\nabla u(t, x)|^2) \nabla u(t, x)) \\ u(0, x) = u_0(x) \end{cases} \quad (3.68)$$

where  $c : [0, +\infty[ \rightarrow ]0, +\infty[$  is a smooth decreasing function. As we saw in Section 3.3.1, equation (3.68) can be rewritten as:

$$\begin{cases} \frac{\partial u}{\partial t}(t, x) = c(|\nabla u(t, x)|^2) u_{TT} + b(|\nabla u(t, x)|^2) u_{NN} \\ u(0, x) = u_0(x). \end{cases}$$

Following our intuition from the 1-D case, if we want to sharpen edges, we need to impose that (3.68) is backward in the normal direction  $N$ , i.e.

$$b(s) = 2s c'(s) + c(s) < 0 \text{ for large } s \geq K, \text{ where } K \text{ is a given threshold.} \quad (3.69)$$

If we want to smooth homogeneous regions, we can impose:

$$c(0) = b(0) = 1,$$

which implies that (3.68) acts as the heat equation for small gradient. Of course, there exist several possible choices for  $c(\cdot)$ . A typical example is:

$$c(s) = \frac{1}{1 + \frac{s}{K}}.$$

Now, what can be said about the existence of a solution for (3.68) with  $b$  satisfying (3.69)? The response is quite clear: hardly nothing. To better understand the difficulty let us examine the 1-D "backward" heat equation:

$$\begin{cases} \frac{\partial u}{\partial t}(t, x) = -u_{xx}(t, x) \text{ on } R \times ]0, T[ \\ u(0, x) = u_0(x). \end{cases} \quad (3.70)$$

By making the change of variable  $\tau = T - t$ , it is easy to see that whenever  $u(t, x)$  is a solution of (3.70) then  $v(\tau, x) = u(t, x - \tau)$  is a solution of

$$\begin{cases} \frac{\partial v}{\partial \tau}(\tau, x) = v_{xx}(\tau, x) \text{ on } R \times ]0, T[ \\ v(t, x) = u_0(x) \end{cases} \quad (3.71)$$

which is exactly the heat equation with the backward datum  $v(t, x) = u_0(x)$ . So, if (3.70) admits a solution, the same goes for (3.71). But, according to the regularizing property of the heat equation, necessarily  $u_0(x)$  should be infinitely differentiable. If not, we deduce that (3.70) does not have a classical (and a weak) solution.

The same conclusion goes for (3.68) in the 1-D case. More precisely, Kichenassamy [143] proved the following result:

**Theorem 3.3.6 (Kichenassamy [143])** *Let us suppose that*

- (i) There exists a constant  $K > 0$  so that  $b(s) > 0$  for  $s < K^2$  and  $b(s) < 0$  for  $s > K^2$ .
- (ii) Both  $c(s)$  and  $b(s)$  tend to zero as  $s \rightarrow +\infty$ .
- (iii) (3.68) has a solution  $u(t, x)$  verifying  $K_1 \leq u_x(t, x) \leq K_2$  for all  $x \in [A, B]$  and all  $t \in [0, T]$ , for some  $A, B$  and  $K_1 > K$ .

Then  $u(t, x)$  is infinitely differentiable at  $t = 0$  and for all  $x \in ]A, B[$ . Therefore, if the initial image is not infinitely differentiable there is no weak solution.

In fact, it results from [143] that a "solution" must consist of regions in which it has a gradient less than  $K$  in absolute value, separated by points of discontinuity where the gradient is infinite. Thus, the notion of solution must be understood in the measure sense.

Regularization of the Perona and Malik model: Catté et al [59]

If we persist to study (3.68) in a backward regime, we have to reconsider our notion of solution. One way to tackle an ill-posed problem as (3.68) is to introduce a regularization that makes the problem well-posed. Then, by reducing the amount of regularization and observing the behaviour of the solution of the regularized problem, one can obtain precious information for the initial one. This way was followed by Catté et al [59]. The idea is to substitute in the diffusion coefficient  $c(|\nabla u|^2)$  the gradient of the image  $\nabla u$  by a smooth version of it  $G_\sigma * \nabla u$ , where  $G_\sigma$  is a smoothing kernel<sup>4</sup>, for example the Gaussian one (3.36). Since  $G_\sigma * \nabla u = \nabla(G_\sigma * u) = \nabla G_\sigma * u$ , Catté et al proposed the regularized model:

$$\begin{cases} \frac{\partial u}{\partial t}(t, x) = \operatorname{div}(c(|(\nabla G_\sigma * u)(t, x)|^2) \nabla u(t, x)) \\ u(0, x) = u_0(x). \end{cases} \quad (3.72)$$

This model has at least two advantages with regard to the Perona and Malik model:

- If the initial data is very noisy (introducing large oscillations in the gradient of  $u$ ), then the Perona and Malik model cannot distinguish between "true" edges and "false" edges created by the noise. The proposed model (3.72) avoids this drawback since now the equation diffuses only if the gradient is estimated to be small. In fact, the model makes the filter insensitive to noise at time  $t\sigma$  since  $(\nabla G_\sigma * u)(t, x)$  is exactly the gradient of the solution at time  $\sigma$  of the solution of the heat equation with initial datum  $u(t, x)$ .

---

<sup>4</sup>Since we write some convolution,  $u$  is in fact prolonged in  $R^2$  as in Section 3.3.1 (by symmetry and periodicity)

- As we prove next, equation (3.72) is now well-posed.

Let us establish that (3.72) is well-posed. Let us note  $\Omega = ]0, 1[ \times ]0, 1[$  and  $g(s) = c(s^2)$ . We have:

**Theorem 3.3.7** [59] *Let  $g : R^+ \rightarrow R_*^+$  smooth, decreasing with  $g(0) = 1$ ,  $\lim_{s \rightarrow +\infty} g(s) = 0$  and  $s \mapsto g(\sqrt{s})$  smooth. If  $u_0 \in L^2(\Omega)$ , then there exists a unique function  $u(t, x) \in C([0, T]; L^2(\Omega)) \cap L^2((0, T); W^{1,2}(\Omega))$  verifying in the distributional sense:*

$$\begin{cases} \frac{\partial u}{\partial t}(t, x) - \operatorname{div}(g(|(\nabla G_\sigma * u)(t, x)|) \nabla u(t, x)) = 0 & \text{on } \Omega \times ]0, T[ \\ \frac{\partial u}{\partial N}(t, x) = 0 & \text{on } \partial\Omega \times ]0, T[ \\ u(0, x) = u_0(x). \end{cases} \quad (3.73)$$

Moreover,  $|u|_{L^\infty((0, T); L^2(\Omega))} \leq |u_0|_{L^2(\Omega)}$  and  $u \in C^\infty(\bar{\Omega} \times ]0, T[)$ .

**Proof** We follow [59].

**Step 1:** Uniqueness of the solution

Let  $u_1$  and  $u_2$  be two solutions of (3.73). For almost every  $t$  in  $[0, T]$  and  $i = 1, 2$ , we have

$$\frac{d}{dt} u_i(t) - \operatorname{div}(\alpha_i(t) \nabla u_i(t)) = 0, \quad \frac{\partial u_i}{\partial N} = 0, \quad u_i(0) = u_0$$

where  $\alpha_i(t) = g(|\nabla G_\sigma * u_i|)$ . Thus:

$$\frac{d}{dt} (u_1 - u_2)(t) - \operatorname{div}(\alpha_1(t) (\nabla u_1 - \nabla u_2)(t)) = \operatorname{div}((\alpha_1 - \alpha_2)(t) \nabla u_2(t)).$$

Then, multiplying the above inequality by  $(u_1 - u_2)$ , integrating over  $\Omega$  and using the Neumann boundary condition, we get a.e.  $t$ :

$$\begin{aligned} \frac{1}{2} \frac{d}{dt} \int_{\Omega} |u_1(t) - u_2(t)|^2 dx + \int_{\Omega} \alpha_1 |\nabla u_1(t) - \nabla u_2(t)|^2 dx &= \\ &= - \int_{\Omega} (\alpha_1 - \alpha_2) \nabla u_2(t) \cdot (\nabla u_1(t) - \nabla u_2(t)) dx. \end{aligned} \quad (3.74)$$

But, since  $u_1$  belongs to  $L^\infty((0, T); L^2(\Omega))$ , then  $|\nabla G_\sigma * u_1|$  belongs to  $L^\infty((0, T); C^\infty(\Omega))$  and there exists a constant  $M = M(G_\sigma, |u_0|_{L^2(\Omega)})$  such that  $|\nabla G_\sigma * u_1| \leq M$  a.e.  $t$ ,  $\forall x \in \Omega$ . As  $g$  is decreasing and positive, it follows that a.e. in  $\Omega \times ]0, T[$

$$\alpha_1(t) = g(|\nabla G_\sigma * u_1|) \geq g(M) = \nu > 0,$$

which implies from (3.74):

$$\begin{aligned} \frac{1}{2} \frac{d}{dt} \left( |(u_1 - u_2)(t)|_{L^2(\Omega)}^2 \right) + \nu |\nabla(u_1 - u_2)(t)|_{L^2(\Omega)}^2 &\leq \\ &\leq |\alpha_1 - \alpha_2|_{L^\infty(\Omega)} |\nabla u_2(t)|_{L^2(\Omega)} |\nabla(u_1 - u_2)(t)|_{L^2(\Omega)}. \end{aligned} \quad (3.75)$$

Moreover, since  $g$  and  $G_\sigma$  are smooth, we have:

$$|\alpha_1(t) - \alpha_2(t)|_{L^\infty(\Omega)} \leq C |u_1(t) - u_2(t)|_{L^2(\Omega)} \quad (3.76)$$

where  $C$  is a constant which only depends on  $g$  and  $G_\sigma$ . From (3.76) and by using the Young's inequality, we obtain:

$$\begin{aligned} \frac{1}{2} \frac{d}{dt} \left( |(u_1 - u_2)(t)|_{L^2(\Omega)}^2 \right) + \nu |\nabla(u_1 - u_2)(t)|_{L^2(\Omega)}^2 &\leq \\ \leq \frac{2}{\nu} C^2 |(u_1 - u_2)(t)|_{L^2(\Omega)}^2 |\nabla u_2(t)|_{L^2(\Omega)}^2 + \frac{\nu}{2} |\nabla(u_1 - u_2)(t)|_{L^2(\Omega)}^2 \end{aligned}$$

from which we deduce:

$$\frac{1}{2} \frac{d}{dt} (|(u_1 - u_2)(t)|_{L^2(\Omega)}^2) \leq \frac{4}{\nu} C^2 |\nabla u_2(t)|_{L^2(\Omega)}^2 |(u_1 - u_2)(t)|_{L^2(\Omega)}^2.$$

To conclude, we need the Gronwall's inequality (see Section 2.5.1) that we recall here:

$$\begin{aligned} \text{If } y(t) \geq 0 \text{ satisfies } \frac{dy}{dt}(t) &\leq c_1(t) y(t) + c_2(t) \\ \text{then } y(t) &\leq \left( y(0) + \int_0^t c_2(s) ds \right) \exp \left( \int_0^t c_1(s) ds \right). \end{aligned}$$

Applying this inequality to  $y(t) = |(u_1 - u_2)(t)|_{L^2(\Omega)}^2$  we get, since  $u_1(0) = u_2(0) = u_0$ :

$$|(u_1 - u_2)(t)|_{L^2(\Omega)}^2 \leq 0, \text{ i.e. } u_1 = u_2.$$

**Step 2:** Existence of a solution

The proof is based on a classical fixed-point argument. Let us define the space:

$$W(0, T) = \left\{ w \in L^2((0, T); W^{1,2}(\Omega)) ; \frac{dw}{dt} \in L^2((0, T); W^{1,2}(\Omega)') \right\}$$

where  $W^{1,2}(\Omega)'$  is the dual of  $W^{1,2}(\Omega)$ .  $W(0, T)$  is an Hilbert space for the norm:

$$|w|_W = |w|_{L^2((0, T); W^{1,2}(\Omega))} + \left| \frac{dw}{dt} \right|_{L^2((0, T); W^{1,2}(\Omega)')}$$

Let  $w \in W(0, T) \cap L^\infty((0, T); L^2(\Omega))$  so that  $|w|_{L^\infty((0, T); L^2(\Omega))} \leq |u_0|_{L^2(\Omega)}$  and let us introduce the variational problem  $(P_w)$ :

$$\left\langle \frac{du(t)}{dt}, v \right\rangle_{W^{1,2}(\Omega)' \times W^{1,2}(\Omega)} + \int_{\Omega} g(|(\nabla G_\sigma * w)(t)|) \nabla u(t) \nabla v dx = 0$$

for all  $v \in W^{1,2}(\Omega)$ , a.e.  $t$  in  $[0, T]$ , which is now linear in  $u$ . As seen in Step 1, there exists a constant  $\nu > 0$  such that  $g(|\nabla G_\sigma * w|) \geq \nu$  a.e.

in  $\Omega \times ]0, T[$ . Therefore, by applying classical results on parabolic equations (see [105], page 356), we prove that the problem  $(P_w)$  has a unique solution  $u_w$  in  $W(0, T)$  satisfying the estimates:

$$\begin{aligned} |u_w|_{L^2((0,T);W^{1,2}(\Omega))} &\leq c_1 \\ |u_w|_{L^\infty((0,T);L^2(\Omega))} &\leq |u_0|_{L^2(\Omega)} \\ \left| \frac{du_w}{dt} \right|_{L^2((0,T);W^{1,2}(\Omega)')} &\leq c_2 \end{aligned} \quad (3.77)$$

where  $c_1$  and  $c_2$  are constants depending only on  $g$ ,  $G_\sigma$  and  $u_0$ . From these estimates, we introduce the subspace  $W_0$  of  $W(0, T)$  defined by:

$$W_0 = \left\{ \begin{array}{l} w \in W(0, T), \quad w(0) = u_0 \\ |w|_{L^2((0,T);W^{1,2}(\Omega))} \leq c_1, \\ |w|_{L^\infty((0,T);L^2(\Omega))} \leq |u_0|_{L^2(\Omega)}, \\ \left| \frac{dw}{dt} \right|_{L^2((0,T);W^{1,2}(\Omega)')} \leq c_2 \end{array} \right\}.$$

By construction,  $w \rightarrow S(w) \equiv u_w$  is a mapping from  $W_0$  into  $W_0$ . Moreover, one can prove that  $W_0$  is no empty, convex and weakly compact in  $W(0, T)$ . Thus, we can apply the Schauder's fixed-point Theorem:

**Theorem 3.3.8 (Schauder's fixed point Theorem)** *If  $E$  is a convex, compact subset of a Banach space and if  $S : E \rightarrow E$  is continuous, then there exists  $x \in E$  such that  $S(x) = x$ .*

So, let us prove that the application  $S : w \rightarrow u_w$  is weakly continuous ( $W_0 \rightarrow W_0$ ). Let  $w_j$  be a sequence, which converges weakly to some  $w$  in  $W_0$  and let  $u_j = u_{w_j}$ . We have to prove that  $S(w_j) = u_j$  converges weakly to  $S(w) = u_w$ . From (3.77) and classical results of compact inclusion in Sobolev spaces [1], we can extract from  $w_j$ , respectively from  $u_j$ , a subsequence (labelled  $w_j$ , respectively  $u_j$ ) such that for some  $u$ , we have

$$\begin{aligned} \frac{du_j}{dt} &\xrightarrow[L^2((0,T);W^{1,2}(\Omega)')]{\longrightarrow} \frac{du}{dt} \\ u_j &\xrightarrow[L^2((0,T);L^2(\Omega))]{\longrightarrow} u \\ \frac{\partial u_j}{\partial x_k} &\xrightarrow[L^2((0,T);L^2(\Omega))]{\longrightarrow} \frac{\partial u}{\partial x_k} \\ w_j &\xrightarrow[L^2((0,T);L^2(\Omega))]{\longrightarrow} w \\ \frac{\partial G_\sigma}{\partial x_k} * w_j &\xrightarrow[L^2((0,T);L^2(\Omega))]{\longrightarrow} \frac{\partial G_\sigma}{\partial x_k} * w \quad \text{and a.e. on } \Omega \times ]0, T[ \\ g(|\nabla G_\sigma * w_j|) &\xrightarrow[L^2((0,T);L^2(\Omega))]{\longrightarrow} g(|\nabla G_\sigma * w|) \\ u_j(0) &\xrightarrow[W^{1,2}(\Omega)']{\longrightarrow} u(0). \end{aligned}$$

The above convergences allow us to pass to the limit in  $(P_w)$  and get  $u = u_w = S(w)$ . Moreover, since the solution is unique, the whole sequence  $u_j = S(w_j)$  converges weakly in  $W_0$  to  $u = S(w)$ , i.e.  $S$  is weakly continuous. Consequently, thanks to the Schauder's fixed-point Theorem, there exists  $w \in W_0$  such that  $w = S(w) = u_w$ . The function  $u_w$  solves (3.73). The regularity follows from the general theory of parabolic equations. ■

**Remark** The Theorem 3.3.7 provides a natural algorithm for the numerical approximation of the solution. Let  $u_0 \in L^2(\Omega)$ , we construct a sequence  $u^n$  by solving the iterative scheme:

$$\begin{cases} \frac{\partial u^n}{\partial t}(t, x) - \operatorname{div}(g(|(\nabla G_\sigma * u^n)(t, x)|)\nabla u^{n+1}(t, x)) = 0 \text{ a.e. on } \Omega \times ]0, T[ \\ \frac{\partial u^{n+1}}{\partial N}(t, x) = 0 \text{ a.e. on } \partial\Omega \times ]0, T[ \\ u^{n+1}(0, x) = u_0(x) \end{cases}$$

It is proven in [59] that  $u^n$  converges in  $C([0, T]; L^2(\Omega))$  to the unique solution of (3.71). ■

Let us mention the existence of other models for regularizing the Perona and Malik equation. For example, Nitzberg-Shiota [194] proposed the coupled system (in 1-D)

$$\begin{cases} \frac{\partial u}{\partial t} = (c(v) u_x)_x \\ \frac{\partial v}{\partial t} = \frac{1}{\tau} (|u_x|^2 - v) \\ u(0, x) = u_0(x) \text{ and } v(0, x) \text{ is a smoothed version of } |u'_0(x)|^2. \end{cases}$$

The function  $v$  plays the role of time-delay regularization, where the parameter  $\tau > 0$  determines the delay. For other models, see Barenblatt et al [22], Chipot et al [75], Alvarez et al [6].

\* *In spite of the lack of a rigorous mathematical theory concerning the Perona and Malik equation, it is successfully used in many numerical experiments. This phenomenon is still unexplained. It is likely that the behaviour of the associated discrete problem does not reflect the ill-posedness of the continuous version but this has to be more investigated.*

A first attempt to justify the Perona and Malik model was done by Kichenassamy who defined in [143] a notion of generalized solution. This direction is promising and should be more investigated.

Once this regularized model is well-defined, a natural question arises: does equation (3.72) approach equation (3.68) as  $\sigma$  tends to zero? This is a difficult question and no mathematical response is available today. Perhaps, a clue to tackle this question would be to find a suitable functional framework for which  $u_\sigma$ , the solution of (3.72), and its gradient would be

uniformly bounded with respect to  $\sigma$ . Then accumulation points could be considered as good candidates.

Another question which is empirical, is the choice of the parameter  $\sigma$ . Here again, there is no satisfying answer. In general, this choice is fixed by the user and is related to other parameters, for example those defining the function  $c(s)$ .

### 3.3.3 Enhancing PDEs

The Osher and Rudin's shock-filters [199]

We close this chapter devoted to image restoration by examining edge enhancement via PDEs. In fact, in a way, enhancing and smoothing are opposite processes. In the former case, we want to create discontinuities at places where they have to appear, while in the latter case we want to remove superfluous features and false discontinuities. A typical example of enhancing is deblurring. In this section, we show how some nonlinear hyperbolic PDEs (called shock filters) can be used for edge enhancement and deblurring. Let us start with the one-dimensional case. Ideally, an edge can be modeled by the step function:

$$u(x) = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x < 0. \end{cases}$$

Let us imagine that some process (a convolution, for example) has blurred this edge, so that we have in hand a smooth version  $u_0(x)$  of  $u(x)$  (see Figure 3.18). The problem is to go back to  $u(x)$ , starting from  $u_0(x)$ .

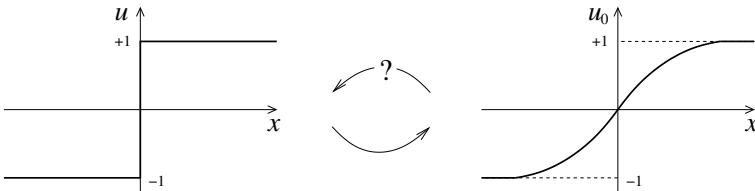


Figure 3.18. Illustration of the one dimensional case

To illustrate the reasoning and cover the different possibilities, we consider in this section the following initial condition:

$$u_0(x) = \cos(x).$$

In this case, as depicted in Figure 3.19, we would like to define a family of evolving curves  $\{u(t, x)\}_{t>0}$  in order to sharpen the edges.

☞ As we can observe, the direction of the motion of  $u(t, x)$  is a function of  $x$  and depends of the sign of the product  $u_x(t, x) u_{xx}(t, x)$ .

The four cases are indicated in Figure 3.19. Notice that at points  $x$  where  $u_{xx}(t, x) = 0$  or  $u_x(t, x) = 0$ , it is desirable that no motion occurs.

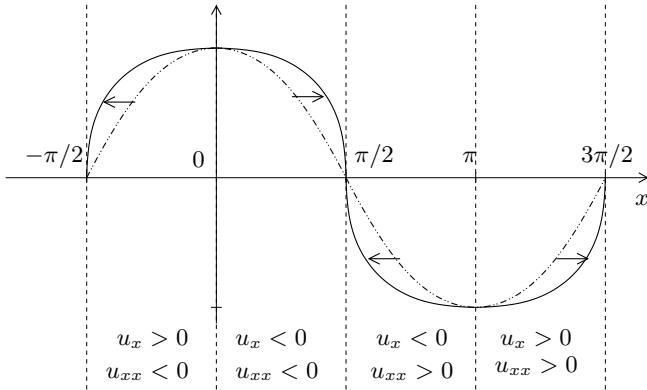


Figure 3.19. Illustration of the deblurring procedure in the 1-D case, for the initial condition  $u_0(x) = \cos(x)$ , represented on  $[-\frac{\pi}{2}, \frac{3\pi}{2}]$ . The dashed line represents the initial condition, and in solid line we display the function after some time. Arrows show the direction of displacement.

Following this idea, Osher and Rudin [199] proposed to solve

$$\begin{cases} u_t(t, x) = -|u_x(t, x)| \operatorname{sign}(u_{xx}(t, x)) \\ u(0, x) = u_0(x) \end{cases} \quad (3.78)$$

where  $\operatorname{sign}(u) = 1$  if  $u > 0$ ,  $\operatorname{sign}(u) = -1$  if  $u < 0$ ,  $\operatorname{sign}(0) = 0$ . For example, at points where  $u_x(t, x) > 0$  and  $u_{xx}(t, x) > 0$ , we can verify that (3.78) behaves like  $u_t(t, x) + u_x(t, x) = 0$  that is a transport equation with speed  $+1$ , which is the desired motion. The same goes for the other cases. Before trying to justify this equation, we are going to consider a simplified version.

A case study: construction of a solution by the method of characteristics

Let us examine more precisely the following simpler case:

$$\begin{cases} u_t(t, x) = -|u_x(t, x)| \operatorname{sign}((u_0)_{xx}(t, x)) \\ u(0, x) = u_0(x) \end{cases} \quad (3.79)$$

with  $u_0(x) = \cos(x)$ . We are going to search for an explicit solution. To do this, we use the *method of characteristics* for which we recall the general formalism.

Method of characteristics [105]

Let  $U$  be an open subset of  $R^N$  and  $\Gamma \subset \partial U$ , a part of the boundary

of  $U$ . Let us consider the nonlinear first-order PDE

$$\begin{cases} F(x, u, Du) = 0 \text{ in } U \\ u = g \text{ on } \Gamma \end{cases} \quad (3.80)$$

where  $F$  and  $g$  are supposed smooth. The idea is to convert the PDE into a system of Ordinary Differential Equations (ODE). Let us fix any point  $x \in U$  and let us suppose that  $u \in C^2$  is a solution of (3.80). We would like to calculate  $u(x)$  by finding some curve lying within  $U$ , connecting  $x$  with a point  $x^0 \in \Gamma$ , and along which we can compute  $u$ . Since  $u = g$  on  $\Gamma$ , we know the value of  $u$  at  $x^0$  and we desire to calculate  $u$  all along the curve, and in particular at  $x$ . Let us suppose that the curve is described parametrically by  $x(s) = (x_1(s), \dots, x_N(s))$ . We define

$$z(s) = u(x(s)) \quad (3.81)$$

$$p(s) = Du(x(s)). \quad (3.82)$$

Now, the question is to choose a good curve  $x(s)$  in such a way that we can compute  $z(s)$  and  $p(s)$ . In practice, we have to find the equations that are satisfied by  $x(s)$ ,  $z(s)$  and  $p(s)$ , and to solve them.

We first differentiate (3.82) with respect to  $s$ :

$$\dot{p}_i(s) = \sum_{j=1}^N u_{x_i x_j}(x(s)) \dot{x}_j(s), \quad i = 1, \dots, N \quad \left( \dot{p}_i = \frac{dp_i}{ds} \right) \quad (3.83)$$

then we differentiate (3.80) with respect to  $x_i$ :

$$\sum_{j=1}^N \frac{\partial F}{\partial p_j}(x, u, Du) u_{x_j x_i} + \frac{\partial F}{\partial z}(x, u, Du) u_{x_i} + \frac{\partial F}{\partial x_i}(x, u, Du) = 0. \quad (3.84)$$

In order to get rid of second derivative terms, let us define  $x(s)$  as the solution of the ODE system:

$$\dot{x}_j(s) = \frac{\partial F}{\partial p_j}(x(s), z(s), p(s)). \quad (3.85)$$

Assuming that  $x(s)$  exists and thanks to (3.81)-(3.82), the equation (3.84) evaluated at  $x = x(s)$  writes as:

$$\begin{aligned} & \sum_{j=1}^N \frac{\partial F}{\partial p_j}(x(s), z(s), p(s)) u_{x_j x_i}(x(s)) + \frac{\partial F}{\partial z}(x(s), z(s), p(s)) p_i(s) + \\ & + \frac{\partial F}{\partial x_i}(x(s), z(s), p(s)) = 0. \end{aligned}$$

If we substitute in this expression  $\frac{\partial F}{\partial p_j}(x(s), z(s), p(s))$  by  $\dot{x}_j(s)$  (cf. (3.85)), we get from (3.83)

$$\dot{p}_i(s) = -\frac{\partial F}{\partial z}(x(s), z(s), p(s)) p_i(s) - \frac{\partial F}{\partial x_i}(x(s), z(s), p(s)) = 0.$$

Finally, differentiating (3.81) with respect to  $s$ , we obtain:

$$\dot{z}(s) = \sum_{j=1}^N \frac{\partial u}{\partial x_j}(x(s)) \dot{x}_j(s) = \sum_{j=1}^N p_j(s) \frac{\partial F}{\partial p_j}(x(s), z(s), p(s))$$

In summary, in vector notation,  $x(s)$ ,  $z(s)$  and  $p(s)$  satisfy the system of  $(2N + 1)$  first-order ODE (called characteristic equations):

$$\begin{cases} \dot{x}(s) = D_p F(x(s), z(s), p(s)) \\ \dot{z}(s) = D_p F(x(s), z(s), p(s)) \cdot p(s) \\ \dot{p}(s) = -D_x F(x(s), z(s), p(s)) - D_z(x(s), z(s), p(s)) p(s) \end{cases} \quad (3.86)$$

Naturally, we need in addition to specify the initial conditions. To make it more clear, let us come back to our initial problem (3.79). Unfortunately, our calculus will be quite formal since in the case of equation (3.79), the function  $F$  is not differentiable! In order to adopt the same notations as previously, we make the change of variables  $x_2 = t$  and  $x_1 = x$ . Therefore, (3.79) writes:

$$\begin{cases} u_{x_2} + \text{sign}(u_0)_{x_1 x_1} |u_{x_1}| = 0 \text{ in } U \\ u(x_1, 0) = u_0(x_1) \text{ on } \Gamma \end{cases} \quad (3.87)$$

i.e.  $F(x_1, x_2, z, p) = p_2 + \text{sign}(u_0)_{x_1 x_1} |p_1|$ . We choose  $u_0(x_1) = \cos(x_1)$ , thus  $(u_0(x_1))_{x_1 x_1} = -\cos(x_1)$ .

Let  $U = ] -\frac{\pi}{2}, \frac{3\pi}{2} [ \times R^+$  and  $\Gamma = \{(x_1, x_2), x_1 \in ] -\frac{\pi}{2}, \frac{3\pi}{2} [, x_2 = 0\}$ . To get rid of the sign function, we split the study into two cases.

**First case:** Equation is studied on  $] -\frac{\pi}{2}, \frac{\pi}{2} [ \times R^+$

In this case,  $\text{sign}(\cos(x_1))_{x_1 x_1} = -1$ . Therefore, we formally have:

$$D_{p_1} F = -\frac{p_1}{|p_1|}, \quad D_{p_2} F = 1, \quad D_z F = D_{x_1} F = D_{x_2} F = 0$$

and (3.86) writes:

$$\begin{cases} \dot{x}_1(s) = -\frac{p_1(s)}{|p_1(s)|}, & \dot{x}_2(s) = 1 \\ \dot{p}_1(s) = \dot{p}_2(s) = 0, \\ \dot{z}(s) = p_2(s) - |p_1(s)| \end{cases} \quad (3.88)$$

For  $s = 0$ , we suppose that  $x_1(0) = a \in ] -\frac{\pi}{2}, \frac{\pi}{2} [$  and  $x_2(0) = 0$ . The integration of (3.88) is immediate. We get for some constants  $p_1^0$  and  $p_2^0$ :

$$\begin{cases} x_1(s) = -\frac{p_1^0}{|p_1^0|} s + a, & x_2(s) = s \\ p(s) = (p_1^0, p_2^0) \\ z(s) = (-|p_1^0| + p_2^0) s + \cos(a). \end{cases}$$

It remains to determine  $(p_1^0, p_2^0)$ . Since  $u(x_1, x_2) = \cos(x_1)$  on  $\Gamma$ , we have  $p_1^0 = u_{x_1}(a, 0) = -\sin a$  and  $p_2^0 = u_{x_2}(a, 0)$ . But from equation (3.87), we

deduce  $p_2^0 = u_{x_2}(a, 0) = |u_{x_1}(a, 0)| = |\sin a|$ . Therefore, the characteristic curve  $x(s)$  is given by

$$\begin{cases} x_1(s) = \frac{\sin a}{|\sin a|} s + a \\ x_2(s) = s \end{cases}$$

and  $u(x_1(s), x_2(s)) = z(s) = (-|\sin a| + |\sin a|) + \cos(a) = \cos(a)$ , that is to say,  $u$  is constant along characteristics. We have two cases (see also Figure 3.20):

(i) if  $\sin a > 0$ , i.e.  $a \in [0, \frac{\pi}{2}[$ , then  $x_1(s) = s + a$ ,  $x_2(s) = s$  and the characteristics are straight lines. Thus, in this case, the solution of (3.87) is  $u(x_1, x_2) = \cos(x_1 - x_2)$  with  $x_2 < x_1 < \pi/2$ .

(ii) if  $\sin a \leq 0$ , i.e.  $a \in ]-\frac{\pi}{2}, 0[$  then  $x_1(s) = -s + a$ ,  $x_2(s) = s$  and  $u(x_1, x_2) = \cos(x_1 + x_2)$  with  $-\pi/2 < x_1 < -x_2$ .

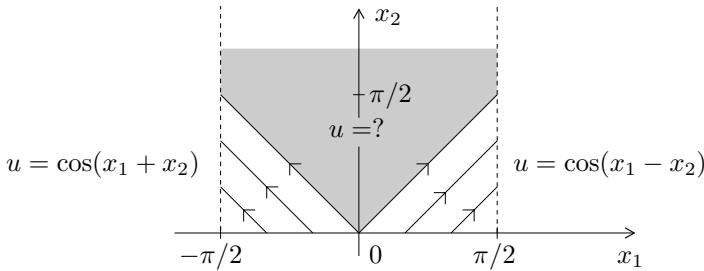


Figure 3.20. Characteristic lines in  $] -\frac{\pi}{2}, \frac{\pi}{2}[ \times R^+$ . Notice that no characteristics go into the gray region.

**Second case:** Equation is studied on  $]\frac{\pi}{2}, \frac{3\pi}{2}[ \times R^+$

In this case the equation becomes  $u_{x_2} + |u_{x_1}| = 0$  and a similar study leads to the solution (see also Figure 3.21):

$$u(x_1, x_2) = \begin{cases} \cos(x_1 + x_2) & \text{if } \frac{\pi}{2} < x_1 < \pi - x_2 \\ \cos(x_1 - x_2) & \text{if } x_2 + \pi < x_1 < \frac{3\pi}{2}. \end{cases}$$

Thanks to these calculi, we can observe that:

- The function  $u(x_1, x_2)$  is discontinuous (a shock) along the line  $x_1 = \frac{\pi}{2}$ , i.e. at a point where characteristics intersect.
- $u(x_1, x_2)$  is not yet defined for  $-x_2 < x_1 < x_2$  and  $\pi - x_1 < x_2 < x_1 - \pi$ .

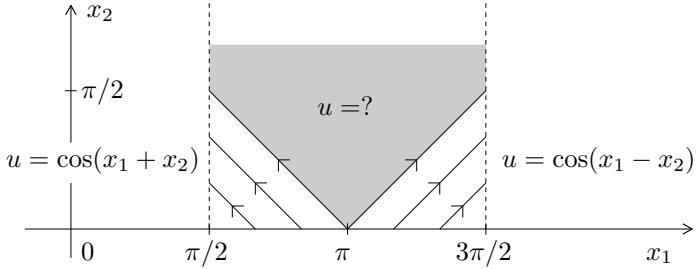


Figure 3.21. Characteristic lines in  $\left]\frac{\pi}{2}, \frac{3\pi}{2}\right[ \times R^+$ . Notice that no characteristics go into the gray region.

- If we do not want to create other discontinuities than those described above, we must set:

$$u(x_1, x_2) = \begin{cases} 1 & \text{if } -x_2 < x_1 < x_2 \\ -1 & \text{if } \pi - x_1 < x_2 < x_1 - \pi. \end{cases}$$

In conclusion, we propose as a solution of

$$\begin{cases} u_{x_2} + \text{sign}(-\cos(x_1)) |u_{x_1}| = 0 & \text{in } \left]-\frac{\pi}{2}, \frac{3\pi}{2}\right[ \times R^+ \\ u(x_1, 0) = \cos(x_1), & \end{cases} \quad (3.89)$$

the piecewise regular function  $u$  depicted in Figure 3.22. It is easy to see, by symmetry, that if  $U = R \times R^+$ , then we can construct a solution of (3.87) (with  $u_0(x_1) = \cos(x_1)$ ) whose discontinuities only develop at  $x_1 = (2k+1)\pi/2$ ,  $k = 0, \pm 1, \pm 2 \dots$

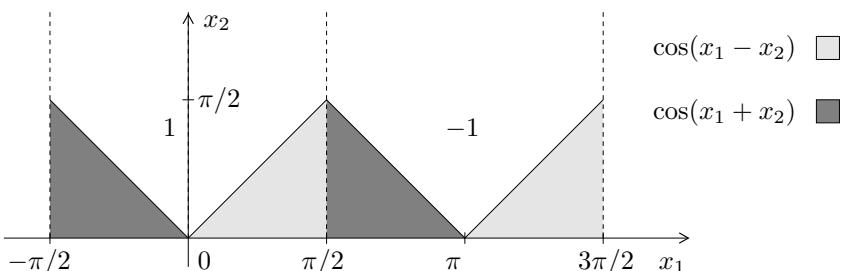


Figure 3.22. Solution proposed for equation (3.89).

This example shows very well why equations like (3.87) acts as edge enhancement filters. Starting from an initial data  $u_0(x) = \cos(x)$  (we return to our previous notations  $(t, x)$ ), we have constructed a family of functions  $\{u(t, x)\}_{t>0}$  so that, as  $t$  increases, the limiting process tends to an ideal one-dimensional edge model: the step function  $u(x) = (-1)^k$  for  $(2k-1)\frac{\pi}{2} < x < (2k+1)\frac{\pi}{2}$ . We illustrate in Figure 3.23 the solution at different

times. We can remark that edge formation and sharpening process occur at the places where  $(u_0(x))_{xx} = 0$ .

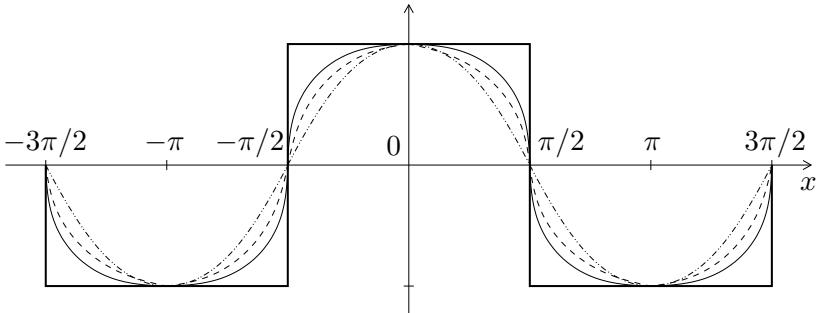


Figure 3.23. Illustration of the function  $u$  solution of (3.79) at different times.

#### Comments on the shock-filter equation

Now, let us come back to more general one-dimensional models. As described in [199], let us consider the equation:

$$\begin{cases} u_t = -|u_x| F(u_{xx}), & x \in R, t > 0 \\ u(0, x) = u_0(x). \end{cases} \quad (3.90)$$

Here,  $F$  is a Lipschitz continuous function satisfying:

$$\begin{cases} F(0) = 0 \\ \text{sign}(s)F(s) > 0, \quad s \neq 0 \end{cases} \quad (3.91)$$

A typical example of (3.90) is:

$$\begin{cases} u_t = -|u_x| u_{xx}, & x \in R, t > 0 \\ u(0, x) = u_0(x) \end{cases} \quad (3.92)$$

which can be written as:

$$\begin{cases} u_t + (u_{xx} \text{sign}(u_x)) u_x = 0, & x \in R, t > 0 \\ u(0, x) = u_0(x). \end{cases} \quad (3.93)$$

The equation 3.93 (or (3.92)) can be considered as a transport equation whose the speed of propagation is locally given by  $c(x) = \text{sign}(u_x) u_{xx}$ . Moreover, since edges are defined as maximum points of  $|u_x|$ , then at these points we have necessarily  $u_{xx} = 0$  and locally  $u_{xx}$  changes of sign. Thus, the speed  $c(x)$  plays the role of an edge-detector. From a mathematical point of view this type of equations is severely ill-posed.

\* As already noticed, up to our knowledge, there is no theoretical justification for this problem. One of the first difficulty is to define the suitable notion of weak solution. We may wonder if the notion of discontinuous

*viscosity solutions* [23, 25] may help for the understanding of this equation.

Nevertheless, Osher and Rudin [199] have performed very satisfying numerical simulations and they conjectured the following result.

**Conjecture [199]** *The evolution equation (3.90), with  $u_0(x)$  continuous, has a unique solution which has jumps only at inflection points of  $u_0(x)$  and for which the total variation in  $x$  of  $u(t, x)$  is invariant in time, as well as the location and value of local extrema.*

The transposition of these ideas to the two-dimensional case is now straightforward. We have to write an equation which is the combination of a propagation term  $|\nabla u|$  and an edge-detection term whose desired behaviour involves changing sign across edges (or singular features) so that the local flow field is directed towards the edges. A good candidate is:

$$\frac{\partial u}{\partial t} = -|\nabla u| F(L(u)) \quad (3.94)$$

where  $F$  satisfies (3.91) and where  $L$  is a nonlinear elliptic operator so that zero crossings define the edges of the processed image. According to the Marr's theory [169], the most classical operator is the Laplacian:

$$L(u) = \Delta u = u_{xx} + u_{yy}.$$

A better choice would be:

$$L(u) = \frac{1}{|\nabla u|^2} (u_x^2 u_{xx} + 2u_x u_y u_{xy} + u_y^2 u_{yy})$$

which corresponds to the second derivative of  $u$  in the direction of the  $\frac{\nabla u}{|\nabla u|}$  (here edges are defined as level curves of  $u$ ).

The efficiency of this approach is demonstrated in Figures 3.24 and 3.25. We refer the reader to Section A.3.3 of the Appendix for the discretization of (3.94). There are usually two main criticisms for this model. The first is that the results obtained is not realistic from a perception point of view. As it can observed in Figure 3.25 the result is a piecewise constant image so that texture and fine details are lost (compare to the clear “Borel building” image in Figure 3.1). However, one cannot expect to recover details not present in the original blurred image... The second criticism is that if we have also some noise present in the image it will be enhanced. To overcome this difficulty Alvarez-Mazorra [7] combine shock filters and anisotropic diffusion and add a smoothing kernel for the estimation of the direction of the edges (see also [152]).

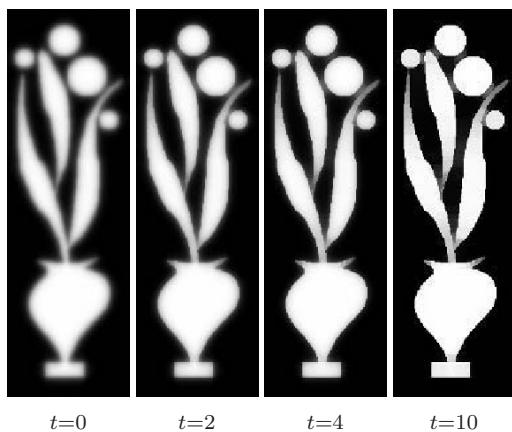


Figure 3.24. Example of shock filters on a blurred image representing a jar with some flowers. The blur has been generated by a convolution with a Gaussian kernel of variance  $\sigma = 10$ . Some iterations until convergence are then shown. This example shows that degradations due to blur cannot be fully recovered for fine structures (see for instance the extremities of the leafs).

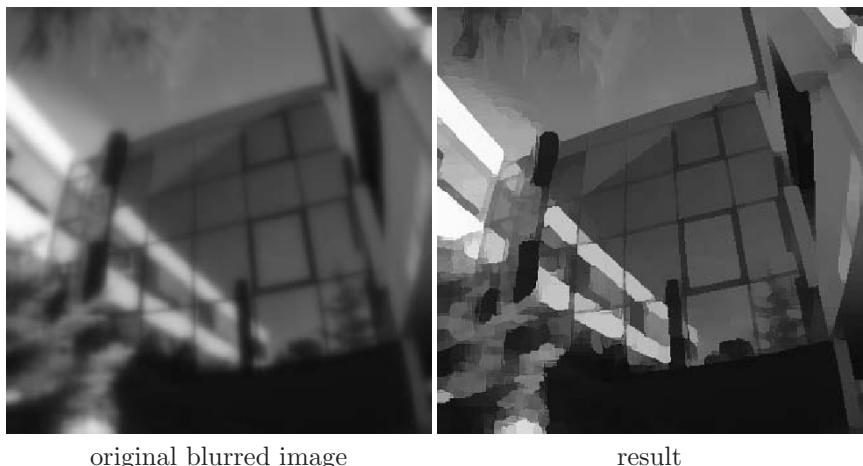


Figure 3.25. Example of shock filters on the blurred “Borel building” image. The blur has been generated by a convolution with a Gaussian kernel of variance  $\sigma = 10$ . One can observe the patch effect produced by this algorithm.

# 4

## The Segmentation Problem

### How to read this chapter?

This chapter is concerned with image segmentation which plays a very important role in many applications. The aim is to find a partition of an image into its constituant parts. As we will see, the main difficulty is that one need to manipulate objects of different nature: functions, domains in  $R^2$  and curves.

- We first try in Section 4.1 to better define what image segmentation is and we briefly survey some classical ideas in image segmentation. In fact the notion of segmentation depends on the kind of image we have to process and what we want to do. In the last decade, two main approaches have been developed: the Mumford and Shah approach and the geodesic active contours method.
- Section 4.2 concerns the Mumford and Shah functional. Here the idea is to find a close image of the initial one compounded of several regions with nearly constant intensity. The difficulty for studying the Mumford and Shah functional is that it involves two unknowns: the intensity function and the set  $K$  of edges. This difficulty is tackled in Sections 4.2.2 and 4.2.3 which are concerned with the mathematical study of this problem (definition of the suitable mathematical framework, optimality conditions, regularity of the edge set). Section 4.2.4 is a survey of some approaches for approximating the Mumford

and Shah functional. Most of them are based on the  $\Gamma$ -convergence theory. Finally, we present in Section 4.2.5 some experimental results.

- Section 4.3 deals with the geodesic active contours and the level sets method. Here, the objective is to find the boundaries of objects in an image. The idea is to modelize those contours as curves that should match the highest gradients. We start in Section 4.3.1 by recalling the Kass, Witkin and Terzopoulos snakes model which is one of the first work in this direction. This model which has some drawbacks has been revisited by Caselles, Kimmel and Sapiro who proposed a geodesic active contour strategy (Section 4.3.2). We clearly establish the connection between these two formulations. One of the main interest of the latter model is that it can be rewritten using a level sets formulation. This is detailed in Section 4.3.3 where we prove the well-posedness of this model in the viscosity sense. This section is rather technical but shows a complete and classical proof using viscosity solutions. We finally illustrate this approach in Section 4.3.4 and we refer to the Appendix for the details regarding the discretization.

## 4.1 Definition and objectives

As a first definition, we could say that segmenting an image means dividing it into its contituant parts. However this definition is rather unsatisfactory and ambiguous. Let us have a look to the images presented in Figure 4.1. In the left-hand side image, every contour (edge) information is important and it would be interesting to have an identification of all the contours separating two regions of different intensities. Equivalently, one would like to have a simplified version of the original image, compounded of homogeneous regions separated by sharp edges. The right-hand side second image illustrates the notion of objects. Some contours (the boundaries of the object) may have more importance (depending on the application) and it would be interesting to find an approach to detect them. These two examples show that the notion of segmentation is not unique and may depend on the kind of image we have to deal with.

Still, in each case, the important features are edges. Edge detection has been an early concern in computer vision. Classical approaches are based on local differential properties of an edge, for instance on the first and second derivatives of the image (see Figure 4.2).

At the first order, the earliest methods were based on the application of some convolution masks to approximate the first derivative, thus enhancing edges [211, 210, 231]. Then, Canny [54] proposed an edge detector that is still widely used. The starting point was to define criteria that an edge detector should satisfy: reliability of the detection, accuracy of the localisation, and the requirement of one response per edge. This leads to an optimal



Figure 4.1. Two examples of images suggesting different notions of segmentation

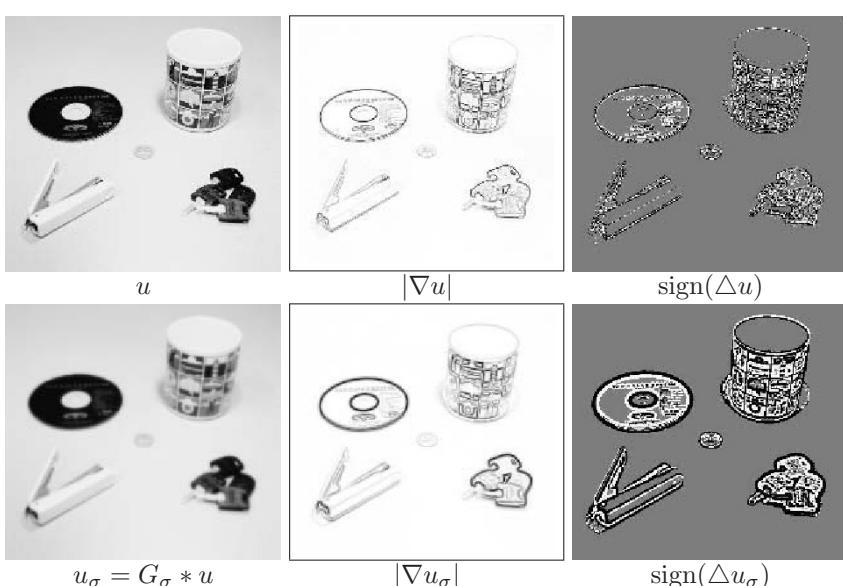


Figure 4.2. Edges and derivatives. First and second order derivatives of the image "objects" and a smoothed version of it are displayed. The gradient is represented using inverted colors for a better visualization. The Laplacian is negative in the black areas, positive in the white ones, zero otherwise. Edges can be seen as the locations where the gradient is locally maximum, or where the Laplacian changes sign. Notice how the smoothing allows to obtain a cleaner description of the edges.

filter which is in fact a very close approximation to the first derivative of a Gaussian. This has been further developed by Deriche [91] and Shen and Castan [230] who proposed sharper filters implemented recursively.

At the second order, the important starting point was the method proposed by Marr and Hildreth [170] based on zero-crossing detection of a Laplacian of a Gaussian, noted LoG (see also [134, 195, 253]). This kind of approach produces closed contours, the corners are rounded and the connectivity at the junctions is poor.

These approaches are local and combine derivatives at different scales. The goal is to identify the edges which are characterized by sharp variations of the intensity. If we consider the examples from Figure 4.1 we may propose two different strategies:

- To segment the “Borel building” image, the dual point of view would be to find a simplified image as a combination of regions of constant intensities. By constructing such an approximation of an image, we would also have the segmentation. Also, as it is local, there is no concern about the smoothness of the contours. These two ideas can be incorporated in a variational framework: starting from an image  $u_0$ , we look for a pair  $(u, K)$  such that  $u$  is a nearly piecewise constant approximation of  $u_0$  and  $K$  corresponds to the set of edges. This was proposed by Mumford and Shah in 1989 and we detail in Section 4.2 this model and its properties.
- Now, if we consider the “objects” image, one would like to have a technique separating the five objects, without any concern of the internal texture. One intuitive idea should be to consider a curve enclosing all the objects and make it evolve until it reaches the boundaries of the objects. Eventually, this curve could shrink or split. This idea has been initially proposed by Kass Witkin and Terzopoulos (*active contours*) and it is based on an energy minimization depending on the curve. This is presented in Section 4.3 as well as further developments like the level sets formulation.

We will work afterwards with the “objects” image which has interesting properties, with regards to the previous discussion. First of all, it is compounded of five different objects with different sizes, shapes or textures (see Figure 4.3), so that both aspects of segmentation can be tested (in terms of edge detection or object segmentation): the mug has covered with several images, one part of the stapler is quite elongated and fine, several similar keys are superimposed, the disk has a hole, and the coin is small and present small contrasts. It will be interesting to observe what kind of results we obtain with the different approaches.

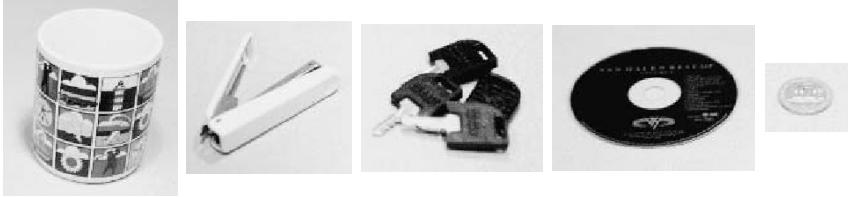


Figure 4.3. Different parts of the “objects” image

## 4.2 The Mumford and Shah functional

### 4.2.1 A minimization problem

Let us present the model introduced by Mumford and Shah in 1989 [187]. In this section  $\Omega$  is a bounded open set of  $R^N$ ,  $N = 2, 3$  and  $u_0(x)$  is the initial image. Without a loss of generality we can always assume that  $0 \leq u_0(x) \leq 1$  a.e.  $x \in \Omega$ . We search for a pair  $(u, K)$ , where  $K \subset \Omega$  is the set of discontinuities, minimizing:

$$F(u, K) = \int_{\Omega - K} (u - u_0)^2 dx + \alpha \int_{\Omega - K} |\nabla u|^2 dx + \beta \int_K d\sigma \quad (\text{MS})$$

where  $\alpha$  and  $\beta$  are non negative constants and  $\int_K d\sigma$  is the length of  $K$ . In their seminal paper [187], the authors conjectured that:

**Conjecture 4.2.1** *There exists a minimizer of  $F$  so that the edges (the discontinuity set  $K$ ) are the union of a finite set of  $C^{1,1}$  embedded curves. Moreover they predicted that each curve may end either as a crack tip (a free extremity, i.e.  $K$  looks like a half line) or in triple junction that is three curves meeting at their endpoints with  $\frac{2\pi}{3}$  angle between each other.*

The purpose of this section is to analyze this conjecture and the recent advances on this subject. We first concentrate on the existence problem (Section 4.2.2) and then study the geometric properties of the set  $K$  (Section 4.2.3). We conclude by giving some details on the approximation of this functional (Section 4.2.4).

### 4.2.2 The mathematical framework for the existence of a solution

Before studying this problem, we need to correctly define the functional  $F(u, K)$  and in particular decide which class of  $K$  to consider. It is clear that we cannot *a priori* impose that  $K$  is made of a finite set of  $C^{1,1}$ -curves since one cannot hope to obtain any compactness property and so any existence theorem with this too restrictive assumption. The regularity

of  $K$  will have to be proven *a posteriori*. This situation is classical in the calculus of variations and one overcomes this difficulty by looking for a solution in a wider class of set of finite length rather than just in a set of  $C^{1,1}$ -curves. This is achieved by defining the length of  $K$  as its  $(N-1)$ -dimensional Hausdorff measure  $\mathcal{H}^{N-1}(K)$  which is the most natural way of extending the notion of length to non-smooth sets (see Definition 2.2.8, Section 2.2). Therefore we rewrite the Mumford and Shah functional as:

$$F(u, K) = \int_{\Omega - K} (u - u_0)^2 \, dx + \alpha \int_{\Omega - K} |\nabla u|^2 \, dx + \beta \mathcal{H}^{N-1}(K)$$

where for the moment  $K \subset \Omega$  is a closed set and  $u$  belongs to the Sobolev space  $W^{1,2}(\Omega - K)$ . We can observe that  $F(u, K)$  is minimal in the sense that removing one of the three terms would imply that  $\inf F(u, K) = 0$  and we could obtain trivial solutions. For example if we drop the first integral in  $F$  then  $u = 0$  and  $K = \emptyset$  are solutions, or if we drop the second term then  $u = u_0$  and  $K = \emptyset$  are solutions. Nevertheless, if we reject trivial solutions it can be interesting to study some of these limiting cases. For example, in the latter case ( $\alpha = 0$ ) we obtain the reduced Mumford and Shah functional:

$$E(u, K) = \int_{\Omega - K} (u - u_0)^2 \, dx + \beta \mathcal{H}^{N-1}(K).$$

It is easy to see that if  $K$  is fixed ( $K \neq \emptyset$ ) then a solution  $u$  is piecewise constant ( $u$  is equal to the mean of  $u_0$  on the connected components of  $\Omega - K$ ) and so  $E$  becomes only a function of  $K$ . A lot of work has been devoted to this particular case. Let us mention contributions from Mumford and Shah [187], Morel and Solimini [185, 183, 184], Massari and Tamanini [173].

► The difficulty for studying  $F$  is that it involves two unknowns  $u$  and  $K$  of different nature:  $u$  is a function defined on a  $N$ -dimensional space, while  $K$  is a  $(N-1)$ -dimensional set.

In order to apply the direct method of the calculus of variations, it is necessary to find a topology, which ensures at the same time lower semi-continuity of  $F$  and compactness of the minimizing sequences. The difficulty comes from  $\mathcal{H}^{N-1}(K)$ . Indeed let  $E$  be a Borel set of  $R^N$  with topological boundary  $\partial E$ . It is easy to convince oneself that:

► the application  $E \rightarrow \mathcal{H}^{N-1}(\partial E)$  is not lower semi-continuous with respect to any compact topology.

Let us consider the following example. Let  $\{x_i\}$  be the sequence of all

rational points in  $R^N$  and let

$$B_i = \{x \in R^N ; |x - x_i| \leq 2^{-i}\}$$

$$E_k = \bigcup_{i=0}^k B_i, \quad E = \bigcup_{i=0}^{\infty} B_i.$$

Denoting by  $|E|$  the  $N$ -dimensional Lebesgue measure of  $E$  and by  $w_N$  the Lebesgue measure of the unit ball in  $R^N$ , we get:

$$|E| \leq \sum_{i=0}^{\infty} |B_i| = w_N \sum_{i=0}^{\infty} 2^{-iN} = \frac{w_N}{1 - 2^{-N}} < \infty.$$

Since rational points are dense in  $R^N$  we have  $\overline{E} = R^N$  and thus  $\partial E = \overline{E} - E = R^N - E$  has infinite Lebesgue measure which implies:

$$\mathcal{H}^{N-1}(\partial E) = +\infty.$$

On the other hand:

$$\begin{aligned} \mathcal{H}^{N-1}(\partial E_k) &\leq \mathcal{H}^{N-1}\left(\bigcup_{i=0}^k \partial B_i\right) = N w_{N-1} \sum_{i=0}^k 2^{-i(N-1)} \leq \\ &\leq N \frac{w_{N-1}}{1 - 2^{-(N-1)}} < +\infty. \end{aligned}$$

Therefore the sequence  $\{\mathcal{H}^{N-1}(\partial E_k)\}$  is bounded,  $E_k \rightarrow E$  ( $k \rightarrow +\infty$ ) in the sense of measures (or equivalently the sequence of characteristic functions  $\chi_{E_k}$  converges in  $L^1$  to  $\chi_E$ , however we do not have:

$$\mathcal{H}^{N-1}(\partial E) \leq \lim_{k \rightarrow +\infty} \mathcal{H}^{N-1}(\partial E_k).$$

This shows the necessity to find another formulation of  $F(u, K)$ . The new formulation involves the space  $BV(\Omega)$  of functions of bounded variation in  $\Omega$  (Section 2.2). The idea is to identify the set of edges  $K$  to the jump set  $S_u$  of  $u$ , which allows to eliminate the unknown  $K$ . So the idea is to consider the functional:

$$G(u) = \int_{\Omega} (u - u_0)^2 dx + \alpha \int_{\Omega} |\nabla u|^2 dx + \beta \mathcal{H}^{N-1}(S_u). \quad (4.1)$$

If we do not have a lower semi-continuity property with sets (see above remark), we are going to show that it can be obtained with functions. Now, it is tempting to minimize  $G$  on the space  $BV(\Omega)$ . Unfortunately the space  $BV(\Omega)$  may contain pathological non-constant functions which are continuous and have approximate differential equal to zero almost everywhere (a well-known example is the Cantor-Vitali function [10]). For such a function

$v$  we have:

$$G(v) = \int_{\Omega} (v - u_0)^2 dx \geq \inf_{u \in BV(\Omega)} G(u) \geq 0$$

and since these pathological functions are dense in  $L^2(\Omega)$  we get:

$$\inf_{u \in BV(\Omega)} G(u) = 0$$

which implies that the infimum of  $G$  cannot be achieved in  $BV(\Omega)$  in general. To avoid this phenomenon we must eliminate these pathological functions which have the particularity that their distributional derivatives are measures concentrated on Cantor sets. Let us recall that the distributional derivative  $Du$  of a  $BV(\Omega)$  function can be splitted into three mutually singular measures:

$$Du = \nabla u dx + (u^+ - u^-) n_u \mathcal{H}_{|S_u}^{N-1} + C_u$$

where  $J(u) = (u^+ - u^-) n_u \mathcal{H}_{|S_u}^{N-1}$  is the jump part and  $C_u$  the Cantor part. Following Di Giorgi [87, 86] we call  $SBV(\Omega)$  the space of special functions of bounded variation as the space of  $BV(\Omega)$  functions such that  $C_u = 0$ . Remark that Cantor-Vitali functions mentioned above do not belong to  $SBV(\Omega)$  since their support are mainly based on Cantor sets. Consequently, the suitable functional space to minimize (4.1) seems to be  $SBV(\Omega)$ . The natural question is now to establish the relation between the two problems:

$$\inf_{u,K} \left\{ \begin{array}{l} F(u, K), \quad u \in W^{1,2}(\Omega - K) \cap L^\infty(\Omega), \\ \quad K \subset \Omega, \quad K \text{ closed}, \quad \mathcal{H}^{N-1}(K) < \infty \end{array} \right\} \quad (P_1)$$

$$\inf_u \left\{ G(u), \quad u \in SBV(\Omega) \cap L^\infty(\Omega) \right\} \quad (P_2)$$

The answer can be found in Ambrosio [9] and is the consequence of the following theorem:

**Theorem 4.2.1** [9] *Let  $K \subset \Omega$  be a closed set so that  $\mathcal{H}^{N-1}(K) < \infty$  and let  $u \in W^{1,2}(\Omega - K) \cap L^\infty(\Omega)$ , then  $u \in SBV(\Omega)$  and  $S_u \subset K \cup L$  with  $\mathcal{H}^{N-1}(L) = 0$ .*

From Theorem 4.2.1 it follows that  $\inf P_2 \leq \inf P_1$ . By using compactness and lower semi-continuity theorems (see below) it can be shown that  $(P_2)$  has a solution  $u$ . For such a minimizer De Giorgi-Carriero-Leaci [88] proved that:

$$\mathcal{H}^{N-1}(\Omega \cap (\overline{S_u} - S_u)) = 0.$$

So by setting  $K = \Omega \cap \overline{S_u}$  we get a solution of  $(P_1)$  and  $\min(P_1) = \min(P_2)$ .

It remains to show that  $(P_2)$  has a solution. This is the direct consequence of the following theorem:

**Theorem 4.2.2 [11]** Let  $u_n \in SBV(\Omega)$  be a sequence of functions so that there exists a constant  $C$  with  $|u_n(x)| \leq C < \infty$  a.e.  $x \in \Omega$  and  $\int_{\Omega} |\nabla u_n|^2 dx + \mathcal{H}^{N-1}(S_{u_n}) \leq C$ , then there exists a subsequence  $u_{n_k}$  converging a.e.  $x$  to a function  $u \in SBV(\Omega)$ . Moreover  $\nabla u_{n_k}$  converges weakly in  $L^2(\Omega)^N$  to  $\nabla u$  and  $\liminf \mathcal{H}^{N-1}(S_{u_{n_k}}) \geq \mathcal{H}^{N-1}(S_u)$ .

We obtain a solution for  $(P_2)$  by applying Theorem 4.2.2 to any minimizing sequence of  $(P_2)$  and by remarking beforehand that we can restrict to minimizing sequences satisfying  $|u_n|_{L^\infty(\Omega)} \leq |u_0|_{L^\infty(\Omega)}$  (using a truncature argument).

► The  $SBV$  cluster points of sequences as defined in Theorem 4.2.2 are solutions of  $(P_2)$ . Remark that no uniqueness result is available. This will be illustrated in the coming section.

As we get the existence of a minimizer we would like now to compute it. The natural way to do that is to search for optimality conditions. Curiously, it is easier to establish them by using  $F(u, K)$  than  $G(u)$ . So, let us suppose that there exists a pair  $(u, K)$  solution of  $(P_1)$  i.e.

$$F(u, K) \leq F(v, K') \quad (4.2)$$

for all  $v \in W^{1,2}(\Omega - K') \cap L^\infty(\Omega)$ ,  $K' \subset \Omega$ ,  $K'$  closed,  $\mathcal{H}^{N-1}(K') < \infty$ . Moreover, let us suppose that  $(u, K)$  satisfies the Mumford and Shah conjecture:

$(C_1)$   $K$  is made of a finite number of  $C^{1,1}$ -curves  $\gamma_i$ , meeting  $\partial\Omega$  and meeting each other only at their endpoints.

$(C_2)$   $u$  is  $C^1$  on each connected component of  $\Omega - K$ .

**Theorem 4.2.3** Let  $(u, K)$  be a solution of  $(P_1)$  satisfying  $(C_1)$  and  $(C_2)$  then

$$\alpha \Delta u = u - u_0 \text{ on } \Omega \quad (4.3)$$

$$\frac{\partial u}{\partial N} = 0 \text{ on } \partial\Omega \text{ and on the two sides } \gamma_i^\pm \text{ of each } \gamma_i \quad (4.4)$$

$$e(u^+) - e(u^-) + \beta \operatorname{curv} \gamma_i = 0 \text{ on } \gamma_i \quad (4.5)$$

$$\text{where } e(u) = (u - u_0)^2 + \alpha |\nabla u|^2,$$

$u^+$  and  $u^-$  are the traces of  $u$  on each side of  $K$  (each side of  $\gamma_i$ ),

$\operatorname{curv} \gamma_i$  is the curvature of  $\gamma_i$ .

**Proof** The proof of (4.3) and (4.4) is standard. We first look at the variations of  $F$  with respect to  $u$ . In (4.2) we choose  $K' = K$  and  $v = u + \theta\varphi$  with  $\theta \in R$  and  $\varphi$  is a test function with compact support. Then:

$$\begin{aligned} 0 \leq F(u + \theta\varphi, K) - F(u, K) &= \theta^2 \int_{\Omega-K} (\varphi^2 + \alpha |\nabla\varphi|^2) dx + \\ &\quad + 2\theta \int_{\Omega-K} (\varphi(u - u_0) + \alpha \nabla\varphi \cdot \nabla u) dx \end{aligned} \quad (4.6)$$

Dividing (4.6) by  $\theta > 0$  (respectively by  $\theta < 0$ ) and letting  $\theta \rightarrow 0^+$  (respectively  $\theta \rightarrow 0^-$ ) we get:

$$0 = \int_{\Omega-K} \varphi(u - u_0) dx + \alpha \int_{\Omega-K} \nabla\varphi \cdot \nabla u dx, \quad \forall \varphi. \quad (4.7)$$

Choosing  $\varphi$  with compact support in  $\Omega - K$  and integrating the second integral by parts in (4.7) we obtain:

$$0 = \int_{\Omega-K} \varphi(u - u_0 - \alpha \Delta u) dx \quad \forall \varphi$$

i.e.

$$u - u_0 - \alpha \Delta u = 0 \text{ on } \Omega - K.$$

Now, multiplying (4.3) by a function  $\varphi \in C^1(\Omega)$  we easily obtain (4.4).

*To prove (4.5), the idea is to look at the variation of  $F$  with respect to  $K$ . We propose to give a slightly different proof than Mumford and Shah one's. The arguments we are going to use will be useful later for active contours. For the sake of clarity, we also look at a simpler version of the Mumford and Shah problem. We suppose that there is only one object in the scene, and that  $K$  is a closed  $C^{1,1}$ -curve.*

Let  $\Omega_{int}$  be the open set enclosed by  $K$  and  $\Omega_{ext} = \Omega - \Omega_{int} - K$ . Our aim is to consider variations of  $K$  according to the flow  $\frac{dx}{dt} = v(t, x)$  where  $v$  is an arbitrary velocity. We denote by  $K(t)$  such a variation,  $t \geq 0$ , with  $K(0) = K$ . Since  $u$  varies as  $K$  moves, we denote  $u(t, x)$  the unique solution of  $\inf_u F(u, K(t))$  and  $u_{int}(t, x) = u(t, x)|_{\Omega_{int}(t)}$ ,  $u_{ext}(t, x) = u(t, x)|_{\Omega_{ext}(t)}$ . So let:

$$f(t) = \int_{\Omega-K(t)} [(u(t, x) - u_0(x))^2 + \alpha |\nabla u(t, x)|^2] dx + \beta \int_{K(t)} d\sigma.$$

Writing  $\Omega = \Omega_{int}(t) \cup \Omega_{ext}(t) \cup K(t)$  we have

$$\begin{aligned} f(t) &= \int_{\Omega_{int}-K(t)} [(u_{int}(t, x) - u_0(x))^2 + \alpha |\nabla u_{int}(t, x)|^2] dx + \\ &+ \int_{\Omega_{ext}-K(t)} [(u_{ext}(t, x) - u_0(x))^2 + \alpha |\nabla u_{ext}(t, x)|^2] dx + \beta \int_{K(t)} d\sigma. \end{aligned}$$

We remark that in the expression of  $f(t)$  both the domain of integration and the integrands depend on  $t$ . As we are interested in the first variation of  $F$  we need to estimate  $f'(t)$ .

- For the first two integrals, we need to use a classical result on *derivative for domain integral*: if  $l(t, x)$  is a regular function defined on a bounded regular domain  $w(t)$  of  $R^N$  and if we set:

$$g(t) = \int_{w(t)} l(t, x) dx \quad (4.8)$$

then:

$$g'(t) = \int_{w(t)} \frac{\partial l}{\partial t}(t, x) dx + \int_{\partial w(t)} l(t, x) v \cdot N d\sigma \quad (4.9)$$

where  $\partial w(t)$  is the boundary of  $w(t)$ ,  $N$  is the unit outward normal to  $\partial w(t)$ , and  $v$  is the velocity of  $\partial w(t)$ .

- Regarding the last term, we also need to know how to estimate the derivative of the length. We can show that:

$$\frac{d}{dt} \left( \int_{K(t)} d\sigma \right) = \int_{K(t)} \text{curv} K(t) v \cdot N d\sigma.$$

The proof can be found in Section 4.3.2 (see (4.22)-(4.23)).

By applying the above results we get:

$$\begin{aligned}
f'(t) = & 2 \int_{\Omega_{int}(t)} (u_{int} - u_0) \frac{\partial u_{int}}{\partial t} dx + \int_{K(t)} (u_{int} - u_0)^2 v \cdot N d\sigma + \quad (4.10) \\
& + 2\alpha \int_{\Omega_{int}(t)} \nabla u_{int} \cdot \nabla \left( \frac{\partial u_{int}}{\partial t} \right) dx + \alpha \int_{K(t)} |\nabla u_{int}|^2 v \cdot N d\sigma + \\
& + 2 \int_{\Omega_{ext}(t)} (u_{ext} - u_0)^2 \frac{\partial u_{ext}}{\partial t} dx - \int_{K(t)} (u_{ext} - u_0)^2 v \cdot N d\sigma + \\
& + 2\alpha \int_{\Omega_{ext}(t)} \nabla u_{ext} \cdot \nabla \left( \frac{\partial u_{ext}}{\partial t} \right) dx - \alpha \int_{K(t)} |\nabla u_{ext}|^2 v \cdot N d\sigma + \\
& + \beta \int_{K(t)} \text{curv } K(t) v \cdot N d\sigma.
\end{aligned}$$

Thanks to Green's formula we have

$$\int_{\Omega_{int}(t)} \nabla u_{int} \cdot \nabla \left( \frac{\partial u_{int}}{\partial t} \right) dx = - \int_{\Omega_{int}} \Delta u_{int} \frac{\partial u_{int}}{\partial t} dx + \int_{K(t)} \frac{\partial u_{int}}{\partial t} \frac{\partial u_{int}}{\partial n} d\sigma,$$

but  $u_{int}(t, x)$  is the solution of

$$\begin{cases} \alpha \Delta u_{int}(t, x) = u_{int}(t, x) - u_0(x) \text{ in } \Omega_{int}(t) \\ \frac{\partial u_{int}}{\partial N} = 0 \text{ on } K(t) \end{cases}$$

Thus

$$\alpha \int_{\Omega_{int}(t)} \nabla u_{int} \cdot \nabla \left( \frac{\partial u_{int}}{\partial t} \right) dx = - \int_{\Omega_{int}(t)} (u_{int}(t, x) - u_0(x)) \frac{\partial u_{int}}{\partial t}(t, x) dx.$$

The same goes for

$$\alpha \int_{\Omega_{ext}(t)} \nabla u_{ext} \cdot \nabla \left( \frac{\partial u_{ext}}{\partial t} \right) dx = - \int_{\Omega_{ext}(t)} (u_{ext}(t, x) - u_0(x)) \frac{\partial u_{ext}}{\partial t}(t, x) dx.$$

Therefore by replacing these last expressions in (4.10) we get:

$$\begin{aligned}
f'(t) = & \int_{K(t)} ((u_{int} - u_0)^2 + \alpha |\nabla u_{int}|^2) v \cdot N d\sigma - \\
& - \int_{K(t)} ((u_{ext} - u_0)^2 + \alpha |\nabla u_{ext}|^2) v \cdot N d\sigma + \beta \int_{K(t)} \text{curv } K(t) v \cdot N d\sigma
\end{aligned}$$

or, with the notations of Theorem 4.2.3:

$$f'(t) = \int_{K(t)} (e(u_{int}) - e(u_{ext}) + \beta \operatorname{curv} K(t)) v \cdot N d\sigma.$$

Until now we have not specified how the variations of  $K(t)$  are. We choose that  $K$  moves along its outward normal according to the following differential equation

$$\begin{aligned} \frac{\partial x}{\partial t} &= v(t, x) = V(x(t)) N \\ x(0) &= K \end{aligned}$$

where  $V$  is an arbitrary velocity. Since  $|n|^2 = 1$ ,  $f'(t)$  writes as:

$$f'(t) = \int_{K(t)} (e(u_{int}(t, x)) - e(u_{ext}(t, x)) + \beta \operatorname{curv} K(t)) V(x(t)) d\sigma.$$

If  $(u, K)$  is a minimizer of the Mumford and Shah functional, we have necessarily  $f'(0) = 0$ , i.e.

$$0 = \int_K (e(u_{int}(x)) - e(u_{ext}(x)) + \beta \operatorname{curv} K) V(x) d\sigma$$

and since  $V$  is arbitrary we obtain:

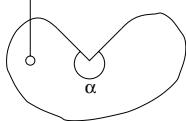
$$e(u_{int}) - e(u_{ext}) + \beta \operatorname{curv} K = 0 \text{ on } K.$$

This proves (4.5) in this simpler case. We let the reader convince himself that the above proof also runs in the general case. ■

**Remark** Before analyzing the edge set  $K$  it would be appropriate to say some words about the regularity of the function  $u$ . If  $K$  is supposed  $C^{1,1}$  and if  $u_0$  is continuous, then the standard theory of elliptic operator [105, 117, 127] implies that:

- $u$  is  $C^1$  on the open set  $\Omega - K$  and at all simple boundary point of  $K$  and  $\partial\Omega$ .
- $u$  extends locally to a  $C^1$ -function on the region plus its boundary.

Connected component of  $\Omega - K$



However problems can arise at corners: if  $P$  is a corner with an angle  $\alpha$  such that  $\pi < \alpha \leq 2\pi$  (including the exterior of crack, i.e. when  $P$  is the end point of a  $C^{1,1}$ -curve which is not continued by any other arc) then  $u$  has the form (in polar coordinates centered at  $P$ ):

$$u(r, \theta) = c r \frac{\pi}{\alpha} \sin\left(\frac{\pi}{\alpha}(\theta - \theta_0)\right) + \hat{v}(r, \theta) \quad (4.11)$$

where  $\hat{v}(r, \theta)$  is  $C^1$  and  $c, \theta_0$  are suitable constants. ■

### 4.2.3 Regularity of the edge set

In this section we look at qualitative properties verified by a minimizer of the Mumford and Shah functional and we examine in particular the structure of the set  $K$ . The first result was given by Mumford and Shah in 1989 [187]. It was a first step to solve their own conjecture (see Section 4.2.1).

**Theorem 4.2.4 [187]** *Let  $N = 2$ . If  $(u, K)$  is a minimizer of  $F(u, K)$  so that  $K$  is a union of simple  $C^{1,1}$ -curve  $\gamma_i$  meeting  $\partial\Omega$  and each other only at their endpoints, then the only vertices of  $K$  are:*

- (i) *Points  $P$  on  $\partial\Omega$  where one  $\gamma_i$  meets  $\partial\Omega$  perpendicularly.*
- (ii) *Triple points  $P$  where three  $\gamma_i$  meet with angles  $\frac{2\pi}{3}$ .*
- (iii) *Crack-tip where a  $\gamma_i$  ends and meets nothing.*

**Proof** (*Sketch of the proof*). As we can imagine, proving (i)-(iii) is not simple, and we do not reproduce all the details. We refer the interested reader to [187] for their instructive and illuminating constructions. Five steps can be distinguished:

- (A) Since we are concerned with the regularity of  $K$ , we deal with a local phenomenon and we just have to consider the energy inside a ball  $B(P, \varepsilon)$ . In fact, we say that  $(u, K)$  minimizes  $F$  if no change altering  $(u, K)$  inside a ball  $B$ , and leaving it unchanged outside, can decrease  $F$ .
- (B) One proves that  $K$  has no kinks, i.e. points  $P$  where two edges  $\gamma_i$  and  $\gamma_j$  meet at angle other than  $\pi$ .
- (C) One shows that  $\gamma_i$  meets  $\partial\Omega$  perpendicularly.
- (D) At triple points that is points  $P$  where three curves  $\gamma_i, \gamma_j, \gamma_k$  meet with angle  $\theta_{i,j}, \theta_{j,k}, \theta_{k,i}$  then we necessarily have  $\theta_{i,j} = \theta_{j,k} = \theta_{k,i} = \frac{2\pi}{3}$ .
- (E) Finally one proves there is no point where four or more  $\gamma_i$  meet at positive angle as well as cuspidal corner, i.e., corners where two arcs are tangent.

*The way to prove (B), ..., or (E) is always the same: by contradiction if (B), ..., or (E) were not satisfied then we can locally construct from  $(u, K)$  another pair  $(u', K')$  which decreases strictly  $F(u, K)$  and so contradicting that  $(u, K)$  is a minimizer in the sense of (A).*

To illustrate the above discussion let us show how (B) can be proved. Let  $P$  be a kink point and let  $B_\varepsilon = B(P, \varepsilon)$  be the ball of center  $P$  and of radius  $\varepsilon$ . Let us suppose that  $\gamma_i \cup \gamma_j$  divides  $B_\varepsilon$  into sectors  $B_\varepsilon^+$  with angle

$\alpha^+, 0 < \alpha^+ < \pi$  and  $B_\varepsilon^-$  with angle  $\alpha^-, \pi < \alpha^- < 2\pi$  (see Figure 4.4).

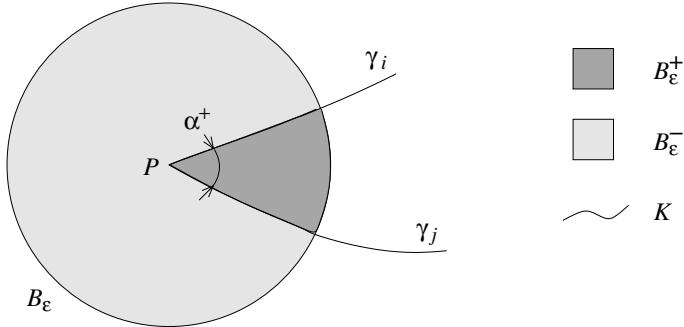


Figure 4.4. Definition of  $B_\varepsilon^+$  and  $B_\varepsilon^-$

Let us define the  $C^\infty$ -function  $\phi(x, y)$ ,  $0 \leq \phi \leq 1$  by:

$$\phi(x, y) = \begin{cases} 1 & \text{if } x^2 + y^2 \geq 1 \\ 0 & \text{if } x^2 + y^2 \leq \frac{1}{2} \end{cases}$$

and let:

$$\phi_\varepsilon(x, y) = \phi\left(\frac{x - x(P)}{\varepsilon}, \frac{y - y(P)}{\varepsilon}\right).$$

Now we are going to construct from  $(u, K)$  another admissible pair  $(u', K')$ . What we are doing is cutting the corner at  $P$  at a distance of  $\frac{\varepsilon}{2}$  shrinking  $B_\varepsilon^+$  and expanding  $B_\varepsilon^-$ . More precisely, the only change we are doing in  $K$  is to remove the curvilinear triangle  $PMN$  from  $B_\varepsilon^+$  and to add it to the set  $B_\varepsilon^-$  leaving unchanged the rest of  $K$  (see Figure 4.5). We call  $K'$  this new set of edges and  $B_\varepsilon^+$  the new  $B_\varepsilon^+$  (respectively  $B_\varepsilon^-$  the new  $B_\varepsilon^-$ ).

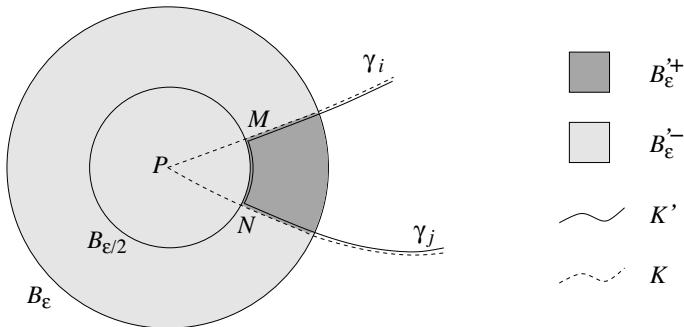


Figure 4.5. Construction of the other solution  $(u', K')$

Then we define  $u'(x, y)$  as follows:

$$u'(x, y) = \begin{cases} u(P) + \phi_\varepsilon(x, y)(u(x, y) - u(P)) & \text{on } B'_\varepsilon^- \\ u(x, y) & \text{otherwise} \end{cases}$$

and without a loss of generality we can suppose that  $u(P) = 0$ . Let us evaluate  $F(u', K') - F(u, K)$ . By construction, this difference reduces to:

$$\begin{aligned} F(u', K') - F(u, K) &= \int_{B'_\varepsilon^-} [(\phi_\varepsilon u - u_0)^2 - (u - u_0)^2] dx + \\ &\quad + \alpha \int_{B'_\varepsilon^-} [|\nabla(\phi_\varepsilon u)|^2 - |\nabla u|^2] dx + \beta \int_{K'} d\sigma - \beta \int_K d\sigma. \end{aligned}$$

We examine the first two terms separately (we will denote by  $c$  a universal constant):

$$\begin{aligned} A_{1\varepsilon} &= \int_{B'_\varepsilon^-} [(\phi_\varepsilon u - u_0)^2 - (u - u_0)^2] dx = \\ &= \int_{B'_\varepsilon^-} \left[ \frac{1 - \phi_\varepsilon}{1 + \phi_\varepsilon} u_0^2 - (1 - \phi_\varepsilon) \left( \sqrt{1 + \phi_\varepsilon} u - \frac{u_0}{\sqrt{1 + \phi_\varepsilon}} \right)^2 \right] dx \leq \\ &\leq \int_{B'_\varepsilon^-} u_0^2 dx \leq c \varepsilon^2 \\ A_{2\varepsilon} &= \alpha \int_{B'_\varepsilon^-} [|\nabla(\phi_\varepsilon u)|^2 - |\nabla u|^2] dx = \\ &= \alpha \int_{B'_\varepsilon^-} \left[ (\phi_\varepsilon^2 - 1) |\nabla u|^2 + u^2 |\nabla \phi_\varepsilon|^2 + 2u\phi_\varepsilon \nabla \phi_\varepsilon \cdot \nabla u \right] dx. \end{aligned}$$

But, thanks to (4.11) we have:

$$u = O\left(r^{\frac{\pi}{\alpha^-}}\right) \quad \text{and} \quad |\nabla u| = O\left(r^{\frac{\pi}{\alpha^-} - 1}\right)$$

and taking into account that  $\phi_\varepsilon^2 \leq 1$ ,  $|\nabla \phi_\varepsilon| \leq \frac{c}{\varepsilon}$ , and that  $\nabla \phi_\varepsilon = 0$  in the ball  $B\left(P, \frac{\varepsilon}{2}\right)$ , we obtain

$$A_{2\varepsilon} \leq c \varepsilon^2 \left( \frac{\varepsilon^{\frac{2\pi}{\alpha^-}}}{\varepsilon^2} + \varepsilon^{\frac{\pi}{\alpha^-}} \frac{1}{\varepsilon} \varepsilon^{\frac{\pi}{\alpha^-} - 1} \right) = c \varepsilon^{\frac{2\pi}{\alpha^-}}.$$

Finally in the third term  $\int_{K'} d\sigma - \int_K d\sigma$ , we are replacing asymptotically as  $\varepsilon \rightarrow 0$  the equal sides of an isosceles triangle with angle  $\alpha^+$  by the third

side, so

$$A_{3\varepsilon} = \int_{K'} d\sigma - \int_K d\sigma \leq \varepsilon \left( \sin \frac{\alpha^+}{2} - 1 \right) \leq 0$$

and we obtain

$$F(u', K') - F(u, K) = A_{1\varepsilon} + A_{2\varepsilon} + A_{3\varepsilon} \leq c \left( \varepsilon^2 + \varepsilon^{\frac{2\pi}{\alpha^-}} + \varepsilon \left( \sin \frac{\alpha^+}{2} - 1 \right) \right). \quad (4.12)$$

Since  $0 < \alpha^+ < \pi$  and  $\pi < \alpha^- < 2\pi$ , (4.12) shows that the energy decreases by order  $\varepsilon$  if  $\varepsilon$  is sufficiently small, which contradicts that  $(u, K)$  is a minimizer. We notice that the data term  $A_{1\varepsilon} = \int_{B_\varepsilon'^-} [(\phi_\varepsilon u - u_0)^2 - (u - u_0)^2] dx$

is of order  $\varepsilon^2$  and thus negligible with respect to the two other terms independently of the data  $u_0$ . ■

**Remark** This regularity conditions are very interesting and constrain the segmentation to verify some properties, at the cost of the fidelity to the image. For instance, if we consider the simple case depicted in Figure 4.6, the Theorem 4.2.4 shows that we can not get the exact segmentation with lines crossing at  $\pi/2$ . Qualitatively, we may obtain one of the two configurations described in Figure 4.6. This is a caricatural illustration of why uniqueness may not be true.

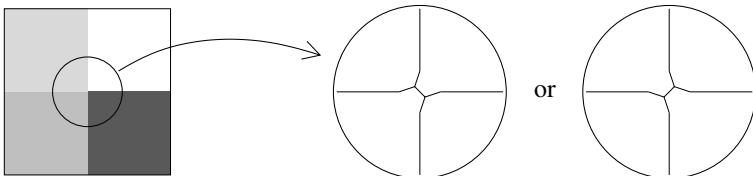


Figure 4.6. Illustration of “equivalent” segmentations for a given image. This is to show why regularity conditions on the edge set may have some influence on the uniqueness of the solution. ■

As said before, Theorem 4.2.4 is a first step in the proof of the Mumford and Shah conjecture. To go further we have to remove the assumption that  $K$  is made of a finite union of simple  $C^{1,1}$ -curves meeting  $\partial\Omega$  and each other only at their endpoints. A. Bonnet carried out an important progress in this direction. We only state his results and we refer the reader to [41, 42, 43] for the proofs.

**Theorem 4.2.5 [41, 43]** *If  $(u, K)$  is a minimizer of  $F$  so that  $K$  is connected, then  $(u, K)$  is one of the following:*

- (i)  *$K$  is empty and  $u$  is constant.*

- (ii)  $K$  is a straight line defining two half-planes and  $u$  is constant on each half-plane.
- (iii)  $K$  is the union of three half lines with  $\frac{2\pi}{3}$  angles and  $u$  is constant on each sector.
- (iv) In a polar set of coordinates  $(r, \theta)$ ,  $u(r, \theta) = \sqrt{\frac{2}{\pi}}\sqrt{r}\cos\frac{\theta}{2}$  for  $\theta \in [0, 2\pi[$  and  $K$  is the half-axis  $\theta = 0$  (a crack-tip).

We notice that conclusions of Theorem 4.2.5 and Theorem 4.2.4 are very similar. The major difference lies in the assumptions. In Theorem 4.2.4,  $K$  is supposed made of a finite union of  $C^{1,1}$ -arcs while in Theorem 4.2.5,  $K$  is supposed connected. Bonnet also proved the following result:

**Theorem 4.2.6** [42, 43] *Every isolated connected component of  $K$  is the union of a finite set of  $C^1$ -arcs. These arcs are  $C^{1,1}$  away from crack-tips and can merge through triple junctions with  $\frac{2\pi}{3}$  angles.*

We notice that Theorem 4.2.6 does not allow a minimizer to have an infinite number of arbitrary small pieces connected to each other. The proof of Theorem 4.2.6 relies on a characterization of the minimizers, which is very similar to the one given in Theorem 4.2.4.

⌘ We conclude this section by saying that the Mumford and Shah conjecture is always an open question in the most general setting, i.e. without connectedness constraint.

#### 4.2.4 Approximations of the Mumford and Shah functional

The lack of differentiability of the functional for a suitable norm does not allow to use, as it is classical, Euler-Lagrange equations. Moreover, the discretization of the unknown discontinuity set is a very complex task. A commonly used method is to approximate  $F(u, K)$  (or  $G(u)$ ) by a sequence  $F_\varepsilon$  of regular functionals defined on Sobolev spaces, the convergence of  $F_\varepsilon$  to  $F$  as  $\varepsilon \rightarrow 0$  being understood in the  $\Gamma$ -convergence framework (see Section 2.1.4). Of course, if we want to get an efficient approximation, the set  $K$  must not appear in  $F_\varepsilon$ . Four classes of approaches may be distinguished:

##### (A) Approximation by elliptic functionals [13].

In this approach the set  $S_u$  (or  $K$ ) is replaced by an auxiliary variable  $v$  (a function) that approximates the characteristic function  $(1 - \chi_{S_u})$  i.e.  $v(x) \approx 0$  if  $x \in S_u$  and  $v(x) \approx 1$  otherwise. Ambrosio and Tortorelli [13] proposed the following sequence of functionals:

$$F_\varepsilon(u, v) = \int_{\Omega} (u - u_0)^2 dx + \int_{\Omega} |v|^2 |\nabla u|^2 dx + \int_{\Omega} \left( \varepsilon |\nabla v|^2 + \frac{1}{4\varepsilon} (v - 1)^2 \right) dx$$

which is discussed afterwards.

- (B) *Approximations by introducing second order singular perturbations* [48] (see also [29, 56])

For instance:

$$F_\varepsilon(u) = \begin{cases} \int\limits_{\Omega} (u - u_0)^2 dx + \frac{1}{\varepsilon} \int\limits_{\Omega} f(\varepsilon |\nabla u|^2) dx + \varepsilon^3 \int\limits_{\Omega} |\nabla^2 u|^2 dx & \text{if } u \in W^{2,2}(\Omega) \\ +\infty \text{ if } u \in L^1(\Omega) - W^{2,2}(\Omega) \end{cases}$$

where  $W^{2,2}(\Omega)$  is the Sobolev space of  $L^2$ -functions whose distributional derivatives up to the second order belong to  $L^2(\Omega)$  and where  $f$  is a lower semi-continuous increasing function from  $[0, +\infty[$  to  $[0, +\infty[$  so that there exist  $\alpha, \beta \in \mathbb{R}$  such that:

$$\alpha = \lim_{s \rightarrow 0^+} \frac{f(s)}{s}, \quad \beta = \lim_{s \rightarrow +\infty} f(s),$$

and  $\nabla^2 u$  denotes the Hessian matrix of  $u$  equipped with the norm  $|A| = \max(\langle A\xi, \xi \rangle, |\xi| = 1)$ . In fact, for this kind of approximation one can prove [48] that  $F_\varepsilon$   $\Gamma$ -converges for the  $L^1$ -topology to a variant of the Mumford and Shah functional:

$$\tilde{G}(u) = \begin{cases} \int\limits_{\Omega} (u - u_0)^2 dx + \alpha \int\limits_{\Omega} |\nabla u|^2 dx + m(\beta) \int\limits_{S_u} \sqrt{u^+ - u^-} d\mathcal{H}^{N-1} & \text{if } u \in GSBV(\Omega) \\ +\infty \text{ otherwise} \end{cases}$$

where  $m(\beta) = \beta^{\frac{3}{4}} \left( 2\sqrt{\frac{3}{2}} + \sqrt{\frac{2}{3}} \right)$  and where  $GSBV(\Omega)$  is the space of  $L^1$ -functions  $u$  for which the truncated function  $u_T = -T \vee u \wedge T$  belongs to  $SBV(\Omega)$  for all  $T > 0$  ( $\vee$ , resp.  $\wedge$ , denotes the sup, resp. inf, operator).

- (C) *Approximation by introducing non-local terms* [49]

A typical example is:

$$F_\varepsilon(u) = \int\limits_{\Omega} (u - u_0)^2 dx + \frac{1}{\varepsilon} \int\limits_{\Omega} f\left(\varepsilon \int\limits_{B(x,\varepsilon)} |\nabla u(y)|^2 dy\right) dx$$

where  $f$  is a suitable non-decreasing continuous function and  $\int\limits_{B(x,\varepsilon)} h(y) dy$  denotes the mean value of  $h$  on the ball  $B(x, \varepsilon)$ .

Perhaps the motivation of introducing non-local approximation comes from the impossibility, as pointed out in [49] (see also [84]), to obtain

a variational approximation by means of local integral functionals of the form:

$$E_\varepsilon(u) = e \int_{\Omega} (u - u_0)^2 dx + \int_{\Omega} f_\varepsilon(\nabla u(x)) dx.$$

Indeed, if such an approximation was existing, the Mumford and Shah functional would be also the  $\Gamma$ -limit of the relaxed sequence:

$$RE_\varepsilon(u) = \int_{\Omega} (u - u_0)^2 dx + \int_{\Omega} f_\varepsilon^{**}(\nabla u(x)) dx$$

where  $f_\varepsilon^{**}$  is the convex envelope of  $f_\varepsilon$  (see Section 2.1.3). Therefore the Mumford and Shah functional would also be convex!

- (D) *Approximation by finite-difference schemes* [61, 62, 121]

This kind of approximation is perhaps the most natural one from a numerical point of view. The method consists in considering  $u(x)$  as a discrete image defined on a mesh of step-size  $h > 0$  and  $F^h$  as a discrete version of the Mumford and Shah functional. To the best of our knowledge, A. Chambolle proposed the first theoretical work in that direction [61, 62] following earlier ideas of Blake and Zissermann [39]. In the 1-D case, Let:

$$\begin{aligned} g_k^h &= \frac{1}{h} \int_{kh}^{(k+1)h} u_0(t) dt \\ u^h &= (u_k^h)_{kh \in \Omega} \text{ a given discrete signal.} \end{aligned}$$

Then Chambolle proposed the following discrete functional:

$$F^h(u^h) = h \sum_k W_h \left( \frac{u_{k+1}^h - u_k^h}{h} \right) + h \sum_k (u_k^h - g_k^h)^2 \quad (4.13)$$

where  $W_h(t) = \min \left( t^2, \frac{1}{h} \right)$ , and proved that  $F^h$   $\Gamma$ -converges to:

$$F(u) = \int_{\Omega} (u - u_0)^2 dx + \int_{\Omega-S_u} u'^2 dx + \text{card}(S_u) \text{ for } u \in SBV(\Omega).$$

This approximation can be adapted in dimension two for:

$$\begin{aligned} F^h(u^h) &= h^2 \sum_{k,l} W_h \left( \frac{u_{k+1,l}^h - u_{k,l}^h}{h} \right) + h^2 \sum_{k,l} W_h \left( \frac{u_{k,l+1}^h - u_{k,l}^h}{h} \right) + \\ &\quad + h^2 \sum_{k,l} (u_{k,l}^h - g_{k,l}^h)^2. \end{aligned} \quad (4.14)$$

A similar  $\Gamma$ -convergence result can be proved but the 1-D Hausdorff measure is changed into an anisotropic 1-D measure which takes into

account the lack of rotational invariance of the natural 2-D extension of (4.14). We refer the reader to [61, 62] for more complete proofs as well as other related works [212, 47, 30].

It is beyond the scope of this book to examine in details these four ways to approximate the Mumford and Shah functional. That would surely make more than one hundred pages! We choose instead to focus on the approximation (A), proposed by Ambrosio and Tortorelli [13], which is the first one that appeared in literature, and which is commonly used in vision. We consider here the case  $N = 2$  and we set the parameters  $\alpha$  and  $\beta$  to 1:

$$F_\varepsilon(u, v) = \int_{\Omega} (u - u_0)^2 dx + \int_{\Omega} v^2 |\nabla u|^2 dx + \int_{\Omega} \left( \varepsilon |\nabla v|^2 + \frac{1}{4\varepsilon} (v - 1)^2 \right) dx.$$

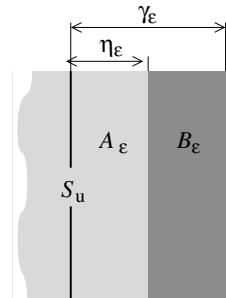
Before stating rigorous mathematical results, let us show thanks to intuitive arguments how  $F_\varepsilon$  approaches the Mumford and Shah functional. We reproduce a good explanation given by March [168]. Since the discontinuity set  $S_u$  is of zero-Lebesgue measure (and so  $v(x)$  would be equal to 1 a.e.), our aim is to construct a sequence of functions  $(u_\varepsilon, v_\varepsilon)$  converging to  $(u, 1)$  so that the sequence  $F_\varepsilon(u_\varepsilon, v_\varepsilon)$  converges to  $G(u)$ .

Let us fix some notations. We denote by:

- $\tau(x)$  the distance of the point  $x$  to  $S_u$ ,
- $A_\varepsilon = \{x ; \tau(x) < \eta(\varepsilon)\}$ ,  $B_\varepsilon = \{x ; \eta(\varepsilon) < \tau(x) < \gamma(\varepsilon)\}$ , with  $\lim_{\varepsilon \rightarrow 0} \eta(\varepsilon) = \lim_{\varepsilon \rightarrow 0} \gamma(\varepsilon) = 0$ . We restrict our construction to functions  $u_\varepsilon$  so that  $u_\varepsilon$  are smooth on  $A_\varepsilon$  and  $u_\varepsilon = u$  outside  $A_\varepsilon$ . For the control functions, since  $\nabla u_\varepsilon$  blows up near  $S_u$ ,  $v_\varepsilon$  have to be small in  $A_\varepsilon$ . We choose  $v_\varepsilon = 0$  on  $A_\varepsilon$ ,  $v_\varepsilon = 1 - w(\varepsilon)$  outside  $A_\varepsilon \cup B_\varepsilon$ , with  $\lim_{\varepsilon \rightarrow 0} w(\varepsilon) = 0$ . Finally we impose that  $v_\varepsilon$  be smooth in the whole domain  $\Omega$ .

It is easy to verify that the first two integrals in  $F_\varepsilon$  converge to the first two integrals in  $G$ . The treatment of the third term is more delicate. By construction we have:

$$\begin{aligned} & \int_{\Omega} \varepsilon |\nabla v_\varepsilon|^2 dx + \int_{\Omega} \frac{1}{4\varepsilon} (1 - v_\varepsilon)^2 dx = \\ &= \frac{1}{4\varepsilon} \left[ \int_{A_\varepsilon} dx + w(\varepsilon)^2 \int_{\Omega - A_\varepsilon \cup B_\varepsilon} dx \right] + \int_{B_\varepsilon} \left( \varepsilon |\nabla v_\varepsilon|^2 + \frac{1}{4\varepsilon} (1 - v_\varepsilon)^2 \right) dx. \end{aligned} \quad (4.15)$$



If we choose  $\eta(\varepsilon)$  and  $w(\varepsilon)$  so that  $\frac{\eta(\varepsilon)}{\varepsilon}$  and  $\frac{w(\varepsilon)^2}{4\varepsilon}$  go to zero as  $\varepsilon \rightarrow 0$ , then the first two integrals in the right-hand side of (4.15) converge to zero.

Thus, it remains to study the limit of  $R_\varepsilon = \int_{B_\varepsilon} \left( \varepsilon |\nabla v_\varepsilon|^2 + \frac{1}{4\varepsilon} (1 - v_\varepsilon)^2 \right) dx$ .

We search for  $v_\varepsilon$  in  $B_\varepsilon$  as a function of the form  $v_\varepsilon(x) = \sigma_\varepsilon(\tau(x))$  with  $\sigma_\varepsilon : R^+ \rightarrow R^+$ . With this choice, since  $|\nabla\tau| = 1$ ,  $R_\varepsilon$  reduces to:

$$R_\varepsilon = \int_{B_\varepsilon} \left( \varepsilon \sigma'_\varepsilon^2(\tau(x)) \right) + \frac{1}{4\varepsilon} (1 - \sigma_\varepsilon(\tau(x))^2) dx.$$

By noting  $t = \tau(x)$  and  $g(t) = \mathcal{H}^1\{x; \tau(x) = t\}$ , we get:

$$R_\varepsilon = \int_{\eta(\varepsilon)}^{\gamma(\varepsilon)} \left( \varepsilon \sigma'_\varepsilon^2(t) + \frac{1}{4\varepsilon} (1 - \sigma_\varepsilon(t)^2) \right) g(t) dt.$$

Then let us define  $\sigma_\varepsilon$  as the solution of the Ordinary Differential Equation:

$$\begin{cases} \sigma'_\varepsilon(t) = \frac{1}{2\varepsilon} (1 - \sigma_\varepsilon(t)) \\ \sigma_\varepsilon(\eta(\varepsilon)) = 0. \end{cases}$$

An elementary calculus gives:

$$\sigma_\varepsilon(t) = 1 - \exp\left(\frac{\eta(\varepsilon) - t}{2\varepsilon}\right)$$

and  $R_\varepsilon$  rewrites as:

$$R_\varepsilon = \frac{1}{2\varepsilon} \int_{\eta(\varepsilon)}^{\gamma(\varepsilon)} \exp\left(\frac{\eta(\varepsilon) - t}{\varepsilon}\right) g(t) dt.$$

Thanks to the mean value Theorem, there exists  $t_0 \in ]\eta(\varepsilon), \gamma(\varepsilon)]$  such that

$$R_\varepsilon = \frac{g(t_0)}{2\varepsilon} \int_{\eta(\varepsilon)}^{\gamma(\varepsilon)} \exp\left(\frac{\eta(\varepsilon) - t}{\varepsilon}\right) g(t) dt = \frac{g(t_0)}{2} \left( 1 - \exp\left(\frac{\eta(\varepsilon) - \gamma(\varepsilon)}{\varepsilon}\right) \right).$$

Choosing  $\gamma(\varepsilon)$  so that  $\lim_{\varepsilon \rightarrow 0} \frac{\gamma(\varepsilon)}{\varepsilon} = +\infty$ , and observing that  $g(t_0)$  converges to  $\mathcal{H}^1(S_u)$ , we get

$$\lim_{\varepsilon \rightarrow 0} R_\varepsilon = \mathcal{H}^1(S_u).$$

To conclude, we have just constructed a sequence  $(u_\varepsilon, v_\varepsilon)$  approaching  $(u, 1)$  and such that  $\lim_{\varepsilon \rightarrow 0} F_\varepsilon(u_\varepsilon, v_\varepsilon) = G(u)$ . Naturally, this does not prove the  $\Gamma$ -convergence of  $F_\varepsilon$  to  $G$  but it gives an idea of the methodology. It is also a way to check how we formally obtain the expected limit.

Now, let us go back to a more rigorous discussion. We first have to prove that  $F_\varepsilon(u, v)$  admits a minimizer and then that  $F_\varepsilon(u, v)$   $\Gamma$ -converges to the Mumford and Shah functional. Let us fix  $\varepsilon > 0$ . Then  $F_\varepsilon(u, v)$  is well-defined on the space  $V = \{(u, v) \in W^{1,2}(\Omega)^2; 0 \leq v \leq 1\}$  and it is weakly (i.e. for the weak topology) lower semi-continuous on this space. To obtain the existence of a minimizer it suffices to bound on  $V$  any minimizing sequence  $(u_\varepsilon^n, v_\varepsilon^n)$  independently of  $n$ . We easily bound in  $L^2(\Omega)$  the sequences

$u_\varepsilon^n$ ,  $v_\varepsilon^n$ ,  $\nabla v_\varepsilon^n$  but a difficulty arises when we want to bound  $\nabla u_\varepsilon^n$  since we have no control on the term  $\int_{\Omega} (v_\varepsilon^n)^2 |\nabla u_\varepsilon^n|^2 dx$ . To bypass this difficulty we slightly modify  $F_\varepsilon(u, v)$  by adding a perturbation:

$$\tilde{F}_\varepsilon(u, v) = F_\varepsilon(u, v) + h(\varepsilon) \int_{\Omega} |\nabla u|^2 dx$$

where  $h(\varepsilon) > 0$  is a suitable constant such that  $\lim_{\varepsilon \rightarrow 0} h(\varepsilon) = 0$ . With this modification, it is now clear that  $\tilde{F}_\varepsilon(u, v)$  is coercive on  $V$  and we have shown the following theorem:

**Theorem 4.2.7** *Let us suppose that  $u_0 \in L^\infty(\Omega)$  then the problem  $\inf_V \tilde{F}_\varepsilon(u, v)$  admits a solution  $(u_\varepsilon, v_\varepsilon)$  with  $|u_\varepsilon|_{L^\infty(\Omega)} \leq |u_0|_{L^\infty(\Omega)}$ .*

When  $\varepsilon \rightarrow 0$ , we have the following  $\Gamma$ -convergence result:

**Theorem 4.2.8 [13, 48]** *Let  $\tilde{F}_\varepsilon : L^1(\Omega) \times L^1(\Omega) \rightarrow [0, +\infty]$  be defined by  $\tilde{F}_\varepsilon(u, v) =$*

$$\begin{cases} \int_{\Omega} (u - u_0)^2 dx + \int_{\Omega} (v^2 + h(\varepsilon)) |\nabla u|^2 dx + \\ \quad + \int_{\Omega} \left( \varepsilon |\nabla v|^2 + \frac{1}{4\varepsilon} (1-v)^2 \right) dx & \text{if } (u, v) \in W^{1,2}(\Omega)^2, 0 \leq v \leq 1 \\ +\infty & \text{otherwise} \end{cases}$$

and let  $G : L^1(\Omega) \times L^1(\Omega) \rightarrow [0, +\infty]$  be defined by

$$G(u) = \begin{cases} \int_{\Omega} (u - u_0)^2 dx + \int_{\Omega} |\nabla u|^2 dx + \mathcal{H}^1(S_u) & \text{if } u \in GSBV(\Omega) \\ +\infty & \text{otherwise.} \end{cases} \quad \text{and } v = 1 \text{ a.e.}$$

If  $h(\varepsilon) = o(\varepsilon)$  then  $\tilde{F}_\varepsilon(u, v)$   $\Gamma$ -converges to  $G(u, v)$  for the  $L^1(\Omega)^2$ -strong topology. Moreover,  $\tilde{F}_\varepsilon$  admits a minimizer  $(u_\varepsilon, v_\varepsilon)$  so that up to subsequences,  $u_\varepsilon$  converges in  $L^1(\Omega)$  to a minimizer of  $G$ ,  $u \in SBV(\Omega)$ , and  $\inf \tilde{F}_\varepsilon \rightarrow \inf G(u)$  ( $\varepsilon \rightarrow 0$ ).

The proof of Theorem 4.2.8 is long and rather technical. We refer the interested reader to Braides [48].

#### 4.2.5 Experimental results

A natural method to compute numerically a solution of the Mumford and Shah functional is to consider one of the approximation (A)-(D) described

in Section 4.2.4, and then to write the discretized version of it. This can be achieved by using a finite difference scheme [62, 61, 121, 237] (see also the Appendix), or a finite element scheme [30, 47].

We present in Figure 4.7 some experimental results obtained with a modified Mumford and Shah functional, namely the piecewise constant model.

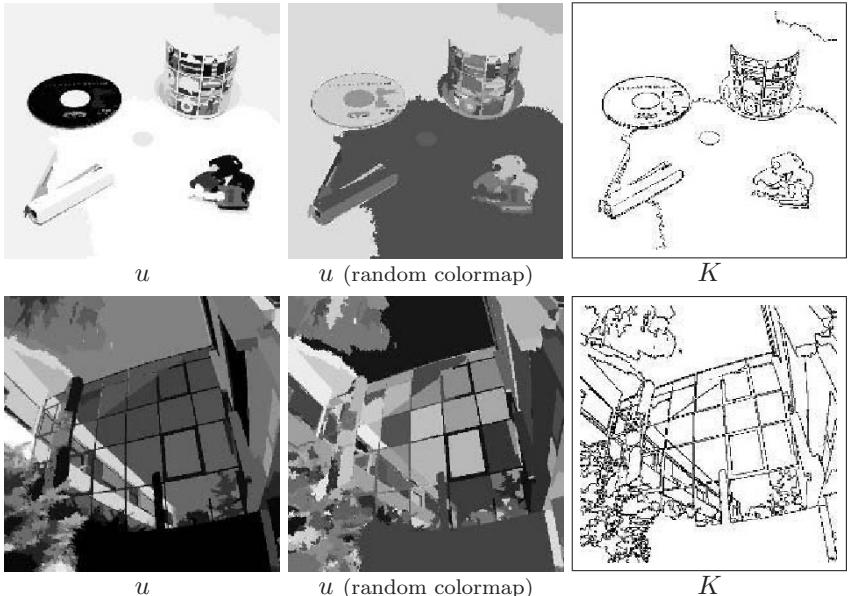


Figure 4.7. Result of the segmentation of the images “objects” and “Borel building” using the region growing approach from *Megawave2*. The solution ( $u, K$ ) is displayed. Notice that the middle image is displayed with a random colormap just to have a better idea of the different regions. Created in 1993 by Jacques Froment at the CEREMADE, University of Paris 9 Dauphine, *MegaWave2* is now directed by the CMLA laboratory of the Ecole Normale Supérieure de Cachan. It can be downloaded from <http://www.cmla.ens-cachan.fr/Cmla/Megawave/>.

The only change with respect to the Mumford and Shah functional is that we only consider two terms in the energy: the fitting term to the data and the measure of the discontinuity set  $K$ :

$$F(u, K) = \int_{\Omega - K} (u - u_0)^2 \, dx + \beta \int_K \, ds.$$

The strategy to minimize  $F$  is relatively simple and belongs to the class of region merging (or region growing) methods (see [213]). It consists in:

- Remarking that, given a segmentation  $\Omega^0 = \bigcup_{i=1}^M (\Omega_i^0 \cup K^0)$ , then the corresponding minimizer  $u$  of  $F(u, K^0)$  is a piecewise constant function where the constants are the averages of  $u_0$  over each  $\Omega_i^0$ .
- Merging recursively all pairs of regions whose merging decreases the energy  $F$ , the coefficient  $\beta > 0$  playing the role of a scale parameter.

## 4.3 Geodesic active contours and the level sets method

In this section we examine the snake and the geodesic active contours models. Unlike the Mumford and Shah functional, the aim is no longer to find a partition of the image but to automatically detect contours of objects. This raises two questions: how a contour may be represented and which criterion would permit to select the true contours. In many methods of image detection one supposes a sharp variation of the image intensity  $I(x)$  between the background of the scene and the objects. Therefore the magnitude of the gradient of  $I$  is high across the boundaries of objects and we may choose  $|\nabla I(x)|$  (or a function of it) as a detector of contours. There exists an extensive literature on snakes and geodesic active contours and the method is by now widely used in image analysis. Though the theory may be applicable both in two and three dimensions, we only develop the 2-D case for the sake of simplicity.

 Active contours, snakes, level sets [38, 229, 198, 223]

### 4.3.1 The Kass-Witkin-Terzopoulos model [142]

We begin by describing the Kass-Witkin-Terzopoulos model [142] which is to the best of our knowledge one of the first work in this direction. We note  $\Gamma$  the set of the image edges (the boundaries of objects). We suppose that  $\Gamma = \bigcup_{j \in J} C_j$ ,  $J$  finite or countable, where each  $C_j$  is a piecewise  $\mathcal{C}^1$  closed curve in  $R^2$ . Concerning the intensity  $I: \Omega \subset R^2 \rightarrow R$  ( $\Omega$  bounded) we assume that the function  $x = (x_1, x_2) \rightarrow |\nabla I(x_1, x_2)|$  belongs to  $W^{1,\infty}(\Omega)$ . In order to characterize edges by zero values rather than by infinite values we define a function  $g: [0, +\infty[ \rightarrow ]0, +\infty[$  satisfying:

- $g$  is regular monotonic decreasing.
- $g(0) = 1$  ,  $\lim_{s \rightarrow +\infty} g(s) = 0$ .

The function  $x \rightarrow g(|\nabla I(x)|)$  is called an edge detector function. A typical choice of  $g$  is  $g(s) = \frac{1}{1+s^2}$  (see Figure 4.8).



Figure 4.8. Example of edge detector function.  $g(s) = \frac{1}{1+s^2}$  has been chosen for the right-hand side image. Noise is reduced and contours are enhanced. Also note the inversion of colors due to the function  $g$ .

In Kass et al [142], boundary detection consists in matching a deformable model to an image by means of energy minimization. Because of the way the contours move while minimizing the energy they called them *snakes*. Let  $C$  be the set of curves of  $R^2$  defined by:

$$C = \{c : [a, b] \rightarrow \Omega, c \text{ piecewise } C^1, c(a) = c(b)\}$$

then for  $c \in C$  let  $J(c)$  be the following energy:

$$J(c) = \underbrace{\int_a^b |c'(q)|^2 dq + \beta \int_a^b |c''(q)|^2 dq}_{\text{internal energy}} + \underbrace{\lambda \int_a^b g^2(|\nabla I(c(q))|) dq}_{\text{external energy}}, \quad (4.16)$$

where  $c(q) = (c_1(q), c_2(q))$ ,  $c'(q) = \left(\frac{dc_1}{dq}, \frac{dc_2}{dq}\right)$ ,  $|c'(q)| = \sqrt{\left(\frac{dc_1}{dq}\right)^2 + \left(\frac{dc_2}{dq}\right)^2}$  and with the same notations for  $c''$ . The first two terms called spline or internal energy are used to impose a smoothness constraint. The first-order term makes the curve act like a membrane and the second-term makes it act like a thin plate. Setting  $\beta = 0$  allows second order discontinuities as corners. The third term, the external energy, attracts the curve towards the edges of the objects. As  $\Omega$  is bounded, it is easy to show that the energy  $J(c)$  admits at least a global minimum in the Sobolev space  $(W^{2,2}(a, b))^2$ . The Euler-Lagrange equations associated to  $J(c)$  are a fourth-order system:

$$\begin{cases} -c'' + \beta c^{(iv)} + \lambda \nabla F|_c(c) = 0 \\ c(a) = c(b) \end{cases} \quad (4.17)$$

where  $c^{(iv)}$  is the fourth order derivative and  $F(c_1, c_2) = g^2(|\nabla I(c_1, c_2)|)$ . Other boundaries conditions can be added.

Unfortunately, since  $J(c)$  is nonconvex (it is lower semi-continuous on  $(W^{2,2}(a,b))^2$ ), no uniqueness result is available and by solving (4.17) (as done in Kass et al [142]) we can only hope to reach a local minimum.

- In this approach, the main idea was to formulate the problem as a minimization one.

However, this approach has real drawbacks:

- The functional  $J(c)$  is not intrinsic since it depends on the parameterization of  $c$ . We could obtain different solutions by changing the parameterization while conserving the same initial curve.
- Because of the regularity constraint, the model does not handle changes of topology. In fact it is not possible to detect more than one object. Moreover, this object has to be convex.
- In practice, to solve numerically the problem we embed (4.17) into a dynamical scheme by making the curve depend on an artificial parameter (the time)  $t \geq 0$ , that is we solve:

$$\begin{cases} \frac{\partial c}{\partial t}(q, t) = -c''(q, t) + \beta c^{(iv)}(q, t) + \lambda \nabla F|_c(c(q, t)) \\ c(q, 0) = c_0(q) \\ c(a, t) = c(b, t) \end{cases} \quad (4.18)$$

where  $c_0(q)$  is an initial curve surrounding the object to be detected. Numerical problems arise when solving (4.18). As we can only reach a local minimum we have to choose  $c_0(q)$  close enough to the object to be detected. Another difficult task is the choice of a set of marker points when discretizing the parameterized evolving curve. The positions of the marker points have to be updated in time according to the approximations in the equations of the motion. For large and complex motions several problems occur. For example, concentration, or on the contrary, void regions can be created causing numerical instabilities and false detections. A good explanation of such phenomenon is given in [227].

We are going now to show how the above difficulties can be overcome with the *geodesic active contour model*.

### 4.3.2 The Caselles-Kimmel-Sapiro geodesic active contours model [58]

In the Kass et al model (4.16), the term  $\beta \int_a^b |c''(q)|^2 dq$  (the elasticity term) is a second order smoothness component which minimizes the squared curvature. As we will see later the model with  $\beta = 0$  also decreases the curvature making this term redundant (see (4.24)). It is then natural to introduce the functional  $J_1$  defined by:

$$J_1(c) = \int_a^b |c'(q)|^2 dq + \lambda \int_a^b g^2(|\nabla I(c(q))|) dq \quad (4.19)$$

on the set  $C = \{c : [a, b] \rightarrow \Omega, c \text{ piecewise } C^1, c(a) = c(b)\}$ . Still, the functional  $J_1$  is not yet satisfactory because it is not intrinsic, that is it depends on the parametrization of  $c$ . So the idea is to introduce the functional  $J_2$  defined by:

$$J_2(c) = 2\sqrt{\lambda} \int_a^b g(|\nabla I(c(q))|) |c'(q)| dq. \quad (4.20)$$

It is easy to see that  $J_2$  is now intrinsic: if we define a new parametrization of the curve via  $q = \phi(r), \phi : [c, d] \rightarrow [a, b], \phi' > 0$ , we obtain:

$$J_2(c) = 2\sqrt{\lambda} \int_c^d g(|\nabla I(\bar{c}(r))|) |\bar{c}'(r)| dr$$

with  $\bar{c}(r) = c(\phi(r))$  i.e. there is no change in the energy. Therefore, if we compare  $J_2$  to the classical length definition of a curve<sup>1</sup>, we observe that  $J_2$  can be seen as a new length by weighting the Euclidian length. The weight is  $g(|\nabla I(c(q))|)$  which contains the information regarding the objects boundaries. In other words we have defined a new metric (a Riemannian metric) for which we search for geodesics. Beyond this geometric argument, we will see that this formulation also enables to apply very efficient numerical schemes.

☞ Starting from  $J_1$ , we have introduced a functional  $J_2$  which is intrinsic and that can be interpreted as a weighted Euclidian length. Now the question is to understand the link between the two minimization problems.

<sup>1</sup>The length of a curve is defined by  $L = \int_a^b |c'(q)| dq$

In [58], Caselles et al have shown, by using concepts of *Hamiltonian theory*, that minimizing  $J_1$  was “equivalent” to minimize  $J_2$ . This idea has been widely re-used in the sequel to justify this choice. However, it is unsatisfactory regarding two concerns. The first is that the notion of equivalence is not even clear. It would be natural to say that two minimization problems are equivalent if they have the same solutions or possibly if they have the same extremals. In our case we cannot apply these notions since it is not clear whether the problem  $\inf_c J_2(c)$  has a solution or not. The edge function  $g$  being degenerated in any neighborhood of an edge, it would be difficult to bound minimizing sequences in any reasonable space. The second is that we may wonder why it is necessary to use concepts from the Hamiltonian theory and if it could be possible instead to use classical techniques of the calculus of variations.

In this direction, Aubert and Blanc-Féraud [15] defined a precise notion of equivalence and proved it in this context. We do not reproduce the whole discussion of [15] but we will only state what can be, in our opinion, a right definition of equivalence. Before setting that definition we need to study the variations of the energies in a neighborhood of a given curve  $c(q)$ . Calculi developed below are set in details since they will be useful afterwards.

Let  $c(q) \in C$  and let  $c(q, t)$  be a family of curves, where  $t \geq 0$  is an exterior parameter (the time), such that  $c(q, 0) = c(q)$ . Let us note  $J_i(t) = J_i(c(q, t))$ ,  $i = 1, 2$ . The first step consists in computing  $J'_i(t)$ ,  $i = 1, 2$ .

- Calculus of  $J'_1(t)$ . We have:

$$J_1(c) = \int_a^b \left| \frac{\partial c}{\partial q}(q, t) \right|^2 dq + \lambda \int_a^b g^2(|\nabla I(c(q, t))|) dq.$$

In order to simplify the notations, we write  $c$  instead of  $c(q, t)$ ,  $g$  for  $g(|\nabla I(c(q, t))|)$  and we suppose that  $\lambda = 1$ . We will also denote  $u \cdot v = \langle u, v \rangle$ . Thus:

$$\frac{1}{2} J'_1(t) = \int_a^b \left\langle \frac{\partial c}{\partial q}, \frac{\partial^2 c}{\partial t \partial q} \right\rangle dq + \int_a^b \left\langle \frac{\partial c}{\partial t}, g \nabla g \right\rangle dq.$$

By integrating by parts with respect to  $q$  the first integral (we assume that  $c(a, t) = c(b, t)$  and  $\frac{\partial c}{\partial q}(a, t) = \frac{\partial c}{\partial q}(b, t)$  for all  $t > 0$ ) we get:

$$\frac{1}{2} J'_1(t) = \int_a^b \left\langle \frac{\partial c}{\partial t}, -\frac{\partial^2 c}{\partial q^2} + g \nabla g \right\rangle dq.$$

Denoting by  $s = s(q) = \int_a^q \left| \frac{\partial c}{\partial q}(\tau) \right| d\tau$  the arc length, then we always have for any curve  $c$ :

$$\frac{\partial^2 c}{\partial q^2} = \left| \frac{\partial c}{\partial q} \right|^2 \frac{\partial^2 c}{\partial s^2} + \langle T, \frac{\partial^2 c}{\partial q^2} \rangle T$$

where  $T$  denotes the unit tangent vector:  $T = \frac{\partial c}{\partial q} / \left| \frac{\partial c}{\partial q} \right|$ . Let  $N$  be the unit normal vector to the curve and let  $\kappa$  be the curvature then we have (see section 2.4):

$$\frac{\partial^2 c}{\partial s^2} = \kappa N$$

thus  $J'_1(t)$  becomes:

$$\frac{1}{2} J'_1(t) = \int_a^b \left\langle \frac{\partial c}{\partial t}, -\kappa \left| \frac{\partial c}{\partial q} \right|^2 N - \langle T, \frac{\partial^2 c}{\partial q^2} \rangle T + g \nabla g \right\rangle dq.$$

If we decompose the vector  $\nabla g$  in the tangential and normal directions:

$$\nabla g = \langle \nabla g, N \rangle N + \langle \nabla g, T \rangle T$$

we obtain

$$\frac{1}{2} J'_1(t) = \int_a^b \left\langle \frac{\partial c}{\partial t}, \left[ -\kappa \left| \frac{\partial c}{\partial q} \right|^2 + \langle g \nabla g, N \rangle \right] N + \left[ \langle g \nabla g, T \rangle - \langle T, \frac{\partial^2 c}{\partial q^2} \rangle \right] T \right\rangle dq$$

therefore thanks to the Cauchy-Schwarz inequality the flow for which  $J_1(t)$  decreases most rapidly is given by:

$$\frac{\partial c}{\partial t} = \left[ \kappa \left| \frac{\partial c}{\partial q} \right|^2 - \langle g \nabla g, N \rangle \right] N - \left[ \langle g \nabla g, T \rangle - \langle T, \frac{\partial^2 c}{\partial q^2} \rangle \right] T. \quad (4.21)$$

- Calculus of  $J'_2(t)$ . We have:

$$\frac{1}{2} J_2(t) = \int_a^b g(|\nabla I(c(q, t))|) \left| \frac{\partial c}{\partial q}(q, t) \right| dq \quad (4.22)$$

then

$$\frac{1}{2} J'_2(t) = \int_a^b g \left\langle \frac{\frac{\partial c}{\partial q}}{\left| \frac{\partial c}{\partial q} \right|}, \frac{\partial^2 c}{\partial t \partial q} \right\rangle dq + \int_a^b \left| \frac{\partial c}{\partial q} \right| \langle \nabla g, \frac{\partial c}{\partial t} \rangle dq.$$

The first integral of the right-hand side is integrated by parts with respect to  $q$ :

$$\frac{1}{2} J'_2(t) = - \int_a^b \left[ \left\langle g \frac{\partial}{\partial q} \left( \frac{\frac{\partial c}{\partial q}}{\left| \frac{\partial c}{\partial q} \right|} \right) + \frac{\frac{\partial c}{\partial q}}{\left| \frac{\partial c}{\partial q} \right|} \langle \nabla g, \frac{\partial c}{\partial q} \rangle, \frac{\partial c}{\partial t} \right\rangle \right] dq + \int_a^b \left| \frac{\partial c}{\partial q} \right| \langle \nabla g, \frac{\partial c}{\partial t} \rangle dq.$$

This equation can be rewritten as:

$$\frac{1}{2} J'_2(t) = \int_a^b \left| \frac{\partial c}{\partial q} \right| \left\langle \frac{\partial c}{\partial t}, \nabla g - \frac{1}{\left| \frac{\partial c}{\partial q} \right|} \frac{\partial}{\partial q} \left( \frac{\frac{\partial c}{\partial q}}{\left| \frac{\partial c}{\partial q} \right|} \right) g - \frac{\frac{\partial c}{\partial q}}{\left| \frac{\partial c}{\partial q} \right|} \langle \nabla g, \frac{\partial c}{\partial q} \rangle \right\rangle dq$$

and remembering the definitions of  $T$ ,  $N$  and  $\kappa$  we get:

$$\frac{1}{2} J'_2(t) = \int_a^b \left| \frac{\partial c}{\partial q} \right| \left\langle \frac{\partial c}{\partial t}, \nabla g - \kappa g N - \langle T, \nabla g \rangle T \right\rangle dq.$$

Decomposing again  $\nabla g$  on the basis  $(N, T)$  we finally obtain:

$$\frac{1}{2} J'_2(t) = \int_a^b \left| \frac{\partial c}{\partial q} \right| \left\langle \frac{\partial c}{\partial t}, \langle \nabla g, N \rangle N - \kappa g N \right\rangle dq \quad (4.23)$$

so the direction for which  $J_2(t)$  decreases most rapidly is given by:

$$\frac{\partial c}{\partial t} = (\kappa g - \langle \nabla g, N \rangle) N. \quad (4.24)$$

**Remark** Note that if  $g \equiv 1$ , then the flow (4.24) reduces as

$$\frac{\partial c}{\partial t} = \kappa N \quad (4.25)$$

which is the well-known mean curvature motion (or shortening flow). This flow decreases the total curvature as well as the number of zero-crossings and the value of maxima/minima curvature. Therefore it has the properties of shortening (an initial curve shrinks under (4.25) to a point in finite time with asymptotically circular shape) as well as smoothing (points with high curvature evolve faster and disappear asymptotically). An example is shown in Figure 4.9 (see also Section A.3.4 in the Appendix). For more geometric details about (4.25) we refer to [104]. ■

We are now in position to state what we mean by saying that the two-minimization problems  $\inf_c J_1(c)$  and  $\inf_c J_2(c)$  are equivalent.

**Definition 4.3.1 (equivalence between  $\inf_c J_1(c)$  and  $\inf_c J_2(c)$ )**

The problems  $\inf_c J_1(c)$  and  $\inf_c J_2(c)$  are equivalent if for any curve  $c \in C$

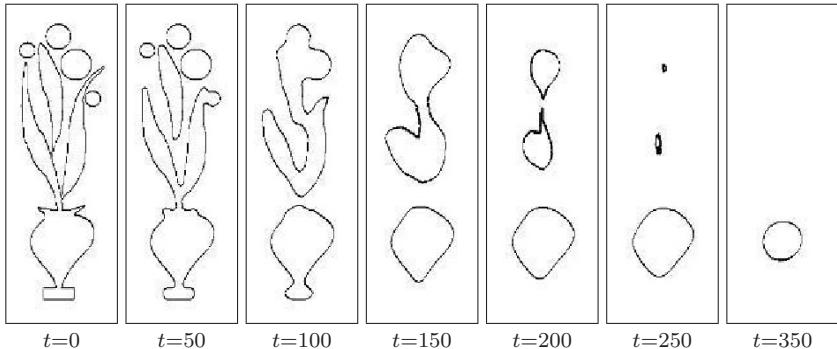


Figure 4.9. Example of the mean curvature motion

there exists a neighborhood  $V(c)$  of  $c$  so that in  $V(c)$  the flow which most decreases  $J_1$  is also a decreasing flow for  $J_2$  and vice et versa.

In order to apply this criterion, we need a more precise definition of an edge.

**Definition 4.3.2 (edge)** Let  $c \in C$ , we say that  $c$  is an edge of the image  $I$  if there exists  $\varepsilon_0$  such that  $\forall \varepsilon < \varepsilon_0, \exists \alpha_\varepsilon, \lim_{\varepsilon \rightarrow 0} \alpha_\varepsilon = 0$ , such that  $|\nabla I(x)| \geq \frac{1}{\varepsilon}$  if  $x \in V_\varepsilon = \{x : d((x); c) \leq \varepsilon\}$ , where  $d$  is the distance function.

**Theorem 4.3.1** Let us assume that the edge detection function  $g$  satisfies: for any edge  $c \in C$  there exist  $l, l' \in Q$ ,  $l < l'$  such that  $\forall \varepsilon < \varepsilon_0, \forall x \in V_\varepsilon$

$$(i) \quad g(|\nabla I(x)|) = O(\varepsilon^l).$$

$$(ii) \quad |\nabla g(|\nabla I(x)|)| = O(\varepsilon^{l'}).$$

Then the two-minimization problems  $\inf_c J_1(c)$  and  $\inf_c J_2(c)$  are equivalent in the sense of the Definition 4.3.1.

**Proof** We have to prove that the flow (4.21) which makes  $J_1$  decrease most rapidly also makes  $J_2$  decrease and conversely the flow (4.24) which makes  $J_2$  decrease most rapidly also makes  $J_1$  decrease. In order to do that we replace  $\frac{\partial c}{\partial t}$  given by (4.21) in the expression of  $J'_2(t)$  and vice et versa. When  $\frac{\partial c}{\partial t}$  is given by (4.21),  $J'_2(t)$  becomes:

$$J'_2(t) = \int_a^b -g(\langle \nabla g, N \rangle - \kappa g) \left( \langle \nabla g, N \rangle - \frac{\kappa}{g} \left| \frac{\partial c}{\partial q} \right|^2 \right) \left| \frac{\partial c}{\partial q} \right| dq. \quad (4.26)$$

In the same way if we replace  $\frac{\partial c}{\partial t}$  by its expression in (4.24) then  $J'_1(t)$  writes as:

$$J'_1(t) = \int_a^b -g(\langle \nabla g, N \rangle - \kappa g) \left( \langle \nabla g, N \rangle - \frac{\kappa}{g} \left| \frac{\partial c}{\partial q} \right|^2 \right) \left| \frac{\partial c}{\partial q} \right| dq. \quad (4.27)$$

In order to know the sign of  $J'_2(t)$  in (4.26) or the sign of  $J'_1(t)$  in (4.27) it suffices to study a.e.  $q$  the sign of the integrand:

$$z(q, t) = -g(\langle \nabla g, N \rangle - \kappa g) \left( \langle \nabla g, N \rangle - \frac{\kappa}{g} \left| \frac{\partial c}{\partial q} \right|^2 \right).$$

By developing this expression we obtain:

$$z(q, t) = -\kappa^2 \left| \frac{\partial c}{\partial q} \right|^2 g + \kappa \left( \left| \frac{\partial c}{\partial q} \right|^2 + g^2 \right) \langle \nabla g, N \rangle - g(\langle \nabla g, N \rangle)^2.$$

Let us recall now that  $g$  stands for  $g(|\nabla I(c(q, t))|)$  and let us assume that  $c(q, t)$  is in a neighborhood  $V_\varepsilon$  of an edge  $c(q, 0)$  of the image  $I$ , then we have thanks to the assumptions (i) and (ii) on  $g$ :

$$z(q, t) \approx -a\varepsilon^l \pm b\varepsilon^{l'} - c\varepsilon^{2l'}$$

with  $a, b, c \geq 0$ . When  $\varepsilon$  is small enough the sign of  $z(q, t)$  is given by the sign of  $-a\varepsilon^l$  and so it is negative (we suppose that  $\kappa$  is bounded otherwise as soon as  $|\kappa| \gg 1$  we have  $\kappa^2 \geq |\kappa|$  and  $z$  would remain nonpositive). ■

The Caselles et al model [58] may be improved by adding to the right-hand side of (4.24) a supplementary term:

$$\frac{\partial c}{\partial t} = (\kappa g - \langle \nabla g, N \rangle + \alpha g)N. \quad (4.28)$$

The main interest of adding  $\alpha g$  to the velocity is that it makes the detection of nonconvex objects easier and it increases the speed of convergence. In fact  $\alpha \geq 0$  must be chosen large enough so that the coefficient  $(\kappa + \alpha)$  remains of constant sign. Consequently, the curvature  $\kappa$  can have a non constant sign and nonconvex shapes can then be detected.

**Remark** It is worth noticing that equation (4.28) does not come from any energy unless the function  $g$  is a constant (and this is not the case since  $g$  is a detector function). As a matter of fact the flow  $\frac{\partial c}{\partial t} = \alpha N$  with  $\alpha \geq 0$  (a constant) is the flow deduced from the area energy:

$$A(t) = -\frac{1}{2} \int_a^b \langle c(q, t), \begin{pmatrix} -\partial c_2 / \partial q \\ \partial c_1 / \partial q \end{pmatrix} \rangle dq.$$

For each  $t > 0$ ,  $A(t)$  is the area enclosed by  $c(q, t) = (c_1(q, t), c_2(q, t))$  and it is easy to verify that:

$$A'(t) = - \int_a^b \left\langle \frac{\partial c}{\partial t}, N \right\rangle \left| \frac{\partial c}{\partial q} \right| dq.$$

Thus the direction in which  $A(t)$  is decreasing most rapidly is  $\frac{\partial c}{\partial t} = N$ . ■

To summarize the situation, starting from the initial formulation of Kass et al (4.16), we introduced the energy  $J_1$  and an intrinsic (geometric) functional  $J_2$ . We clearly defined the link between these two optimization problems. From a numerical point of view, we can wonder which formulation is the best to choose. As we are going to see in the next section, the Euler-Lagrange equations associated to  $J_2$  can be written in an Eulerian formulation by using a *level sets approach* (which is not the case for  $J_1$ ). The level sets approach is based on the description of the curve as the zero-crossing of a higher dimensional function and allows major simplifications.

### 4.3.3 The level sets method

The aim of this section is to find an efficient algorithm to solve (4.28). Naturally, one could parametrize the curve  $c$  and discretize the equation but this direct approach faces difficulties that we will emphasize later. More generally, we are interested in flows governed by equations of the form:

$$\begin{cases} \frac{\partial c}{\partial t} = F N \\ c(q, 0) = c_0(q). \end{cases} \quad (4.29)$$

The equation (4.29) says that the curve  $c(q, t)$  moves along its normal with a speed  $F$  which may depend on  $t, c, c', c''$ . The *level sets formulation* is based on the following observation due to Osher-Sethian [200]:

- A curve can be seen as the zero-level of a function in higher dimension.

For example, a curve in  $R^2$  can be represented as the zero-level line of function  $R^2 \rightarrow R$  (see Figure 4.10). More precisely, let us suppose that there exists a function  $u : R^+ \times R^2 \rightarrow R$  so that:

$$u(t, c(t, q)) = 0 \quad \forall q, \forall t \geq 0. \quad (4.30)$$

Then if  $u$  is sufficiently regular, by differentiating (4.30) with respect to  $t$  we obtain:

$$\frac{\partial u}{\partial t} + \left\langle \nabla u, \frac{\partial c}{\partial t} \right\rangle = 0$$

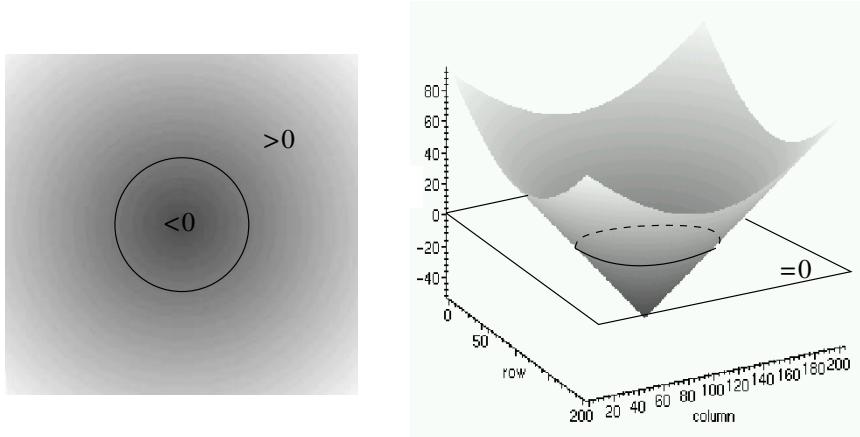


Figure 4.10. Basis of the level sets approach: a closed curve can be seen as the zero-level of a function in higher dimension. For instance, the function can be the signed distance to the curve, negative inside and positive outside.

and by replacing the expression of the speed given in (4.29) we get:

$$\frac{\partial u}{\partial t} + \langle \nabla u, F N \rangle = 0. \quad (4.31)$$

But recalling that the unit inward normal to the front defined by (4.30) is given by  $N = -\frac{\nabla u}{|\nabla u|}$  (we suppose that  $u$  is negative inside the curve and positive outside) then (4.31) rewrites as:

$$\frac{\partial u}{\partial t}(t, c(t, q)) = F |\nabla u(t, c(t, q))|. \quad (4.32)$$

According to the way we have established (4.32), this equation is *a priori* only valid for the zero-level set of  $u$ . But one of the advantages of the method is that  $u$  may be regarded as defined on the whole domain  $R^+ \times \Omega$ . So, we can solve the PDE:

$$\frac{\partial u}{\partial t}(t, x) = F |\nabla u(t, x)|$$

for  $t \geq 0$  and  $x \in \Omega$  as soon as  $F$  is well-defined off the front, i.e. on the whole space. Then once  $u$  is calculated on  $R^+ \times \Omega$  we just need to extract the zero level set of  $u$  to get the curve. We will come back on this question later. Of course we have to add:

- (i) A boundary condition: one generally chooses that the normal derivative vanishes on  $\partial\Omega$  i.e.  $\frac{\partial u}{\partial N} = 0$  on  $\partial\Omega$ .

- (ii) An initial condition at  $t = 0$ . A good candidate is the signed distance function to an initial given curve  $c_0(q)$  surrounding the objects:

$$u(0, x) = \bar{d}(x, c_0) = \begin{cases} +d(x, c_0) & \text{if } x \text{ is outside } c_0 \\ -d(x, c_0) & \text{if } x \text{ is inside } c_0 \end{cases}$$

where  $d(x, c_0)$  is the Euclidian distance to  $c_0$ .

Therefore the final model is

$$\begin{cases} \frac{\partial u}{\partial t}(t, x) = F |\nabla u(t, x)| \text{ for } (t, x) \in ]0, +\infty[ \times \Omega \\ u(0, x) = \bar{d}(x, c_0) \\ \frac{\partial u}{\partial N} = 0 \text{ for } (t, x) \in ]0, +\infty[ \times \partial\Omega. \end{cases} \quad (4.33)$$

The equation (4.33) is called an Hamilton-Jacobi equation (see also Section 2.3.2). There are many advantages to work with this Eulerian formulation:

- The first is that the evolving function  $u(t, x)$  always remains a function as long as  $F$  is smooth. But, if we only consider the level set  $u = 0$  (and so the front  $c(t, q)$ ), it may change topology, break, merge as  $u$  evolves. We illustrate this in Figure 4.11. This is a main advantage of this representation, since we do not need to take these topology changes into account numerically.
- A second important interest concerns the numerical approximation: we can use a fixed discrete grid in the spatial domain and choose finite differences approximations for the spatial and temporal derivatives. We refer to Sections 4.3.4 and A.3.4 for more details.
- Another advantage is that intrinsic geometric elements of the front such as the normal vector or the curvature can be easily expressed with respect to  $u$ . Notice that This is a necessary condition for any representation to be useful.
- Finally, this above level set formulation can be extended and applied in any dimension. For instance a surface can be represented implicitly by the zero level set of a function defined in a volume.

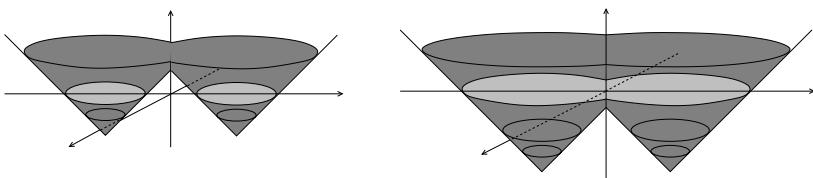


Figure 4.11. Illustration of the change of topology

- ☞ This representation is useful as soon as the evolution of an hypersur-

face is involved and that the motion can be expressed as a velocity along the normal.

This approach has been extensively used in computer vision as well as in other domains.

### Level sets method [229, 198, 223]

In image analysis, [57, 166, 58, 165, 144] are some of the papers that first appeared in this direction. All these papers rely on the same ideas. The major difference between them is that [58, 144, 165] start from an energy concept (the minimization of the weighted length) while [57, 166, 165] formulate directly their problem in terms of level sets.

To give an example, let us come back to the segmentation problem. We can show that the level sets expression of (4.28) is:

$$\frac{\partial u}{\partial t} = \left( (\kappa + \alpha) g + \langle \nabla g, \frac{\nabla u}{|\nabla u|} \rangle \right) |\nabla u|$$

or, remembering that the curvature  $\kappa$  is given by  $\kappa = \operatorname{div} \left( \frac{\nabla u}{|\nabla u|} \right)$ :

$$\frac{\partial u}{\partial t} = g(|\nabla I|) \left( \operatorname{div} \left( \frac{\nabla u}{|\nabla u|} \right) + \alpha \right) |\nabla u| + \langle \nabla g, \nabla u \rangle, \quad (4.34)$$

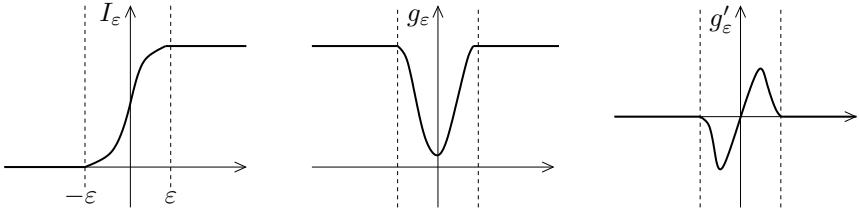
with the boundary and initial conditions of (4.33). In the first term, the coefficient  $g(|\nabla I|)$  permits to stop the evolving curve when it arrives to the object boundaries<sup>2</sup>. The action of the second term,  $\langle \nabla g, \nabla u \rangle$ , is less obvious. To better understand its contribution, let us consider the following one-dimensional example. Let  $I(x)$  be the Heaviside function:  $I(x) = 1$  if  $x \geq 0$  and  $I(x) = 0$  otherwise, and let  $I_\varepsilon$  be a regularization of  $I$  by a cubic function:

$$I_\varepsilon(x) = \begin{cases} 1 & \text{if } x \geq \varepsilon \\ -\frac{x^3}{4\varepsilon^3} + \frac{3x}{4\varepsilon} + \frac{1}{2} & \text{if } -\varepsilon \leq x \leq \varepsilon \\ 0 & \text{if } x \leq -\varepsilon \end{cases}$$

Then, if we note  $g_\varepsilon(x) = \frac{1}{1 + |I'_\varepsilon(x)|^2}$ , it is easy to verify that in a neighborhood of  $x = 0$  we have  $g_\varepsilon(x) \approx \varepsilon^2$  and  $g'_\varepsilon(x) \approx \frac{9}{4\varepsilon^4}x$  (see Figure 4.12). Therefore, the leading term in (4.34) is (in this one-dimensional case) the transport term  $g'_\varepsilon(x) u'(x)$ . Thus, the front evolves from the right to the left for  $x > 0$  and from the left to the right for  $x < 0$ . The point  $x = 0$

---

<sup>2</sup>Notice that in practice, because of the presence of additional noise in the image, we use a smoothed version of  $I$

Figure 4.12. The initial signal  $I_\varepsilon$  and the functions  $g_\varepsilon(x)$ ,  $g'_\varepsilon(x)$ 

(the discontinuity front) can be seen as an attractive point. This effect is the same for images.

- the term  $\langle \nabla g, \nabla u \rangle$  increases the attraction of the deforming contour towards the boundary of objects.

Before proceeding with the approximation we now present results regarding the existence and the uniqueness of a solution for (4.34), using the theory of viscosity solutions (see Section 2.3).

According to the identity:

$$\operatorname{div} \left( \frac{\nabla u}{|\nabla u|} \right) = \frac{1}{|\nabla u|^2} [(u_{x_1})^2 u_{x_2 x_2} + (u_{x_2})^2 u_{x_1 x_1} - 2u_{x_1} u_{x_2} u_{x_1 x_2}],$$

where  $u_{x_i} = \frac{\partial u}{\partial x_i}$ , the equation (4.34) rewrites as:

$$\frac{\partial u}{\partial t} = g(x) \sum_{i,j=1}^2 a_{i,j}(\nabla u) u_{x_i x_j} + H(x, \nabla u) \quad (4.35)$$

where:

$$a_{i,j}(p) = \delta_{i,j} - \frac{p_i p_j}{|p|^2} \quad \text{if } p \neq 0 \quad (\delta_{i,j} \text{ is the Kronecker symbol}) \quad (4.36)$$

$$H(x, p) = \alpha g(x) |p| + \sum_{i=1}^2 \frac{\partial g}{\partial x_i}(x) p_i \quad \text{with } g(x) = g(|\nabla I(x)|). \quad (4.37)$$

To avoid some tedious technicalities at corners, instead of the Neumann boundary condition  $\frac{\partial u}{\partial N} = 0$  on  $]0, +\infty[ \times \partial\Omega$ , we will work with periodic solutions (see Section 3.3.1 to extend the function  $u_0$  defined on  $\Omega$ , to a periodic function defined on  $R^2$ ). Of course, we also suppose that the initial condition  $u(0, x) = u_0(x)$  is periodic.

Let us recall the definition of viscosity solutions, adapting the Definition 2.3.1 for the parabolic equation (4.35):

**Definition 4.3.3 (viscosity subsolution, supersolution, solution)** Let  $u$  in  $C([0, T] \times R^2)$ ,  $0 < T < \infty$ , then  $u$  is a viscosity subsolution of (4.35) if for  $\varphi \in C^2([0, T] \times R^2)$  the following condition holds: at any point  $(t_0, x_0)$

in  $]0, T[ \times R^2$  which is a local maximum of  $(u - \varphi)$  then:

$$\begin{aligned} \frac{\partial \varphi}{\partial t}(t_0, x_0) - g(x_0) \sum_{i,j=1}^2 a_{i,j}(\nabla \varphi(t_0, x_0)) \varphi_{x_i x_j}(t_0, x_0) - H(x_0, \nabla \varphi(t_0, x_0)) &\leq 0 \\ \text{if } |\nabla \varphi(t_0, x_0)| &\neq 0, \\ \frac{\partial \varphi}{\partial t}(t_0, x_0) - g(x_0) \overline{\lim}_{p \rightarrow 0} \sum_{i,j=1}^2 a_{i,j}(p) \varphi_{x_i x_j}(t_0, x_0) &\leq 0 \quad \text{if } |\nabla \varphi(t_0, x_0)| = 0. \end{aligned}$$

Similarly, we define the notion of viscosity supersolution changing local maximum by local minimum,  $\leq 0$  by  $\geq 0$ , and limsup by liminf. A viscosity solution is a continuous function which is both a sub and a supersolution.

If  $V$  is a Banach space, we recall that  $L^\infty((0, T); V)$  is the space defined by:

$$L^\infty((0, T); V) = \left\{ \begin{array}{l} f : (0, T) \rightarrow V \text{ such that} \\ |f|_{L^\infty((0, T); V)} = \inf \{c; |f|_V \leq c, \text{ a.e. } t\} < \infty \end{array} \right\}$$

We may now state the main result.

**Theorem 4.3.2** Let us assume that  $g \geq 0$ ,  $g$  and  $\sqrt{g}$  are Lipschitz continuous.

- (i) Let  $u_0(x)$  be the initial condition, be Lipschitz continuous. Then (4.35) has a unique viscosity solution in  $C([0, \infty[ \times R^2) \cap L^\infty((0, T); W^{1,\infty}(R^2))$  for any  $T < \infty$ . Moreover:

$$\inf_{R^2} u_0(x) \leq u(t, x) \leq \sup_{R^2} u_0(x).$$

- (ii) Let  $v$  be a viscosity solution of (4.35) with  $u_0$  replaced by  $v_0$ . Then for all  $T$  in  $[0, +\infty[$  we have:

$$\sup_{0 \leq t \leq T} |u(t, \cdot) - v(t, \cdot)|_{L^\infty(R^2)} \leq |u_0 - v_0|_{L^\infty(R^2)}. \quad (4.38)$$

**Proof** The proof follows [57]. It is rather technical and long and we divide it into three steps.

#### Step 1: Stability and uniqueness

We begin with the stability estimate (4.38) from which we will deduce the uniqueness of a solution. Let  $u$  and  $v$  be two (viscosity) solutions associated respectively to  $u_0$  and  $v_0$ . We are interested in the maximum of  $(u(t, x) - v(t, x))$  and in showing that it is nonpositive. If  $u$  and  $v$  are smooth we can proceed as described in Section 2.3.4. Otherwise we need a trick which is the duplication of variables. Let us define:

$$l(t, x, y) = u(t, x) - v(t, y) - \frac{1}{4\varepsilon} |x - y|^2 - \lambda t, \quad x, y \in R^2, t \in [0, T] \quad (4.39)$$

where  $\varepsilon, \lambda \in ]0, +\infty[$  will be determined later. Now we are interested in the maximum of  $l$ . Let  $(t_0, x_0, y_0)$  be a maximum point of  $l$  on  $[0, T] \times R^2 \times R^2$  ( $\varepsilon$  fixed).

We claim that  $t_0 = 0$ . Otherwise the maximum would be attained at some point  $(t_0, x_0, y_0)$  with  $t_0 > 0$ . In that case thanks to the Crandall-Ishii's Lemma (Lemma 2.3.2), we can find for any  $\mu > 0$  two real numbers  $a, b$  and two symmetric  $(2 \times 2)$  matrices  $X$  and  $Y$  such that:

$$a - b = \lambda, \quad \begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix} \leq \begin{pmatrix} B + \mu B^2 & -B - \mu B^2 \\ -B - \mu B^2 & B + \mu B^2 \end{pmatrix} \quad (4.40)$$

with:

$$B_{i,j} = \varepsilon^{-1} |x_0 - y_0|^2 \delta_{i,j} + 2\varepsilon^{-1} (x_0 - y_0)_i (x_0 - y_0)_j.$$

Moreover, if  $x_0 \neq y_0$  one has:

$$\begin{aligned} a - g(x_0) a_{i,j} (\varepsilon^{-1} |x_0 - y_0|^2 (x_0 - y_0)) X_{i,j} - \\ - H(x_0, \varepsilon^{-1} |x_0 - y_0|^2 (x_0 - y_0)) \leq 0, \end{aligned} \quad (4.41)$$

$$\begin{aligned} b - g(x_0) a_{i,j} (\varepsilon^{-1} |x_0 - y_0|^2 (x_0 - y_0)) Y_{i,j} - \\ - H(y_0, \varepsilon^{-1} |x_0 - y_0|^2 (x_0 - y_0)) \geq 0. \end{aligned} \quad (4.42)$$

In fact if  $x_0 = y_0$  the two last inequalities have to be interpreted in terms of suitable limits. In that case  $B = 0$ ,  $X \leq 0$ ,  $Y \geq 0$  and (4.41), (4.42) writes as

$$\begin{aligned} a - g(x_0) \overline{\lim_{p \rightarrow 0}} a_{i,j}(p) X_{i,j} &\leq 0 \\ b - g(x_0) \underline{\lim_{p \rightarrow 0}} a_{i,j}(p) Y_{i,j} &\geq 0 \end{aligned}$$

Hence, in particular  $a \leq 0$ ,  $b \geq 0$  which contradicts  $a - b = \lambda > 0$ . So we have  $x_0 \neq y_0$ .

Next, we choose  $\mu = \varepsilon |x_0 - y_0|^{-2}$  (which is now possible) and deduce from (4.40):

$$\begin{pmatrix} X & 0 \\ 0 & -Y \end{pmatrix} \leq 2\varepsilon^{-1} \begin{pmatrix} C & -C \\ -C & C \end{pmatrix} \quad (4.43)$$

where  $c_{i,j} = |x_0 - y_0|^2 \delta_{i,j} + 5(x_0 - y_0)_i (x_0 - y_0)_j$ . Then we set:

$$G = \begin{pmatrix} g(x_0)A & \sqrt{g(x_0)g(y_0)}A \\ \sqrt{g(x_0)g(y_0)}A & g(y_0)A \end{pmatrix}$$

where  $A = a_{i,j} (\varepsilon^{-1} |x_0 - y_0|^2 (x_0 - y_0))$ .  $G$  is a non-negative symmetric matrix so that multiplying (4.43) to the left by  $G$  and taking the trace we get:

$$g(x_0) \sum_{i,j} a_{i,j} X_{i,j} - g(y_0) \sum_{i,j} a_{i,j} Y_{i,j} \leq 2\varepsilon^{-1} (\sqrt{g(x_0)} - \sqrt{g(y_0)})^2 \text{trace}(AC) \quad (4.44)$$

Next combining (4.40), (4.41), (4.42), (4.44) and carrying out some manipulations we obtain:

$$\lambda \leq \frac{c_1}{\varepsilon} \leq |x_0 - y_0|^4 \quad (4.45)$$

where  $c_1$  is a constant depending only on  $a_{i,j}(p)$  and  $g$ . We now estimate  $|x_0 - y_0|$ . According to the definition of  $(t_0, x_0, y_0)$  we have:

$$u(t_0, x_0) - v(t_0, y_0) - \frac{1}{4\varepsilon} |x_0 - y_0|^4 - \lambda t_0 \geq u(t_0, y_0) - v(t_0, y_0) - \lambda t_0$$

and thus:

$$\frac{1}{4\varepsilon} |x_0 - y_0|^4 \leq c_2 |x_0 - y_0| \quad (4.46)$$

where  $c_2$  is the Lipschitz constant of  $u(t_0, \cdot)$  on  $[0, T] \times R^2$ . Therefore from (4.45):

$$\lambda \leq c_3 \varepsilon^{\frac{1}{3}} \text{ with } c_3 = c_1 4^{\frac{4}{3}} c_2^{\frac{4}{3}}.$$

Now, recall that  $\lambda$  and  $\varepsilon$  are arbitrary. Without a loss of generality we may suppose that  $\sup_{[0, T] \times R^2} |u - v| \neq 0$  (otherwise we conclude) and we choose:

$$\varepsilon^{\frac{1}{3}} = \delta \sup_{[0, T] \times R^2} |u - v| \quad (\delta > 0), \quad (4.47)$$

$$\lambda = 2\delta c_3 \sup_{[0, T] \times R^2} |u - v|. \quad (4.48)$$

This choice contradicts (4.46) and so  $t_0 = 0$ .

Next let us estimate  $\sup_{[0, T] \times R^2} |u - v|$ . We fix  $\lambda$  and  $\varepsilon$  as before. Since  $t_0 = 0$  we have for all  $(t, x, y)$ :

$$u(t, x) - v(t, y) - \frac{1}{4\varepsilon} |x - y|^4 - \lambda t \leq u_0(x_0) - v_0(y_0) - \frac{1}{4\varepsilon} |x_0 - y_0|^4 \quad (4.49)$$

but observing that:

$$\begin{aligned} u_0(x_0) - v_0(y_0) &= u_0(y_0) - v_0(y_0) + u_0(x_0) - u_0(y_0) \leq \\ &\leq |u_0(y_0) - v_0(y_0)| + c_2 |x_0 - y_0| \end{aligned}$$

and letting  $x = y$  in (4.49), we get:

$$\sup_{[0, T] \times R^2} (u(t, x) - v(t, x)) - \lambda t \leq \sup_{y \in R^2} (u_0(y) - v_0(y)) + \sup_{r > 0} \left( c_2 r - \frac{1}{4\varepsilon} r^4 \right)$$

i.e.

$$\sup_{[0, T] \times R^2} (u(t, x) - v(t, x)) - \lambda t \leq \sup_{y \in R^2} (u_0(y) - v_0(y)) + \frac{3}{4} c_2^{\frac{4}{3}} \varepsilon^{\frac{1}{3}}.$$

By (4.47) and (4.48) we find:

$$\sup_{[0, T] \times R^2} (u - v) \leq \sup_y |u_0(y) - v_0(y)| + \frac{3}{4} c_2^{\frac{4}{3}} \delta \sup_{[0, T] \times R^2} |u - v| + 2\delta c_3 \sup_{[0, T] \times R^2} |u - v| T.$$

Exchanging the role of  $u$  and  $v$  and letting  $\delta \rightarrow 0$ , we obtain:

$$\sup_{[0,T] \times R^2} (u - v) \leq \sup_y |u_0(y) - v_0(y)|.$$

This proves the part (ii) of Theorem 4.3.2 and the uniqueness in part (i). Remark that by using the definition of viscosity solutions we may also obtain the following bound:  $\inf u_0(x) \leq u(t, x) \leq \sup u_0(x)$  for all  $(t, x)$ .

We now prove the existence part of the theorem. We begin by setting an *a priori* estimate on  $\nabla u$ .

**Step 2:** Estimate of  $\nabla u$  on  $L^\infty(R^2)$

In this step we suppose that all the coefficients of (4.35) have been smoothed as much as necessary (notations are unchanged) and that (4.35) admits a regular solution (we will come back on this question in step 3). Thus let  $u$  be a regular solution of:

$$\frac{\partial u}{\partial t} = g(x) \sum_{i,j} a_{i,j}(\nabla u) u_{x_i x_j} + H(x, \nabla u).$$

We are going to apply the classical *Bernstein method* and derive a parabolic inequality for  $w = |\nabla u|^2$ . To this end we differentiate (4.35) with respect to  $x_l$ :

$$\begin{aligned} \frac{\partial u_{x_l}}{\partial t} - \frac{\partial g}{\partial x_l} \sum_{i,j} a_{i,j}(\nabla u) u_{x_i x_j} - g \left[ \sum_{i,j} \sum_k \frac{\partial a_{i,j}(\nabla u)}{\partial p_k} u_{x_k x_l} u_{x_i x_j} + \right. \\ \left. + a_{i,j}(\nabla u) u_{x_l x_i x_j} \right] - \frac{\partial H}{\partial x_l} - \sum_k \frac{\partial H}{\partial p_k} u_{x_k x_l} = 0 \end{aligned}$$

We multiply this equation by  $2u_{x_l}$  and sum over  $l$  to find:

$$\begin{aligned} \frac{\partial w}{\partial t} - g \left[ \sum_{i,j} a_{i,j}(\nabla u) w_{x_i x_j} + \sum_{i,j} \sum_k \frac{\partial a_{i,j}(\nabla u)}{\partial p_k} u_{x_i x_j} w_{x_k} \right] - \sum_k \frac{\partial H}{\partial p_k} w_{x_k} = \\ = 2 \sum_l \frac{\partial g}{\partial x_l} u_{x_l} \sum_{i,j} a_{i,j}(\nabla u) u_{x_i x_j} - 2 \sum_l \frac{\partial H}{\partial x_l} u_{x_l} - 2g \sum_l \sum_{i,j} a_{i,j}(\nabla u) u_{x_l x_i} u_{x_l x_j} \end{aligned} \tag{4.50}$$

where  $w = |\nabla u|^2$ . We note  $Lw$  the left-hand side of (4.50). On the other hand for the right-hand side of (4.50), since  $g$  is smooth we have:

$$(i) \sum_l u_{x_l} \frac{\partial H}{\partial x_l} = \sum_l u_{x_l} \left[ \alpha \frac{\partial g}{\partial x_l} |\nabla u| + \sum_i \frac{\partial^2 g}{\partial x_i \partial x_l} u_{x_i} \right] \leq k_1 |\nabla u|^2 = k_1 w,$$

where  $k_1$  is a constant depending only on  $g$ .

(ii) Since the matrix  $a_{i,j}$  is symmetric semi-definite we can reduce it to its principal axes and we easily get:

$$\left( \sum_{i,j} a_{i,j} u_{x_i} u_{x_j} \right)^2 \leq k_2 \sum_k \sum_{i,j} a_{i,j} u_{x_i x_k} u_{x_j x_k} \quad \text{for some constant } k_2.$$

Using the Cauchy-Schwarz inequality, the obvious inequality  $2ab \leq a^2 + b^2$  and again that  $a_{i,j}$  is symmetric semi-definite, we obtain:

$$2 \sum_k \frac{\partial g}{\partial x_k} u_{x_k} \sum_{i,j} a_{i,j} u_{x_i} u_{x_j} - 2g \sum_k \sum_{i,j} a_{i,j} u_{x_i x_k} u_{x_j x_k} \leq k_3 w.$$

Inserting these bounds in (4.50) we have just proved the existence of a constant  $k_3 > 0$  so that  $Lw \leq k_3 w$ . Unfortunately, since  $k_3 > 0$  we cannot directly apply the maximum principle to get a result of the kind:

$$|w(t,.)|_{L^\infty(R^2)} \leq |w_0(.)|_{L^\infty(R^2)} \quad \text{a.e.}$$

However we can still show that (see the remark at the end this proof):

$$|w(t,.)|_{L^\infty(R^2)} \leq (1 + ct e^{ct}) |w_0(.)|_{L^\infty(R^2)} \quad \text{a.e.}$$

where  $c$  is a constant depending only on  $k_3$ . For  $u$  this means:

$$|\nabla u(t,.)|_{L^\infty(R^2)} \leq e^{ct} |\nabla u_0|_{L^\infty(R^2)} \leq e^{cT} |\nabla u_0|_{L^\infty(R^2)} \quad (4.51)$$

where the constant  $c$  only depends on  $|g|_{L^\infty(R^2)}$  and  $|\nabla g|_{L^\infty(R^2)}$ . In particular,  $c$  does not depend on the way the coefficients have been regularized. This inequality is proved in the remark following the proof. Notice that this result is only interesting because we consider  $t$  bounded.

**Step 3:** Approximation and existence of a viscosity solution

In order to conclude, we are going to approximate (4.35) by a similar equation for which we are able to prove the existence of a smooth solution satisfying (4.51). To this end we consider a periodic- $C^\infty$  function  $u_0^\varepsilon$  such that  $u_0^\varepsilon \rightarrow u_0$  uniformly satisfying:

$$|\nabla u_0^\varepsilon|_{L^\infty(R^2)} \leq |\nabla u_0|_{L^\infty(R^2)}, \quad |u_0^\varepsilon|_{L^\infty(R^2)} \leq |u_0|_{L^\infty(R^2)}.$$

We also replace  $a_{i,j}$ ,  $g$ , and  $H$  respectively by:

$$a_{i,j}^\varepsilon(p) = \varepsilon \delta_{i,j} + \delta_{i,j} - \frac{p_i p_j}{|p|^2 + \varepsilon}$$

$$g_\varepsilon = g + \varepsilon$$

$$H_\varepsilon(x, p) = \alpha g_\varepsilon(x) \sqrt{|p|^2 + \varepsilon} + \sum_i \frac{\partial g_\varepsilon}{\partial x_i}(x) p_i.$$

According to the general theory of quasilinear parabolic equations [156], we know there exists a smooth solution  $u^\varepsilon$  of:

$$\begin{cases} \frac{\partial u}{\partial t} = g_\varepsilon(x) \sum_{i,j} a_{i,j}^\varepsilon(\nabla u) u_{x_i x_j} + H_\varepsilon(x, \nabla u) \\ u(0, x) = u_0^\varepsilon(x). \end{cases} \quad (4.52)$$

Thanks to (4.51) we have:

$$|\nabla u^\varepsilon(t,.)|_{L^\infty(R^2)} \leq e^{cT} |\nabla u_0^\varepsilon|_{L^\infty(R^2)} \leq e^{cT} |\nabla u_0|_{L^\infty(R^2)} \equiv c_T.$$

This means:

$$|u^\varepsilon(t, x) - u^\varepsilon(t, y)| \leq c_T |x - y| \text{ for all } x, y \in R^2 \text{ and } t \in [0, T] \quad (4.53)$$

We can also show from (4.52) and (4.53) that:

$$|u^\varepsilon(t, x) - u^\varepsilon(s, x)| \leq c_T |t - s|^{\frac{1}{2}} \text{ for all } x \in R^2 \text{ and } t, s \in [0, T]. \quad (4.54)$$

Inequalities (4.53) and (4.54) together with  $|u|_{L^\infty((0,T) \times R^2)} \leq cte$  allow us to conclude by means of Arzelà-Ascoli theorem (see Section 2.5.4) that there exists a subsequence of  $u^\varepsilon$  converging uniformly on  $[0, T] \times R^2$  to a function  $u \in C([0, T) \times R^2] \cap L^\infty([0, T); W^{1,\infty}(R^2))$  for  $T < \infty$ . Then by applying a stability result for viscosity solutions (see Lemma 2.3.1) we conclude that  $u$  is a viscosity of (4.35). ■

**Remark (About the maximum principle in the parabolic case)**

Most uniqueness results for linear parabolic (or elliptic) PDEs follow from maximum or comparison principles. Roughly speaking, they state that if  $u$  is a solution of a parabolic PDE on  $U_T = (0, T) \times U$  then the maximum and the minimum of  $u$  are attained on the parabolic boundary of  $U_T$  defined by  $\Gamma_T = \partial(0, T) \times U \cup U \times \{t = 0\}$ . More precisely, let us consider the operator:

$$Lu = - \sum_{i,j} a_{i,j} u_{x_i x_j} + \sum_i b_i u_{x_i}$$

with  $\sum_{i,j} a_{i,j} \xi_i \xi_j \geq \theta |\xi|^2 \quad \forall (t, x) \in U_T, a_{i,j}, b_i$  continuous and  $U$  bounded. If  $u$  is a regular solution of  $u_t + Lu \leq 0$ , then we have [105]

$$\max_{\overline{U}_T} u = \max_{\Gamma_T} u.$$

Likewise, if  $u_t + Lu \geq 0$ , then we have [105]

$$\min_{\overline{U}_T} u = \min_{\Gamma_T} u.$$

So it is clear that if two solutions coincide on  $\Gamma_T$ , they coincide on  $\overline{U}_T$ . From the above result, we can deduce the following corollaries:

- Let  $\tilde{u}$  such that:

$$\begin{cases} \tilde{u}_t + L\tilde{u} = g(t) \geq 0 \\ \tilde{u}(0, x) = u_0(x) \end{cases}$$

then  $\tilde{u} = v + \int_0^t g(s) ds$  where  $v$  is the solution of:

$$\begin{cases} v_t + Lv = 0 \\ v(0, x) = u_0(x). \end{cases}$$

So we have  $\min_x u_0(x) \leq v(t, x) \leq \max_x u_0(x)$ .

- Let us consider:

$$\begin{cases} u_t + Lu \leq g(t) = \tilde{u}_t + L\tilde{u} \\ \tilde{u}(0, x) = u_0(x). \end{cases}$$

This can be rewritten:

$$\begin{cases} (u - \tilde{u})_t + L(u - \tilde{u}) \leq 0 \\ (u - \tilde{u})(0, x) = 0. \end{cases}$$

Thanks to the maximum principle, we have

$\max_{\bar{U}_T} (u - \tilde{u}) = \max_{\Gamma_T} (u - \tilde{u}) = 0$  that is  $u \leq \tilde{u}$ . But:

$$\tilde{u} = v + \int_0^t g(s) ds \leq |u_0|_{L^\infty(U_T)} + \int_0^t g(s) ds$$

so we have:

$$u \leq |u_0|_{L^\infty(U_T)} + \int_0^t g(s) ds.$$

- Finally, let us consider:

$$\begin{cases} w_t + Lw \leq cw \text{ with } c \geq 0 \\ w(0, x) = u_0(x). \end{cases}$$

The classical maximum principle can no longer be applied (because  $c \geq 0$ ). However, we have  $w_t + Lw \leq c \max_x w(t, x) = g(t)$ , so:

$$w(t, x) \leq |u_0|_{L^\infty(U)} + c \int_0^t \max_x w(x, s) ds$$

and then:

$$|w|_{L^\infty(U)}(t) \leq |u_0|_{L^\infty(U)} + c \int_0^t |w|_{L^\infty(U)}(s) ds.$$

By applying the Gronwall inequality, (see Section 2.5), we obtain:

$$|w|_{L^\infty(U)}(t) \leq |u_0|_{L^\infty(U)} (1 + ct e^{ct}) \leq k |u_0|_{L^\infty(U)}$$

with  $k$  depending on  $c$  and  $T$ . Remark that on  $\partial(0, T) \times U$ , we can choose Neumann or periodic boundary conditions. ■

Finally, we would like to check the correctness of the geometric model that is to show that the zero level set of  $u(t, x)$  asymptotically fits the desired contour  $\Gamma = \{x \in [0, 1]^2; g(x) = 0\}$ . Let us recall the result proven in [57]. We assume that  $\Gamma$  is smooth and separates  $[0, 1]^2$  into two regions, the inside  $I(\Gamma)$  and the outside  $E(\Gamma)$  (i.e.  $\Gamma$  is a Jordan curve). We recall that  $\mathcal{P}(R^N)$  the set of all subsets of  $R^N$  can be equipped with the Hausdorff

metric:

$$d(A, B) = \max \left( \sup_{x \in A} d(x, B), \sup_{x \in B} d(x, A) \right),$$

where  $d(x, A)$  denotes as usually the distance of the point  $x$  to the set  $A$ :

$$d(x, A) = \inf_{y \in A} d(x, y).$$

**Theorem 4.3.3 [57]** *Let  $\Gamma = \{x \in [0, 1]^2; g(x) = 0\}$  be a Jordan curve of class  $C^2$  and  $\nabla g(x) = 0$  on  $\Gamma$ . Assume that  $u_0(x)$  is smooth and bounded and that the set  $\{x \in R^2; u_0(x) \leq 0\}$  contains  $\Gamma$  and its interior. Let  $u(t, x)$  be the viscosity solution of (4.35) and  $\Gamma(t) = \{x; u(t, x) = 0\}$ . Then if  $\alpha$  (the constant component of the velocity) is sufficiently large,  $\Gamma(t) \rightarrow \Gamma$  as  $t \rightarrow \infty$  in the Hausdorff metric.*

#### 4.3.4 Experimental results

This section concerns the experimental results that can be obtained with (4.34). In fact, we will only show the result obtained on the “objects” image which clearly illustrates the behaviour of the method.

As far as the discretization is concerned, it is classical to use finite differences schemes. As it is recalled in the Appendix, these schemes are well-adapted to the structure of digital images since we can associate a natural regular grid. Now, coming back to our problem we can rewrite (4.34) as the sum of three separated terms:

$$\frac{\partial u}{\partial t} = g(|\nabla I|) |\nabla u| \operatorname{div} \left( \frac{\nabla u}{|\nabla u|} \right) + \alpha g(|\nabla I|) |\nabla u| + \langle \nabla g, \nabla u \rangle. \quad (4.55)$$

This reveals two kinds of terms:

- The first term in (4.55) acts as a parabolic term.
- The second and the third terms are hyperbolic terms. The second term describes motion in the normal direction to the front while the third (linear) term corresponds to pure advection.

As one would imagine, this difference must be taken into account at the discrete level. So discretizing equation (4.55) is not straightforward. To better understand, we refer the reader to the Appendix where main ideas of finite differences are explained and more precisely to Section A.3.4 where the discretization of (4.55) is presented. Given the discrete scheme, one can implement and test this approach. An example of result is presented in Figure 4.13 on the “objects” image. During the evolution, the curve is shrinking, stopping as soon as it is close from an object boundary (high gradients) and splitting in order to detect the five objects. However, it can be noticed that:



Figure 4.13. Segmentation of the “objects” image using the geodesic active contour model (4.55). From top left to bottom right, different iterations (solution as time evolves) are displayed. Initialized by the boundaries of the images, the curve is shrinking and splitting until it isolates each object.

- The interior of the objects is not segmented (for instance the interior of the disk). Once the curve has detected a contours it stops.
- Because of the level sets description, we always have closed curves. This is not really a problem in this image but one may think about situations where open contours are present.

We review in the next section recent work which consider these two issues.

#### 4.3.5 About some recent advances

##### Global stopping criterion

We have previously presented active contour models and snakes which use the gradient as a criterion to stop the curve. However, there are some objects whose boundaries cannot be defined or are badly defined through the gradient. This include for example smeared boundaries, or cognitive contours (boundaries of larger objects defined by grouping smaller ones, see G. Kanizsa [141]) as shown in Figure 4.14.

We present here a different active contour model called “without edges” [66, 69]. The main idea is to consider also the information inside the regions, and not only at their boundaries. Let us describe the model. Let  $u_0$  be the original image to be segmented,  $c$  denote the evolving curve and  $i_1, i_2$  be two unknown constants. In [66, 69] the authors introduced the following minimization problem:

$$\inf_{i_1, i_2, c} F(i_1, i_2, c) = \mu|c| + \int_{\text{inside}(c)} |u_0 - i_1|^2 dx + \int_{\text{outside}(c)} |u_0 - i_2|^2 dx,$$

where  $\mu$  is a positive parameter. This model looks for the best approximation of the image  $u_0$  as a set of regions with only two different intensities ( $i_1$  and  $i_2$ ). Typically, one of the regions represents the objects to be detected ( $\text{inside}(c)$ ) and the second region corresponds to the background ( $\text{outside}(c)$ ). The snake  $c$  will be the boundary between these two regions. We see that this model is closely related to a binary segmentation. This model has many advantages. It allows to detect both contours with or without gradient, automatically detects interior contours (think about the interior of the CD in the “objects” image), and it is robust in the presence of noise.

This approach has been implemented using the level sets method that we presented in the Section 4.3.3. Using the Heaviside function  $H$  the energy  $F$  can be rewritten as:

$$\tilde{F}(i_1, i_2, \phi) = \mu \int_{\Omega} |\nabla H(\phi)| + \int_{\Omega} |u_0 - i_1|^2 H(\phi) dx + \int_{\Omega} |u_0 - i_2|^2 (1 - H(\phi)) dx,$$

where  $\phi$  is the level sets function. To find the minimum, we need to consider the problem that the functional  $\tilde{F}$  is not Gâteaux differentiable with

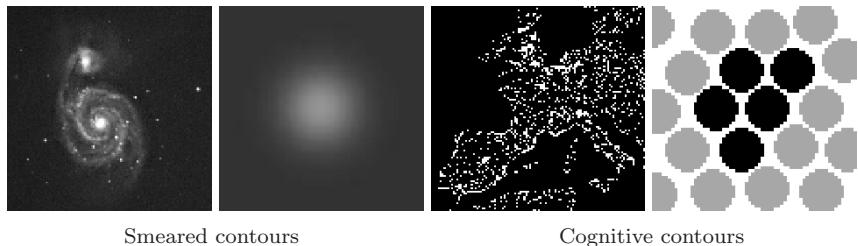


Figure 4.14. The gradient is not always adapted to segment...

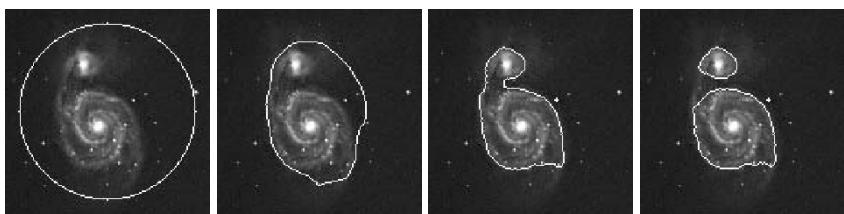


Figure 4.15. Smeared contours of a galaxy

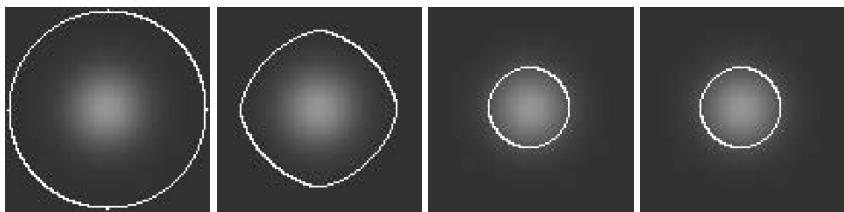


Figure 4.16. Contour of a blurred circular object

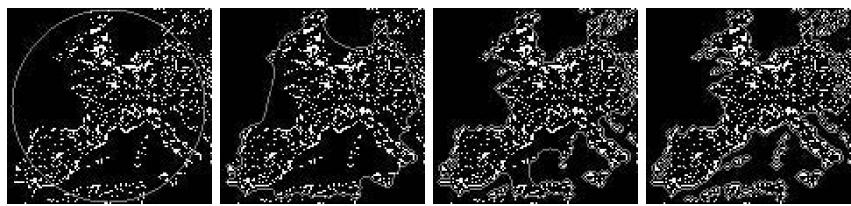


Figure 4.17. Cognitive contour for an image representing Europe night-lights

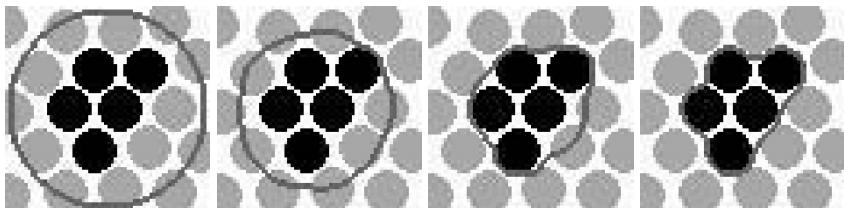


Figure 4.18. Cognitive contours. Another example which illustrates grouping based on chromatic identity

respect to the third variable. The reason is simply because the Heaviside function is not differentiable. It is then classical to regularize the problem by changing  $H$  into  $H_\varepsilon$  (a  $C^1$ -approximation of  $H$ ) to be able to compute the derivatives<sup>3</sup>. So, to minimize  $\tilde{F}$  with respect to  $i_1, i_2$  and  $\phi$ , we need to solve the equations:

$$i_1 = \frac{\int_{\Omega} u_0 H(\phi) dx}{\int_{\Omega} H(\phi) dx}, \quad i_2 = \frac{\int_{\Omega} u_0 (1 - H(\phi)) dx}{\int_{\Omega} (1 - H(\phi)) dx},$$

$$\frac{\partial \phi}{\partial t} = \delta_\varepsilon(\phi) \left( \mu \operatorname{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) - |u_0 - i_1|^2 + |u_0 - i_2|^2 \right),$$

where  $\delta_\varepsilon = H'_\varepsilon$ . Notice that we do not need  $H_\varepsilon$  for the first two equations. These equations can then be implemented using standard finite difference. We display in Figure 4.15 to 4.18 some numerical results from [66, 69] illustrating the possibilities of the approach.

On a theoretical point of view, we can prove the existence of minimizers of the energy  $\tilde{F}$ , but the convergence of the algorithm is an open problem.

To conclude we mention that this kind of model is also used in other segmentation problems like Mumford-Shah [67, 68], vector-valued images [65], texture [204, 206], classification [221], etc

### Toward more general shape representation

As seen in previous sections, level sets are a very convenient way to describe and implement the evolution of hypersurfaces, and it has been extensively used in numerous applications. Unfortunately, this description has two main limitations. The first is that level sets do not permit to represent shapes with a codimension different than one or open shapes (see Figure 4.19). The second is that level sets do not permit to describe the motion of self intersecting interfaces.

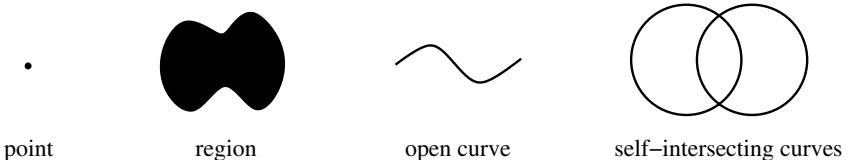


Figure 4.19. Examples of shapes in the 2D case that cannot be described using level sets

---

<sup>3</sup>This problem is closely related to the classification problem from [221], presented in Section 5.2.2. In particular, the reader will find more details regarding the functionals regularizations.

To overcome these difficulties, we mention some recent contributions in which an alternative of the level sets is proposed [219, 124]. Instead of considering the distance function  $d(x, \Gamma) = |x - y|$  for representing and evolving objects, the idea is to keep the vector distance function  $u(x, \Gamma) = x - y$  (see Figure 4.20 for some examples). This vectorial function has remarkable properties and related PDEs satisfied by  $u(x)$  allow to envisage more complicated motions than those treated with the classical level sets method (see [124]).

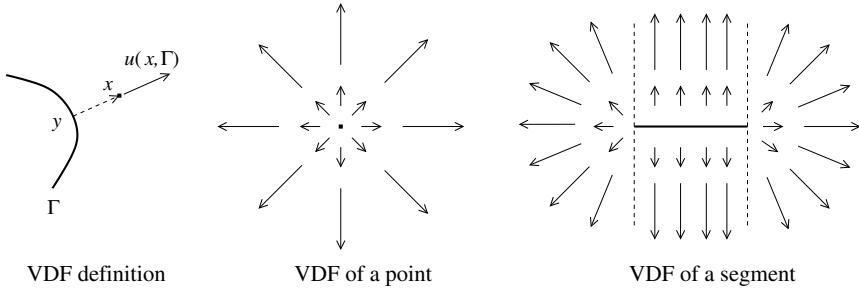


Figure 4.20. Vector distance function (VDF). Definition and examples

This direction is very promising and should be more investigated in particular from a numerical point of view.

# 5

## Other Challenging Applications

### How to read this chapter?

In this last chapter, we would like to present some recent extensions of the ideas developed previously. Two directions are considered.

The first is the case of sequences of images (Section 5.1).

- As it is shown in Section 5.1.1, a sequence contains many informations like motion, objects, depth, etc. Our aim is to show how some of these informations may be recovered only from the sequence.
- We start with studying the problem of motion estimation. We show in Section 5.1.2 how it can be estimated using variational formulations.
- Another task which is becoming more and more important is the segmentation of the sequence, which consists in describing it by its different elements or objects (typically background and foreground). Little work exists in this direction using PDEs and we present in Section 5.1.3 an approach developed in [154] which has interesting mathematical properties.
- Finally, if image restoration is quite a well-known problem, sequence restoration is still an important challenge. In Section 5.1.4 we try to analyze the variety of problems involved, propose a classification of defects and emphasize on the importance of perception.

The second subject that we would like to consider in this chapter is the problem of classification of aerial images, which presents some similarities

with the segmentation problem (Section 5.2). We present two different and original approaches from Samson et al [221, 222].

- In the first one (Section 5.2.2) the evolution of a given number of curves (level sets) is used to describe a partition of the image. The difficulty comes from the opposite actions imposed on the curves (to cover the whole image without overlapping).
- In the second one (Section 5.2.3), an optimization problem is considered where classification is performed with restoration. The form of the energy is borrowed from the Van der Walls, Cahn-Hilliard theory of phase transitions in mechanics. This is also the occasion to give a complete proof of a  $\Gamma$ -convergence theorem.

Although some theoretical results are given and proved, the goal of this chapter is to show the variety of domains for which PDEs can be useful. There are naturally many other applications of PDEs in image analysis that we do not study here. Some examples include diffusion on vector-valued images and diffusion on nonflat manifolds (see [223]), shape from shading, desocclusion (inpainting), image interpolation, shapes interpolation, stereovision, etc.

## 5.1 Sequence analysis

### 5.1.1 *Introduction*

When dealing with digital sequences, we already have to deal with all the characteristics of static digital images. As it is mentioned in Section 1.2 of the introduction, we have to consider problems like low resolution, low contrasts and the same variety of gray level information like graduated shadings, sharp transitions and fine elements (see Figure 1.3). It is naturally more complex because of the temporal dimension which intrinsically contains a lot of informations like motion, depth, the different objects in the scene, etc. To illustrate this, let us comment on the real sequence presented in Figure 5.1. From this sequence, we may remark the following:

- As one can observe, this sequence has a static background and three objects are moving (two people and a car) with motions of different ranges. This velocity is naturally linked with the sampling in time of the sequence (the number of frame per second), or equivalently, the time  $\Delta t$  between the acquisition of two consecutive frames. Naturally, the more images we have, the easier motion can be understood because the variations between two consecutive images will be small.
- The nature of the motions is also quite different. The car has a smooth trajectory, the two people can change quickly of direction, and the

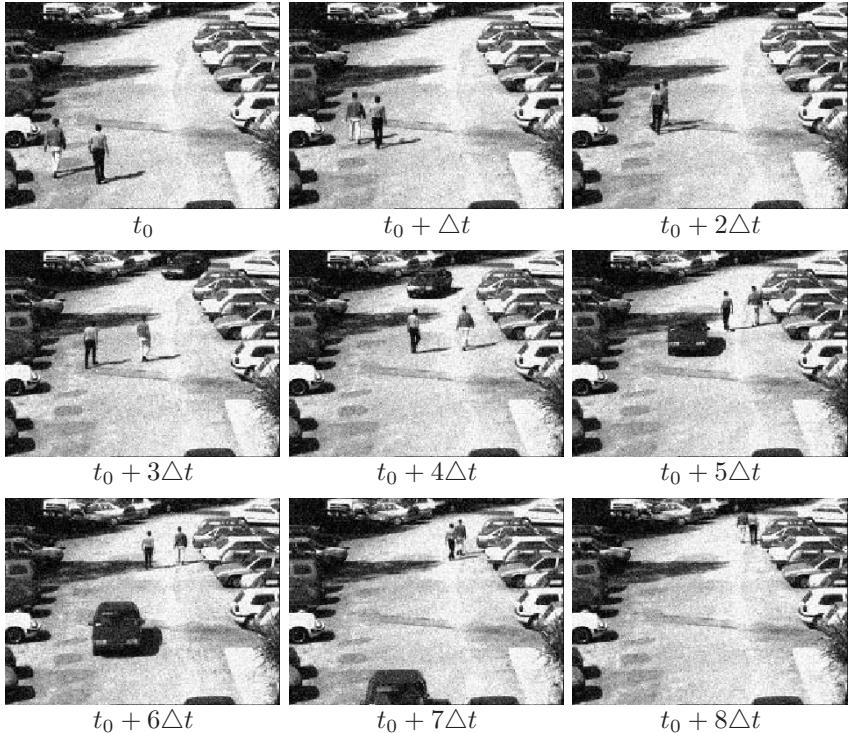


Figure 5.1. The “Street” sequence shows different ranges of motion, spurious motions (from the bushes or from the shadows), occlusions, the effect of noise, etc. Notice that the noise visible in the sequence has been added afterwards (it is a Gaussian additive noise of variance  $\sigma = 20$ ).

bushes on the bottom right-hand side corner have a “random” motion (due to the wind).

- As we will discuss in the next section, the motion that we perceive is based on intensity variations. Thus we can already wonder which interpretation of the motion that may be recovered. For instance, should we consider the shadows as a motion? This question is actually unclear and we see that to answer it, we see need all our “life experience” and knowledge about the sequence to know the difference between a real object and its shadow. Another concern will be the sensitivity of the motion perception with respect to noise.
- There is also partially contained in the sequence the notion of depth. For instance one can distinguish elements from the background and the moving objects as foreground. This information is obtained by analyzing the occlusions and we will see in Section 5.1.3 how this can be used. More generally, if we have a sequence taken with a moving

camera, we can imagine that more precise depth informations from the scene may be recovered.

Beyond these remarks, we can also imagine that analyzing a weather forecast sequence or a soccer game may involve different problems and models to interpret them. In the case of clouds motion for instance, it is especially important to better take into account the physics and have an appropriate notion of smoothness for the motion, while for the soccer game sequence motions are discontinuous and occlusions are very important.

### 5.1.2 *The optical flow: an apparent motion*

As soon as we consider a sequence, there is the idea of motion. Coming from displacements in the physical world, we can only observe a projection of it. This is illustrated in Figure 5.2.

Unfortunately, we are not able to measure the *2D motion field* (the projection on the image plane of the 3D velocity of the scene). As it is mentioned in the title of this section, what we are able to perceive is just an apparent motion also called the *optical flow*. By apparent, we mean that this 2D motion is only observable through intensity variations. Unfortunately, the optical flow and the 2D motion field are in general quantitatively different, unless very special conditions are satisfied. We refer to the discussion of Verri and Poggio [243, 244] for more details. Just to illustrate this difference, we show in Figure 5.3 two caricatural examples.

Still, if the optical flow and the 2D motion field are quantitatively different, they often share the same qualitative properties, like for instance motion discontinuities. The optical flow is then a rich source of information about the 3D kinematic behaviour of objects (see for instance [112]) or the geometrical structure of the world. It is also used in many other applications: segmentation, time to collision, earth sciences,...

In this last decade, numerous methods have been proposed to compute optical flow and it is still an active field of research. Three different strategies can be distinguished. *Correlation-based* techniques compare parts of the first image with parts of the second in terms of the similarity in brightness patterns in order to determine the motion vectors. Correlation is generally used to aid the matching of image features or to find image motion once features have been determined by alternative methods. *Feature-based* approaches aim at computing and analyzing the optic flow at a small number of well-defined image features (such as corners, edges, blobs) in a scene. *Gradient-based* methods (also called differential techniques) make use of spatio-temporal partial derivatives to estimate the image flow at each point in the image.

There is also a wide range of methodologies: wavelets, Markov Random Fields, Fourier analysis and naturally partial differential equations. In [26], Barron, Fleet and Beauchemin present the main different classes

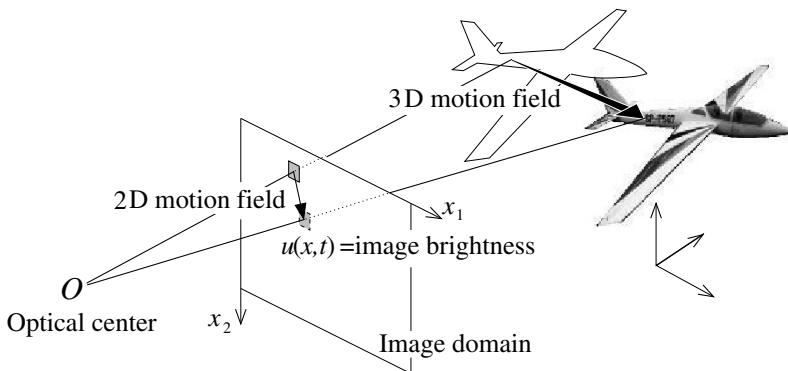


Figure 5.2. Simplified illustration of a camera (the pinhole camera model)

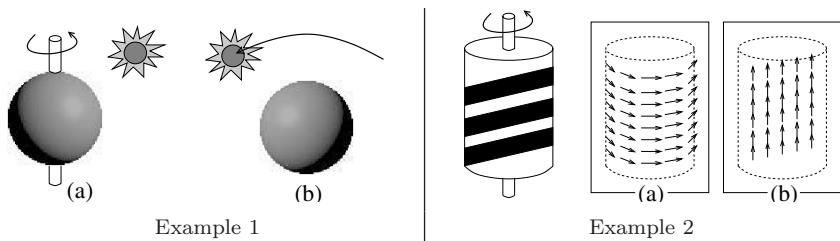


Figure 5.3. The optical flow is not always the motion field. Example 1 (see also [137]): In (a), no motion is perceived because intensity keeps constant while in (b) a static sphere is illuminated by a moving source, producing intensity variations. Example 2: the barber's pole. (a) shows the flow field and (b) the optical flow that is the motion perceived.

of techniques and perform numerical quantitative experiments to compare them. We also mention to the interested reader other interesting reviews: [197, 178, 242, 201, 113, 17].

According to the scope of the book, we now focus on differential techniques for which variational formulations can be proposed.

### The Optical Flow Constraint (OFC)

One of the first point to clarify and to formalize is the link between the intensity variations and the motion. A common and widely used assumption is that the intensity of a point keeps constant along its trajectory. We can consider it as reasonable for small displacements for which changes of the light source are small, and as long as there is no occlusion. More precisely, let  $u(t, x)$  denote the intensity of the pixel  $x = (x_1, x_2)$  at time  $t$ . Starting from a point  $x_0$  at the time  $t_0$ , we define the trajectory:

$$t \mapsto (t, x(t))$$

such that:

$$u(t, x(t)) = u(t_0, x_0) \quad \forall t \quad (5.1)$$

$$(t_0, x(t_0)) = (t_0, x_0). \quad (5.2)$$

By formally differentiating (5.1) with respect to  $t$ , we obtain at  $t = t_0$ :

$$\frac{dx}{dt}(t_0) \cdot \nabla u(t_0, x_0) + \frac{\partial u}{\partial t}(t_0, x_0) = 0. \quad (5.3)$$

So we will search the optical flow as the velocity field  $\sigma(x_0) = \frac{dx}{dt}(t_0)$  verifying (5.3).

To summarize, for a given sequence  $u(t, x)$  and a time of observation  $t_0$ , the aim is to find the instantaneous apparent velocity  $\sigma(x)$  such that:

$$\sigma(x) \cdot \nabla u(t, x_0) + u_t(t, x_0) = 0. \quad (5.4)$$

This equation is called the *optical flow constraint*, also noted OFC<sup>1</sup>.

- Unfortunately, one scalar equation is not enough to find both components of the velocity field. It only gives the component in the direction of  $\nabla u$ , that is the normal to the isophotes of the images. It is called the normal flow. This problem is usually designated by the aperture problem.

Solving the aperture problem

As we just saw, the equation (5.4) is not sufficient to compute the optical flow. Several ideas have been proposed to overcome this difficulty.

- Use second order derivative constraints. For instance, one could impose the conservation of  $\nabla u(t, x)$  along trajectories that is to say:

$$\frac{d\nabla u}{dt}(t, x) = 0.$$

This is a stronger restriction than (5.4) on permissible motion fields. This implies that rigid deformations are not considered. This condition can be rewritten in the following form:

$$\begin{bmatrix} u_{x_1 x_1} & u_{x_2 x_1} \\ u_{x_1 x_2} & u_{x_2 x_2} \end{bmatrix} \begin{pmatrix} \sigma_1 \\ \sigma_2 \end{pmatrix} + \begin{pmatrix} u_{x_1 t} \\ u_{x_2 t} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (5.5)$$

These equations can be used alone or together with the optical flow constraint. Several possibilities are then proposed (see [202, 240]. However, this kind of method is often noise sensitive because we need to compute second order derivatives.

---

<sup>1</sup>Notice that (5.4) is just an approximation at the first order of (5.1) and is valid only for small time differences. This will be commented next

- Another possibility is to solve the problem using a weighted least square approach [163, 164]. The central point of this method is a model of constant velocities in a small spatial neighborhood. For instance, to compute the velocity  $\sigma$  at point  $x_0$ , the idea is to minimized:

$$\inf_{\sigma(x_0)} \int_{B(x_0, r)} w^2(x) (\sigma(x_0) \cdot \nabla u + u_t)^2 dx,$$

where  $B(x_0, r)$  is the ball of center  $x_0$  and radius  $r$  (the neighborhood), and  $w(x)$  is a window function that gives more influence to the constraint at the center of the neighborhood than at the periphery. This approach which gives good results, has been extended and it is still often used to compute the optical flow. However, it is local and there is no notion of global regularity for the resulting flow.

- Using parametric models of velocity that respect as much as possible the optical flow constraint. In the affine case, one looks for  $\sigma$  such that:

$$\sigma(x) = \sigma_\theta(x) = \begin{pmatrix} \theta_1 + \theta_2 x_1 + \theta_3 x_2 \\ \theta_4 + \theta_5 x_1 + \theta_6 x_2 \end{pmatrix}$$

where the unknown parameter vector  $\theta \in R^6$  is determined by minimizing:

$$E(\theta) = \int_{\Omega} \phi(\sigma_\theta \cdot \nabla u + u_t) dx$$

where  $\phi$  is a suitable given function. In general this minimization leads to non quadratic (and possibly nonconvex) optimization problems that we can solve by half quadratic techniques if  $\phi$  satisfying hypotheses of Section 3.2.4. Naturally, other models may be proposed [139, 196].

- Regularizing the velocity field is another possibility. The idea is to consider a minimization problem of the form:

$$\inf_{\sigma} (A(\sigma) + S(\sigma)) \quad (5.6)$$

where  $A(\sigma)$  is the fidelity attach term for instance based on (5.4) or (5.1), and  $S(\sigma)$  is the smoothing term.

Amongst the first ones, Horn and Schunck [138] (see also [224]) proposed to solve the following problem:

$$\inf_{\sigma} \underbrace{\int_{\Omega} (\sigma \cdot \nabla u + u_t)^2 dx}_{A(\sigma)} + \alpha \underbrace{\sum_{j=1}^2 \int_{\Omega} |\nabla \sigma_j|^2 dx}_{S(\sigma)} \quad (5.7)$$

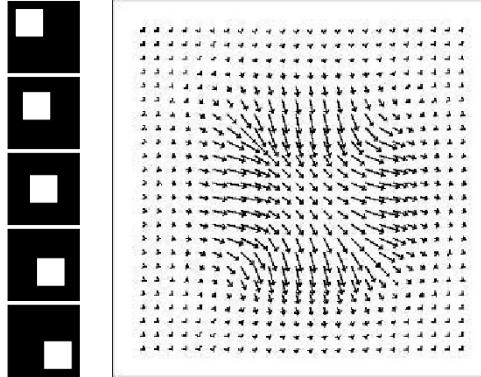


Figure 5.4. The method of Horn and Schunck applied on a synthetic example. Notice that this example is very caricatural since we do not have any texture information on the background and on the moving object. One may observe that the discontinuities near the edges are lost.

where  $\alpha$  is a constant. However, this kind of penalty term introduced by Tikhonov and Arsenin [239] is well-known to smooth isotropically without taking into account the discontinuities (see Figure 5.4 and also Section 3.2.2 where a similar term arise for image restoration). Unfortunately, as it has been mentioned, the discontinuities of the optical flow field are a very important cue for sequence analysis. Since then, many research have been carried out to compute discontinuous optical flow fields by changing the smoothing term  $S(\sigma)$ . We describe below some of the most significant ones:

- Modifying the Horn and Schunck functional was pioneered by Black et al [34, 37]. The idea was to change the regularization term into:

$$\sum_{j=1}^2 \int_{\Omega} \phi(|\nabla \sigma_j|) dx \quad (5.8)$$

where the function  $\phi$  would permit noise removal and edge conservation. Some examples include Cohen [77], Kumar, Tannenbaum and Balas [155] with the  $L^1$  norm (i.e. the total variation,  $\phi(s) = s$ ) or Aubert, Deriche et al [92, 17, 18] and Blanc-Féraud, Barlaud et al [40] with the  $\phi$  functions traditionally used for image restoration to preserve discontinuities. This was also proposed at the same time in a statistical framework where  $\phi$  functions are called robust estimators (see for instance [35, 177]).

- Suter [236], Gupta and Prince [132] or Guichard and Rudin [131] add some penalty terms based on the divergence and the

rotational of the flow field:

$$\int_{\Omega} \varphi(\operatorname{div}(\sigma), \operatorname{rot}(\sigma)) \, dx \quad (5.9)$$

where several possibilities for  $\varphi$  may be proposed. For instance, in [131] the authors choose  $\varphi(\operatorname{div}(\sigma), \operatorname{rot}(\sigma)) = |\operatorname{div}(\sigma)|$  which is adapted to rigid 2D objects with a “2D” motion.

- Nagel and Enkelmann [188, 103] propose an oriented smoothness constraint in which smoothness is not imposed across steep intensity gradients (edges) in an attempt to handle occlusions. So the penalty term is of the form:

$$\int_{\Omega} \frac{1}{|\nabla u|^2 + 2\lambda^2} \operatorname{trace}((\nabla \sigma)^T D(\nabla u)(\nabla \sigma)) \, dx$$

with  $D(\nabla u) = n \, n^T + \lambda^2 Id$  and  $n = \begin{pmatrix} u_{x_2} \\ -u_{x_1} \end{pmatrix}$

$$(5.10)$$

where  $\lambda$  is a constant. The idea is to attenuate the blurring of the flow across the boundaries of the intensity, when  $|\nabla u| \gg \lambda$ . In this case, the smoothing is essentially in the direction tangent to the isophotes. Otherwise, the smoothing is isotropic. This is a major difference with (5.8) or (5.9) since the characteristics of the smoothing depend here on the intensity and not on the motion itself. However, one can wonder about the action of this term (5.10) for highly textured scenes where the gradient is varying a lot and is not very representative of objects boundaries

- Nési [193] adapts the formulation of Horn and Schunck introducing the length of the discontinuity set of  $\sigma$  (noted  $|S_\sigma|$ ). We recall that this kind of idea has been introduced by Mumford and Shah [187] for image segmentation (see Chapter 4). The regularization term is of the form:

$$\sum_{j=1}^2 \int_{\Omega} |\nabla \sigma_j|^2 \, dx + \alpha |S_\sigma|,$$

where  $\alpha$  is a constant. Numerically, like for the image segmentation problem, the main difficulty is to approximate the last term. One possible solution is to use the notion of  $\Gamma$ -convergence (see Section 2.1.4). We introduce a sequence of functionals so that the sequence of minimizers converge to the unique minimum of the initial functional. Typically, the way to approximate the regularization term is (see Section 4.2.4 and [13] for more detail):

$$\sum_{j=1}^2 \int_{\Omega} z^2 |\nabla \sigma_j|^2 \, dx + \alpha \int_{\Omega} \left( \frac{|\nabla z|^2}{k} + \frac{k(1-z)^2}{4} \right) \, dx$$

where  $z$  is an additional function and  $k$  is a parameter which is destined to tend to infinity. The function  $z$  can be considered as a control variable which is equal to zero near discontinuities and close to 1 in homogeneous regions.

### Overview of a discontinuity preserving variational approach

Although many models have been proposed to find the optical flow, little work considered its mathematical analysis. We summarize below some results presented in [17, 18, 151], where the smoothing term is the same as in the restoration problem of Section 3.2.3.

Given a sequence  $u(t, x)$  we search for the velocity field  $\sigma$  which realizes the minimum of the energy:

$$E(\sigma) = \underbrace{\int_{\Omega} |\sigma \cdot Du + u_t|}_{A(\sigma)} + \underbrace{\alpha^s \sum_{j=1}^2 \int_{\Omega} \phi(D\sigma_j)}_{S(\sigma)} + \underbrace{\alpha^h \int_{\Omega} c(Du) |\sigma|^2}_{H(\sigma)} \quad (5.11)$$

where  $\alpha^s, \alpha^h$  are positive constants. From now on, unless specified otherwise all the derivative are written in a formal setting (in the distributional sense). Since we look for discontinuous optical flows the suitable theoretical background to study this problem will be  $\mathbf{BV}(\Omega)$ , the space of bounded variation (see Section 2.2). The energy is compounded of three terms:

- $A(\sigma)$  is the “ $L^1$ ”-norm of the OFC (5.4). In fact it is formal and has to be interpreted as a measure.
- $S(\sigma)$  is the smoothing term. Like for image restoration (see Section 3.2), one would like to find conditions on  $\phi$  so that discontinuities may be kept. We recall the assumptions of Section 3.2.3:

$$\phi \text{ is a strictly convex, nondecreasing function from } R^+ \text{ to } R^+, \text{ with } \phi(0) = 0 \text{ (without a loss of generality)} \quad (5.12)$$

$$\lim_{s \rightarrow +\infty} \phi(s) = +\infty \quad (5.13)$$

$$\begin{aligned} &\text{There exist two constants } c > 0 \text{ and } b \geq 0 \text{ such that} \\ &cs - b \leq \phi(s) \leq cs + b \quad \forall s \geq 0. \end{aligned} \quad (5.14)$$

Notice that these conditions will guarantee the well-posedness of the theoretical problem. Regarding the edge preservation properties, other qualitative conditions should be added (see Section 3.2.2, conditions (3.10) and (3.12)). Under these assumptions,  $S(\sigma)$  has to be interpreted as a convex function of measures (see Section 2.2.4). We recall that this term is l.s.c. for the  $\text{BV} - w^*$  topology.

- $H(\sigma)$  is related to homogeneous regions. The idea is that if there is no texture that is to say no gradient, there is no way to estimate

correctly the flow field. Then one may force it to be zero. This is done through a weighted  $L^2$ -norm, where the function  $c$  is such that:

$$\lim_{s \rightarrow 0} c(s) = 1 \quad \text{and} \quad \lim_{s \rightarrow \infty} c(s) = 0.$$

Since  $u$  (and then  $Du$ ) is given, we will simply denote  $c(x)$  instead of  $c(Du(x))$ . Without any loss of generality, we assume that:

$$c \in C^\infty(\Omega) \tag{5.15}$$

$$\text{There exists } m_c > 0 \text{ such that } c(x) \in [m_c, 1] \text{ for all } x \in \Omega \tag{5.16}$$

Although this term may be criticized on a modelization point of view, it is necessary for the coercivity of the functional. Let us also remark that this term is well-defined on  $\mathbf{BV}(\Omega)$  thanks to the inclusion of  $\mathbf{BV}(\Omega)$  into  $\mathbf{L}^2(\Omega)$  ( $N = 2$ ).

Now that we have stated the problem, let us consider its theoretical study. Up to now nothing has been told about the regularity of the data  $u$ . Interestingly this will influence very importantly the nature of the problem. As a first example, let us consider that the data is Lipschitz in space and time [17, 151]:

$$u \in W^{1,\infty}(R \times \Omega). \tag{5.17}$$

The derivatives in  $A(\sigma)$  are then functions and  $A(\sigma)$  is simply the  $L^1$ -norm of the OFC. Notice that this assumption is realistic from a numerical point of view because a pre-smoothing is usually carried out to diminish noise effects, remove small amounts of temporal aliasing and improve the subsequent derivative estimates<sup>2</sup>. Then, we have the following result:

**Theorem 5.1.1** [17] *Under the hypotheses (5.12)-(5.14), (5.15)-(5.16) and (5.17), the minimization problem:*

$$\inf_{\sigma \in \mathbf{BV}(\Omega)} E(\sigma) = \int_{\Omega} |\sigma \cdot \nabla u + u_t| \, dx + \sum_{j=1}^2 \int_{\Omega} \phi(D\sigma_j) + \int_{\Omega} c(x)|\sigma|^2 \, dx \tag{5.18}$$

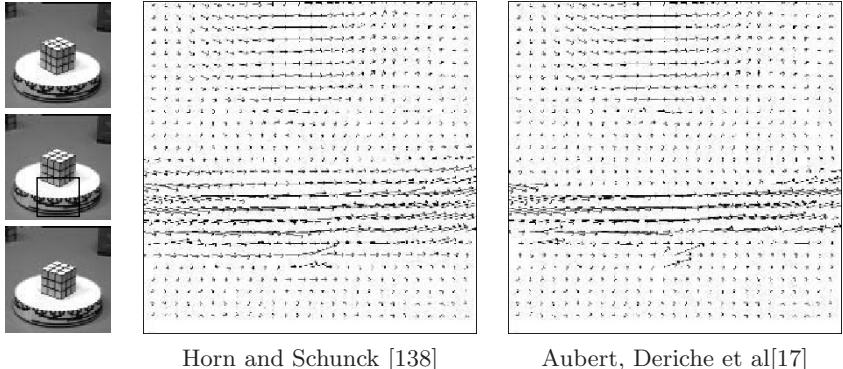
*admits a unique solution in  $\mathbf{BV}(\Omega)$ .*

In this case, the proof follows from classical arguments. According to (5.14)-(5.16), the functional  $E$  is coercive on  $BV(\Omega)$ . Thus, we can uniformly bound the minimizing sequences and extract a converging subsequence for the  $BV - w^*$  topology. Since  $E$  is lower semi-continuous (l.s.c.) for this topology, we easily deduce the existence of a minimum.

---

<sup>2</sup>As mentioned in [26] differential techniques are naturally very sensitive to the quality of the estimation of the spatiotemporal derivative. Along the same lines, on a discrete point of view the method of numerical differentiation is very important.

Now the problem is to get an approximation of the solution. As in the case of image restoration, an algorithm based on  $\Gamma$ -convergence and half-quadratic minimization can be proposed (see Section 3.2.4 and [17]). We show in Figure 5.5 a typical result where one can observe the qualitative differences with the original Horn and Schunck's model.



Horn and Schunck [138]      Aubert, Deriche et al[17]

Figure 5.5. Example of result obtained on the rotating cube sequence. A close-up of the lower part of the plate is displayed to highlight the qualitative differences. One can observe that the diffusion is controloed in the right-hand side case

Unfortunately, the regularity assumption (5.17) on the data  $u$  may not always be verified and one can wonder what would be the problem if we only assume that  $u$  is a function of bounded variation. This is considered in [18, 151] and the fact that  $u$  may have jumps induces not trivial theoretical questions. The first difficulty is to give a sense to the first term  $A(\sigma)$ . It is now a measure. To be more explicit, one needs to find an integral representation of this term<sup>3</sup>. The precise assumptions on  $u$  are:

$$u \in SBV(R \times \Omega) \cap L^\infty(R \times \Omega) \quad (5.19)$$

$$\begin{aligned} &\text{There exists } h_1 \in L^1(\Omega) \text{ and } h_2 \in L^1_{\mathcal{H}^1}(S_u) \\ &\text{such that } u_t = h_1 dx + h_2 \mathcal{H}^1|_{S_u} \end{aligned} \quad (5.20)$$

where  $SBV(R \times \Omega)$  is the space of special functions of bounded variation (no Cantor part) and  $\mathcal{H}^1$  is the one-dimensional Hausdorff measure. Notice that (5.20) means that the measure  $u_t$  is absolutely continuous with respect to  $|Du|$ . This is physically correct since when there is no texture (no gradient) no intensity variation should be observed. Then, under these assumptions, it can be established that the energy  $E$  defined in (5.11) can be rewritten

---

<sup>3</sup>By integral representation, we mean finding a measure  $\mu$  and a function  $h \in L^1_\mu(\Omega)$  such that  $A(\sigma) = \int_\Omega h d\mu$ .

as:

$$\begin{aligned} E(\sigma) = & \int_{\Omega} |\sigma \cdot \nabla u + h_1| dx + \int_{S_u} |\dot{\sigma} \cdot n_u (u^+ - u^-) + h_2| ds + \\ & + \sum_{j=1}^2 \int_{\Omega} \phi(D\sigma_j) + \int_{\Omega} c(x)|\sigma|^2 dx \end{aligned}$$

where  $S_u$  is the jump set of  $u$  and  $\dot{\sigma}$  is the precise representation of  $\sigma$  (see Section 2.2.3). We recall that  $\dot{\sigma}$  belongs to the same class as  $\sigma$  but is now defined  $\mathcal{H}^1$ -a.e. by:

$$\dot{\sigma}(x) = \lim_{r \rightarrow 0} \frac{1}{|B(x, r)|} \int_{B(x, r)} \sigma(y) dy$$

where  $B(x, r)$  is the ball of center  $x$  and radius  $r$ . We also recall that:

$$\dot{\sigma}(x) = \sigma(x) \text{ } dx \text{ - a.e. and } \dot{\sigma}(x) = \frac{\sigma^+(x) + \sigma^-(x)}{2} \text{ } \mathcal{H}^1 \text{ a.e. on } S_{\sigma}.$$

So, we can observe that the OFC is now splitted into two parts: an absolutely continuous part and a length part on the jump set of  $u$ . To study the existence of a solution for this problem the difficulty is that the functional is no longer l.s.c. for the  $\mathbf{BV}-w^*$  topology because of the term on  $S_u$ . Basically, the problem is that the trace function is l.s.c. for the strong topology of  $BV(\Omega)$  (even continuous, see Section 2.2.3) but not for the  $\mathbf{BV}-w^*$  topology. If  $\sigma^n$  is a minimizing sequence of  $E$  such that  $\sigma^n \xrightarrow{\mathbf{BV}-w^*} \sigma$  then in general:

$$\text{tr}(\sigma^n) \rightarrow \nu \in \mathcal{M}(\Omega) \text{ and } \nu \neq \text{tr}(\sigma),$$

where  $\text{tr}(\cdot)$  is the trace operator. It is then necessary to compute the relaxed functional. This leads to long and technical calculus and we refer to [18] for more details.

### Alternatives of the OFC

Is the OFC unavoidable? Even if it is widely used to compute optical flow, several reasons may invite us to look for something different.

The first one is that one may wonder about the validity of using the OFC (5.4) in case of large displacements. If we denote by  $\Delta t$  the time interval between two consecutive images (in the discrete temporal case), the differentiation (5.1) is only valid for “small”  $\Delta t$  or equivalently for “small” displacements. To deal with large displacements there are two possibilities:

- One can use a multiresolution approach with computations at each resolution level according to a coarse to fine strategy (see for instance [103, 176]). It is then a modified OFC which is considered at each resolution level.

- One may prefer to keep the conservation equation (5.1) without differentiating it i.e. the displacement frame difference. So the fidelity attach term in (5.6) is of the form [188, 131, 8]:

$$A(\sigma) = \int_{\Omega} (u(x + \sigma \Delta t, t + \Delta t) - u(t, x))^2 \, dx,$$

which is now nonlinear with respect to  $\sigma$ . It is then more similar to a correlation problem. If this approach gives satisfying numerical results it is not clear on a theoretical point of view (how to give a sense to a possibly discontinuous function  $u$  depending on another discontinuous function  $\sigma$ ?).

The second is related to the *a priori* knowledge that we may have on the origins of the sequence. Typical examples include fluid flow estimation [252, 78] or weather forecast sequence analysis [31, 257] (see also [94, 2, 189, 225, 226]). For instance, for fluid motion, one should impose a conservation of mass:

$$\operatorname{div}(\rho V) + \frac{\partial \rho}{\partial t} = 0$$

where  $\rho(t, x)$  is the density of the fluid at position  $X \in R^3$  and time  $t$ , and  $V(t, x) \in R^3$  is the velocity. Now the problem is to know the link between the density  $\rho$  and the image brightness. For example, in the case of 2D transmittance image of a 3D fluid flow the intensities are proportional to the density integrated along the path of impinging energy. Then, it can be shown [108, 252] that the mass conservation equation implies that:

$$\operatorname{div}(u\sigma) + u_t = 0$$

which is different from the classical optical flow constraint.

The third is that we may simply want to relax the brightness consistency assumption, which is true when the scene surface is Lambertian and is either stationary as the camera moves or moves parallel to the image plane. It is also a good approximation as soon as the surface has rich texture, such that the brightness change due to shadings or surface lightning conditions are negligible relative to that due to the motion effects. Unfortunately, this is not the case in many applications [190] and one needs to relax this assumption. For instance, the models proposed by Negahdaripour and Yu [192] or Mattavelli and Nicoulin [175] permit an affine variation of the intensity during time (and not a conservation). We also refer the interested reader to [135, 36, 133, 191] for other possibilities.

\* The optical flow problem is not near to being completely solved. The choice of the data term is not yet well-established and there is not a unique possibility. As far as the regularity is concerned, when presenting the short overview of possible regularization terms, we may have noticed that:

- Most of them are intrinsic: the diffusion is only controlled by the flow itself.
- Few of them are extrinsic: the diffusion is controlled by the intensity image itself (see (5.10)).

*It is not clear actually which solution is “the best” in terms of modelization and numerical results.*

*Another point which would need further developments is to better understand the link between  $u$  and  $\sigma$  in terms of their discontinuity sets. If it seems clear that the set  $S_\sigma$  should be contained in  $S_u$ , this should be taken into account in the model and one should do a finer analysis of these relationships.*

*Finally one can not ignore the coupling between the fidelity attach term and the regularization term.*

### 5.1.3 Sequence segmentation

#### Introduction

As suggested in the introduction, another important task in sequence analysis is segmentation. Here segmentation means finding the different objects in a scene, and this is naturally in relation with velocity estimation or optical flow. Two kinds of approaches based on motion estimation can be distinguished: either they detect flow discontinuities (local operators) or they extract patches of self consistant motion (global measurements). In any case, this is dependent on the quality of the flow that can be obtained. As suggested in the previous section, this estimation may be hard or impossible to obtain (for instance it is necessary to have a reasonable time sampling, with a limited amount of noise). Unfortunately, this is not always the case for many real applications. Just think about video-surveillance where sensors are often of poor quality and for which low images rates are usually considered because of storage capacity.

To avoid these difficulties, another possibility is to consider that the sequence is compounded of layers, typically a background and a foreground. To make this point more clear, let us focus on the case of sequences with static background (see for example Figure 5.1 or the synthetic sequence in Figure 5.6). The idea is that the background has a “persistancy” and then the objects can be seen as occluding it, “being on top of it” (see Figure 5.7). Naturally, this means that only the objects with a different color than the background can be detected. However as this is not motion based an object stopping for a while will still be detected.

This idea of comparing a reference image with the current image is very intuitive but not always applicable in real applications, especially when noise is present. This is illustrated for the sequence presented in Figure 5.8. If we simply compute the difference between one image of the sequence and

the ideal background (see Figure 5.9) we obtain an image which enhances the objects but also the noise. A threshold can be used to turn this image into a binary one. Examples are shown in Figure 5.10. It can be observed that the choice of this threshold is not an easy task: either noise is kept or objects are partially lost. Another important question is how a reference image (i.e. the background image) can be obtained? A simple idea is to compute the temporal mean of the sequence<sup>4</sup> (see Figure 5.10). It can be noticed that some shades of the objects are still present and thus making the difference with an image of the sequence will not be satisfying. This effect is even stronger given that the sequence is short. It is then necessary to have a robust technique to estimate the background and classical approaches compute statistical background models [107, 126, 254, 205, 203].

Something important which comes out of this discussion is that in the case of noisy sequences obtaining a reference image is as difficult as segmenting the sequence. In fact...

 Having a reference image means that we have previously extracted the objects. Conversely, being able to extract the objects means having somewhere a reference image.

This remark suggests strong links in the estimation of background and of moving objects. It should then be more efficient to estimate both at the same time.

### A variational formulation (the time-continuous case)

According to the previous discussion, let us present the variational approach proposed by Kornprobst, Deriche and Aubert [154]. Let  $N(t, x)$  denotes the given noisy sequence ( $t \in [0, t_{\max}], x \in \Omega$ ) for which the background is assumed to be static. We look simultaneously for (see Figure 5.11):

- The restored background  $B(x)$ .
- The sequence  $C(t, x)$  which indicates the moving regions. Typically,  $C(t, x) = 0$  if the pixel  $x$  belongs to a moving object, and 1 otherwise.

---

<sup>4</sup>In this example where the sequence is given by 5 images noted  $(N_i)_{i=1..5}$ , the temporal mean is  $M(x) = \frac{1}{5} \sum_{i=1}^5 N_i(x)$ ,  $x \in \Omega$ .

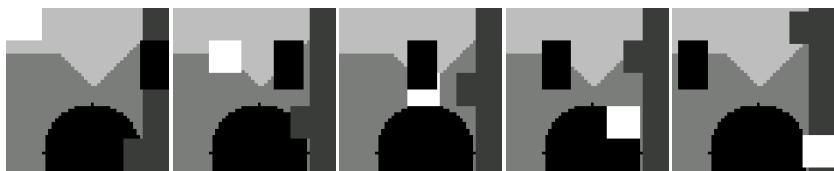


Figure 5.6. Example of synthetic sequence of 5 images with 3 moving objects

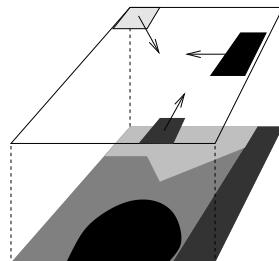


Figure 5.7. Interpretation of the sequence presented in Figure 5.6 as two layers: background and foreground

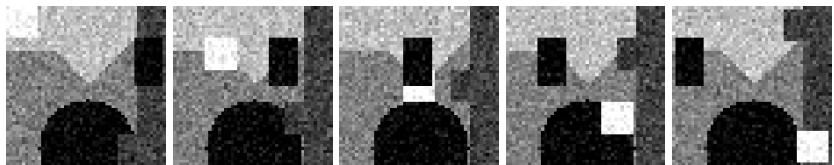


Figure 5.8. Synthetic sequence with Gaussian additive noise ( $\sigma = 20$ )

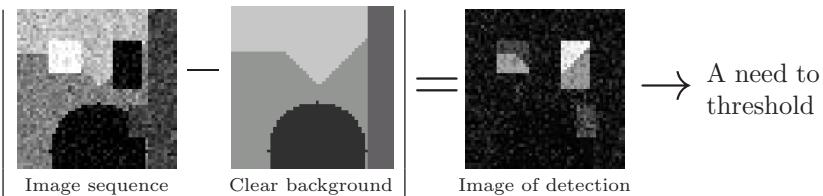


Figure 5.9. Example of background subtraction for the second image (The resulting image needs to be thresholded to extract the objects)

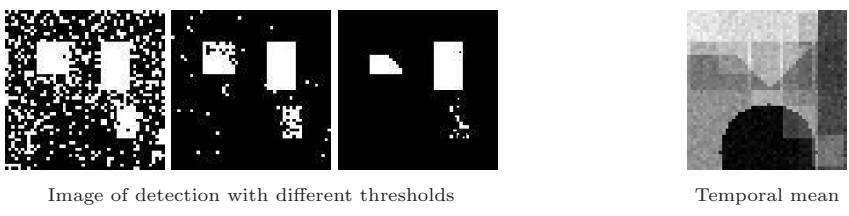


Figure 5.10. Difficulties of background subtraction method

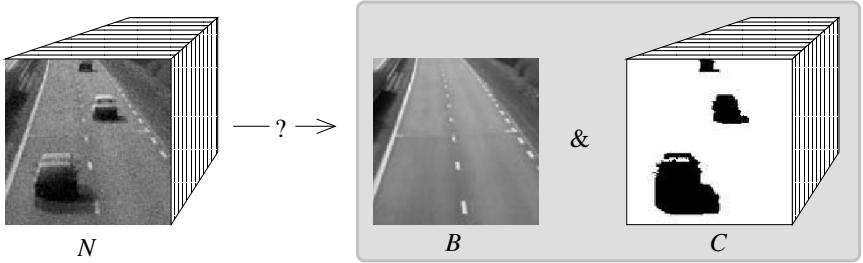


Figure 5.11. Objectives of the approach [154] on a typical example

To solve this problem it is proposed in [154] to minimize with respect to  $B$  and  $C$ :

$$\underbrace{\iint_V \left[ C^2(B - N)^2 + \alpha_c(C - 1)^2 \right] dxdt}_{A(\sigma)} + \underbrace{\alpha_b^r \int_{\Omega} \phi_1(DB) + \alpha_c^r \iint_V \phi_2(DC)}_{S(\sigma)} \quad (5.21)$$

where  $V = [0, t_{\max}] \times \Omega$  and  $\alpha_c, \alpha_b^r, \alpha_c^r$  are positive constants. As usual all the derivative are written in a formal setting (in the distributional sense). The energy is compounded of two kinds of terms:

- $A(\sigma)$  realizes the coupling between the two variables  $B$  and  $C$ . The second term forces the function  $C$  to be equal to 1, which corresponds to the background. However if the current image  $N$  is too different from the background (meaning that an object is present) then the first term will be too high, which will force  $C$  to be 0.
- $S(\sigma)$  is the smoothing term. Like for image restoration (see Section 3.2) one would like to find conditions on  $(\phi_i)_{i=1,2}$  so that discontinuities may be kept (for the background as well as for the images of detection). We recall the assumptions of Section 3.2.3:

$\phi_i$  is a strictly convex, nondecreasing function from  $R^+$  to  $R^+$ , with  $\phi_i(0) = 0$  (without a loss of generality) (5.22)

$$\lim_{s \rightarrow +\infty} \phi_i(s) = +\infty \quad (5.23)$$

There exist two constants  $c > 0$  and  $b \geq 0$  such that (5.24)

$$cs - b \leq \phi_i(s) \leq cs + b \quad \forall s \geq 0.$$

As noticed previously, these conditions will guarantee the well-posedness of the theoretical problem. Regarding the edge preservation properties, other qualitative conditions should be added (see Section 3.2.2, conditions (3.10) and (3.12)). Under these assumptions,  $S(\sigma)$  has to be interpreted as a convex function of measures (see Section

2.2.4). We recall that this term is l.s.c. for the  $BV - w^*$  topology. In the sequel, as there is no reason to choose,  $\phi_1 \neq \phi_2$  we will simply denote  $\phi = \phi_1 = \phi_2$ .

Mathematical study of the time sampled energy

In fact we do not have a continuum of images. A sequence is represented by a finite number of images. From the continuum  $[0, t_{\max}]$  we have  $T$  images noted  $N_1(x), \dots, N_T(x)$ . For the theoretical framework it is assumed that:

$$N_h \in BV(\Omega) \cap L^\infty(\Omega) \quad \forall h = 1..T. \quad (5.25)$$

The bounds of the gray-value levels over the sequence are denoted by:

$$\begin{cases} m_N = \text{ess-inf}_{h \in [0..T], x \in \Omega} N_h(x) \\ M_N = \text{ess-sup}_{h \in [0..T], x \in \Omega} N_h(x). \end{cases} \quad (5.26)$$

where ess-inf (resp. ess-sup) is the essential infimum (resp. essential supremum), that is the infimum up to Lebesgue measurable sets.

Again, the unknowns are  $B$  (the image of the restored background) and  $C_1, \dots, C_T$  (the  $T$  images of detection). The suitable functional space for studying this problem is the space of bounded variation (see Section 2.2) since both  $B$  and  $C_h$  are likely to have discontinuities across some contours. The time discretized version of (5.21) is then to search for the solution of:

$$\inf_{(B, C_1, \dots, C_T) \in BV(\Omega)^{T+1}} E(B, C_1, \dots, C_T) \quad \text{with} \quad (5.27)$$

$$E(B, C_1, \dots, C_T) = \quad (5.28)$$

$$= \sum_{h=1}^T \int_{\Omega} \left[ C_h^2 (B - N_h)^2 + \alpha_c (C_h - 1)^2 \right] dx + \alpha_b^r \int_{\Omega} \phi(DB) + \alpha_c^r \sum_{h=1}^T \int_{\Omega} \phi(DC_h).$$

We are now interested in proving the existence and the uniqueness for this problem. Before going further let us point out two main difficulties:

- The functional  $E$  is degenerated because of the first term. As a consequence, applying the direct method of the calculus of variations as described in Section (2.1.2) does not give any result: if we choose a minimizing sequence  $(B_n)_{n \in N}, (C_n^h)_{n \in N}$ , we can easily bound the sequence  $(C_n^h)_{n \in N}$  thanks to the second term but nothing can be said as for  $(B_n)_{n \in N}$  because the functions  $(C_n^h)_{n \in N}$  have no lower bound.
- Though the functional  $E$  is convex with respect to each variable, it is nonconvex globally. This is naturally an issue as far as uniqueness is concerned.

To overcome the first difficulty, let us consider the problem (5.27)-(5.28) set over the constrained space:

$$\begin{aligned} \bar{\mathcal{B}}(\Omega) = & \left\{ (B, C_1, \dots, C_T) \in BV(\Omega)^{T+1} \text{ such that } m_N \leq B \leq M_N \right. \\ & \left. \text{and } 0 \leq C_h \leq 1 \quad \forall h \right\}. \end{aligned}$$

One can remark that these constraints are quite natural: it is not expected that  $B$  would have values never reached in the original sequence and considering the functions  $C_h$  bounded is *a priori* reasonable. Since the variables are now uniformly bounded it is now clear that the problem:

$$\inf_{(B, C_1, \dots, C_T) \in \bar{\mathcal{B}}(\Omega)} E(B, C_1, \dots, C_T) \quad (5.29)$$

admits a solution in  $\bar{\mathcal{B}}(\Omega)$ . However this result is not satisfying as soon as we are interested in finding a numerical solution: the optimality conditions are now inequations instead of equations, and one would need to use Lagrange multipliers.

► Interestingly, if there is a solution of the problem (5.27) then the solution belongs to  $\bar{\mathcal{B}}(\Omega)$ . As a consequence, we can prove the existence of a solution for the problem (5.27).

The following technical lemma is helpful to prove the main result:

**Lemma 5.1.1** Let  $u \in BV(\Omega)$ ,  $\phi$  a function verifying hypotheses (5.22), (5.23), (5.24) and  $\varphi_{\alpha, \beta}$  the cut-off function defined by:

$$\varphi_{\alpha, \beta}(x) = \begin{cases} \alpha & \text{if } x \leq \alpha \\ x & \text{if } \alpha \leq x \leq \beta \\ \beta & \text{if } x \geq \beta \end{cases} \quad (5.30)$$

Then we have:

$$\int_{\Omega} \phi(D\varphi_{\alpha, \beta}(u)) \leq \int_{\Omega} \phi(Du).$$

**Proof** This lemma is very intuitive but the proof requires some attention. Let us first recall the Lebesgue decomposition of the measure  $\phi(Du)$ :

$$\int_{\Omega} \phi(Du) = \underbrace{\int_{\Omega} \phi(|\nabla u|) dx}_{\text{term 1}} + \underbrace{\int_{S_u} |u^+ - u^-| \mathcal{H}^{N-1}}_{\text{term 2}} + \underbrace{\int_{\Omega - S_u} |C_u|}_{\text{term 3}}.$$

We are going to show that cutting the function  $u$  using the function  $\varphi_{\alpha,\beta}$  permits to reduce each term. To simplify the notations we will denote  $\hat{u}$  instead of  $\varphi_{\alpha,\beta}(u)$ .

**Term 1:** let  $\Omega_c = \{x \in \Omega; u(x) \leq \alpha \text{ or } u(x) \geq \beta\}$  and  $\Omega_i = \Omega - \Omega_c$ .

Thanks to [146], we have  $\int_{\Omega_i} \phi(|\nabla \hat{u}|) dx = \int_{\Omega_i} \phi(|\nabla u|) dx$ . Consequently:

$$\int_{\Omega} \phi(|\nabla \hat{u}|) dx = \int_{\Omega_i} \phi(|\nabla u|) dx + \int_{\Omega_c} \phi(|\nabla \hat{u}|) dx \stackrel{(\equiv 0)}{\leq} \int_{\Omega} \phi(|\nabla u|) dx. \quad (5.31)$$

**Term 2:** using results proved in [9], we know that:

$$S_{\hat{u}} \subset S_u \text{ and } \hat{u}^+ = \varphi_{\alpha,\beta}(u^+), \hat{u}^- = \varphi_{\alpha,\beta}(u^-).$$

Since  $\varphi_{\alpha,\beta}$  is Lipschitz continuous with a constant equals to 1, we then have:

$$\int_{S_{\hat{u}}} |\hat{u}^+ - \hat{u}^-| \mathcal{H}^{N-1} \leq \int_{S_{\hat{u}}} |u^+ - u^-| \mathcal{H}^{N-1} \leq \int_{S_u} |u^+ - u^-| \mathcal{H}^{N-1}. \quad (5.32)$$

**Term 3:** we need to understand how is the Cantor part of the distributional derivative of the composed function  $\varphi_{\alpha,\beta}(u)$ . Vol'pert [247] first proposed a *chain rule formula* for functions  $v = \varphi(u)$  for  $u \in BV(\Omega)$  and when  $\varphi$  is continuously differentiable. Ambrosio and Dal Maso [12] gave extended results for functions  $\varphi$  uniformly Lipschitz continuous. Since  $u$  is scalar, it is demonstrated in [12] that we can write:

$$C(\varphi_{\alpha,\beta}(u)) = \varphi'_{\alpha,\beta}(\tilde{u}) C(u) \quad |Du|\text{-a.e. on } \Omega - S_u \quad (5.33)$$

where  $\tilde{u}$  is the approximate limit (see Section 2.2.3) of  $u$  defined by:

$$\lim_{r \rightarrow 0^+} \frac{1}{|B(x, r)|} \int_{B(x, r)} |u(y) - \tilde{u}(x)| dy = 0$$

where  $B(x, r)$  is the closed ball with center  $x$  and radius  $r$ . Moreover, we have:

$$\int_{\Omega - S_{\hat{u}}} |C_{\hat{u}}| = \int_{\Omega - S_u} |C_{\hat{u}}| + \int_{S_u / S_{\hat{u}}} |C_{\hat{u}}| \quad (5.34)$$

where the last term over  $S_u/S_{\hat{u}}$  is zero<sup>5</sup>. Then, using the chain rule formula (5.33), we have:

$$\int_{\Omega - S_{\hat{u}}} |C_{\hat{u}}| \leq |\varphi'_{\alpha,\beta}|_{L^\infty} \int_{\Omega - S_u} |C_u| \leq \int_{\Omega - S_u} |C_u|. \quad (5.35)$$

The inequalities (5.31), (5.32) and (5.35) conclude the proof. ■

Then, the following result can be established:

**Theorem 5.1.2** *The problem (5.27) admits a solution on  $BV(\Omega)^{T+1}$ .*

*If moreover:*

$$\alpha_c \geq 3(M_N - m_N)^2, \quad (5.36)$$

*then the solution is unique.*

**Proof** Existence is proven showing that minimizing (5.28) over  $\bar{\mathcal{B}}(\Omega)$  is equivalent to the same problem posed over  $BV(\Omega)^{T+1}$ , that is to say without any constraint (this is a direct consequence of Lemma 5.1.1) for which we apply the direct method of the calculus of variations

As far as uniqueness is concerned, the difficulty comes from the nonconvexity of the function:

$$(B, C_1, \dots, C_T) \rightarrow \sum_{h=1}^T C_h^2 (B - N_h)^2 + \alpha_c \sum_{h=1}^T (C_h - 1)^2.$$

However, if  $\alpha_C$  is large enough, it can be proved that this functional is strictly convex over  $\bar{\mathcal{B}}$  which permits to conclude (see [154] for more details). ■

**Remark** The condition (5.36) is in fact quite natural. It means that the background must be sufficiently taken into account. ■

## Experiments

An important consequence of Theorem 5.1.2 is that it permits to consider the minimization problem over all  $BV(\Omega)^{T+1}$  without any constraint. Consequently, on a numerical point of view, the difficulties and techniques are the same as for the previous variational approaches studied in this book, for image restoration (Section 3.2.4) or optical flow (Section 5.1.2). Although optimality equations may be written, they remain hard to handle (see Section 3.2.4). Similarly, an algorithm based on  $\Gamma$ -convergence and half-quadratic minimization can be proposed. We refer to [154] where convergence results are proved.

---

<sup>5</sup>We recall that for any  $v \in BV(\Omega)$  and any set  $S$  of Hausdorff dimension at most  $N - 1$ , we have  $C_v(S) = 0$ .

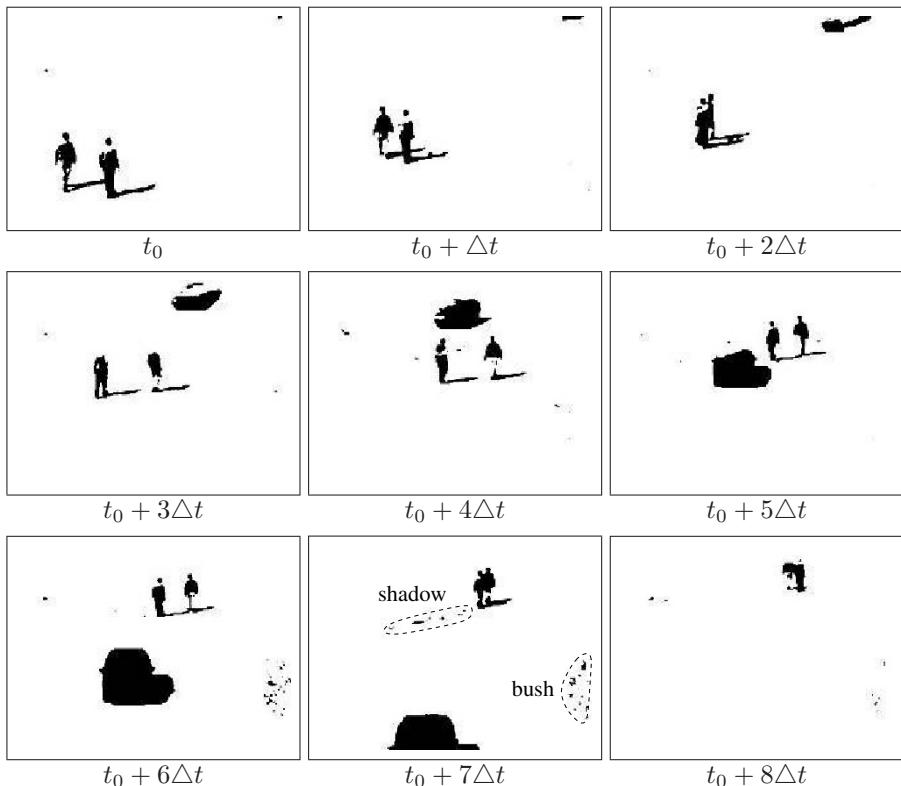


Figure 5.12. “Street” sequence (see also Figure 5.1): sequence of detections.

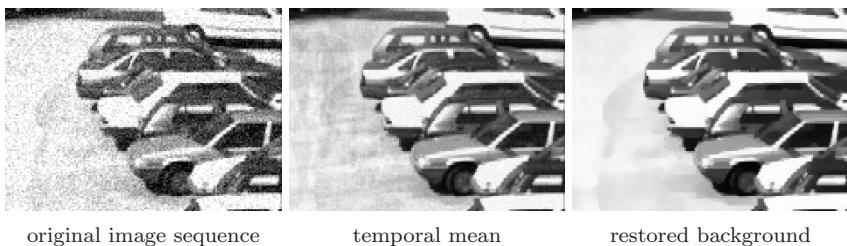


Figure 5.13. “Street” sequence (see also Figure 5.1). Close-ups on the same area (top right-hand corner) for the initial sequence, the temporal mean of the sequence and the restored background.

To illustrate this approach, let us present some results on real sequences. We first show some results on the “Street” sequence presented in Section 5.1.1 (Figure 5.1). Figure 5.12 shows an example of detection. Notice that the motions of the bush and the shadow are detected. We illustrate in Figure 5.13 the restoration aspect. The temporal mean of the sequence is still noisy (because the motion of the persons is taken into account and

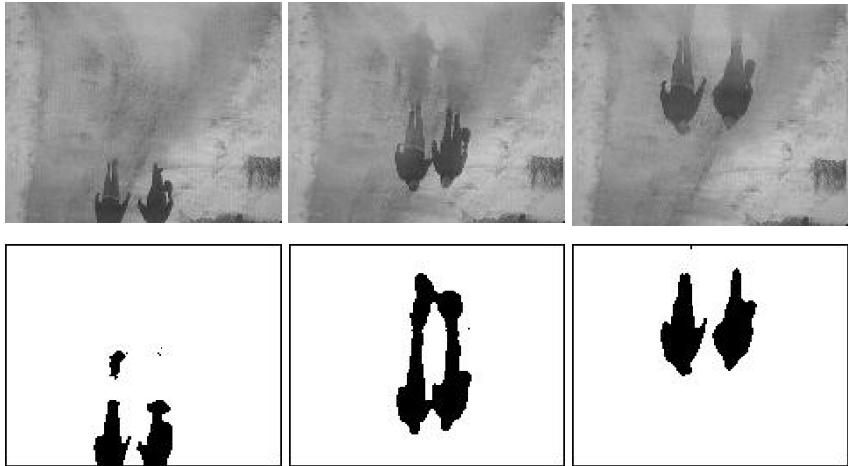


Figure 5.14. “Walking in Finland” sequence, from University of Oulu, 55 images. Three images of the sequence  $N$  the corresponding  $C$  functions.

the noise is not of zero mean), while the restored background has a very good quality.

We show in Figure 5.14 another real sequence where the reflections on the ground are detected as “motion”.

We also recall that this method permits to detect objects by comparison with a reference image. As it is not based on motion, a person stopping for some time and walking again will be detected. We refer the interested reader to [154] for more results and for quantitative experiments on the noise influence with respect to the results.

Finally, let us mention that this approach can be extended in the case of video-streams instead of batch (that is post-processing a set of given images). This is the case of most real applications, for instance in video-surveillance applications. There is a continuous flux of images to be analyzed. In this case, one can only update the last image  $C_i$  and consider all the others as fixed. If computations have been done for  $t = 1, \dots, T$  and that a new image  $N_{T+1}$  is available, it is enough to minimize:

$$\tilde{E}(B, C_{T+1}) = E(B, \overline{C_1}, \dots, \overline{C_T}, C_{T+1})$$

where  $\overline{C_i}$  are the detections previously computed. We refer to [151] for more details.

#### 5.1.4 Sequence restoration

This section concerns the problem of restoration as applied to sequences. Unlike static images, very little research has been carried out on this subject. The aim of this section is essentially to show the difficulties and

specificities of this problem by focusing on the problem of the restoration of degraded movies. If sound track restoration now is pretty well resolved as a problem, it is not the same case as far as image is concerned.

There are two main reasons at the origin of this lack of research. The first is economic. The decision to restore a movie may either come from political institutions or from the broadcasting market companies. In both cases, budgets for this are limited as these investments do not always pay. The second is simply because it is a very difficult problem...

To begin with there are the technical problems associated to the storage and processing of the data. With 24 or 25 frames per second (US) size  $1920 \times 1080$  (new High Definition Progressive Video Format), we let the reader see for himself the memory size necessary for the storage of a 90 or 120 minutes long movie...

Then, by simply trying to define what movie restoration is, this becomes quite impossible because of the numerous types of degradation. First defects may affect the base of the film or the emulsion side or both. They may be mechanical, due to lack of precaution while handling manually the film or caused by cameras, printers, developing tanks, all the various equipments used. It is the case for scratches (on one frame, caused by manual handling), vertical scratches on many contiguous frames (caused by mechanical devices), dirt or dust spots, hairs, emulsion tearing off, water marks. They may also be chemical, degradation of the nitrate base for old film or of the acetate base for more recent ones, or some kind of mushrooms, or irregular shades. They may also be a combination of mechanical and chemical like the vertical blue scratches caused by obstructed orifices in the development tank. The splices that tie pieces of film together may be deteriorated and the film may be shranked because of too much drying. The original film may be missing, leaving only a copy that may exhibit part or all the preceding defects plus some photographed defects. That is, dust spots and scratches that have been present on the original have been transferred to the copy in a blurred form which render their detection less evident. During the transfer operation a badly adjusted equipment may leave "hot spot" due to improper focusing of the illumination. Some of these degradations are illustrated in Figures 5.15 to 5.19. These images have been provided by the company *Dust Restauration*<sup>6</sup> which is in particular specialized in digital restoration of sequences at film resolution.

- From this quick overview alone, it is clear that movie restoration cannot be defined as being a unique and simple problem.

Faced with this variety of defects, it may be interesting to classify them. As presented in a report by "Commission Supérieure Technique de l'image et

---

<sup>6</sup><http://www.dust.fr/>

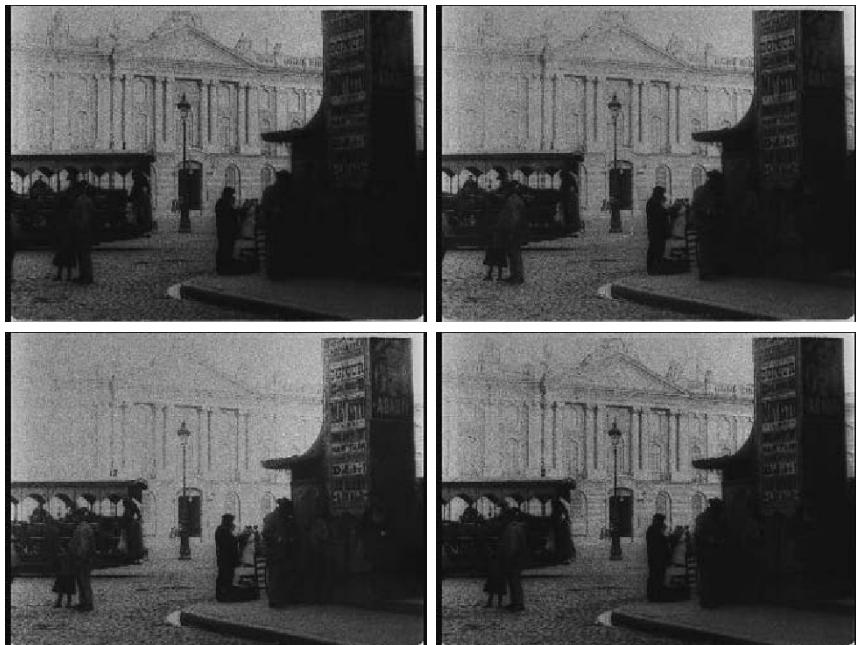


Figure 5.15. By courtesy of “association des frères Lumière”. These images show general intensity variations throughout the sequence (observe the posters at the right-hand side of the images, or the background brightness). These variations are due to the nonuniform aperture time since this was manually controlled....



Figure 5.16. Images of Duke Ellington, “document amateur”, by courtesy of “Cinémathèque de la Danse (Paris)”. In the left-hand side image, the film has burnt. In the right-hand side image, there are many spots and rays.

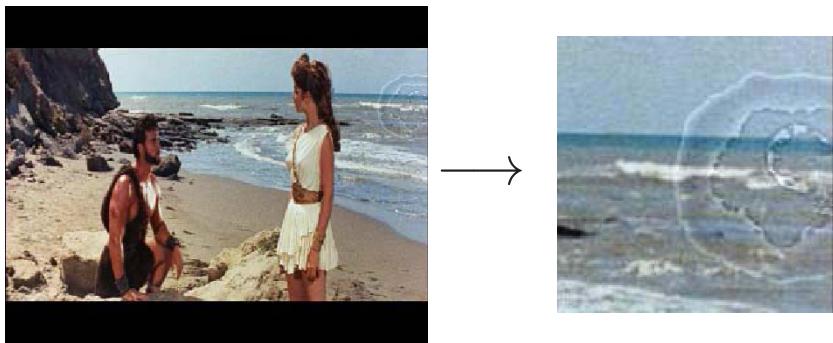


Figure 5.17. By courtesy of DUST Restauration<sup>7</sup>. Example of chemical stain.



Figure 5.18. By courtesy of DUST Restauration<sup>7</sup>. In this image there is an example of a ray and there is also clearly a problem with the grain.

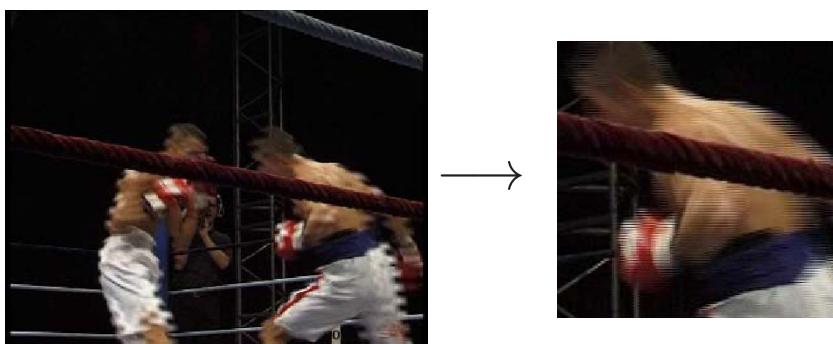


Figure 5.19. By courtesy of DUST Restauration<sup>7</sup>. Example of interlaced video.

---

<sup>7</sup>Original images in color

du son” [93], they can be characterized by the amount of human interaction needed to detect and correct them. This can be summarized in table 5.1. Let us comment:

		Detection	
		Interactive	Automatic
Correction	Interactive	Defects who might be identified as an element of the image and correction of which could not be done without contextual information.	Emulsion tearing off or spots of every kind, non stationary and wide (mushrooms, irregular shades,...)
	Automatic	Static defects of small area (hairs, static spots), stable irregularities, “hot spots”, contrast, color or sharpness differences in the sequence.	Defects whose mathematical model cannot be confused with an image element (dusts, scratches, small instabilities in the position or the exposition)

Table 5.1. Proposal of a classification for defects [93]

- In the left-hand side column, can be found everything that cannot be detected automatically, that is to say elements that could be identified as image elements like shadow effects or reflexions that may be wanted in the scenario. The discrimination of such defects seems to be beyond reach for a long time. Any correction requiring an esthetic choice, like color grading, will also come into this category. Conversely, the right-hand side column presents all the defects that can be detected automatically.
- On the upper line, are all the defects that cannot be corrected automatically. This may be due to:
  - The impossibility to measure a variation precisely, that is to define a norm.
  - The impossibility to define an automatic rule for the correction.
  - The definitive loss of information that cannot be reconstructed without human intervention.

Interestingly, this degree of interaction/automation can also be found in the different systems that are proposed nowadays:

- Paint system is the standard fully manual system, which is also widely used in the special effects industry. Still, it does not allow us to correct all the problems like some colorimetry defects for instance. If subtle color variation occurs in a sequence, it is impossible for the human eye to perceive it and correct it frame by frame, while this is perfectly visible at the projection rate.

- A different strategy is to focus on degradation, to detect and to correct them as automatically as possible, but with a manual preselection. For each kind of degradation, one needs to develop a suitable tool. Some such systems are Limelight or Da Vinci. Limelight is a Eureka project (1993 to 1997) whose purpose was the development of a prototype restoration system. Dedicated software were developed by University of La Rochelle in France and by the Joanneum research in Graz, Austria. A combination of mathematical morphology and temporal comparison with adjacent frames is used to detect dirt and dust. Da Vinci provides digital film, HDTV and SDTV color enhancement technology and is now distributing “Revival” a software developed initially within the Singapore University. This software focus also on degraded area of the picture.
- Some fully automatic systems also exist and even real time. They are based on a temporal diffusion process of the sequence. Such methods have been developed using wavelets or Markov Random Fields (see for instance [150, 149]). Within this framework, we can mention the system Archangel proposed by Snell & Wilcox. This system follow the research effort of BBC Research, Snell & Wilcox, INA, the University of Cambridge and the University of Delft inside the Eureka project AURORA. A hardware embeds all the necessary memory to allow temporal filtering on many adjacent frames. However, even if the temporal filtering is adaptive, since it acts on all the image domain it is unavoidable to have some blur effect, loss of grain and loss of resolution for fine textures.

So we can see that from fully manual to fully automatic, nobody can pretend to have a “perfect” restoration tool. But what would “perfect” restoration mean?...

⌘ *This is one of the main and probably most difficult questions: how can a restoration be judged? Clearly, one cannot be satisfied with the results obtained on one image, even if it looks good. One needs to see the result at a rate of 24 frames per second. A typical example is vertical scratches correction. Imagine that a ray is well corrected for each frame except at some locations which differ from frame to frame. The result is that the ray will still be visible when playing the sequence. This example clearly shows the temporal aspect but also the importance of perception. In other words, the main objective of movie restoration is not to reconstruct but to focus on perception: it will be sufficient for a defect to no longer be perceived even if it has not been completely removed. More than any other domain in computer vision, movie restoration should benefit from advances and studies of human perception (see [141] for instance).*

*Could PDEs help with this problem? In fact, very little research has been carried out up to now. Some approaches have recently been proposed for the problem of disocclusion [172, 171, 32], which consists of reconstructing the image in a specified lost area (a ray, a spot,...). They provide satisfying results when the regions to be reconstructed have smooth isophotes. There is in general a loss of grain and the case of textures has not yet been considered. As far as diffusion is concerned few approaches have been developed (see for instance [128, 181, 162]) and, by observing the success of PDEs in image restoration, it will certainly be very interesting to continue work in this direction, taking perception more into account.*

## 5.2 Image classification

### 5.2.1 Introduction

In this section, we present two supervised classification models for satellite images. We show in Figure 5.20 some typical images to be analyzed. Classification aims at finding in the image some classes that have been previously defined, in terms of intensity. This intensity usually corresponds to different ground natures. This kind of techniques is especially useful to study forests evolutions, ground natures, city developments, etc...



Figure 5.20. Examples of satellite images. By courtesy of the CNES (Centre National d'Etudes Spatiales).

The classification problem is closely related to the segmentation one, in the sense that we want to get a partition composed of homogeneous regions. The main difference is that the number of classes and their characteristics are fixed. Many models can be found in the field of stochastic approaches, with the use of Markov Random Field (MRF) theory [33, 258]. Hereafter, we present two different variational models. The first one is only concerned with classification and is based on a level sets formulation [221]. The second one is coupled with a restoration process [222] and is inspired by work about phase transition in mechanics. The latter is purely variational and relies

on approximation principles via the  $\Gamma$ -convergence theory. For the sake of simplicity, we will make for the two models the following assumptions:

- The discriminating criterion is based on the intensity level of pixels.
- Each class  $C_i$  has a Gaussian distribution of intensity  $N(\mu_i, \sigma_i)$  where  $\mu_i$  and  $\sigma_i$  are respectively the mean and the standard deviation of the class  $C_i$ .
- The number  $K$  of classes and the parameters  $(\mu_i, \sigma_i)$  are known (it is a *supervised* classification).

### 5.2.2 A level sets approach for image classification [221]

The classification procedure consists in two steps:

- Defining the classes according to discriminating features. In our case, according to our hypotheses, we choose the parameters of the Gaussian distribution  $\mu_i$  and  $\sigma_i$ . Of course, other discriminant attributes, as textures parameters for example, could be chosen.
- Defining a partitioning process that:
  - Takes into account the first step.
  - Penalizes overlapping regions (pixels with two labels) and the formation of vacuum.
  - Exhibits regular interfaces between classes, i.e. interfaces with minimal perimeter.

These three properties need to be taken into account in the model. Let us write the precise mathematical formulation. Let  $\Omega$  be an open bounded domain of  $R^2$  and let  $u_0 : \Omega \rightarrow R$  the observed data function (the gray level intensity). Let  $\Omega_i$  be the region defined as:<sup>8</sup>

$$\Omega_i = \{x \in \Omega; x \text{ belongs to the } i^{\text{th}} \text{ class}\}. \quad (5.37)$$

A partitioning of  $\Omega$  consists in finding a family of sets  $\{\Omega_i\}_{i=1,\dots,K}$  so that:

$$\Omega = \bigcup_{i=1}^K (\Omega_i \cup \Gamma_i) \quad \text{and} \quad \Omega_i \cap \Omega_j = \emptyset, \quad i \neq j$$

where  $\Gamma_i = \partial\Omega_i \cap \Omega$  is the intersection of the boundary of  $\Omega_i$  with  $\Omega$  and  $\Gamma_{ij} = \Gamma_{ji} = \Gamma_i \cap \Gamma_j$ ,  $i \neq j$ , the interface between  $\Omega_i$  and  $\Omega_j$ . Of course, we have  $\Gamma_i = \bigcup_{i \neq j} \Gamma_{ij}$  (eventually  $\Gamma_{ij} = \emptyset$ ). We note  $|\Gamma_i|$  the one-dimensional Hausdorff measure of  $\Gamma_i$ . We have:  $|\Gamma_i| = \sum_{i \neq j} |\Gamma_{ij}|$ , ( $|\emptyset| = 0$ ).

The classification model we propose for an observed image  $u_0$  consists in searching for a family of sets  $\{\Omega_i\}_{i=1,\dots,K}$  defined by (5.37) and satisfying:

---

<sup>8</sup> $\Omega_i$  can actually be a set of non connected regions

- (A)  $\{\Omega_i\}_{i=1,\dots,K}$  is a partition of  $\Omega$ , i.e.  $\Omega = \bigcup_{i=1}^K \Omega_i \cup \Gamma_i$  and  $\Omega_i \cap \Omega_j = \emptyset$ ,  $i \neq j$ .
- (B) The partition  $\{\Omega_i\}_{i=1,\dots,K}$  takes into account the Gaussian distribution property of the classes (data term):
- $$\Omega_i = \left\{ x \in \Omega; \text{ the intensity } u_0(x) \text{ has a Gaussian distribution of mean } \mu_i \text{ and of standard deviation } \sigma_i \right\}.$$
- (C) The classification is regular, in the sense that the length of each interface  $\Gamma_{ij}$  is minimal.

Conditions (B) and (C) can be expressed in terms of energy minimization:

- (B) Minimize with respect to  $\Omega_i$ :

$$\sum_i \int_{\Omega_i} \frac{(u_0(x) - \mu_i)^2}{\sigma_i^2} dx. \quad (5.38)$$

In fact, in a probabilistic framework, (5.38) means that we want to maximize the conditional probability  $\Pr(u_0(x)/x \in \Omega_i)$ .

- (C) Minimize with respect to  $\Gamma_{ij}$ :

$$\sum_{i,j} \xi_{ij} |\Gamma_{ij}| \quad (5.39)$$

the parameter  $\xi_{ij} \in R^+$  being fixed and permitting to take into account an eventual information about the length of contours.

The main difficulty in the above formulation comes from the fact that the unknowns are sets and not functions. To overcome this difficulty we propose to use a level sets method inspired by the work of Zhao et al [256] concerning multiphase evolution in fluid dynamics.

Let us suppose that for each  $i = 1, \dots, K$ , there exists a Lipschitz function  $\phi_i$  such that:

$$\begin{cases} \phi_i(x) > 0 & \text{if } x \in \Omega_i \\ \phi_i(x) = 0 & \text{if } x \in \Gamma_i \\ \phi_i(x) < 0 & \text{otherwise} \end{cases} \quad (5.40)$$

i.e. the region  $\Omega_i$  is entirely described by the function  $\phi_i$ . Now, let us look at the writing of conditions (A), (B) and (C) in terms of an energy functional involving  $\{\phi_i\}_{i=1,\dots,K}$ . This functional will have to contain three terms:

- A term related to condition (A) (partition condition):

$$F^A(\phi_1, \dots, \phi_K) = \frac{\lambda}{2} \int_{\Omega} \left( \sum_{i=1}^K H(\phi_i(x)) - 1 \right)^2 dx, \quad \lambda \in R^+$$

where  $H(s)$  is the Heaviside function:  $H(s) = 1$  if  $s > 0$  and  $H(s) = 0$  if  $s < 0$ . The minimization of  $F^A$  with respect to  $\{\phi_i\}_{i=1,\dots,K}$ , leads to a solution where the formation of vacuum (pixels with no labels) and regions overlapping (pixels with more than one label) are penalized.

- A term related to condition (B):

$$F^B(\phi_1, \dots, \phi_K) = \sum_{i=1}^K e_i \int_{\Omega} H(\phi_i(x)) \frac{(u_0(x) - \mu_i)^2}{\sigma_i^2} dx, \text{ with } e_i \in R^+$$

where  $\{e_i\}_{i=1,\dots,K}$  are constants which could be useful to take into account, for instance, a bad estimation of the statistics for one of the  $K$  classes.

- A third term related to condition (C) (length shortening of interface set):

$$F^C(\phi_1, \dots, \phi_K) = \sum_{i=1}^K \gamma_i \int_{\phi_i=0} ds, \quad \text{with } \gamma_i \in R^+.$$

Therefore the complete functional is:

$$F(\phi_1, \dots, \phi_K) = F^A(\phi_1, \dots, \phi_K) + F^B(\phi_1, \dots, \phi_K) + F^C(\phi_1, \dots, \phi_K). \quad (5.41)$$

**Remark** In fact, the functional  $F$  is closely related to the Mumford and Shah functional (see Section 4.2.2) for which solutions are expected to be piecewise constant. ■

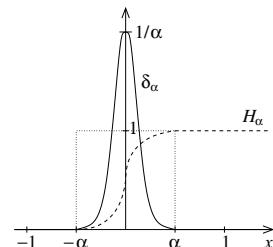
Unfortunately, stated as above the functional  $F$  has still some drawbacks from a practical point of view:  $F$  is not Gâteaux-differentiable and the length term is not easy to handle numerically. So we have to regularize  $F$ . To do this, let  $\delta_\alpha$  and  $H_\alpha$  be respectively the following approximations of the Dirac and Heaviside distributions:

$$\delta_\alpha(s) = \begin{cases} \frac{1}{2\alpha} (1 + \cos(\frac{\pi s}{\alpha})) & \text{if } |s| \leq \alpha \\ 0 & \text{if } |s| \geq \alpha \end{cases}$$

$$H_\alpha(s) = \begin{cases} \frac{1}{2} \left(1 + \frac{s}{\alpha} + \frac{1}{\pi} \sin(\frac{\pi s}{\alpha})\right) & \text{if } |s| \leq \alpha \\ 1 & \text{if } s > \alpha \\ 0 & \text{if } s < -\alpha \end{cases}$$

Then, we approximate  $F$  by:

$$F_\alpha(\phi_1, \dots, \phi_K) = F_\alpha^A(\phi_1, \dots, \phi_K) + F_\alpha^B(\phi_1, \dots, \phi_K) + F_\alpha^C(\phi_1, \dots, \phi_K) \quad (5.42)$$



where:

$$\begin{aligned} F_\alpha^A(\phi_1, \dots, \phi_K) &= \frac{\lambda}{2} \int_{\Omega} \left( \sum_{i=1}^K H_\alpha(\phi_i(x)) - 1 \right)^2 dx \\ F_\alpha^B(\phi_1, \dots, \phi_K) &= \sum_{i=1}^K e_i \int_{\Omega} H_\alpha(\phi_i(x)) \frac{(u_0(x) - \mu_i)^2}{\sigma_i^2} dx \\ F_\alpha^C(\phi_1, \dots, \phi_K) &= \sum_{i=1}^K \gamma_i \int_{\Omega_i} \delta_\alpha(\phi_i(x)) |\nabla \phi_i(x)| dx. \end{aligned}$$

If the definition of the approximated functionals  $F_\alpha^A$  and  $F_\alpha^B$  comes very naturally, the one of  $F_\alpha^C$  is less immediate and relies upon the following lemma.

**Lemma 5.2.1** *Let  $\phi : R^N \rightarrow R$  be Lipschitz continuous then:*

$$\lim_{\alpha \rightarrow 0} \int_{\Omega} \delta_\alpha(\phi(x)) |\nabla \phi(x)| dx = \int_{\phi=0} ds$$

**Proof** Thanks to the coarea formula (Section 2.5.2) we have:

$$L_\alpha(\phi) = \int_{\Omega} \delta_\alpha(\phi(x)) |\nabla \phi(x)| dx = \int_R \left[ \delta_\alpha(t) \int_{\phi=t} ds \right] dt.$$

By setting  $h(t) = \int_{\phi=t} ds$ , we get:

$$L_\alpha(\phi) = \int_R \delta_\alpha(t) h(t) dt = \frac{1}{2\alpha} \int_{-\alpha}^{+\alpha} \left( 1 + \cos\left(\frac{\pi t}{\alpha}\right) \right) h(t) dt.$$

If we take  $\theta = \frac{t}{\alpha}$ , then:

$$L_\alpha(\phi) = \frac{1}{2} \int_{-1}^{+1} (1 + \cos(\pi \theta)) h(\alpha \theta) d\theta$$

and when  $\alpha \rightarrow 0$ , we obtain:

$$\lim_{\alpha \rightarrow 0} L_\alpha(\phi) = \frac{1}{2} h(0) \int_{-1}^{+1} (1 + \cos(\pi \theta)) d\theta = \int_{\phi=0} ds.$$

■

By construction,  $F_\alpha$  is Gâteaux-differentiable with respect to  $\{\phi_i\}_{i=1,\dots,K}$  and all the integrals are defined over  $\Omega$  (which is fixed).

Another improvement to avoid oversmoothing of the interfaces between classes (where the gradient of  $u_0$  is high) is to introduce into  $F_\alpha^C$  the stopping function:

$$g(u_0) = \frac{1}{1 + |\nabla G_\rho * u_0|^2}$$

where  $G_\rho$  is the usual Gaussian kernel. This is particularly useful if the data is very noisy since in this case the parameters  $\gamma_i$  are chosen large enough to ensure a good smoothing inside the classes. So the final minimization problem is:

$$\inf_{\phi_i} F_\alpha(\phi_1, \dots, \phi_K) \quad (5.43)$$

with:

$$\begin{aligned} F_\alpha(\phi_1, \dots, \phi_K) &= \\ &= \frac{\lambda}{2} \int_{\Omega} \left( \sum_{i=1}^K H_\alpha(\phi_i(x)) - 1 \right)^2 dx + \sum_{i=1}^K e_i \int_{\Omega} H_\alpha(\phi_i(x)) \frac{(u_0(x) - \mu_i)^2}{\sigma_i^2} dx + \\ &\quad + \sum_{i=1}^K \gamma_i \int_{\Omega_i} g(u_0(x)) \delta_\alpha(\phi_i(x)) |\nabla \phi_i(x)| dx. \end{aligned}$$

And the classification problem we are interested in, can be formulated as: We do not know whether or not (5.43) has a solution. The question is open and is under investigation. Nevertheless, we may write formally the associated Euler-Lagrange equations. We get a system of K-coupled PDEs:

$$\begin{aligned} \delta_\alpha(\phi_i) \left[ e_i \frac{(u_0 - \mu_i)^2}{\sigma_i^2} - \gamma_i g(u_0) \operatorname{div} \left( \frac{\nabla \phi_i}{|\nabla \phi_i|} \right) - \frac{\nabla g \cdot \nabla \phi_i}{|\nabla \phi_i|} \right. \\ \left. + \lambda \left( \sum_{i=1}^K H_\alpha(\phi_i) - 1 \right) \right] = 0 \end{aligned} \quad (5.44)$$

for  $i = 1, \dots, K$ , with Neumann boundary conditions. Remark that unlike the classical active contour equation (see Chapter 4, equation (4.34)), the divergence term is multiplied by  $\delta_\alpha(\phi_i)$  and not by  $|\nabla \phi_i|$  which clearly give to this equation a different nature than (4.34). If its theoretical study remains an open question, it is an advantage numerically to have the term  $\delta_\alpha(\phi_i)$  which delimits a natural narrow band (where the  $i^{\text{th}}$  is non zero).

To solve numerically (5.44), we embed it into a dynamical process:

$$\frac{\partial \phi_i}{\partial t} = \delta_\alpha(\phi_i) \left[ e_i \frac{(u_0 - \mu_i)^2}{\sigma_i^2} - \gamma_i g(u_0) \operatorname{div} \left( \frac{\nabla \phi_i}{|\nabla \phi_i|} \right) - \frac{\nabla g \cdot \nabla \phi_i}{|\nabla \phi_i|} + \lambda \left( \sum_{i=1}^K H_\alpha(\phi_i) - 1 \right) \right]. \quad (5.45)$$

We discretize (5.45) by using similar finite difference schemes than for the discretization of (4.34) and we refer to Section A.3.4 for more details.

To illustrate this approach, we present in Figures 5.21 and 5.22 two examples on a synthetic image and on the SPOT image<sup>8</sup>.

- In the first example (Figure 5.21) we illustrate how regions interact to avoid overlapping while covering all the space.
- The second example (Figure 5.22) illustrates the algorithm on a real image. Notice in particular the initialization. We use an automatic method for the initialization of the  $\{\phi_i\}_{i=1,\dots,K}$  that we call *seed initialization*. This method consists in cutting the data image  $u_0$  into  $n$  windows  $w_p$ ,  $p = 1, \dots, n$  of predefined size. We compute the mean  $m_p$  and the standard deviation  $s_p$  for the Gaussian distribution of  $u_0$  over each window  $w_p$ . Then we select the index  $k$  such that  $k = \underset{j}{\operatorname{argmin}} d_B(N(m_p, s_p), N(\mu_j, \sigma_j))$  where  $d_B$  is the

Bhattacharyya distance<sup>8</sup> which measures the distance between two Gaussian distributions  $N(m_p, s_p)$  and  $N(\mu_j, \sigma_j)$ . Finally we initialize the corresponding signed distance functions  $\phi_k$  on each  $w_p$ . Windows are not overlapping and each of them is supporting one and only one function  $\phi_k$ . The size of the windows is related to the smallest details we expect to detect. The major advantages of this initialization are: it is automatic (only the size of the windows has to be fixed), it accelerates the speed of convergence, and it is less sensitive to noise.

### 5.2.3 A variational model for image classification and restoration [222]

The objectives are the same as those described in Section 5.2.2, but in addition we want to add a restoration process. In [222], it is proposed to minimize:

$$J_\varepsilon(u) = \int_{\Omega} (u(x) - u_0(x))^2 dx + \lambda^2 \varepsilon \int_{\Omega} \varphi(|\nabla u(x)|) dx + \frac{\eta^2}{\varepsilon} \int_{\Omega} W(u(x), \mu, \sigma) dx$$

---

<sup>8</sup>Results provided by the INRIA project ARIANA: <http://www.inria.fr/ariana/>

<sup>8</sup> $d_B(N(\mu_a, \sigma_a), N(\mu_b, \sigma_b)) = \frac{(\mu_a - \mu_b)^2}{4(\sigma_a^2 + \sigma_b^2)} + \frac{1}{2} \log \frac{\sigma_a^2 + \sigma_b^2}{2\sigma_a \sigma_b}$

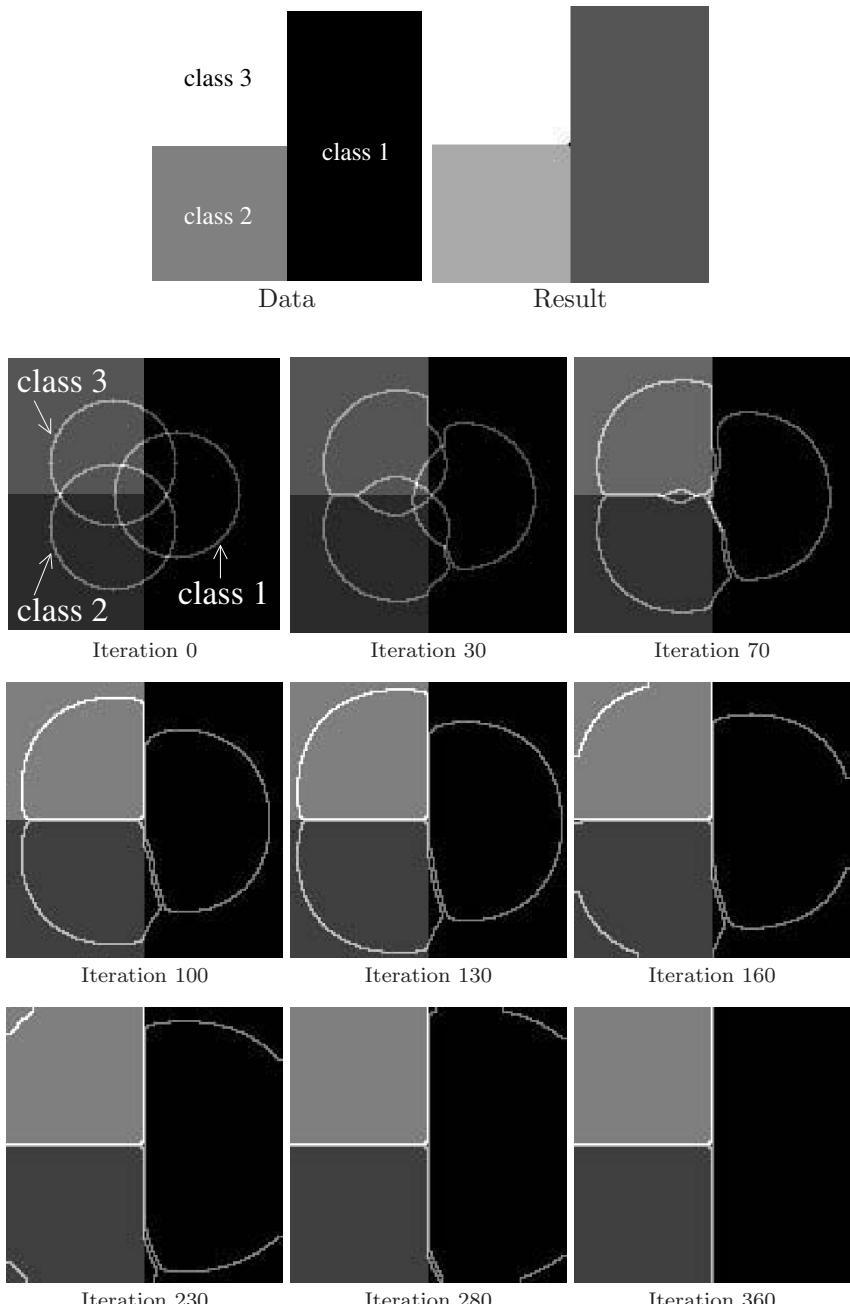


Figure 5.21. Synthetic example (3 classes). Upper row shows the initial image and the result. Iterations of the algorithm are displayed below. This simple example illustrates what is required on the regions: no overlap and full coverage.

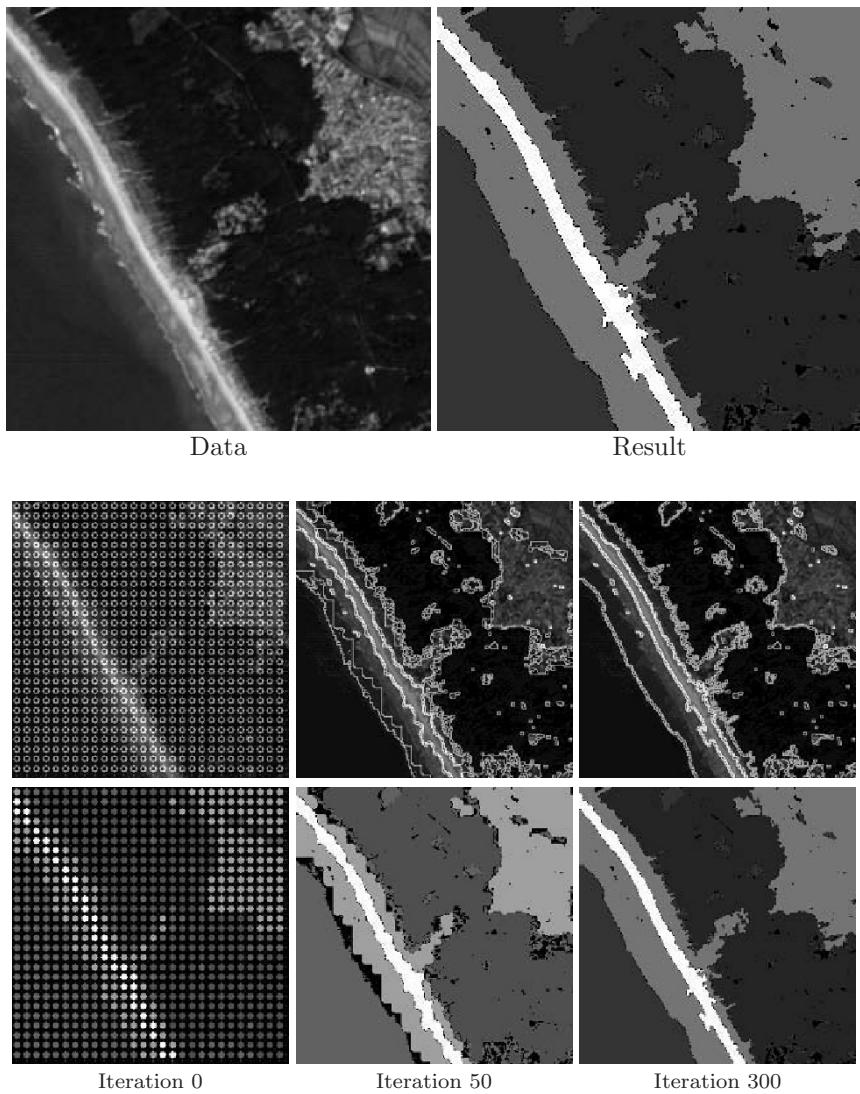


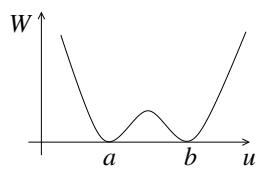
Figure 5.22. The SPOT image (4 classes). Upper row shows the initial image and the result. Iterations of the algorithm are displayed below with two representations: the boundaries of the actual regions and below regions are colored according to their class.

where:

- $u_0$  is the initial image (the data) and  $u(x)$  the image we want to restore and segment into  $K$  homogeneous and disjoint classes.
- The function  $\varphi(\cdot)$  is a regularization function (with the same role and properties than those defined in Section 3.2 and used several times in this book).
- The last term  $W$  is a potential inducing a classification constraint. It takes into account the intensity and the parameters  $(\mu_i, \sigma_i)$  of the classes<sup>8</sup>.  $W$  has as many minima as the number of classes. It has to attract the values of  $u$  toward the label of classes. We will come back further on the precise description of  $W$ .
- The parameters  $\lambda \geq 0$  and  $\eta \geq 0$  permit to adjust the weight of each term.
- $\varepsilon > 0$  is a parameter to be destined to tend to zero. During the first steps of convergence the weight of the third term in  $J_\varepsilon$  is negligible and the restoration process (with the two first terms) is predominant. As  $\varepsilon$  tends to zero, we progressively get a weakened diffusion while raising the classification since the third term becomes preponderant.

The form of the energy  $J_\varepsilon$  is borrowed from the Van der Walls, Cahn-Hilliard theory of phase transitions in mechanics (see for example [3, 53, 111, 179, 235]). To better understand the model, let us recall some results about phase transitions.

Let us consider a mechanical system made of two unstable components (or phases). These components may be liquids having different levels of density distribution. The problem is to describe stable configurations and to characterize the interface between the two phases while the system reaches its stability. For the sake of clarity, let us consider a single fluid whose energy per unit of volume is a function  $W$  of the density distribution  $u$ . The function  $W$  is supposed to be non negative having two minima in  $a$  and  $b$  such that  $W(a) = W(b) = 0$ . Moreover, it is assumed that  $W$  is quadratic around  $a$  and  $b$  and is growing at least linearly at infinity.  $W$  is known as a double-well potential. The stable configurations of the system are obtained by solving the following variational




---

<sup>8</sup>We recall that each class  $C_i$  is characterized by a Gaussian distribution  $N(\mu_i, \sigma_i)$

problem (see for instance [111, 179, 235]):

$$\mathcal{P}_\varepsilon \left\{ \begin{array}{l} \inf_u E_\varepsilon(u) \text{ with} \\ E_\varepsilon(u) = \varepsilon \int_{\Omega} |\nabla u(x)|^2 dx + \frac{1}{\varepsilon} \int_{\Omega} W(u(x)) dx \\ \text{subject to the constraint } \int_{\Omega} u(x) dx = m \end{array} \right.$$

where  $\Omega$  is a bounded open subset of  $R^N$  (the region occupied by the fluid) and  $m$  is the total mass of the fluid.

**Remark** The introduction of the perturbation term permits to solve the uniqueness problem for:

$$\inf_u \left\{ \int_{\Omega} W(u(x)) dx; \int_{\Omega} u(x) dx = m \right\}.$$

■

The asymptotic behaviour as  $\varepsilon$  tends to zero of the model allows to characterize stable configurations. This rely on the  $\Gamma$ -convergence theory.

**Theorem 5.2.1** If  $W$  verifies the previously described conditions, then:

(i)  $E_\varepsilon$   $\Gamma$ -converges to  $E_0$  (for the  $L^1$ -strong topology) with:

$$E_0(u) = \begin{cases} K \operatorname{Per}_{\Omega}(R_1) & \text{if } u(x) \in \{a, b\} \text{ a.e.} \\ +\infty & \text{otherwise} \end{cases}$$

with  $R_1 = \{x \in \Omega; u(x) = a\}$ ,  $K$  defined by:

$$K = 2 \inf_{\gamma} \left\{ \int_{-1}^1 \sqrt{W(\gamma(s))} |\gamma'(s)| ds; \gamma \text{ piecewise } C^1, \gamma(-1) = a, \gamma(1) = b \right\} \quad (5.46)$$

and  $\operatorname{Per}_{\Omega}(R_1)$  stands for the perimeter of  $R_1$ :

$$\operatorname{Per}_{\Omega}(R_1) = \sup \left\{ \int_{\Omega} \chi_{R_1}(x) \operatorname{div}(\varphi) dx; \varphi \in C_0^1(\Omega)^N, |\varphi|_{L^\infty(\Omega)} \leq 1 \right\}$$

which is the total variation of the characteristic function  $\chi_{R_1}(x)$ .

(ii) If  $u_\varepsilon$  is a sequence of minimizers of  $E_\varepsilon$  such that  $u_\varepsilon$  converges to  $\bar{u}$  in  $L^1(\Omega)$ , then  $\bar{u}$  is a solution of the problem:

$$\inf_{u \in BV(\Omega)} \left\{ \operatorname{Per}_{\Omega}(x \in \Omega; u(x) = a); W(u(x)) = 0 \text{ a.e.}; \int_{\Omega} u(x) dx = m \right\}.$$

- (iii) Any sequence  $(v_\varepsilon)$  such that  $E_\varepsilon(v_\varepsilon) \leq cte < \infty$ ,  $\forall \varepsilon$ , admits a subsequence converging in  $L^1(\Omega)$ .

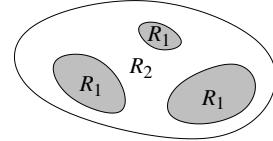
We will give the proof of Theorem 5.2.1 (which is rather technical) at the end of this section.

As a consequence of this result, the action of the term  $W$  is quite clear: it forces the solution to take one of the two values  $a$  and  $b$ . Moreover, since the perimeter of the interface between the two phases  $\{u(x) = a\}$  and  $\{u(x) = b\}$  is minimal, it follows that this interface is not too irregular.

The transposition of the previous ideas in image analysis is straightforward. Let  $u : \Omega \rightarrow R$  be a function that represents the intensity of each pixel and let us consider a feature criterion of classification, only based upon the distribution of intensity. Let us assume that the image is compounded of two regions  $R_1 = \{x \in \Omega; u(x) = a\}$  and  $R_2 = \{x \in \Omega; u(x) = b\}$ . Let  $u_0$  be the observed data corrupted by an additive white Gaussian noise. Following the previous discussion, let us define the energy:

$$E_\varepsilon(u) = \int_{\Omega} (u(x) - u_0(x))^2 dx + \int_{\Omega} \left[ \varepsilon |\nabla u|^2 + \frac{1}{\varepsilon} W(u(x)) \right] dx$$

with  $W$  as before. According to Theorem 5.2.1, as  $\varepsilon \rightarrow 0$ , there exists a subsequence of minimizers  $u_\varepsilon$  of  $E_\varepsilon$  converging to a smooth segmented image<sup>8</sup>, closed to  $u_0$ , and whose pixels belong to  $R_1$  or  $R_2$  a.e. (these two regions being separated by sharp regularized i.e. minimal edges). When  $\varepsilon > 0$  is fixed (not too small), the second term in  $E_\varepsilon$  induces an isotropic smoothing. As  $\varepsilon$  decreases, the third term in  $E_\varepsilon$  forces the solution to choose the two characterizing regions  $R_1$  and  $R_2$ . The two values  $a$  and  $b$  are the labels of  $R_1$  and  $R_2$ .



Until now, we have presented the model with two phases (or labels). The extension to multiple-wells (and even vectorial wells) exists but it is not quite obvious from a mathematical point of view. The most complete result is due to Baldo:

**Theorem 5.2.2 [21]** *Let  $\Omega \subset R^N$ ,  $N \geq 2$  be open and bounded with Lipschitz boundary. Let  $W : R_+^n = \{u = (u_1, \dots, u_n), u_i > 0\} \rightarrow [0, +\infty[$  be a continuous function such that  $W(u) = 0 \iff u = \mu_1, \mu_2, \dots, \mu_k$ ,  $\mu_i \in R_+^n$ . We also assume the technical hypothesis:  $\exists \alpha_1, \alpha_2$ ,  $0 < \alpha_1 < \alpha_2$ , such that  $W(u) \geq \sup \{W(v); v \in [\alpha_1, \alpha_2]^n\}$  for every  $u \notin [\alpha_1, \alpha_2]^n$ . For*

---

<sup>8</sup>The introduction of the additional term  $\int_{\Omega} (u(x) - u_0(x))^2 dx$  has no consequence on the conclusions of the Theorem 5.2.1.

$u_0 \in L^2(\Omega)^n$ , let us define the functional:

$$E_\varepsilon(u) = \int_{\Omega} |u(x) - u_0(x)|_{R^n}^2 dx + \int_{\Omega} \left[ \varepsilon |\nabla u|^2 + \frac{1}{\varepsilon} W(u(x)) \right] dx.$$

For each  $\varepsilon$ , let  $u_\varepsilon$  be a solution of  $\inf_{u \in W^{1,2}(\Omega)^n} E_\varepsilon(u)$ . Let us suppose that  $u_\varepsilon \rightarrow u \in L^2(\Omega)$ . Then:

$$u(x) = \sum_{i=1}^k \mu_i \chi_{\Omega_i}(x)$$

where  $\Omega_1, \Omega_2, \dots, \Omega_k$  is a partition of  $\Omega$  which minimizes the energy:

$$\sum_{i=1}^k \int_{\Omega_i} |u(x) - u_0(x)|_{R^n}^2 dx + \sum_{i=1}^k d(\mu_i, \mu_j) \mathcal{H}^{N-1}(\partial\Omega_i \cap \partial\Omega_j)$$

with:

$$d(\mu_i, \mu_j) = \inf_g \left\{ \int_0^1 \sqrt{W(g(s))} |g'(s)|; g \in C^1(0, 1)^n, g(0) = \mu_i, g(1) = \mu_j \right\}.$$

Baldo's result allows to solve classification problems when the number of labels is greater than two. The discriminating function  $W(u)$  is such that the label of a class  $\Omega_i$  is the corresponding mean  $\mu_i$ , i.e.  $\Omega_i = \{x \in \Omega; u(x) = \mu_i\}$ ,  $i = 1..k$ . So,  $W$  necessarily verifies  $W(\mu_i) = 0$ ,  $i = 1..k$ , and  $W(v) > 0$  for  $v \neq \mu_i$ ,  $i = 1..k$ . There exists several ways of constructing such a potential  $W$ . A piecewise quadratic potential can be used (see [222] for more details).

As far as the regularization is concerned, it has been shown previously that the quadratic smoothing term was too strong (see Section 3.2.2). We can then propose a modified version of the Baldo functional:

$$E_\varepsilon(u) = \int_{\Omega} (u(x) - u_0(x))^2 dx + \int_{\Omega} \left[ \lambda \varepsilon \varphi(|\nabla u|) + \frac{\eta}{\varepsilon} W(u(x)) \right] dx$$

with  $\varphi(\cdot)$  to be chosen. Let us comment some experiments<sup>8</sup>:

- To show the interest of changing the smoothing term, the original Baldo functional (i.e.  $\varphi(s) = s^2$ ) has been compared with the following regularization functions:  $\varphi(s) = \log \cosh(s)$  and  $\varphi(s) = \frac{s^2}{1+s^2}$  (see Figure 5.23). Of course, for the two last functions  $\varphi(\cdot)$ , no mathematical result of convergence does exist. Results are purely

---

<sup>8</sup>Results provided by the INRIA project ARIANA: <http://www.inria.fr/ariana/>

experimental and numerical algorithms are based on half-quadratic minimization, as described in Section 3.2.4. It can be observed that more points are misclassified using  $\varphi(s) = s^2$  which is due to the oversmoothing of the image.

- The convergence as  $\varepsilon \rightarrow 0$  is illustrated in the synthetic example from Figure 5.24. For various  $\varepsilon$ , the minimizer and its associated dual variable are displayed. This example permits to observe the transition between the restoration process and the classification. As for the decay of  $\varepsilon$ , it is usually chosen  $\varepsilon^n = \varepsilon_0^n$ , where  $n$  is the iteration number in  $\varepsilon$ , and  $\varepsilon_0$  may depend on the amount of noise in the sequence. Typical values lie in the interval 0.9 to 0.98 (strong noise). We refer to [220] for more details.
- Finally, we show in Figure 5.25 the results obtained on the SPOT image, already processed using the previous approach (see Figure 5.22). It can be noticed that this approach permits to keep smaller details.

We close this section by showing how asymptotical results (as  $\varepsilon$  tend to zero) can be rigorously established via the  $\Gamma$ -convergence theory. We do not prove the general case given in Theorem 5.2.2 but we only examine the two phases case and the  $\Gamma$ -convergence part (i) of Theorem 5.2.1.

Let  $W : R \rightarrow R$  satisfying the following properties:

$$W \in C^2(R), W \geq 0 \quad (5.47)$$

$$\begin{aligned} W &\text{ has exactly two roots, that we label } a \text{ and } b (a < b). \\ &\text{We suppose } W'(a) = W'(b) = 0, W''(a) > 0, W''(b) > 0 \end{aligned} \quad (5.48)$$

$$\begin{aligned} &\text{There exist positive constants } c_1, c_2 \text{ and } m, \text{ and an integer} \\ &p \geq 2 \text{ such that } c_1|u|^p \leq W(u) \leq c_2|u|^p \text{ for } |u| \geq m. \end{aligned} \quad (5.49)$$

Now, let us define the functionals:

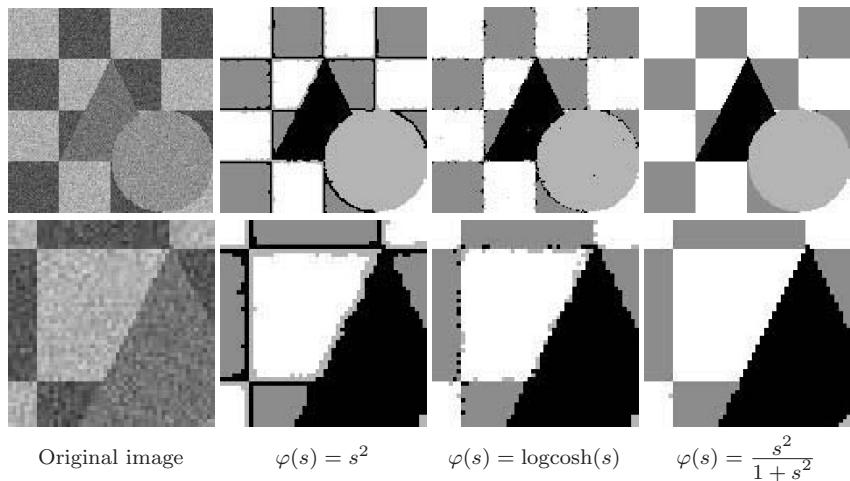
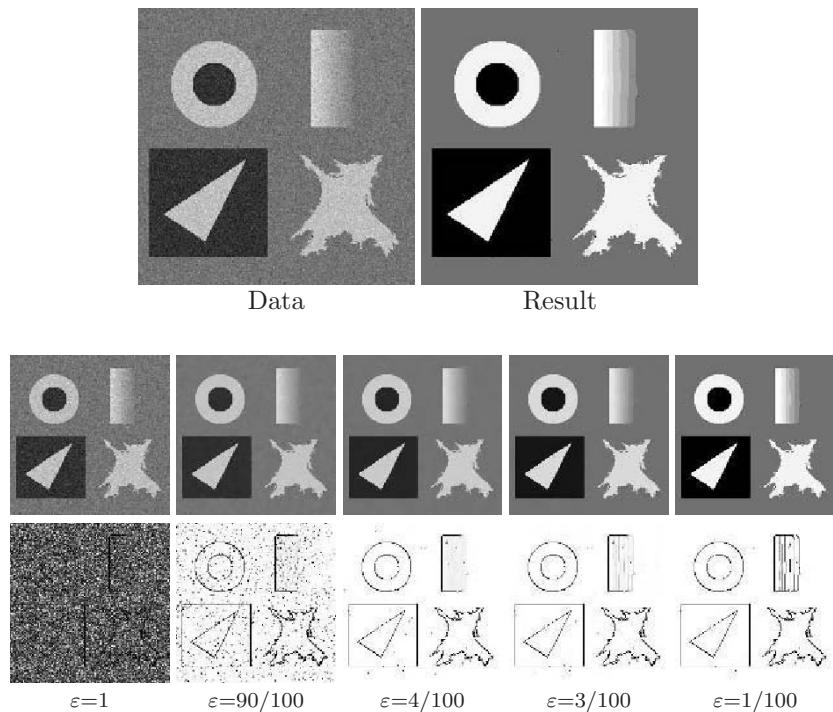
$$E_\varepsilon(u) = \begin{cases} \frac{1}{\varepsilon} \int_{\Omega} W(u) dx + \varepsilon \int_{\Omega} |\nabla u|^2 dx & \text{if } u \in H^1(\Omega) \\ +\infty & \text{otherwise} \end{cases}$$

and:

$$E_0(u) = \begin{cases} K \operatorname{Per}_{\Omega}(\{u = a\}), & u \in BV(\Omega), W(u(x)) = 0 \text{ a.e.} \\ +\infty & \text{otherwise} \end{cases}$$

where  $K$  is defined by (5.46). To prove that  $E_\varepsilon$   $\Gamma$ -converges to  $E_0$ , we need to define an auxiliary function  $g(\cdot)$  which will play a crucial role afterwards:

$$g(u) = \inf_{\substack{\gamma(-1) = a \\ \gamma(1) = u}} \int_{-1}^1 \sqrt{W(\gamma(s))} |\gamma'(s)| ds$$

Figure 5.23. Influence of the function  $\varphi(\cdot)$ . The lower line shows a close-up.Figure 5.24. Illustration of the convergence as  $\varepsilon \rightarrow 0$ . In this example 6 classes are searched even if a continuous regions is present in the image. From  $\varepsilon = 1$  to  $\varepsilon = 4/100$  the restoration acts whereas from  $\varepsilon = 3/100$  to the end the classification seems predominant (regions appear). This is especially visible on the edge indicator functions (the dual variable) in the lower row.

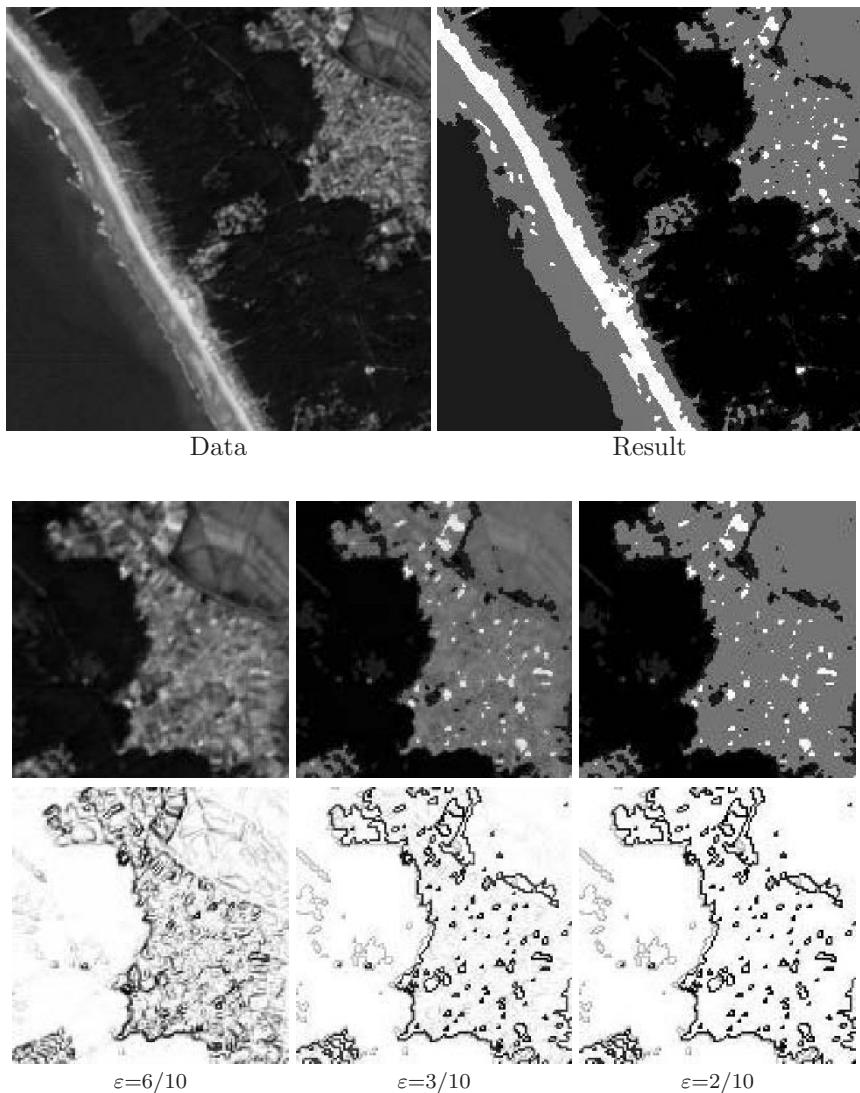


Figure 5.25. The SPOT image (4 classes). Upper row shows the initial image and the result. Iterations of the algorithm as  $\varepsilon$  tend to zero are displayed below with the associated dual variables. A close-up of the upper left corner is shown. Similar behaviour as in Figure 5.24 may be observed.

where the infimum is taken over functions  $\gamma(\cdot)$  which are Lipschitz-continuous. We summarize the properties of  $g(\cdot)$  in the following lemma:

**Lemma 5.2.2** *For every  $u \in R$ , there exists a function  $\gamma_u : [-1, 1] \rightarrow R$  such that  $\gamma(-1) = a$ ,  $\gamma(1) = u$  and:*

$$g(u) = \int_{-1}^1 \sqrt{W(\gamma_u(s))} |\gamma'_u(s)| \, ds. \quad (5.50)$$

The function  $g$  is Lipschitz continuous and satisfies:

$$|g'(u)| = \sqrt{W(u)} \text{ for a.e. } u. \quad (5.51)$$

There exists a smooth increasing function  $\beta : ]-\infty, +\infty[ \rightarrow [-1, 1]$  such that the function  $\xi(\tau) = \gamma_b(\beta(\tau))$  satisfies:

$$2g(b) = \int_{-\infty}^{+\infty} [W(\xi(\tau)) + |\xi'(\tau)|^2] \, d\tau \quad (5.52)$$

$$\lim_{\tau \rightarrow -\infty} \xi(\tau) = a, \lim_{\tau \rightarrow +\infty} \xi(\tau) = b \quad (5.53)$$

with these limits being attained at an exponential rate.

We refer to Sternberg [234, 235] for the proof.

**Proof of (i) of Theorem 5.2.1.**  $E_\varepsilon$   $\Gamma$ -converges to  $E_0$  for the  $L^1$ -strong topology. We follow the proof of Sternberg [234]. According to the definition of the  $\Gamma$ -convergence, we have to show:

- If  $v_\varepsilon \rightarrow v_0$  as  $\varepsilon \rightarrow 0$ , then:

$$\varliminf_{\varepsilon \rightarrow 0} E_\varepsilon(v_\varepsilon) \geq E_0(v_0). \quad (5.54)$$

- For any  $v_0 \in L^1(\Omega)$ , there exists a sequence  $\rho_\varepsilon$  with  $\rho_\varepsilon \rightarrow v_0$  in  $L^1(\Omega)$  and:

$$\varlimsup_{\varepsilon \rightarrow 0} E_\varepsilon(v_\varepsilon) \leq E_0(v_0). \quad (5.55)$$

**Step 1:** Proof of (5.54). One need only to consider  $v_0$  of the form:

$$v_0(x) = \begin{cases} a & \text{if } x \in A \\ b & \text{if } x \in B \end{cases} \quad (5.56)$$

for two disjoint sets  $A$  and  $B$  with  $A \cup B = \Omega$ . Otherwise  $v_\varepsilon \rightarrow v_0$  in  $L^1(\Omega)$  implies  $F_\varepsilon(v_\varepsilon) \rightarrow +\infty$  (and (5.54) is obvious). So, let us suppose that  $v_\varepsilon \rightarrow v_0$  in  $L^1(\Omega)$  and define  $h_\varepsilon(x) = g(v_\varepsilon(x))$ . It follows from Lemma 5.2.2:

$$|\nabla h_\varepsilon(x)| = |\nabla v_\varepsilon(x)| \sqrt{W(v_\varepsilon(x))}. \quad (5.57)$$

But it is clear that:

$$h_\varepsilon \xrightarrow[L^1(\Omega)]{} g(v_0) = \begin{cases} 0 & \text{if } x \in A \\ g(b) & \text{if } x \in B. \end{cases}$$

Then, from the inequality  $a^2 + b^2 \geq 2ab$ , (5.57) and the lower semi-continuity of the  $BV(\Omega)$ -norm for the  $L^1$ -strong convergence, we obtain:

$$\underline{\lim}_{\varepsilon \rightarrow 0} E_\varepsilon(v_\varepsilon) \geq \underline{\lim}_{\varepsilon \rightarrow 0} 2 \int_{\Omega} \sqrt{W(v_\varepsilon)} |\nabla v_\varepsilon| dx = \underline{\lim}_{\varepsilon \rightarrow 0} 2 \int_{\Omega} |\nabla h_\varepsilon(x)| dx \geq 2 \int_{\Omega} |\nabla g(v_0)|.$$

But an easy computation gives:

$$\int_{\Omega} |\nabla g(v_0)| = g(b) \operatorname{Per}_{\Omega}(\{x; v_0(x) = a\}).$$

Thus, we get the desired inequality (5.54):

$$\underline{\lim}_{\varepsilon \rightarrow 0} E_\varepsilon(v_\varepsilon) \geq K \operatorname{Per}_{\Omega}(\{x; v_0(x) = a\})$$

with  $K = 2g(b)$ .

**Step 2:** Proof of (5.55). One may assume  $v_0 \in BV(\Omega)$  and again take  $v_0$  of the form (5.56), or otherwise the trivial choice  $\rho_\varepsilon \equiv v_0$  for each  $\varepsilon$  is suitable. Let  $\Gamma = \partial A \cup \partial B$  and assume  $\Gamma \in C^2$  without a loss of generality (since one can always approximate a set of finite perimeter by a sequence of sets having smooth boundary [120]). Then, let us define the signed distance function  $D : \Omega \rightarrow R$  by:

$$d(x) = \begin{cases} -\operatorname{dist}(x, \Gamma) & \text{if } x \in A \\ +\operatorname{dist}(x, \Gamma) & \text{if } x \in B. \end{cases}$$

Near  $\Gamma$ ,  $d(\cdot)$  is smooth and verifies:

$$|\nabla d| = 1, \quad \lim_{s \rightarrow 0} \mathcal{H}^{N-1}\{x; d(x) = s\} = \mathcal{H}^{N-1}(\Gamma) = \operatorname{Per}_{\Omega} A. \quad (5.58)$$

Finally, let us define the sequence  $\rho_\varepsilon(x)$  by:

$$\rho_\varepsilon(x) = \begin{cases} \xi \left( \frac{-1}{\sqrt{\varepsilon}} \right) & \text{if } d(x) < -\sqrt{\varepsilon} \\ \xi \left( \frac{d(x)}{\varepsilon} \right) & \text{if } |d(x)| \leq \sqrt{\varepsilon} \\ \xi \left( \frac{1}{\sqrt{\varepsilon}} \right) & \text{if } d(x) \geq \sqrt{\varepsilon} \end{cases}$$

where the function  $\xi(\cdot)$  is the one defined in Lemma 5.2.2. The  $L^1$  convergence of  $\rho_\varepsilon$  to  $v_0$  follows directly from (5.52). Then, thanks to (5.52)-(5.53),

(5.58) and the coarea formula, we get:

$$\begin{aligned}
 \overline{\lim}_{\varepsilon \rightarrow 0} E_\varepsilon(\rho_\varepsilon) &= \overline{\lim}_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_{\{|d(x)| \leq \sqrt{\varepsilon}\}} W\left(\xi\left(\frac{d(x)}{\varepsilon}\right)\right) + \left|\xi'\left(\frac{d(x)}{\varepsilon}\right)\right|^2 dx = \\
 &= \overline{\lim}_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_{-\sqrt{\varepsilon}}^{\sqrt{\varepsilon}} W\left(\xi\left(\frac{s}{\varepsilon}\right)\right) + \left|\xi'\left(\frac{s}{\varepsilon}\right)\right|^2 \mathcal{H}^{N-1}\{x; d(x) = s\} ds = \\
 &= \overline{\lim}_{\varepsilon \rightarrow 0} \int_{-1/\sqrt{\varepsilon}}^{1/\sqrt{\varepsilon}} W(\xi(\tau)) + |\xi'(\tau)|^2 \mathcal{H}^{N-1}\{x; d(x) = \varepsilon\tau\} d\tau \leq \\
 &\leq 2g(b) \left( \overline{\lim}_{\varepsilon \rightarrow 0} \max_{|s| \leq \sqrt{\varepsilon}} \mathcal{H}^{N-1}\{x; d(x) = s\} \right) = E_0(v_0)
 \end{aligned}$$

i.e.  $\overline{\lim}_{\varepsilon \rightarrow 0} E_\varepsilon(\rho_\varepsilon) \leq E_0(v_0)$ , and (5.55) is proven. ■

**Remark** The part (ii) and (iii) of Theorem 5.2.1 are not difficult to prove. We refer to Sternberg [234] for the complete proof. ■

# Appendix A

## Introduction to Finite Difference

### How to read this chapter?

This chapter concerns the problem of solving numerically the partial differential equations that we have encountered in this book. Although several kinds of approaches can be considered (like finite elements or spectral methods), the success of finite differences in image analysis is due to the structure of digital images for which we can associate a natural regular grid. This chapter is an introduction to the main notions that are commonly used when one wants to solve a Partial Differential Equation. From Section A.1 to Section A.2, we will only consider the one dimensional case and focus on the main ideas of finite differences. Section A.3 will be more applied: it will concern the discretization of certain approaches detailed in this book. More precisely:

- Section A.1 introduces the main definitions and theoretical considerations about finite difference schemes (convergence, consistency and stability, Lax Theorem). Every notion is illustrated by developing explicit calculus for the case of the one dimensional heat equation. Besides the precise definitions, this will help the reader to understand them in a simple situation.
- Section A.2 concerns hyperbolic equations. We start with the linear case and show that if we do not choose an upwind scheme, then the scheme is always unstable. We then investigate the nonlinear case by focusing on the Burgers equation.

- The purpose of Section A.3 is to show how finite difference schemes can be used in image analysis. We first introduce in Section A.3.1 the main notations and consider the 2-D heat equation. The remainder of Section A.3 is concerned with the discretization of certain PDEs studied in this book:
  - Restoration by energy minimization (Section A.3.2): we detail the discretization of the divergence term which can also be found for the Perona and Malik equation.
  - Enhancement by Osher and Rudin’s shock filters (Section A.3.3): the main interest is to use a flux limiter called minmod.
  - Curves evolution with level sets and especially segmentation with geodesic active contours (Section A.3.4). For the sake of simplicity, we examine separately each term of the model (mean curvature motion, constant speed motion, advection equation). We essentially write their discretization and give some experimental results.

To complete this introduction, we refer the interested reader to [238, 76, 136] for general presentation and to [122, 157] for more details on hyperbolic equations.

## A.1 Definitions and theoretical considerations illustrated by the 1-D parabolic heat equation

### A.1.1 Getting started

There are many approaches that are used for discretizing a partial differential equation. Amongst the most important ones, we can mention finite differences, finite elements and spectral methods.

We focus here on finite differences which are widely used in image processing. This is due to the structure of a digital image as a set of uniformly distributed pixels (see Section A.3).

To present the main ideas we will consider the following well-posed initial-value problem, written in the one dimensional case:<sup>8</sup>

$$\begin{cases} \mathcal{L}v = F & x \in R, t > 0 \\ v(0, x) = f(x) & x \in R \end{cases} \quad (\text{A.1})$$

where  $v$  and  $F$  are defined on  $R$  and  $\mathcal{L}$  is a differential *linear* operator. The function  $v$  denotes the exact solution of (A.1).

---

<sup>8</sup>This is a initial-value problem, which means that there is no boundary condition. For initial-boundary-value problems, the discussion that follows needs to be slightly adapted and we refer to [238] for more details.

**Example:** One of the easiest equation that we may consider is the 1-D heat equation:

$$\frac{\partial v}{\partial t} = \nu \frac{\partial^2 v}{\partial x^2} \quad x \in R, t > 0 \quad (\text{A.2})$$

where  $\nu > 0$  is a constant, which is equivalent to:

$$\mathcal{L}v = 0 \quad \text{with} \quad \mathcal{L}v = \frac{\partial v}{\partial t} - \nu \frac{\partial^2 v}{\partial x^2}.$$

The initial condition is  $v(0, x) = f(x)$ . From now on, we shall use this equation to illustrate the different notions to be defined. ■

Our aim is to solve the PDE (A.1) numerically. We begin by discretizing the spatial domain by placing a grid over the domain. For convenience, we will use a uniform grid, with grid spacing  $\Delta x$ . Likewise the temporal domain can be discretized and we denote by  $\Delta t$  the temporal grid spacing. The resulting grid in the time space domain is illustrated in Figure A.1. Solving the problem numerically means finding a function  $u$  defined at the

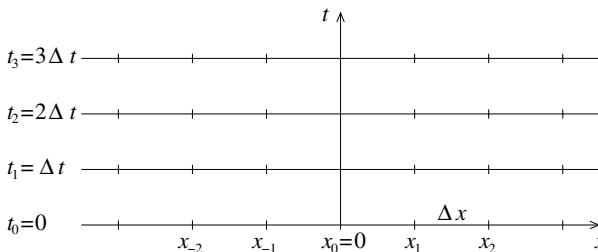


Figure A.1. Grid on time-space domain

points  $(n\Delta t, i\Delta x)$  (we will denote by  $u_i^n$  the value of  $u$  at these points) which is a “good approximation” of  $v$ . The function  $u$  will be obtained as the solution of a discrete equation which will be an approximation of (A.1):

$$\begin{cases} L_i^n u_i^n = G_i^n & i = -\infty, \dots, +\infty \\ u_i^0 = f(i\Delta x) \end{cases} \quad (\text{A.3})$$

where  $L_i^n$  (resp.  $G_i^n$ ) corresponds to the discrete approximation of  $\mathcal{L}$  (resp.  $F^8$ ). Notice that both spatial and temporal derivatives have to be approximated.

**Example:** Let us show on the 1-D heat equation (A.2) how the discrete equation can be obtained. In fact, the starting point for writing any finite difference scheme are Taylor

---

<sup>8</sup>Because of discretization,  $G_i^n$  is *a priori* different from  $F_i^n$  which is simply the value of  $F$  in  $(n\Delta t, i\Delta x)$

expansions. For  $\Delta t$  and  $\Delta x$  small, we have:

$$v((n+1)\Delta t, i\Delta x) = \left( v + \Delta t \frac{\partial v}{\partial t} + \frac{\Delta t^2}{2} \frac{\partial^2 v}{\partial t^2} \right) (n\Delta t, i\Delta x) + \mathcal{O}(\Delta t^3) \quad (\text{A.4})$$

$$v(n\Delta t, (i+1)\Delta x) = \left( v + \Delta x \frac{\partial v}{\partial x} + \frac{\Delta x^2}{2} \frac{\partial^2 v}{\partial x^2} \right) (n\Delta t, i\Delta x) + \mathcal{O}(\Delta x^3) \quad (\text{A.5})$$

$$v(n\Delta t, (i-1)\Delta x) = \left( v - \Delta x \frac{\partial v}{\partial x} + \frac{\Delta x^2}{2} \frac{\partial^2 v}{\partial x^2} \right) (n\Delta t, i\Delta x) + \mathcal{O}(\Delta x^3). \quad (\text{A.6})$$

We recall that  $g = \mathcal{O}(\phi(s))$  for  $s \in S$  if there exists a constant  $C$  such that  $|f(s)| \leq C|\phi(s)|$  for all  $s \in S$ . We say that  $g(s)$  is a “big  $\mathcal{O}$ ” of  $\phi(s)$  or that  $g(s)$  is of order  $\phi(s)$ . In the previous Taylor expansions, notice that the constant  $C$  naturally depends on the high-order derivatives of  $v$ .

By Equation (A.4), we have

$$\frac{\partial v}{\partial t}(n\Delta t, i\Delta x) = \frac{v_i^{n+1} - v_i^n}{\Delta t} + \mathcal{O}(\Delta t)$$

where we noted  $v_i^n = v(n\Delta t, i\Delta x)$ . Similarly, by using previous Taylor expansions, we may propose an approximation of the second spatial derivative. By adding (A.5) and (A.6), we have:

$$\frac{\partial^2 v}{\partial x^2}(n\Delta t, i\Delta x) = \frac{v_{i+1}^n - 2v_i^n + v_{i-1}^n}{\Delta x^2} + \mathcal{O}(\Delta x^2). \quad (\text{A.7})$$

Consequently, if we consider the differential operator  $\mathcal{L}$  from (A.2), we have:

$$\frac{\partial v}{\partial t}(n\Delta t, i\Delta x) - \nu \frac{\partial^2 v}{\partial x^2}(n\Delta t, i\Delta x) = \frac{v_i^{n+1} - v_i^n}{\Delta t} - \nu \frac{v_{i+1}^n - 2v_i^n + v_{i-1}^n}{\Delta x^2} + \mathcal{O}(\Delta t) + \mathcal{O}(\Delta x^2). \quad (\text{A.8})$$

So a reasonable approximation of the equation (A.2) is

$$L_i^n u = 0 \quad \text{with} \quad L_i^n = \frac{u_i^{n+1} - u_i^n}{\Delta t} - \nu \frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{\Delta x^2}. \quad (\text{A.9})$$

This difference Equation (A.9) can also be rewritten in the following form:

$$u_i^{n+1} = (1 - 2r)u_i^n + r(u_{i+1}^n + u_{i-1}^n) \quad \text{where} \quad r = \nu \Delta t / \Delta x^2. \quad (\text{A.10})$$

This shows clearly that this scheme is explicit which means that the values at time  $(n+1)\Delta t$  are obtained only from the values at time  $n\Delta t$ . We will mention how to write implicit schemes in the end of this section. ■

☞ It is important to realize that the discretized equation replaces the original equation with a new one, and that even an exact solution of the discretized problem will lead to an approximate solution of the original PDE, since we introduce a discretization error (the error of replacing a continuous equation by a discrete one).

For a given approximation (A.3), we would like to know precisely the relations between the discrete equation with the PDE and their respective solutions. In other words, what does it mean that  $u$  is an approximation of  $v$  and can we quantify it? Are there any conditions requested on the grid size ( $\Delta t, \Delta x$ ) to have a “good” approximation? To answer these questions

we define precisely the notions of convergence, consistence and stability in the next sections.

### A.1.2 Convergence

The first notion which is essential is to understand what it means that the discrete solution  $u$  of (A.3) is an approximation of, or converges to the solution  $v$  of (A.1). To be more precise, we can define the pointwise convergence as follows:

**Definition A.1.1 (pointwise convergent scheme)** *The scheme (A.3) approximating the partial differential equation (A.1) is pointwise convergent if for any  $x$  and  $t$ , as  $((n+1)\Delta t, i\Delta x)$  converges to  $(t, x)$ , then  $u_i^n$  converges to  $v(t, x)$  as  $\Delta x$  and  $\Delta t$  converge to 0.*

**Example:** Let us show that the solution of the difference scheme (A.10):

$$\begin{cases} u_i^{n+1} = (1 - 2r)u_i^n + r(u_{i+1}^n + u_{i-1}^n) & x \in R \\ u_i^0 = f(i\Delta x) \end{cases} \quad (\text{A.11})$$

where  $r = \nu\Delta t/\Delta x^2$ , converges pointwise to the solution of the initial-value problem (A.2):

$$\begin{cases} \frac{\partial v}{\partial t} = \nu \frac{\partial^2 v}{\partial x^2} & x \in R \\ v(0, x) = f(x) \end{cases} \quad (\text{A.12})$$

We will assume that  $0 \leq r \leq 1/2$  in order to have all the coefficients positive in the difference equation. We need to estimate:

$$z_i^n = u_i^n - v(n\Delta t, i\Delta x).$$

where  $v$  is the exact solution of the initial-value problem (A.12). Equation (A.8) becomes

$$v_i^{n+1} = (1 - 2r)v_i^n + r(v_{i+1}^n + v_{i-1}^n) + \mathcal{O}(\Delta t^2) + \mathcal{O}(\Delta t\Delta x^2). \quad (\text{A.13})$$

Then by subtracting equation (A.13) from (A.11), we have:

$$z_i^{n+1} = (1 - 2r)z_i^n + r(z_{i+1}^n + z_{i-1}^n) + \mathcal{O}(\Delta t^2) + \mathcal{O}(\Delta t\Delta x^2), \quad (\text{A.14})$$

and then (since we assumed  $0 \leq r \leq 1/2$ ):

$$|z_i^{n+1}| \leq (1 - 2r)|z_i^n| + r|z_{i+1}^n| + r|z_{i-1}^n| + C(\Delta t^2 + \Delta t\Delta x^2), \quad (\text{A.15})$$

where  $C$  is a constant associated to the “big  $\mathcal{O}$ ” terms and depends on the assumed bounds of the higher order derivatives of  $v$ , in space and time. In fact we will assume that the derivatives  $v_{tt}$  and  $v_{xxxx}$  (which would appear in the subsequent terms of the Taylor expansion of (A.8)) are uniformly bounded on  $[0, t] \times R$ . So, by taking the supremum with respect to  $i$  in (A.15) we obtain:

$$Z^{n+1} \leq Z^n + C(\Delta t^2 + \Delta t\Delta x^2) \quad \text{with} \quad Z^n = |z^n|_{\ell^\infty} \equiv \sup_{i \in Z} \{|z_i^n|\}. \quad (\text{A.16})$$

Applying (A.16) repeatedly yields

$$\begin{aligned} Z^{n+1} &\leq Z^n + C(\Delta t^2 + \Delta t\Delta x^2) \leq Z^{n-1} + 2C(\Delta t^2 + \Delta t\Delta x^2) \leq \\ &\leq \dots \leq Z^0 + (n+1)C(\Delta t^2 + \Delta t\Delta x^2). \end{aligned}$$

Since  $Z^0 = 0$ , the previous inequality implies:

$$|u_i^{n+1} - v((n+1)\Delta t, i\Delta x)| \leq (n+1)\Delta t C(\Delta t + \Delta x^2). \quad (\text{A.17})$$

Thus we see that the right-hand side of (A.17) goes to zero as  $(n+1)\Delta t \rightarrow t$  and  $\Delta t, \Delta x \rightarrow 0$  which means that  $u$  converges to  $v$  pointwise. Notice that in fact we have just proven a stronger result than the pointwise convergence

$$Z^{n+1} = |z^{n+1}|_{\ell^\infty} \rightarrow 0 \quad (\text{A.18})$$

as  $(n+1)\Delta t \rightarrow t$  and  $\Delta t, \Delta x \rightarrow 0$ . ■

The pointwise convergence is in general difficult to prove. So we shall instead use a definition of convergence in terms of a  $l^p$ -norm ( $p < \infty$ ) of a difference between the solution of the PDE and the solution of the difference equation. In the following definition, we will use the notation:

$$\begin{aligned} u^{n+1} &= (\dots, u_{-1}^n, u_0^n, u_1^n, \dots) \\ v^{n+1} &= (\dots, v_{-1}^n, v_0^n, v_1^n, \dots). \end{aligned}$$

**Definition A.1.2 (convergent scheme)** *The scheme (A.3) approximating the partial differential equation (A.1) is a convergent scheme at time  $t$  if, as  $(n+1)\Delta t \rightarrow t$*

$$|u^{n+1} - v^{n+1}|_* \rightarrow 0 \quad (\text{A.19})$$

as  $\Delta x \rightarrow 0$  and  $\Delta t \rightarrow 0$ , and where  $|\cdot|_*$  is a norm to be specified.

This definition shows that whenever the convergence is being discussed, the norm that is used must be specified. Its choice depends on the problem to be solved. For  $z = (\dots, z_{-1}, z_0, z_1, \dots)$ , typical examples include:

$$|z|_{\ell^\infty} = \sup_{i \in Z} \{|z_i|\}, \quad |z|_{\ell^2} = \sqrt{\sum_{i=-\infty}^{i=+\infty} |z_i|^2} \quad \text{or} \quad |z|_{\ell^2, \Delta x} = \sqrt{\sum_{i=-\infty}^{i=+\infty} |z_i|^2 \Delta x}. \quad (\text{A.20})$$

Another important information that we may be interested in is the rate of convergence, i.e. how fast the solution of the difference equation converges to the solution of the PDE. This order of convergence is defined by:

**Definition A.1.3 (order of convergence)** *A difference scheme (A.3) approximating the partial differential equation (A.1) is a convergent scheme of order  $(p, q)$  if for any  $t$ , as  $(n+1)\Delta t$  converges to  $t$ ,*

$$|u^{n+1} - v^{n+1}| = \mathcal{O}(\Delta x^p) + \mathcal{O}(\Delta t^q) \quad (\text{A.21})$$

as  $\Delta x$  and  $\Delta t$  converge to 0.

**Example:** For the approximation (A.11) of the heat equation, we have in fact proven its convergence for the  $\ell^\infty$  norm (A.18). Moreover, we can verify that this scheme is of order (2,1). ■

The convergence is usually something difficult to prove. Most of the time, its proof is based on the Lax Theorem that we present in the next section.

### A.1.3 The Lax Theorem

This theorem gives a sufficient condition for a two-level difference<sup>8</sup> scheme to be convergent:

**Theorem A.1.1 (Lax)** *A consistent, two level difference scheme for a well-posed linear initial value problem is convergent if and only if it is stable.*

In this theorem, we introduced two new notions:

- *Consistence*: it concerns the error introduced by the discretization of the equation. This error should tend to zero as  $\Delta t$  and  $\Delta x$  go to zero.
- *Stability*: the intuitive idea is that small errors in the initial condition should cause small errors in the solution. This is similar to the definition of well-posedness of a PDEs.

Practically, most of the schemes that are used are consistent. The major problem will be to prove their stability.

The two next sections define precisely these two notions and give the main ideas to ensure that they are satisfied.

### A.1.4 Consistency

As in the case of convergence, we can first define the property of a scheme to be *pointwise* consistent with the PDE:

**Definition A.1.4 (pointwise consistent)** *The scheme (A.3) approximating the partial differential equation (A.1) is pointwise consistent at point  $(t, x)$  if for any smooth function  $\phi = \phi(t, x)$ ,*

$$(\mathcal{L}\phi - F)|_i^n - [L_i^n \phi(n\Delta t, i\Delta x) - G_i^n] \rightarrow 0 \quad (\text{A.22})$$

as  $\Delta x, \Delta t \rightarrow 0$  and  $((n+1)\Delta t, i\Delta x) \rightarrow (t, x)$ .

**Example:** Notice that from equality (A.8), we have in fact just proven that the scheme (A.10) is pointwise consistent with the PDE (A.2). ■

As in the case of convergence, it is usually more interesting to have a definition in terms of norms and not only pointwise. If we write the two-level scheme by:

$$u^{n+1} = Qu^n + \Delta t \, G^n \quad (\text{A.23})$$

---

<sup>8</sup> A two-level difference scheme is a scheme where only two different levels of time are present in the difference equation, typically  $H(u^{n+1}, u^n) = 0$ .

where  $u^n = (\dots, u_{-1}^n, u_0^n, u_1^n, \dots)$ ,  $G^{n+1} = (\dots, G_{-1}^n, G_0^n, G_1^n, \dots)$  and  $Q$  is an operator acting on the appropriate space, then a stronger definition of consistency can be given as follows:

**Definition A.1.5 (consistent)** *The scheme (A.3) is consistent with the partial differential equation (A.1) in a norm  $|\cdot|_*$  if the solution of the partial differential equation,  $v$ , satisfies:*

$$v^{n+1} = Qv^n + \Delta t G^n + \Delta t \tau^n$$

where  $\tau^n$  is such that

$$|\tau^n|_* \rightarrow 0$$

as  $\Delta x, \Delta t \rightarrow 0$ .

The term  $\tau^n$  is called the truncature term. We may be more precise and define also the order in which  $\tau^n$  goes to 0.

**Definition A.1.6 (truncature error, order of accuracy)** *The difference scheme (A.3) is said to be accurate of order  $(p, q)$  if*

$$|\tau^n|_* = \mathcal{O}(\Delta x^p) + \mathcal{O}(\Delta t^q).$$

**Remark** It is easy to see that if a scheme is of order  $(p, q)$ ,  $p, q \geq 1$ , then it is a consistent scheme. Also, it can be verified that if a scheme is either consistent or accurate of order  $(p, q)$ , the scheme is pointwise consistent. ■

**Example:** Let us discuss the consistency of the scheme:

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} = \nu \frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{\Delta x}$$

with the PDE  $\frac{\partial v}{\partial t} = \nu \frac{\partial^2 v}{\partial x^2}$ ,  $x \in R$ ,  $t > 0$ . If we denote by  $v$  the solution of the PDE, then the Equation (A.8) becomes:

$$\frac{v_i^{n+1} - v_i^n}{\Delta t} - \nu \frac{v_{i+1}^n - 2v_i^n + v_{i-1}^n}{\Delta x^2} = \mathcal{O}(\Delta t) + \mathcal{O}(\Delta x^2).$$

As we can see, we need to be more precise to apply Definitions A.1.6 and A.1.5. In particular, we need to know what are exactly the terms in  $\mathcal{O}(\Delta t) + \mathcal{O}(\Delta x^2)$ . In fact, similar calculations have to be done but using Taylor expansions with remainder instead of standard Taylor expansions.

After rewriting the difference scheme in the form of (A.23):

$$u_i^{n+1} = (1 - 2r)u_i^n + r(u_{i+1}^n + u_{i-1}^n) \quad \text{where } r = \nu \Delta t / \Delta x^2,$$

we can define the truncature error by:

$$\Delta t \tau_i^n = v_i^{n+1} - \{(1 - 2r)v_i^n + r(v_{i+1}^n + v_{i-1}^n)\}. \quad (\text{A.24})$$

where  $v$  is a solution of the PDE. Then we need to develop the right-hand side term of (A.24) by using Taylor expansions with remainder. After some calculation, there exists

$t_1 \in ]n\Delta t, (n+1)\Delta t[, x_1 \in ](i-1)\Delta x, i\Delta x[$  and  $x_2 \in ]i\Delta x, (i+1)\Delta x[$  such that:

$$\Delta t \tau_i^n = \frac{\partial^2 v}{\partial t^2}(t_1, i\Delta x) \frac{\Delta t}{2} - \nu \left( \frac{\partial^4 v}{\partial x^4}(n\Delta t, x_1) + \frac{\partial^4 v}{\partial x^4}(n\Delta t, x_2) \right) \frac{\Delta x^2}{24} \quad (\text{A.25})$$

Notice that, as we mentioned it, when we write simply  $\mathcal{O}(\Delta t) + \mathcal{O}(\Delta x^2)$ , we have to be aware that the coefficients involved are not constants but depend on certain derivatives of the solution. This also means that as soon as we will talk about consistency, we will need to make some smoothness assumptions.

To apply Definition A.1.5, we need to choose a norm. If we assume that

$$\frac{\partial^2 v}{\partial t^2} \text{ and } \frac{\partial^4 v}{\partial x^4} \text{ are uniformly bounded on } [0, T] \times R \text{ for some } T,$$

then we can then choose the sup-norm to get that this scheme is accurate of order  $(2, 1)$  with respect to this norm. Otherwise, if we assume

$$\sum_{i=-\infty}^{i=+\infty} \left| \left( \frac{\partial^2 v}{\partial t^2} \right)_i^n \right|^2 \Delta x < A < \infty \text{ and } \sum_{i=-\infty}^{i=+\infty} \left| \left( \frac{\partial^4 v}{\partial x^4} \right)_i^n \right|^2 \Delta x < B < \infty,$$

for any  $\Delta x$  and  $\Delta t$ , then the difference scheme is accurate order  $(2, 1)$  with respect to the  $\ell^{2, \Delta x}$  norm. ■

One important remark which comes out from the previous example is that as soon as one considers the problem of consistency, one needs to choose a norm. It is also important to note that this choice is in fact related to some smoothness assumptions on the solution.

Finally, we would like to mention that proving consistency can be very difficult, especially for implicit schemes. We refer the interested reader to [238] for more details.

### A.1.5 Stability

To conclude this section, we need to discuss the problem of stability, which is necessary to apply the Lax Theorem. Though stability is much easier to establish than convergence, it is still often difficult to prove that a given scheme is stable. Many definitions of stability can be found in the literature, and we present below one which is commonly used:

**Definition A.1.7 (stable scheme)** *The two-level difference scheme*

$$\begin{cases} u^{n+1} = Qu^n, n \geq 0 \\ u^0 \text{ given} \end{cases} \quad (\text{A.26})$$

where  $u^n = (\dots, u_{-1}^n, u_0^n, u_1^n, \dots)$  is said to be stable with respect to the norm  $| \cdot |_*$  if there exist positive constants  $\Delta x_0$  and  $\Delta t_0$ , and non-negative constants  $K$  and  $\beta$  such that

$$|u^{n+1}|_* \leq K e^{\beta t} |u^0|_* \quad (\text{A.27})$$

for  $0 \leq t = (n+1)\Delta t$ ,  $0 < \Delta x \leq \Delta x_0$  and  $0 < \Delta t \leq \Delta t_0$ .

**Remarks** From Definition A.1.7, we may remark the following:

- This definition has been established for homogeneous schemes (A.26). If we have a nonhomogeneous scheme, it can be proved that the stability of the associated homogeneous scheme, along with the convergence, is enough to prove its convergence.
- As for convergence and consistency, we will need to define which norm is used.
- This definition of stability does allow the solution to grow with time. ■

As we already mentioned, there are other definitions for stability. In particular, another common definition is one that does not allow for exponential growth. The inequality (A.27) then becomes:

$$|u^{n+1}|_* \leq K |u^0|_* \quad (\text{A.28})$$

which clearly implies (A.27). The interest of (A.27) is that it permits to include more general situations.

**Example:** Let us show that the scheme

$$u_i^{n+1} = (1 - 2r)u_i^n + r(u_{i+1}^n + u_{i-1}^n) \quad (\text{A.29})$$

is stable for the sup-norm. If we assume that  $r \leq 1/2$ , (A.29) yields:

$$|u_i^{n+1}| \leq (1 - 2r)|u_i^n| + r|u_{i+1}^n| + r|u_{i-1}^n| \leq |u^n|_{\ell^\infty}.$$

If we take the supremum over the right-hand side, we get

$$|u^{n+1}|_{\ell^\infty} \leq |u^n|_{\ell^\infty}.$$

Hence inequality (A.27) is satisfied with  $K = 1$  and  $\beta = 0$ . Notice that in order to prove the stability, we have assumed that  $r = \nu \Delta t / \Delta x^2 \leq 1/2$ . In this case we say that the scheme is conditionally stable. In the case where there is no restriction on  $\Delta x$  and  $\Delta t$ , we say that the scheme is unconditionally stable. ■

The previous example was a simple case where we were able to prove directly stability, i.e. inequalities (A.27) or (A.28). In fact, there are several tools that can be used to prove it. The one which is probably the most commonly used is the Fourier analysis which is used for linear difference schemes with constant coefficients. We recall in Table A.1 the definitions of the Fourier transform and the inverse Fourier transform, for the continuous and discrete setting (i.e. for a vector  $u^n = (\dots, u_{-1}^n, u_0^n, u_1^n, \dots) \in \ell^2$ ). We also recall an important property of Fourier transform which is the Parseval's identity (see [110] for more details).

Interestingly, the discrete transform has similar properties than the continuous one. In particular, to prove the stability of a difference scheme, we will use two main ideas:

	Continuous setting	Discrete setting
Fourier transform	$\hat{v}(t, \omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-i\omega x} v(t, x) dx$	$\hat{u}(\xi) = \frac{1}{\sqrt{2\pi}} \sum_{m=-\infty}^{m=+\infty} e^{-im\xi} u_m$
inverse Fourier transform	$v(t, x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{i\omega x} \hat{v}(t, \omega) d\omega$	$u_m = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{+\pi} e^{im\xi} \hat{u}(\xi) d\xi$
Parseval's identity	$ v _{L^2(R)} =  \hat{v} _{L^2(R)}$	$ u _{L^2(R)} =  \hat{u} _{L^2(-\pi, \pi)}$

Table A.1. Some recalls about Fourier transform ( $i^2 = 1$ )

- The first is that taking the Fourier transform of a PDE turns it into an EDO. Spatial derivatives are turned into products. For example, we can easily verify that:

$$\widehat{v_{xx}}(t, \omega) = -\omega^2 \hat{v}(t, \omega).$$

Analogous idea is valid in the discrete case. Let us consider the “standard” approximation of the second order derivative that we have been using until now (with  $\Delta x = 1$  just to simplify notations):

$$u_{xx}|_k = u_{k+1} - 2u_k + u_{k-1}. \quad (\text{A.30})$$

Then the Fourier tranform of  $\{u_{xx}\}$  in the discrete setting is:

$$\begin{aligned} \widehat{(u_{xx})} &= \frac{1}{\sqrt{2\pi}} \sum_{k=-\infty}^{k=+\infty} e^{-ik\xi} u_{xx}|_k \stackrel{(\text{A.30})}{=} \\ &= \frac{1}{\sqrt{2\pi}} \sum_{k=-\infty}^{k=+\infty} e^{-ik\xi} u_{k+1} - 2 \underbrace{\frac{1}{\sqrt{2\pi}} \sum_{k=-\infty}^{k=+\infty} e^{-ik\xi} u_k}_{\hat{u}(\xi)} + \frac{1}{\sqrt{2\pi}} \sum_{k=-\infty}^{k=+\infty} e^{-ik\xi} u_{k-1}. \end{aligned} \quad (\text{A.31})$$

By doing suitable changes of variable in the previous expression for the first sum ( $m = k + 1$ ) and third one ( $m = k - 1$ ), we have:

$$\begin{aligned} \widehat{(u_{xx})} &= \frac{1}{\sqrt{2\pi}} \sum_{m=-\infty}^{m=+\infty} e^{-i(m-1)\xi} u_m - 2\hat{u}(\xi) + \frac{1}{\sqrt{2\pi}} \sum_{m=-\infty}^{m=+\infty} e^{-i(m+1)\xi} u_m \\ &= (e^{-i\xi} - 2 + e^{+i\xi}) \hat{u}(\xi) = -4 \sin^2\left(\frac{\xi}{2}\right) \hat{u}(\xi). \end{aligned} \quad (\text{A.32})$$

- The second concerns the Parseval's identity. The main interest of this identity is that it is equivalent to prove the inequality (A.27) in the transform space or in the solution space. As a matter of fact, in Definition A.1.7 of stability, the inequality that was required in terms

of energy norm was of the form:

$$|u^{n+1}|_{\ell^2} \leq K e^{\beta(n+1)\Delta t} |u^0|_{\ell^2} \quad (\text{A.33})$$

But since  $|u|_{\ell^2, \Delta x} = \sqrt{\Delta x} |u|_{\ell^2} = \sqrt{\Delta x} |\hat{u}|_{\ell^2}$ , if we can find a  $K$  and a  $\beta$  such that:

$$|\hat{u}^{n+1}|_{\ell^2} \leq K e^{\beta(n+1)\Delta t} |\hat{u}^0|_{\ell^2}, \quad (\text{A.34})$$

then the same  $K$  and  $\beta$  will also satisfy (A.33). So the sequence  $\{u^n\}$  will be stable if and only if the sequence  $\{\hat{u}^n\}$  is stable in  $L^2(-\pi, \pi)$ .

These ideas are applied in the following example where we show how to prove the stability of the discrete scheme associated to the 1-D heat equation.

**Example:** Let us prove the stability of the difference scheme

$$u_k^{n+1} = r u_{k+1}^n + (1 - 2r) u_k^n + r u_{k-1}^n \quad (\text{A.35})$$

where  $r = \nu \Delta t / \Delta x^2 \leq 1/2$ . By doing similar computations as in (A.31)-(A.32) taking the Fourier transform of  $u^{n+1}$  with (A.35) leads to:

$$\begin{aligned} \hat{u}^{n+1}(\xi) &= r \frac{1}{\sqrt{2\pi}} \sum_{k=-\infty}^{k=+\infty} e^{-ik\xi} u_{k+1}^{n+1} + (1 - 2r) \hat{u}(\xi) + r \frac{1}{\sqrt{2\pi}} \sum_{k=-\infty}^{k=+\infty} e^{-ik\xi} u_{k-1}^{n+1} = \\ &= \left( 1 - 4r \sin^2 \left( \frac{\xi}{2} \right) \right) \hat{u}^n(\xi). \end{aligned} \quad (\text{A.36})$$

The coefficient of  $\hat{u}^n$  in the right-hand side of (A.36) is called the symbol of difference scheme (A.35). We denote it by  $\rho(\xi)$ . Then if we apply the result of (A.36)  $n+1$  times, we get

$$\hat{u}^{n+1}(\xi) = (\rho(\xi))^{n+1} \hat{u}^0(\xi).$$

So the condition (A.34) will be verified with  $K = 1$  and  $\beta = 0$  as soon as  $|\rho(\xi)| \leq 1$  for all  $\xi \in [-\pi, \pi]$ , i.e.

$$\left| 1 - 4r \sin^2 \left( \frac{\xi}{2} \right) \right| \leq 1. \quad (\text{A.37})$$

This condition implies that  $r \leq 1/2$ . Thus  $r \leq 1/2$  is a sufficient condition for stability (and along with the consistency, for convergence). It is also necessary indeed. If  $r > 1/2$ , then at least for some  $\xi$ ,  $|\rho(\xi)| > 1$  and then  $|\rho(\xi)|^{n+1}$  will be greater than  $K e^{\beta(n+1)\Delta t}$  for any  $K$  and  $\beta$ <sup>8</sup>. ■

Another approach which is often used is to consider a discrete Fourier mode for the problem:

$$u_k^n = \xi^n e^{ijk\pi\Delta x} \quad (\text{A.38})$$

---

<sup>8</sup>This is true since for any sequence of  $\Delta t$  (chosen so that  $(n+1)\Delta t \rightarrow t$ ) and choice of  $\Delta x$  (so that  $r$  remains constant), the expression  $|\rho(\xi)|^{n+1}$  becomes unbounded while for sufficiently large values of  $n$ ,  $K e^{\beta(n+1)\Delta t}$  will be bounded by  $K e^{\beta(t_0+1)}$  for some  $t_0 > t$ ,  $t_0$  near  $t$ .

where  $0 \leq j \leq M$  and the superscript on the  $\xi$  term is a multiplicative exponent. The idea is then to insert this general Fourier mode into the difference scheme and find the expression for  $\xi$ . A necessary condition for stability is obtained by restricting  $\Delta x$  and  $\Delta t$  so that  $|\xi| \leq 1$  (the  $\xi^n$  term will not grow without bound). This method is usually referred to as the discrete von Neumann criterion to stability (see [238] for more details).

**Example:** Let us apply the discrete von Neumann criterion for stability for the difference scheme

$$u_k^{n+1} = ru_{k+1}^n + (1 - 2r)u_k^n + ru_{k-1}^n. \quad (\text{A.39})$$

By inserting the general Fourier mode

$$u_k^n = \xi^n e^{ijk\pi\Delta x}$$

in the difference scheme, we easily obtain:

$$\xi^{n+1} e^{ijk\pi\Delta x} = \xi^n e^{ijk\pi\Delta x} \left( r e^{-ijk\pi\Delta x} + (1 - 2r) + r e^{ijk\pi\Delta x} \right).$$

Thus if we divide both sides of the above equation by  $\xi^n e^{ijk\pi\Delta x}$ , we get

$$\xi = r e^{-ijk\pi\Delta x} + (1 - 2r) + r e^{ijk\pi\Delta x} = 1 - 4r \sin^2 \left( \frac{j\pi\Delta x}{2} \right).$$

By saying that  $|\xi| \leq 1$ , we recover previous result (A.37). ■

**Example:** As we already mentioned, the discretization (A.9) initially proposed for the one dimensional heat equation was explicit. The values of  $u$  at time  $(n+1)\Delta t$  were fully determined by the values of  $u$  at time  $n\Delta t$ . One may investigate more general schemes like:

$$u_i^{n+1} = u_i^n + \frac{\nu\Delta t}{h^2} \left( \lambda(u_{i+1}^{n+1} - 2u_i^{n+1} + u_{i-1}^{n+1}) + (1 - \lambda)(u_{i+1}^n - 2u_i^n + u_{i-1}^n) \right)$$

If  $\lambda \neq 0$  the scheme is now implicit: one needs to solve a linear system in order to know the solution at time  $(n+1)\Delta t$ . For  $\lambda = 1$  the scheme is fully implicit and for  $\lambda = 0.5$  the so-called Crank-Nicholson scheme is obtained. This difference between explicit and implicit schemes can simply be represented as in Figure A.2

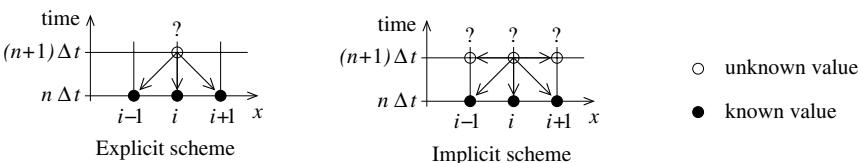


Figure A.2. Relations of dependency: the estimation of  $u_i^{n+1}$  depends on the neighbors indicated by an arrow

Depending on the value of  $\lambda$ , we have:

- For  $\lambda = 0$ , the scheme is explicit, of order  $\mathcal{O}(\Delta t, h^2)$  and stable under the condition  $\Delta t \leq \frac{h^2}{2\nu}$ . This condition implies that the time step has to be chosen small enough which will naturally slow down the resolution of the equation.
- For  $\lambda = 1/2$ , the scheme is implicit, of order  $\mathcal{O}(\Delta t^2, h^2)$  and unconditionally stable.

- For  $\lambda > 1/2$ , the scheme is implicit, of order  $\mathcal{O}(\Delta t, h^2)$  and unconditionally stable.

We leave it as an exercise to the reader to verify these results, that is proving consistence<sup>8</sup> and stability. ■

## A.2 Hyperbolic equations

Let us consider the one-dimensional linear advection equation (also called transport or wave equation):

$$\begin{cases} \frac{\partial v}{\partial t}(t, x) + a \frac{\partial v}{\partial x}(t, x) = 0 & x \in R \\ v(0, x) = v_0(x) \end{cases} \quad (\text{A.40})$$

where  $a$  is a constant. It can be easily verified that the solution is

$$v(t, x) = v_0(x - at) \quad (\text{A.41})$$

Consequently  $v(t, x)$  is constant on lines of slope  $a$  which are called characteristics. It means that the information is propagated in the direction of the sign of  $a$ , for example from the left to the right if  $a$  is positive.

In order to solve (A.40) numerically we have to approximate the temporal and spatial derivatives of  $v$ . As for the case of the heat equation (see Section A.1), the method is based on the Taylor expansions of  $v$  that we recall here:

$$v((n+1)\Delta t, i\Delta x) = \left( v + \Delta t \frac{\partial v}{\partial t} + \frac{\Delta t^2}{2} \frac{\partial^2 v}{\partial t^2} \right) (n\Delta t, i\Delta x) + \mathcal{O}(\Delta t^3) \quad (\text{A.42})$$

$$v(n\Delta t, (i+1)\Delta x) = \left( v + \Delta x \frac{\partial v}{\partial x} + \frac{\Delta x^2}{2} \frac{\partial^2 v}{\partial x^2} \right) (n\Delta t, i\Delta x) + \mathcal{O}(\Delta x^3) \quad (\text{A.43})$$

$$v(n\Delta t, (i-1)\Delta x) = \left( v - \Delta x \frac{\partial v}{\partial x} + \frac{\Delta x^2}{2} \frac{\partial^2 v}{\partial x^2} \right) (n\Delta t, i\Delta x) + \mathcal{O}(\Delta x^3). \quad (\text{A.44})$$

The temporal derivative can be approximated using (A.42) by

$$\frac{\partial v}{\partial t}(n\Delta t, i\Delta x) = \frac{v_i^{n+1} - v_i^n}{\Delta t} + \mathcal{O}(\Delta t).$$

---

<sup>8</sup>In the example considered throughout this section, we have expanded the functions about the index point  $(n\Delta t, i\Delta x)$ , and it was reasonably obvious that this was the correct point about which to expand. However, in some situations, the consistency cannot be proved if the point about which to extend is not adapted. The decision about which point to expand must be made by carefully considering how we expect the difference scheme to approximate the PDE. Typically, to prove the consistency of the Crank-Nicholson scheme, it is logical to consider the consistency of the scheme at the point  $(n+1/2\Delta t, i\Delta x)$ .

As far as the spatial derivative is concerned, there are several possibilities:

- From (A.43) we have  $\frac{\partial v}{\partial x}(n\Delta t, i\Delta x) = \frac{v_{i+1}^n - v_i^n}{\Delta x} + \mathcal{O}(\Delta x)$  (forward difference).
- From (A.44) we have  $\frac{\partial v}{\partial x}(n\Delta t, i\Delta x) = \frac{v_i^n - v_{i-1}^n}{\Delta x} + \mathcal{O}(\Delta x)$  (backward difference).
- By subtracting (A.44) to (A.43) we have  $\frac{\partial v}{\partial x}(n\Delta t, i\Delta x) = \frac{v_{i+1}^n - v_{i-1}^n}{2\Delta x} + \mathcal{O}(\Delta x^2)$  (centered difference).

Consequently, there are three different possibilities for the discrete scheme of (A.40):

$$u_i^{n+1} = u_i^n + a\Delta t \begin{cases} \delta_x^+ u_i^n & \left( \equiv \frac{u_{i+1}^n - u_i^n}{\Delta x} \right) \text{ foreward scheme} \\ \delta_x u_i^n & \left( \equiv \frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x} \right) \text{ centered scheme} \\ \delta_x^- u_i^n & \left( \equiv \frac{u_i^n - u_{i-1}^n}{\Delta x} \right) \text{ backward scheme} \end{cases}$$

Let us first consider the centered approximation which is of order (2, 1):

$$u_n^{n+1} = u_i^n - a\Delta t \delta_x u_i^n = u_i^n - \frac{a\Delta t}{2\Delta x} (u_{i+1}^n - u_{i-1}^n). \quad (\text{A.45})$$

Let us examine its stability. As explained in Section A.1.5, a common approach is to use the discrete von Neumann criterion which consists in inserting the general Fourier mode

$$u_k^n = \xi^n e^{ijk\pi\Delta x} \quad (0 \leq j \leq M) \quad (\text{A.46})$$

into the difference scheme. We recall that the superscript on the  $\xi$  term is a multiplicative exponent. By doing analogous calculations as for the difference scheme (A.39), we find that

$$\xi = 1 - i \frac{a\Delta t}{\Delta x} \sin(j\pi\Delta x) \text{ and then } |\xi|^2 = 1 + \left( \frac{a\Delta t}{\Delta x} \right)^2 \sin^2(j\pi\Delta x) \geq 1 \quad \forall j,$$

which means that this scheme is always unstable.

☞ The reason for this is that we did not take into account the nature of the equation. As there is a propagation, that is the information is propagated in a certain direction, the numerical scheme should take it into account.

To take this observation into account, one may propose the following scheme:

$$u_i^{n+1} = u_i^n - \begin{cases} a \Delta t \delta_x^- u_i^n & \text{if } a > 0 \\ a \Delta t \delta_x^+ u_i^n & \text{if } a < 0 \end{cases}$$

This can be rewritten as

$$u_i^{n+1} = u_i^n - \Delta t [\max(0, a) \delta_x^- u_i^n + \min(0, a) \delta_x^+ u_i^n]. \quad (\text{A.47})$$

We call (A.47) an upwind scheme because it uses values in the direction of information propagation. Let us see again the stability.

- Case  $a > 0$ : by replacing (A.46) in (A.47), we let the reader see for himself that we get

$$|\xi| = 1 - 2C(1 - C)(1 - \cos(j\pi\Delta x)) \quad \text{with} \quad C = \frac{a\Delta t}{\Delta x} > 0.$$

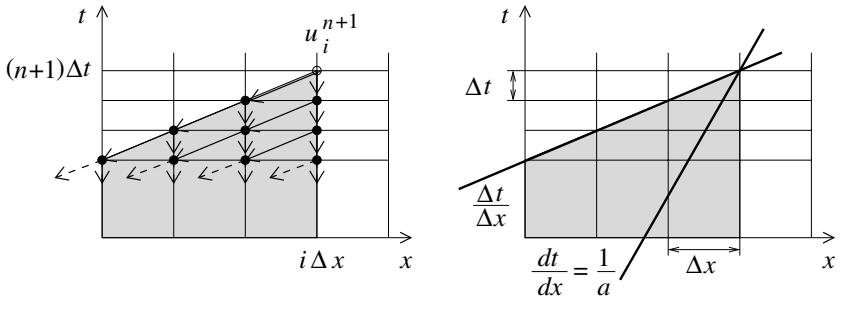
It will be less than or equal to 1 if and only if  $C \leq 1$ .

- Case  $a < 0$ : if we denote  $C = \frac{a\Delta t}{\Delta x} < 0$ , similar calculus yield to  $-C \leq 1$ .

To summarize, the stability condition is

$$|a| \frac{\Delta t}{\Delta x} \leq 1. \quad (\text{A.48})$$

It is usually called CFL, as a reference to their authors Courant-Friedrichs-Lowy (in 1928).



(a) Discrete domain of dependence of  $u_i^{n+1}$  for (A.47) with  $a > 0$

(b) Interpretation of the CFL condition

Figure A.3. Definition and interpretation of the discrete domain of dependence

This condition may be interpreted in terms of domain of dependence. In the continuous case, as it has been mentioned, the information is propagated along the characteristics, and their equation is  $\frac{dt}{dx} = \frac{1}{a}$ . In the discrete case, if  $a > 0$  then (A.47) becomes:

$$u_i^{n+1} = \left(1 - \frac{a\Delta t}{\Delta x}\right) u_i^n + \frac{a\Delta t}{\Delta x} u_{i-1}^n.$$

This allows to define the discrete domain of dependence of  $u_i^{n+1}$ :  $u_i^{n+1}$  depends on  $u_i^n$  and  $u_{i-1}^n$ ,  $u_i^n$  depends on  $u_i^{n-1}$  and  $u_{i-1}^{n-1}$ , etc (see Figure A.3). Moreover, the CFL condition means:

$$\frac{\Delta t}{\Delta x} \leq \frac{1}{a} = \frac{dt}{dx}$$

which signifies that the characteristic line has to be included in the discrete domain of dependence.

 *The discrete domain of dependence must contain the exact continuous domain of dependence*

In the nonlinear case the study is of course more complicated. For example, let us examine the nonlinear Burgers equation:

$$\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} = 0 \quad (\text{A.49})$$

with the initial condition:

$$v(0, x) = v_0(x) = \begin{cases} 1 & \text{if } x \leq 0 \\ 1 - x & \text{if } 0 < x < 1 \\ 0 & \text{if } x \geq 1. \end{cases} \quad (\text{A.50})$$

Here the propagation speed depends on the value of  $u$  itself. We may try as in the linear case to get an explicit solution of (A.49)-(A.50). As it is classical for hyperbolic equations, the method of characteristics can be used. Let us suppose that  $u$  is a smooth solution of (A.49)-(A.50) and let  $x(t)$  be an integral curve of the differential equation:

$$\frac{dx}{dt}(t) = v(t, x(t)), \quad x(0) = x_0.$$

We claim that  $u$  is constant along the characteristic curve  $x(t)$ . Indeed, since  $u$  is a solution of (A.49), we have

$$\begin{aligned} \frac{d}{dt}(v(t, x(t))) &= \frac{\partial v}{\partial t}(t, x(t)) + \frac{dx}{dt}(t) \frac{\partial v}{\partial x}(t, x(t)) \\ &= \frac{\partial v}{\partial t}(t, x(t)) + v(t, x(t)) \frac{\partial v}{\partial x}(t, x(t)) = 0 \end{aligned}$$

Therefore:

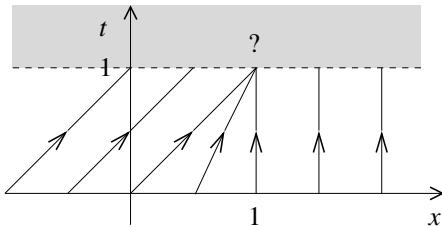
$$\frac{dx}{dt}(t) = v(0, x(0)) = v_0(x_0)$$

and:

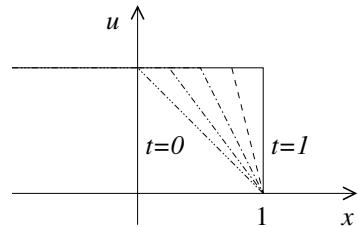
$$x(t) = x_0 + t v_0(x_0).$$

According to the definition of  $u_0$  we deduce that:

$$x(t) = \begin{cases} x_0 + t & \text{if } x_0 \leq 0 \\ x_0 + (1 - x_0)t & \text{if } 0 \leq x_0 \leq 1 \\ x_0 & \text{if } x_0 \geq 1. \end{cases}$$



(a) Characteristics



(b) Solution at different times

Figure A.4. Behavior of the Burgers Equation (A.49). The left-hand side figure represents the characteristics while the right-hand side one shows some solution at different times. As we can observe, characteristics intersect at  $t = 1$  and after this time, it is not clear to define the solution.

For  $t < 1$  the characteristics do not intersect. Hence, given a point  $(t, x)$ ,  $t < 1$ , we draw the characteristic passing through this point and we determine the corresponding point  $x_0$ :

$$x_0 = \begin{cases} x - t & \text{if } x \leq t \\ \frac{x-t}{1-t} & \text{if } t \leq x \leq 1 \\ x & \text{if } x \geq 1 \end{cases}$$

and we get the following continuous solution for  $t < 1$

$$u(t, x) = \begin{cases} 1 & \text{if } x \leq t \\ \frac{1-x}{1-t} & \text{if } t \leq x \leq 1 \\ 0 & \text{if } x \geq 1 \end{cases}$$

It consists of a front moving from the left to the right for  $t < 1$ . At  $t = 1$ , the characteristics collide and beyond this collision time it is not clear how to define the solution uniquely. This discontinuity phenomenon is known as shock. What about for  $t \geq 1$ ? Is it possible to define a unique solution? We must devise some way to interpret a less regular notion of solution. Let  $\varphi : [0, +\infty[ \times \mathbb{R} \rightarrow \mathbb{R}$  be smooth with compact support. We call  $\varphi$  a test function. We first observe that (A.49) can be written in a conservative

form:

$$\frac{\partial v}{\partial t} + \frac{1}{2} \frac{\partial}{\partial x}(v^2) = 0$$

then multiplying the above equality by  $\varphi$ , we deduce:

$$\int_0^\infty \int_{-\infty}^\infty \left( \frac{\partial v}{\partial t} + \frac{1}{2} \frac{\partial}{\partial x}(v^2) \right) \varphi dx dt = 0$$

and by integrating by parts this last equation:

$$\begin{aligned} \int_0^\infty \int_{-\infty}^\infty v(t, x) \frac{\partial \varphi}{\partial t}(t, x) dx dt &+ \int_{-\infty}^\infty v_0(x) \varphi(0, x) dx + \\ &+ \int_0^\infty \int_{-\infty}^\infty \frac{1}{2} v^2(t, x) \frac{\partial \varphi}{\partial x}(t, x) dx dt = 0. \end{aligned} \quad (\text{A.51})$$

We derive (A.51) supposing  $v$  to be smooth but it is still valid if  $v$  is only bounded.

**Definition A.2.1** We say that  $v \in L^\infty((0, \infty) \times R)$  is a weak solution of (A.49) provided equality (A.51) holds for each test function  $\varphi$ .

So according to this definition we are going now to search for discontinuous solutions. But before doing that what can be deduced from (A.51)? Let us suppose in some open domain  $\Omega \subset (0, \infty) \times R$  that  $v$  is smooth on either side of a smooth curve  $x = \xi(t)$ . Let us denote by  $\Omega_L$  (respectively  $\Omega_R$ ) the part of  $\Omega$  on the left (respectively on the right) of  $x = \xi(t)$ . We assume that  $v$  has limits  $v_-$  and  $v_+$  on each side of  $x = \xi(t)$ :  $v_\pm = \lim_{\varepsilon \rightarrow 0} u((t, \xi(t)) \pm \varepsilon N)$

where  $N$  is the normal vector to  $x = \xi(t)$  given by  $N = \begin{pmatrix} -\xi'(t) \\ 1 \end{pmatrix}$ .

Now by choosing a test function  $\varphi$  with compact support in  $\Omega$  but which does not vanish along  $x = \xi(t)$  we get:

$$\begin{aligned} \iint_{\Omega_L} \left( v(t, x) \frac{\partial \varphi}{\partial t}(t, x) + \frac{1}{2} v^2(t, x) \frac{\partial \varphi}{\partial x}(t, x) \right) dx dt + \\ \iint_{\Omega_R} \left( v(t, x) \frac{\partial \varphi}{\partial t}(t, x) + \frac{1}{2} v^2(t, x) \frac{\partial \varphi}{\partial x}(t, x) \right) dx dt = 0. \end{aligned}$$

But since  $\varphi$  has compact support within  $\Omega$  we have:

$$\begin{aligned} & \iint_{\Omega_L} \left( v(t, x) \frac{\partial \varphi}{\partial t}(t, x) + \frac{1}{2} v^2(t, x) \frac{\partial \varphi}{\partial x}(t, x) \right) dx dt = \\ &= - \iint_{\Omega_L} \left( \frac{\partial v}{\partial t} + \frac{1}{2} \frac{\partial}{\partial x} (v^2) \right) \varphi(t, x) dx dt + \int_{x=\xi(t)} \left( -\xi'(t) v_- + \frac{v_-^2}{2} \right) \varphi(t, x) dl \\ &= \int_{x=\xi(t)} \left( -\xi'(t) v_- + \frac{v_-^2}{2} \right) \varphi dl \end{aligned}$$

since  $v$  is a smooth solution satisfying (A.49) in  $\Omega_L$ . Similarly we have:

$$\begin{aligned} & \iint_{\Omega_R} \left( v(t, x) \frac{\partial \varphi}{\partial t}(t, x) + \frac{1}{2} v^2(t, x) \frac{\partial \varphi}{\partial x}(t, x) \right) dx dt = \\ &= - \int_{x=\xi(t)} \left( -\xi'(t) v_+ + \frac{v_+^2}{2} \right) \varphi(t, x) dl. \end{aligned}$$

Adding these two last identities we obtain:

$$\int_{x=\xi(t)} \left( -\xi'(t) v_- + \frac{v_-^2}{2} \right) \varphi(t, x) dl - \int_{x=\xi(t)} \left( -\xi'(t) v_+ + \frac{v_+^2}{2} \right) \varphi(t, x) dl = 0. \quad (\text{A.52})$$

Since  $\varphi$  is arbitrary, we easily deduce from (A.52):

$$\xi'(t) (v_+ - v_-) = \frac{1}{2} (v_+^2 - v_-^2). \quad (\text{A.53})$$

For a general equation of the form  $\frac{\partial v}{\partial t} + \frac{\partial}{\partial x}(f(v)) = 0$ , we should have obtained:

$$\xi'(t) (v_+ - v_-) = (f(v_+) - f(v_-)). \quad (\text{A.54})$$

Identity (A.54) is known as the *Rankine-Hugoniot* condition. It may be read as:

$$\text{Speed of discontinuity} \times \text{jump of } v = \text{jump of } f(v).$$

Unfortunately if (A.53) or (A.54) are necessary conditions for the existence of discontinuous solutions they are not sufficient to ensure the uniqueness. For example if we consider again (A.49) with the initial condition

$$v_0(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases} \quad \text{then it is easy to check that:}$$

$$v(t, x) = \begin{cases} 0 & \text{if } x < t/2 \\ 1 & \text{if } x > t/2 \end{cases}$$

is a weak solution of (A.49) satisfying the Rankine-Hugoniot condition. However we can find another such solution

$$\hat{u}(t, x) = \begin{cases} 1 & \text{if } x > t \\ x/t & \text{if } 0 < x < t \\ 0 & \text{if } x < 0 \end{cases}$$

Thus, in general, weak solutions are not unique and we have to find further criterion that ensures uniqueness. Such a condition exists, it is called *entropy*.

We do not continue the investigation of the theoretical difficulties of hyperbolic equations of conservation laws since the general theory is complex and it is far beyond the scope of this Appendix to review it. We refer the interested reader to [122, 157] for the complete theory. Of course, these difficulties are still present when trying to discretize these equations. For example, let us consider again Burgers equation:

$$\begin{cases} \frac{\partial v}{\partial t} + \frac{1}{2} \frac{\partial}{\partial x} (v^2) = 0 \\ v(0, x) = \begin{cases} 1 & \text{if } x < 0, \\ 0 & \text{otherwise.} \end{cases} \end{cases} \quad (\text{A.55})$$

If we rewrite (A.55) in the quasilinear form:

$$\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} = 0 \quad (\text{A.56})$$

then a natural finite difference scheme inspired from the upwind method for (A.40) and assuming that  $v \geq 0$  is:

$$\begin{cases} u_i^{n+1} = u_i^n - \frac{\Delta t}{\Delta x} u_i^n (u_i^n - u_{i-1}^n) \\ u_i^0 = \begin{cases} 1 & \text{if } j < 0, \\ 0 & \text{otherwise.} \end{cases} \end{cases} \quad (\text{A.57})$$

Then it is easy to verify that  $u_i^n = u_i^0$  for all  $i$  and  $n$  regardless of the step sizes  $\Delta t$  and  $\Delta x$ . Therefore as  $\Delta t$  and  $\Delta x$  tend to zero, the numerical solution converges to the function  $\bar{v}(t, x) = v_0(x)$ . Unfortunately,  $\bar{v}(t, x)$  is not a weak solution. The reason is that discretizing Burgers equation written in the form (A.56) is not equivalent for non smooth solutions. For non smooth solutions, the product  $vv_x$  has not necessarily a meaning (even weakly). So, studying Burgers equation written in a conservative form is the right approach. But in this case we have to define the numerical schemes that agrees with this form. These schemes does exist and are called *conservative schemes*.

Let us consider a general hyperbolic equation of conservation laws:

$$\begin{cases} \frac{\partial v}{\partial t} + \frac{1}{2} \frac{\partial}{\partial x} (f(v)) = 0 \\ v(0, x) = v_0(x). \end{cases} \quad (\text{A.58})$$

We say that the numerical scheme is in conservation form if it writes as:

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{\Delta x} [F(u_{i-p}^n, u_{i-p+1}^n, \dots, u_{i+q}^n) - F(u_{i-p-1}^n, u_{i-p}^n, \dots, u_{i+q-1}^n)] \quad (\text{A.59})$$

for some function  $F$  of  $(p+q+1)$  arguments.  $F$  is called the numerical flux function. Of course, some consistency relations between  $F$  and  $f$  have to be satisfied. For example, if  $p=0$  and  $q=1$ , then (A.59) becomes:

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{\Delta x} [F(u_i^n, u_{i+1}^n) - F(u_{i-1}^n, u_i^n)]. \quad (\text{A.60})$$

In fact, for hyperbolic equations, it is often preferable to view  $u_i^n$  as an approximation of an average of  $v(n\Delta t, x)$  defined by:

$$u_i^n = \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} v(n\Delta t, x) dx \quad (\text{A.61})$$

where  $x_{i\pm\frac{1}{2}} = (i \pm 1/2)\Delta x$ . From the definition of a weak solution of (A.58) and by choosing a particular test function  $\varphi$ , we can show that if  $u$  is a weak solution then:

$$\begin{aligned} & \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} v((n+1)\Delta t, x) dx = \\ &= \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} v(n\Delta t, x) dx - \int_{n\Delta t}^{(n+1)\Delta t} [f(v(t, x_{i+\frac{1}{2}})) - f(v(t, x_{i-\frac{1}{2}}))] dt \end{aligned}$$

Then dividing by  $\Delta x$ , we get from (A.61):

$$u_i^{n+1} = u_i^n - \frac{1}{\Delta x} \int_{n\Delta t}^{(n+1)\Delta t} [f(v(t, x_{i+\frac{1}{2}})) - f(v(t, x_{i-\frac{1}{2}}))] dt \quad (\text{A.62})$$

So, comparing (A.60) and (A.62) it is natural to choose:

$$F(u_i^n, u_{i+1}^n) = \frac{1}{\Delta t} \int_{n\Delta t}^{(n+1)\Delta t} f(v(t, x_{i+\frac{1}{2}})) dt \quad (\text{A.63})$$

and then the scheme defined by (A.60) will be consistent with the original conservation law if  $F$  reduces to  $f$  for the case of constant solution, i.e. if  $v(t, x) \equiv c$  then necessarily:

$$F(c, c) = f(c) \quad \forall c \in R. \quad (\text{A.64})$$

This is the definition of a consistent scheme. This notion is very important since, according to the Lax-Wendroff Theorem [122, 157], if the numerical

scheme is consistent and in a conservative form, and if the resulting sequence of approximated solutions converges then necessarily the limiting function is a weak solution of the conservation law.

Unfortunately, consistence and conservative form are not sufficient in general to capture the correct discontinuous solution. For example schemes might develop undesirable oscillations. These conditions are related to the entropy condition mentioned above. Since we have not developed at all this notion, we will say no more on the numerical approximation of hyperbolic equations and we refer to [122, 157] for more development.

We summarize all the numerical concerns by saying that a monotone (which means that the numerical flux function  $F$  is a monotone increasing function of each of its arguments), consistent and conservative scheme always captures the solution we would like to get (the unique entropic weak solution).

## A.3 Difference schemes in image analysis

### A.3.1 Getting started

In this section we would like to show how certain PDEs studied in this book can be discretized. The generic form of these PDEs is:

$$\begin{cases} \mathcal{L}v = F & (x, y) \in \Omega \subset R^2 \\ \frac{\partial v}{\partial N}(t, x, y) = 0 & \text{on } \partial\Omega \quad (\text{Neumann boundary condition}) \\ v(0, x, y) = f(x, y) & \quad (\text{initial condition}) \end{cases} \quad (\text{A.65})$$

where  $\Omega$  is the image domain and  $N$  is the normal to the boundary of  $\Omega$  noted  $\partial\Omega$ .  $\mathcal{L}$  is generally a second order differential operator like (see for example Section 3.3):

$$\frac{\partial v}{\partial t}(t, x, y) + H(x, y, v(t, x, y), \nabla v(t, x, y), \nabla^2 v(t, x, y)) = 0.$$

**Example:** One of the simplest PDE which is presented in image analysis is the heat equation (see Section 3.3.1 where it is analyzed):

$$\begin{cases} \frac{\partial v}{\partial t} = \nu \Delta v = \nu \left( \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} \right) & (x, y) \in \Omega, \quad t \geq 0 \\ \frac{\partial v}{\partial N}(t, x, y) = 0 & \text{on } \partial\Omega \\ v(0, x, y) = f(x, y) \end{cases} \quad (\text{A.66})$$

where  $\nu$  is a positive constant. ■

As already mentioned, finite differences are widely used in image processing which is due to the digital structure of an image, as a set of pixels uniformly distributed (see Section 1.2). It is then very easy and natural to associate to an image a uniform grid, as presented in Figure A.5. Since there is no reason to choose it differently, the grid spacing in the  $x$  and  $y$  directions is

usually equal:

$$\Delta x = \Delta y = h.$$

Notice that in many articles from the computer vision literature, it is even chosen  $h = 1$  which means that the pixel size is chosen as the unit of reference. We will call the positions  $(ih, jh)$  vertices, nodes or pixels equivalently. We will denote by  $v_{i,j}^n$  (resp.  $u_{i,j}^n$ ) the value of the exact solution (resp. the discrete solution) at location  $(ih, jh)$  and time  $n\Delta t$ .

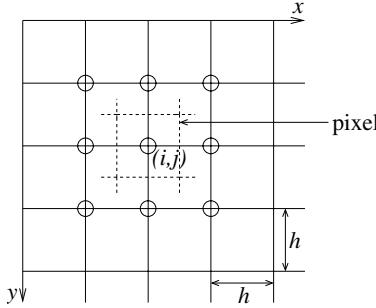


Figure A.5. Grid on the space domain. The circles indicates the vertices which belong to the  $3 \times 3$  neighborhood of the vertex  $(i, j)$ .

**Example:** The PDE (A.66) is a initial-boundary value problem. To discretize it, we need to consider:

- The equation. To find the difference scheme associated to the heat equation (A.66), we can proceed as in the one dimensional case (see Section A.1.1), that is by using Taylor expansions expanded about the index point  $(n\Delta t, ih, jh)$ . Naturally, the simplest method is to consider separately the discretization of each second order derivative in  $x$  and  $y$ , which is equivalent to use the one dimensional approximation. By doing so, we obtain:

$$\begin{aligned} \frac{\partial v}{\partial t} - \nu \Delta v \Big|_{i,j}^n &= \frac{v_{i,j}^{n+1} - v_{i,j}^n}{\Delta t} - \\ &- \nu \frac{v_{i+1,j}^n + v_{i-1,j}^n + v_{i,j+1}^n + v_{i,j-1}^n - 4v_{i,j}^n}{h^2} + \mathcal{O}(\Delta t) + \mathcal{O}(h^2). \end{aligned}$$

Then the difference scheme that we can propose is:

$$u_{i,j}^{n+1} = u_{i,j}^n + \frac{\nu \Delta t}{h^2} (u_{i+1,j}^n + u_{i-1,j}^n + u_{i,j+1}^n + u_{i,j-1}^n - 4u_{i,j}^n) \quad (\text{A.67})$$

It is of order  $(2, 1)$ .

- The boundary condition. The Neumann boundary condition can be taken into account by a symmetry procedure. If the value of a pixel (vertex) which is outside the domain is needed, we use the value of the pixel which is symmetric with respect to the boundaries.
- The initial condition is simply:  $u_{i,j}^0 = g_{i,j}$  where  $g$  is the discretization of  $f$ .

To illustrate this algorithm, we show in figure A.6 some iterations as applied to a very simple image.

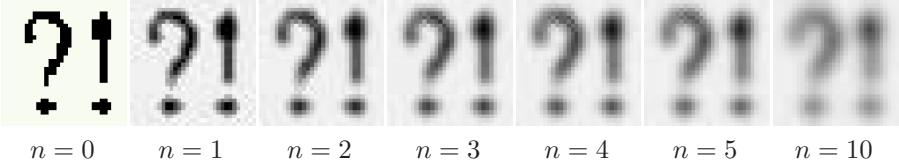


Figure A.6. Example of results with the scheme (A.67) at different times (iterations), as applied to a simple and small size image ( $32 \times 32$ )

Notice that this example shows clearly the propagation of the information as the number of iterations increases. ■

**Remark** The scheme (A.67) has been obtained by discretizing the Laplacian as a sum of the second order derivatives in the  $x$  and  $y$  directions:

$$\Delta v|_{i,j} \approx \frac{v_{i+1,j}^n + v_{i-1,j}^n + v_{i,j+1}^n + v_{i,j-1}^n - 4v_{i,j}^n}{h^2}. \quad (\text{A.68})$$

Clearly, this discretization does not take into account the 2-D nature and properties of this operator. To illustrate what we mean by “2-D nature and properties”, we can remark that the Laplacian operator is rotationally invariant. If we apply a rotation of center  $(x, y)$  to the image  $v$  (with any angle  $\theta \in [0, 2\pi[$ ), then  $\Delta v(x, y)$  keeps constant for all  $\theta$ . So should it be for the discretization. Naturally, as we consider a discrete domain, we may just ask that  $\Delta v|_{i,j}$  keeps constant under rotations of  $\pi/4$ , as depicted in Figure A.7. This is not the case of discretization (A.68) since we obtain 1 or 2 (for  $h = 1$ ) depending on the situation.

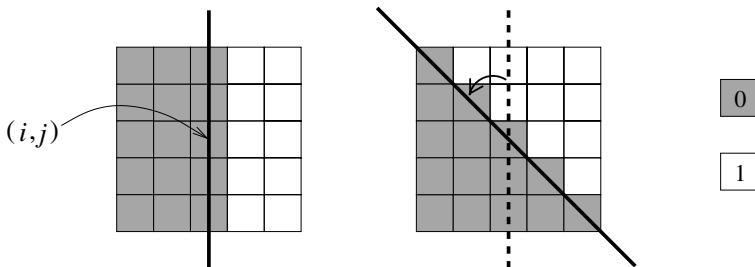


Figure A.7. Example of binary image representing a vertical edge and the same image after a rotation of  $\pi/4$  radians. Rotationally invariant operators estimated at the point in the middle (indicated by a dotted circle) should yield the same value in both situations.

To overcome this difficulty, we need to use the complete  $3 \times 3$  neighborhood.

We may propose the following approximation:

$$\begin{aligned}\Delta v|_{i,j} \approx & \lambda \frac{v_{i+1,j} + v_{i-1,j} + v_{i,j+1} + v_{i,j-1} - 4v_{i,j}}{h^2} + \\ & + (1-\lambda) \frac{v_{i+1,j+1} + v_{i-1,j+1} + v_{i+1,j-1} + v_{i-1,j-1} - 4v_{i,j}}{2h^2}\end{aligned}\quad (\text{A.69})$$

where  $\lambda \in [0, 1]$  is a constant to be chosen. We can verify that this approximation is consistent. By applying this operator (A.69) in the two situations from Figure A.7 and saying that both results should be equal yields  $\lambda = 1/3$ , hence the approximation.

Similarly, we can propose a discretization for the first order derivatives in  $x$  and  $y$  which is coherent with the fact that the norm of the gradient is invariant under rotation. As we will see further, a second order centered approximation of the first derivative in  $x$  is:

$$\frac{\partial v}{\partial x}|_{i,j} \approx \delta_x v_{i,j} = \frac{v_{i+1,j} - v_{i-1,j}}{2h} \quad (\text{A.70})$$

which can also be written in the  $y$  direction. The vertices involved in the estimation (A.70) are represented in Figure A.9.

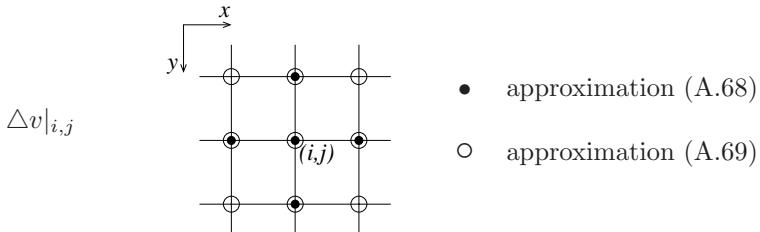


Figure A.8. Representation of the vertices involved in the finite difference schemes

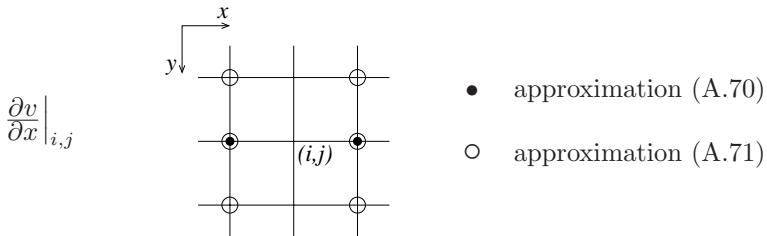


Figure A.9. Representation of the vertices involved in the finite difference schemes

As for the case of the Laplacian, these approximations are in fact “one dimensional” and do not really take an advantage that the data is of dimension 2. This is visible if we consider the value of the norm of the gradient of  $u$  in the two situations described in Figure A.7: we obtain either  $1/2$  or  $1/\sqrt{2}$ . The solution is to use more pixels in the estimation of the derivatives.

In particular, we may suggest the following approximation:

$$\begin{aligned} \frac{\partial v}{\partial x} \Big|_{i,j} &\approx \lambda \frac{v_{i+1,j} - v_{i-1,j}}{2h} + \\ &+ \frac{(1-\lambda)}{2} \left( \frac{v_{i+1,j+1} - v_{i-1,j+1}}{2h} + \frac{v_{i+1,j-1} - v_{i-1,j-1}}{2h} \right) \end{aligned} \quad (\text{A.71})$$

where  $\lambda$  is a parameter to be chosen. By applying the operator (A.71) in the two situations of Figure A.9 and by saying that both results should be equal yields  $\lambda = \sqrt{2} - 1$ , hence the approximation.

Finally, we would like to mention that using more points in the approximations is not only good for rotation invariance properties, but, practically, the result is also less sensitive to noise. The reason is that it is equivalent to perform a smoothing of the data before the estimation. ■

### A.3.2 Image restoration by energy minimization

We first consider the image restoration problem as presented in Section 3.2. By introducing a dual variable  $b$ , the problem became to minimize with respect to  $v$  and  $b$  the functional:

$$J_\varepsilon(v, b) = \frac{1}{2} \int_{\Omega} |Rv - v_0|^2 dx + \lambda \int_{\Omega} (b |\nabla v|^2 + \psi_\varepsilon(b)) dx.$$

The so-called half-quadratic minimization algorithm consists in minimizing successively  $J_\varepsilon$  with respect to each variable. The algorithm is (see Section 3.2.4 for more detail):

---

For  $(v^0, b^0)$  given

---

- $v^{n+1} = \underset{v}{\operatorname{argmin}} J_\varepsilon(v, b^n)$  i.e.  $\begin{cases} R^* R v^{n+1} - \operatorname{div}(b^n \nabla v^{n+1}) = 0 \text{ in } \Omega \\ b^n \frac{\partial v^{n+1}}{\partial N} = 0 \text{ on } \partial \Omega \end{cases}$
- $b^{n+1} = \underset{b}{\operatorname{argmin}} J_\varepsilon(v^{n+1}, b)$  i.e.  $b^{n+1} = \frac{\phi'(|\nabla v^{n+1}|)}{2|\nabla v^{n+1}|}$

- Go back to first step until convergence.

---

The limit  $(v^\infty, b^\infty)$  is the solution

---

As far as discretization is concerned, the only term which may be difficult to approximate is the divergence operator. So for  $b \geq 0$  and  $v$  given at nodes  $(i, j)$  the problem is to find an approximation at the node  $(i, j)$  for  $\operatorname{div}(b \nabla v)$ . This kind of term is naturally present as soon as there is a regularization with a  $\phi$  function (see for instance optical flow, Section 5.1.2, or sequence segmentation, Section 5.1.3). The diffusion operator used in

the Perona and Malik model is also of the same kind (see Sections 3.3.1 and 3.3.2).

Since this divergence operator may be rewritten:

$$\operatorname{div}(b\nabla v) = \frac{\partial}{\partial x} \left( b \frac{\partial v}{\partial x} \right) + \frac{\partial}{\partial y} \left( b \frac{\partial v}{\partial y} \right)$$

we can use previous one-dimensional approximation and combine them. For example, if we use the central finite difference approximation (A.70), we have:

$$\begin{aligned} \operatorname{div}(b\nabla v)|_{i,j} &\approx \delta_x(b_{i,j}\delta_x v_{i,j}) + \delta_y(b_{i,j}\delta_y v_{i,j}) = \\ &= \frac{1}{4h^2} \left( b_{i+1,j}v_{i+2,j} + b_{i-1,j}v_{i-2,j} + b_{i,j+1}v_{i,j+2} + b_{i,j-1}v_{i,j-2} - \right. \\ &\quad \left. - (b_{i+1,j} + b_{i-1,j} + b_{i,j+1} + b_{i,j-1})v_{i,j} \right). \end{aligned} \quad (\text{A.72})$$

The main drawback of this representation is that it involves only the points  $((i \pm 2)h, (j \pm 2)h)$ , and none of the  $3 \times 3$  neighborhood (see also Figure A.10). This may be non robust for noisy data or when there is a lot of variations in this region. Another possibility is to combine forward and backward differences (see Section A.2)

$$\begin{aligned} \operatorname{div}(b\nabla v)|_{i,j} &\approx \delta_x^+(b_{i,j}\delta_x^- v_{i,j}) + \delta_y^+(b_{i,j}\delta_y^- v_{i,j}) = \\ &= \frac{1}{h^2} \left( b_{i+1,j}v_{i+1,j} + b_{i,j}v_{i-1,j} + b_{i,j+1}v_{i,j+1} + b_{i,j}v_{i,j-1} - \right. \\ &\quad \left. - (b_{i+1,j} + b_{i,j+1} + 2b_{i,j})v_{i,j} \right). \end{aligned}$$

This approximation now involves the  $3 \times 3$  neighborhood, but it introduces a dissymmetry: the values of  $b$  at  $((i-1)h, jh)$  and  $(ih, (j-1)h)$  are not used. A solution is to use the following approximation for the derivatives:

$$\delta_x^* v_{i,j} = \frac{v_{i+\frac{1}{2},j} - v_{i-\frac{1}{2},j}}{h} \quad \text{and} \quad \delta_y^* v_{i,j} = \frac{v_{i,j+\frac{1}{2}} - v_{i,j-\frac{1}{2}}}{h}.$$

where  $v_{i\pm\frac{1}{2},j\pm\frac{1}{2}}$  is the value of  $v$  at location  $((i \pm \frac{1}{2})h, (j \pm \frac{1}{2})h)$  which can be obtained by interpolation. As for (A.70) it is a second order approximation. Then we have:

$$\begin{aligned} \operatorname{div}(b\nabla v)|_{i,j} &\approx \delta_x^*(b_{i,j}\delta_x^* v_{i,j}) + \delta_y^*(b_{i,j}\delta_y^* v_{i,j}) = \\ &= \frac{1}{h^2} \left( b_{+0}v_{i+1,j} + b_{-0}v_{i-1,j} + b_{0+}v_{i,j+1} + b_{0-}v_{i,j-1} - \right. \\ &\quad \left. - (b_{+0} + b_{-0} + b_{0+} + b_{0-})v_{i,j} \right) \end{aligned} \quad (\text{A.73})$$

where  $b_{\pm 0} = b_{i\mp\frac{1}{2},j}$ ,  $b_{\pm\mp} = b_{i\pm\frac{1}{2},j\mp\frac{1}{2}}$ , etc. Notice that since we applied twice the operators  $\delta_x^*$  and  $\delta_y^*$  this approximation only uses the values of  $v$

at  $((i \pm 1)h, (j \pm 1)h)$  (see Figure A.10). However, interpolation is needed for  $b$ .

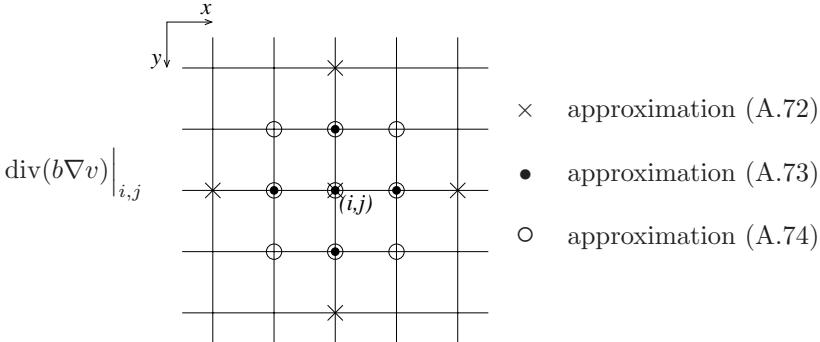


Figure A.10. Vertices involved in the approximation of the divergence term for the different schemes

As mentioned previously for the estimation of the Laplacian, it would also be interesting to take into account also the diagonal values. Then, we can look for an approximation such that:

$$\text{div}(b\nabla v)|_{i,j} \approx \quad (\text{A.74})$$

$$\begin{aligned} & \approx \frac{\lambda_p}{h^2} (b_{+0}v_{i+1,j} + b_{-0}v_{i-1,j} + b_{0+}v_{i,j+1} + b_{0-}v_{i,j-1} - \beta_p v_{i,j}) \\ & + \frac{\lambda_d}{h^2} (b_{++}v_{i+1,j+1} + b_{--}v_{i-1,j-1} + b_{-+}v_{i-1,j+1} + b_{+-}v_{i+1,j-1} - \beta_d v_{i,j}) \\ \text{with } & \left\{ \begin{array}{l} \beta_p = b_{0+} + b_{0-} + b_{+0} + b_{-0} \\ \beta_d = b_{++} + b_{--} + b_{-+} + b_{+-} \end{array} \right. \end{aligned}$$

where  $\lambda_p$  and  $\lambda_d$  are two weights to be chosen. The first condition is that the scheme must be consistent and it can be verified that this implies:

$$\lambda_p + 2\lambda_d = 1. \quad (\text{A.75})$$

Now, there remains one degree of freedom. Two possibilities can be considered:

- The first is to choose  $(\lambda_p, \lambda_d)$  constant, and for instance equal to  $(1/2, 1/4)$ , giving a privilege to the principal directions.
- The second is to choose  $(\lambda_p, \lambda_d)$  by taking into account the orientation of the gradient of  $v$ , as described in Figure A.11.

We tested these different discretizations as applied to a simple image with geometric structures (see Figure A.12). From left to right, an improvement in the results can be perceived (by observing the restoration of the horizontal and vertical edges). It is the adaptative choice that gives the best

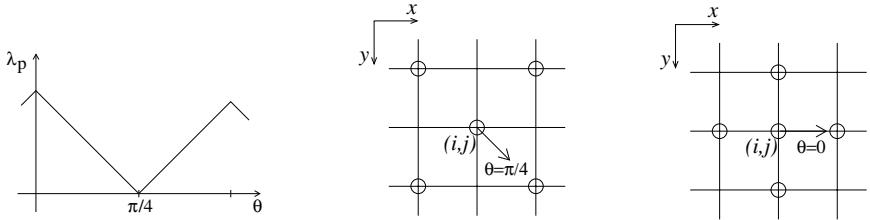


Figure A.11. Adaptative choice of the coefficients  $(\lambda_p, \lambda_d)$  as a function of  $\theta$ , the orientation of  $\nabla v$ . The two right-hand side figures show which points will be used in the discretization of the divergence term in two specific situations

result.

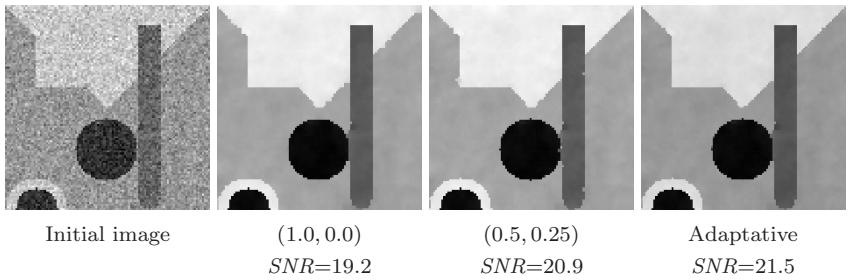


Figure A.12. Numerical tests for the different discretizations of the divergence term (the choice of  $(\lambda_p, \lambda_d)$  is indicated below the images).

### A.3.3 Image enhancement by the Osher and Rudin's shock-filters

This section concerns the shock-filters equation discussed in Section 3.3.3 and proposed by Osher and Rudin [199].

$$\frac{\partial v}{\partial t} = -|\nabla v| F(L(v)) \quad (\text{A.76})$$

where:

- $F$  is a Lipschitz function satisfying  $F(0) = 0$ ,  $\text{sign}(s)F(s) > 0$  ( $s \neq 0$ ), for example  $F(s) = \text{sign}(s)$ .
- $L$  is a second-order edge-detector, for example

$$L(v) = \Delta v = v_{xx} + v_{yy} \quad \text{or} \quad L(v) = \frac{1}{|\nabla v|^2} (v_x^2 v_{xx} + 2v_x v_y v_{xy} + v_y^2 v_{yy})$$

which corresponds to the second derivative of  $v$  in the direction of the normal to the isophotes.

Equation (A.76) involves two kinds of terms: a first order term  $|\nabla v|$  and a second order term  $F(L(v))$ :

- $L$  is discretized with central finite differences.
- $|\nabla v|$  has to be approximated with more care.  $v_x$  and  $v_y$  are approximated using the minmod operator  $m(\alpha, \beta)$ . For instance

$$v_x|_i = m(\delta_x^- v_i, \delta_x^+ v_i)$$

where

$$m(\alpha, \beta) = \begin{cases} \text{sign}(\alpha) \min(|\alpha|, |\beta|) & \text{if } \alpha\beta > 0 \\ 0 & \text{if } \alpha\beta \leq 0. \end{cases}$$

This function is usually called a flux limiter. As shown in Figure A.13, it permits to choose the lowest slope, or zero in case of a local extremum (this prevents instabilities due to noise).

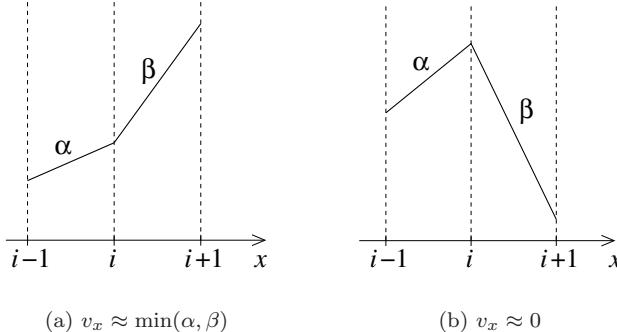


Figure A.13. Approximation of the first derivative using the minmod function

To summarize, the approximation of (A.76) is then given by

$$u_{i,j}^{n+1} = u_{i,j}^n - \frac{\Delta t}{h} \sqrt{(m(\delta_x^+ u_{i,j}^n, \delta_x^- u_{i,j}^n))^2 + (m(\delta_y^+ u_{i,j}^n, \delta_y^- u_{i,j}^n))^2} F_{i,j}(L(u^n))$$

where  $F_{i,j}(L(u)) = F(L_{i,j}(u))$ . We show in Figure A.14 an example of result and refer to Section 3.3.3 for more detail.

#### A.3.4 Curves evolution with the level sets method

In this section we briefly discuss the discretization of the PDEs governing curve evolutions. We only examine the case when these curves are identified

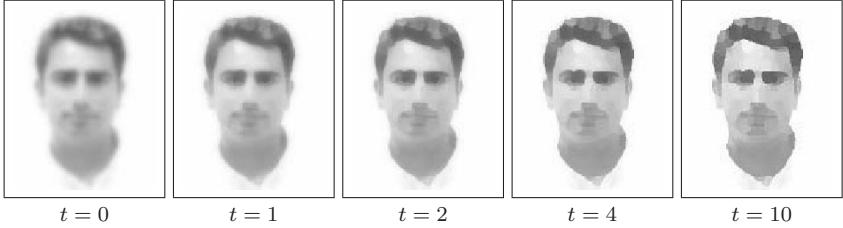


Figure A.14. Example of the shock filter on a blurred image of a face. It shows clearly that this filter reconstructs a piecewise constant image which is not satisfying on a perception point of view (the result does not look like a real image).

as level sets of the same function  $u(t, x)$  (Eulerian formulation). Of course, we have in mind the geodesic active contours model given by (see Section 4.3.3):

$$\frac{\partial v}{\partial t} = g(|\nabla I|) |\nabla v| \operatorname{div} \left( \frac{\nabla v}{|\nabla v|} \right) + \alpha g(|\nabla I|) |\nabla v| + \nabla g \cdot \nabla v. \quad (\text{A.77})$$

As mentioned before, (A.77) involves two kinds of terms: a parabolic term (the first one) and hyperbolic terms (the two last ones). One can easily imagine that the discretization of each term needs an appropriate treatment, according to its nature (parabolic or hyperbolic). The main idea is that parabolic terms can be discretized by central finite differences while hyperbolic terms need to be approximated by non-oscillatory upwind schemes. For the sake of clarity we start by examining the evolution driven by each of these terms.

For a detailed description of the above schemes and other numerical questions not developed here, we refer the reader to [229, 200].

### Mean curvature motion

Let us consider:

$$\begin{cases} \frac{\partial v}{\partial t} = |\nabla v| \operatorname{div} \left( \frac{\nabla v}{|\nabla v|} \right) \\ v(0, x, y) = v_0(x, y). \end{cases} \quad (\text{A.78})$$

Equation (A.78) is a parabolic equation and has diffusive effects (like the heat equation). So, the use of upwind schemes is inappropriate and classical central differences are used:

$$u_{i,j}^{n+1} = u_{i,j}^n + \Delta t \sqrt{(\delta_x u_{i,j}^n)^2 + (\delta_y u_{i,j}^n)^2} K_{i,j}^n$$

where  $K_{i,j}^n$  is the central finite difference approximation of the curvature:

$$K = \operatorname{div} \left( \frac{\nabla v}{|\nabla v|} \right) = \frac{v_{xx} v_y^2 + v_{yy} v_x^2 - 2v_x v_y v_{xy}}{(v_x^2 + v_y^2)^{3/2}}.$$

We let the reader write the expression of  $K_{i,j}^n$ . Unfortunately, the discretization of (A.78) is not as easy as it may appear. At some points  $(t, x)$ ,  $\nabla v$  can be undefined, or  $|\nabla v| = 0$ , or  $|\nabla v| = +\infty$ . This situation can occur even in very simple cases [208]. For example, let us consider the shrinking of a unit circle in 2-D which corresponds to:

$$v(0, x, y) = \sqrt{x^2 + y^2} - 1 \quad (\text{A.79})$$

i.e.  $v_0$  is the signed distance to the unit circle. This equation is rotationally invariant and if we search for the solution of the form:

$$v(t, x, y) = \phi(t, \sqrt{x^2 + y^2})$$

we easily get  $v(t, x, y) = \sqrt{x^2 + y^2 + 2t} - 1$  from which we deduce:

$$\begin{aligned} \nabla v &= \frac{1}{\sqrt{x^2 + y^2 + 2t}} \begin{pmatrix} x \\ y \end{pmatrix}, \quad |\nabla v| = \frac{\sqrt{x^2 + y^2}}{\sqrt{x^2 + y^2 + 2t}}, \\ \frac{\nabla v}{|\nabla v|} &= \frac{1}{\sqrt{x^2 + y^2}} \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{and} \quad \operatorname{div} \left( \frac{\nabla v}{|\nabla v|} \right) = \frac{1}{\sqrt{x^2 + y^2}} \end{aligned}$$

so the two last quantities are not defined at the origin and effectively a spike occurs at the origin (see [208]). Moreover, the interface  $\Gamma(t) = \{(x, y); u(t, x, y) = 0\}$  is the circle  $x^2 + y^2 = 1 - 2t$  and on  $\Gamma(t)$  we have  $|\nabla v|(t) = \sqrt{1 - 2t}$ . Therefore  $v(t, x, y)$  becomes more and more flat as the interface evolves and disappears at  $t = 1/2$ . To circumvent this type of problem, we have to find a numerical trick that prevents the gradient norm to vanish (or blow up). This can be realized by reinitializing the function  $v$  from time to time to a signed distance function.

More precisely, we run (A.78) until some step  $n$ , then we solve the auxiliary PDE:

$$\begin{cases} \frac{\partial \phi}{\partial t} + \operatorname{sign}(\phi)(|\nabla \phi| - 1) = 0 \\ \phi(0, x, y) = v(n\Delta t, x, y). \end{cases} \quad (\text{A.80})$$

The resulting solution (as  $t$  tend to infinity), noted  $\phi^\infty$ , is a signed distance function whose zero level set is the same as the function  $v(n\Delta t, x, y)$ . Then we can run again (A.78) with the initial data  $v(0, x, y) = \phi^\infty(x, y)$ . Practically, this reinitialization has to be done every  $n = 20$  iterations of the curve evolution equation and it is usually performed about 5 to 10 iterations of (A.80)<sup>8</sup>.

**Remark** There exists another way to avoid doing the reinitialization step. It consists in considering a modified equation (A.78) which has the property of maintaining the norm of the gradient of the solution equal to one.

---

<sup>8</sup>These numbers are just an indication and naturally depend on the kind of equation to be solved and on the time steps.

For further details see [125, 208] ■

An example of mean curvature motion is shown in Figure A.15. Notice that if we let the evolution run until convergence, any curve transforms into a circle and then collapses.

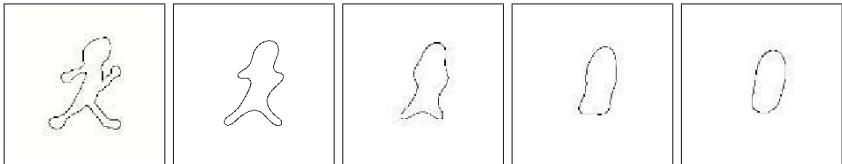


Figure A.15. Example of mean curvature motion

Constant speed evolution

The second example is given by:

$$\begin{cases} \frac{\partial v}{\partial t} = c |\nabla v| \\ v(0, x, y) = v_0(x, y). \end{cases} \quad (\text{A.81})$$

where  $c$  is a constant. This equation describes a motion in the direction normal to the front (the corresponding Lagrangian formulation of (A.81) is  $\frac{\partial \Gamma}{\partial t}(t, p) = c N(t, p)$  where  $N$  is the normal to  $\Gamma(t)$ ). For  $c = 1$ , it is also referred to as *grass fire* since it simulates a grass fire wavefront propagation.

Equation (A.81) is approximated by a non-oscillatory upwind scheme:

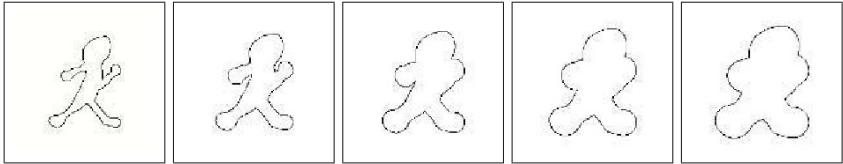
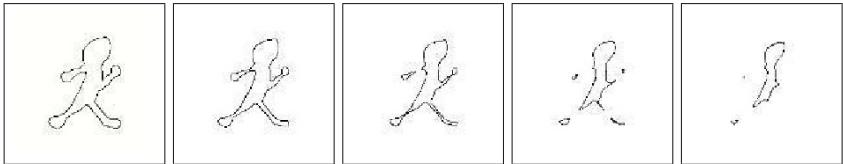
$$u_{i,j}^{n+1} = u_{i,j}^n + \Delta t \nabla^+ u_{i,j}^n$$

where:

$$\begin{aligned} \nabla^+ u_{i,j}^n &= \left[ \max(\delta_x^- u_{i,j}^n, 0)^2 + \min(\delta_x^+ u_{i,j}^n, 0)^2 + \right. \\ &\quad \left. + \max(\delta_y^- u_{i,j}^n, 0)^2 + \min(\delta_y^+ u_{i,j}^n, 0)^2 \right]^{\frac{1}{2}} \end{aligned}$$

We show in Figures A.16 and A.17 two examples of constant speed motions.

**Remark** Motions like equation (A.81) and more generally with a monotone speed have the following property: every point is crossed once and only once by the curve during its evolution. Notice that this is not the case for mean curvature motion. This property can be used to derive an efficient numerical approach called *fast marching algorithm* [241, 228]. It is beyond the scope of this Appendix to explain this method and we refer to the original articles and to [229] for more detail. ■

Figure A.16. Example constant speed motion ( $c = 1$ , grass fire)Figure A.17. Example constant speed motion ( $c = -1$ )

The pure advection equation

We consider here the equation:

$$\begin{cases} \frac{\partial v}{\partial t} = A(x, y) \cdot \nabla v \\ v(0, x, y) = v_0(x, y). \end{cases} \quad (\text{A.82})$$

where  $A(x, y) = (A_1(x, y), A_2(x, y))$ . For (A.82) we use a simple upwind scheme i.e. we check the sign of each component of  $A$  and construct a one-side upwind difference in the appropriate direction:

$$u_{i,j}^{n+1} = u_{i,j}^n + \Delta t \left[ \max((A_1)_{i,j}^n, 0) \delta_x^- u_{i,j}^n + \min((A_1)_{i,j}^n, 0) \delta_x^+ u_{i,j}^n + \right. \\ \left. \max((A_2)_{i,j}^n, 0) \delta_y^- u_{i,j}^n + \min((A_2)_{i,j}^n, 0) \delta_y^+ u_{i,j}^n \right].$$

Image segmentation by the geodesic active contour model

Now, we can consider the geodesic active contour model (A.77) which can be seen as the sum of the previous discretization. So the discrete scheme is:

$$u_{i,j}^{n+1} = u_{i,j}^n + \Delta t \left[ \begin{array}{l} g_{i,j} K_{i,j}^n [(\delta_x u_{i,j}^n)^2 + (\delta_y u_{i,j}^n)^2]^{\frac{1}{2}} + \\ + \alpha [\max(g_{i,j}, 0) \nabla^+ + \min(g_{i,j}, 0) \nabla^-] u_{i,j} + \\ + \max((g_x)_{i,j}^n, 0) \delta_x^- u_{i,j}^n + \min((g_x)_{i,j}^n, 0) \delta_x^+ u_{i,j}^n + \\ + [\max((g_y)_{i,j}^n, 0) \delta_y^- u_{i,j}^n + \min((g_y)_{i,j}^n, 0) \delta_y^+ u_{i,j}^n]. \end{array} \right]$$

where  $\nabla^- u_{i,j}^n$  is obtained from  $\nabla^+ u_{i,j}^n$  by inverting the signs plus and minus.

Figure A.18 shows a typical example of result (see Figure 4.13 for the complete evolution).

**Remark** On a numerical point of view, all the equations presented in

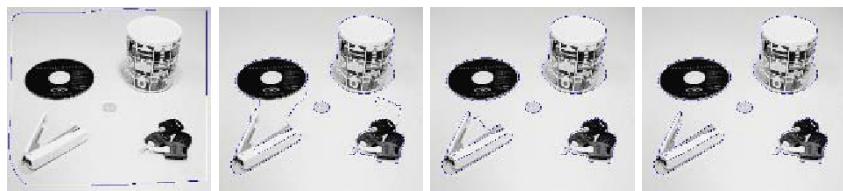


Figure A.18. Example of segmentation. Different iterations are displayed

this section involve local operations. As we are only interested in the curve, it is enough to update the values in a band around the current position of the curve, also called *narrow band*. Naturally, this region (band) has to be updated as the curve evolves. See for instance [229] for more details. ■

# References

- [1] R. Adams. *Sobolev spaces*, volume 65 of *Pure and Applied Mathematics, Series of Monographs and Textbooks*. Academic Press, Inc., New York - San Francisco - London, 1975.
- [2] A.A. Alatan and L. Onural. Gibbs random field model based 3-D motion estimation by weakened rigidity. *International Conference on Image Processing*, II:790–794, 1994.
- [3] S. Allen and J. Cahn. A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening. *Acta Metallurgica*, 27:1085–1095, 1979.
- [4] L. Alvarez, F. Guichard, P.L. Lions, and J.M. Morel. Axiomatisation et nouveaux opérateurs de la morphologie mathématique. *C.R. Acad. Sci. Paris*, pages 265–268, 1992. t. 315, Série I.
- [5] L. Alvarez, F. Guichard, P.L. Lions, and J.M. Morel. Axioms and fundamental equations of image processing. *Archive for Rational Mechanics and Analysis*, 123(3):199–257, 1993.
- [6] L. Alvarez, P.L. Lions, and J.M. Morel. Image selective smoothing and edge detection by nonlinear diffusion (II). *SIAM Journal of numerical analysis*, 29:845–866, 1992.
- [7] L. Alvarez and L. Mazorra. Signal and image restoration using shock filters and anisotropic diffusion. *SIAM Journal of Numerical Analysis*, 31(2):590–605, April 1994.
- [8] L. Alvarez, J. Weickert, and J. Sanchez. A scale-space approach to nonlocal optical flow calculations. *Second International Conference on Scale Space, Lecture Notes in Computer Sciences*, 1682:235–246, 1999.

- [9] L. Ambrosio. A compactness theorem for a new class of functions of bounded variation. *Bollettino della Unione Matematica Italiana*, VII(4):857–881, 1989.
- [10] L. Ambrosio. Variational problems in  $SBV$  and image segmentation. *Acta Applicandae Mathematicae*, 17:1–40, 1989.
- [11] L. Ambrosio. Existence theory for a new class of variational problems. *Archive for Rational Mechanics and Analysis*, 111:291–322, 1990.
- [12] L. Ambrosio and G. Dal Maso. A general chain rule for distributional derivatives. *Proceedings of the American Mathematical Society*, 108(3):691–702, 1990.
- [13] L. Ambrosio and V.M. Tortorelli. Approximation of functionals depending on jumps by elliptics functionals via  $\Gamma$ -convergence. *Communications on Pure and Applied Mathematics*, XLIII:999–1036, 1990.
- [14] M. Attouch. Epi-convergence and duality. Convergence of sequences of marginal and Lagrangians functions. Applications to homogenization problems in mechanics. *Lecture Notes Mathematics*, 1190:21–56, 1986.
- [15] G. Aubert and L. Blanc-Féraud. Some remarks on the equivalence between 2D and 3D classical snakes and geodesic active contours. *The International Journal of Computer Vision*, 34(1):19–28, 1999.
- [16] G. Aubert, L. Blanc-Féraud, M. Barlaud, and P. Charbonnier. A deterministic algorithm for edge-preserving computed imaging using Legendre transform. *International Conference on Pattern Recognition*, III:188–191, 1994.
- [17] G. Aubert, R. Deriche, and P. Kornprobst. Computing optical flow via variational techniques. *SIAM Journal of Applied Mathematics*, 60(1):156–182, 1999.
- [18] G. Aubert and P. Kornprobst. A mathematical study of the relaxed optical flow problem in the space  $BV$ . *SIAM Journal on Mathematical Analysis*, 30(6):1282–1308, 1999.
- [19] G. Aubert and R. Tahraoui. Résultats d’existence en optimisation non convexe. *Applicable Analysis*, 18:75–100, 1984.
- [20] G. Aubert and L. Vese. A variational method in image recovery. *SIAM Journal of Numerical Analysis*, 34(5):1948–1979, October 1997.
- [21] S. Baldo. Minimal interface criterion for phase transitions in mixtures of Cahn-Hilliard fluids. *Annales de l’Institut Henri Poincaré, Analyse Non Linéaire*, 7(2):67–90, 1990.
- [22] G.I. Barenblatt, M. Bertsch, R. Dal Passo, and M. Ughi. A degenerate pseudoparabolic regularization of a nonlinear forward-backward heat equation arising in the theory of heat and mass exchange in stably stratified turbulent shear flow. *SIAM Journal on Mathematical Analysis*, 24:1414–1439, 1993.
- [23] G. Barles. Discontinuous viscosity solutions of first order Hamilton-Jacobi equations: a guided visit. *Nonlinear Analysis TMA*, 20(9):1123–1134, 1993.
- [24] G. Barles. *Solutions de viscosité des équations de Hamilton-Jacobi*. Springer-Verlag, 1994.

- [25] E.N. Barron and R. Jensen. Optimal control and semicontinuous viscosity solutions. *Proceedings of the American Mathematical Society*, 113:49–79, 1991.
- [26] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. *The International Journal of Computer Vision*, 12(1):43–77, 1994.
- [27] A.C. Barroso, G. Bouchitté, G. Buttazzo, and I. Fonseca. Relaxation of bulk and interfacial energies. *Archive for Rational Mechanics and Analysis*, 135:107–173, 1996.
- [28] Bart M. ter Haar Romeny. *Geometry-driven diffusion in computer vision. Computational imaging and vision*. Kluwer Academic Publishers, 1994.
- [29] G. Bellettini and A. Coscia. Approximation of a functional depending on jumps and corners. *Bollettino della Unione Matematica Italiana*, 8(1):151–181, 1994.
- [30] G. Bellettini and A. Coscia. Discrete approximation of a free discontinuity problem. *Numerical Functional Analysis and Optimization*, 15(3–4):201–224, 1994.
- [31] D. Béréziat, I. Herlin, and L. Younes. A generalized optical flow constraint and its physical interpretation. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 487–492, Hilton Head Island, South Carolina, June 2000. IEEE Computer Society.
- [32] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. *SIGGRAPH*, pages 417–424, 2000.
- [33] M. Berthod, Z. Kato, S. Yu, and J. Zerubia. Bayesian image classification using Markov Random Fields. *Image and Vision Computing*, 14(4):285–293, 1996.
- [34] M.J. Black. *Robust incremental optical flow*. PhD thesis, Yale University, Department of Computer Science, 1992.
- [35] M.J. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *CVGIP: Image Understanding*, 63(1):75–104, 1996.
- [36] M.J. Black, D.J. Fleet, and Y. Yacoob. Robustly estimating changes in image appearance. *Computer Vision and Image Understanding*, 78:8–31, 2000.
- [37] M.J. Black and P. Rangarajan. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *The International Journal of Computer Vision*, 19(1):57–91, 1996.
- [38] A. Blake and M. Isard. *Active contours*. Springer-Verlag, 1998.
- [39] A. Blake and A. Zisserman. *Visual reconstruction*. MIT Press, 1987.
- [40] L. Blanc-Féraud, M. Barlaud, and T. Gaidon. Motion estimation involving discontinuities in a multiresolution scheme. *Optical Engineering*, 32(7):1475–1482, July 1993.

- [41] A. Bonnet. Caractérisation des minima globaux de la fonctionnelle de Mumford-Shah en segmentation d'images. *Comptes Rendus de l'Académie des Sciences*, 321(I):1121–1126, 1995.
- [42] A. Bonnet. Sur la régularité des bords de minima de la fonctionnelle de Mumford-Shah. *Comptes Rendus de l'Académie des Sciences*, 321(I):1275–1279, 1995.
- [43] A. Bonnet. On the regularity of the edge set of Mumford-Shah minimizers. *Progress in nonlinear differential equations*, 25:93–103, 1996.
- [44] G. Bouchitté, A. Braides, and G. Buttazzo. Relaxation results for some free discontinuity problems. *J. Reine Angew. Math.*, 458:1–18, 1995.
- [45] G. Bouchitté, I. Fonseca, and L. Mascarenhas. A global method for relaxation. Technical report, Université de Toulon et du Var, May 1997.
- [46] C. Bouman and K. Sauer. A generalized Gaussian model for edge-preserving MAP estimation. *IEEE Transactions on Image Processing*, 2(3):296–310, 1993.
- [47] B. Bourdin. Image segmentation with a finite element method. *M2AN, Math. Model. Numer. Anal.*, 33(2):229–244, 1999.
- [48] A. Braides. *Approximation of free-discontinuity problems*, volume 1694 of *Lecture Note in Mathematics*. Springer Verlag, 1998.
- [49] A. Braides and G. Dal Maso. Non-local approximation of the Mumford-Shah functional. *Calc. Var. Partial Differ. Equ.*, 5(4):293–322, 1997.
- [50] H. Brezis. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*. North-Holland Publishing Comp, Amsterdam-London, 1973.
- [51] H. Brezis. *Analyse fonctionnelle. Théorie et applications*. Masson, 1983.
- [52] G. Buttazzo. *Semicontinuity, relaxation and integral representation in the calculus of variations*. Longman Scientific & Technical, 1989.
- [53] J. Cahn and J.E. Hilliard. Free energy of a nonuniform system. I. interfacial free energy. *Journal of chemical physics*, 28(1):258–267, 1958.
- [54] J.F. Canny. Finding edges and lines in images. Technical Report AI-TR-720, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, June 1983.
- [55] F. Cao. Partial differential equations and mathematical morphology. *J. Math. Pures Appl., IX.*, 77(9):909–941, 1998.
- [56] M. Carriero, A. Leaci, and F. Tomarelli. A second order model in image segmentation: Blake & Zisserman functional. *Nonlinear Differ. Equ. Appl.*, 25:57–72, 1996.
- [57] V. Caselles, F. Catte, T. Coll, and F. Dibos. A geometric model for active contours. *Numerische Mathematik*, 66:1–31, 1993.
- [58] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *The International Journal of Computer Vision*, 22(1):61–79, 1997.
- [59] F. Catte, P.L. Lions, J.M. Morel, and T. Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM Journal of Numerical Analysis*, 29(1):182–193, February 92.

- [60] T. Cazenave and A. Haraux. *Introduction aux problèmes d'évolution semi-linéaires. (Introduction to semilinear evolution problems)*. Mathématiques & Applications. Ellipses, 1990.
- [61] A. Chambolle. Image segmentation by variational methods: Mumford and shah functional and the discrete approximation. *SIAM Journal of Applied Mathematics*, 55(3):827–863, 1995.
- [62] A. Chambolle and G. Dal Maso. Discrete approximations of the Mumford-Shah functional in dimension two. Technical Report 9820, Université Paris Dauphine, Ceremade, 1998.
- [63] A. Chambolle, R.A. DeVore, N.Y. Lee, and B.J. Lucier. Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Transactions on Image Processing*, 7(3):319–334, 1998.
- [64] A. Chambolle and P.L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(2):167–188, 1997.
- [65] T. Chan, B.Y. Sandberg, and L. Vese. Active contours without edges for vector-valued images. *Journal of Visual Communication and Image Representation*, 11:130–141, 2000.
- [66] T. Chan and L. Vese. Active contours without edges. Technical Report 98-53, UCLA CAM Report, 1999.
- [67] T. Chan and L. Vese. Image segmentation using level sets and the piecewise-constant Mumford-Shah model. Technical Report 00-14, UCLA CAM Report, 2000.
- [68] T. Chan and L. Vese. A level set algorithm for minimizing the Mumford-Shah functional in image processing. Technical Report 00-13, UCLA CAM Report, 2000.
- [69] T. Chan and L. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, February 2001.
- [70] P. Charbonnier. *Reconstruction d'image: régularisation avec prise en compte des discontinuités*. PhD thesis, Université de Nice-Sophia Antipolis, 1994.
- [71] P. Charbonnier, G. Aubert, M. Blanc-Féraud, and M. Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. *International Conference on Image Processing*, II:168–172, November 1994.
- [72] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing*, 6(2):298–311, 1997.
- [73] R. Chellappa and A. Jain. *Markov Random Fields: theory and application*. Academic Press, Boston, 1993.
- [74] Y.G. Chen, Y. Giga, and S. Goto. Uniqueness and existence of viscosity solutions of generalized mean curvature flow equations. *Journal on Differential Geometry*, 33:749–786, 1991.
- [75] M. Chipot, R. March, M. Rosati, and G. Vergara Caffarelli. Analysis of a nonconvex problem related to signal selective smoothing. *Mathematical Models and Methods in Applied Science*, 7(3):313–328, 1997.

- [76] P.G. Ciarlet and J.L. Lions, editors. *Handbook of numerical analysis. Volume I. Finite difference methods. Solution of equations in  $R^n$ .* North-Holland, 1990.
- [77] I. Cohen. Nonlinear variational method for optical flow computation. In *Scandinavian Conference on Image Analysis*, volume 1, pages 523–530, 1993.
- [78] T. Corpetti, E. Mémin, and P. Pérez. Estimating fluid optical flow. *International Conference on Pattern Recognition*, 3:1045–1048, 2000.
- [79] M. G. Crandall and H. Ishii. The maximum principle for semicontinuous functions. *Differential and Integral Equations*, 3(6):1001–1014, 1990.
- [80] M.G. Crandall. Viscosity solutions of Hamilton-Jacobi equations. In *Non-linear problems: present and future, Proc. 1st Los Alamos Conf., 1981*, volume 61, pages 117–125. North-Holland Math. Stud., 1982.
- [81] M.G. Crandall, H. Ishii, and P.L. Lions. User’s guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Soc.*, 27:1–67, 1992.
- [82] M.G. Crandall and P.L. Lions. Condition d’unicité pour les solutions généralisées des équations de Hamilton-Jacobi du premier ordre. *Comptes Rendus de l’Académie des Sciences*, (292):183–186, 1981.
- [83] B. Dacorogna. *Direct methods in the calculus of variations*. Number 78 in Applied Mathematical Sciences. Springer Verlag, 1989.
- [84] G. Dal Maso. *An introduction to  $\Gamma$ -convergence*. Progress in Nonlinear Differential Equations and their Applications. Birkhauser, 1993.
- [85] E. De Giorgi. Convergence problems for functionals and operators. *Recent methods in non-linear analysis, Proc. Int. Meet., Rome 1978*, pages 131–188, 1979.
- [86] E. De Giorgi. Free discontinuity problems in calculus of variations. *Frontiers in pure and applied mathematics, Coll. Pap. Ded. J.L. Lions Occas. 60th Birthday*, pages 55–62, 1991.
- [87] E. De Giorgi and L. Ambrosio. Un nuovo tipo di funzionale del calcolo delle variazioni. *Att. Accad. Naz. Lincei, Rend. Cl Sci. Fis. Mat. Nat.*, 82:199–210, 1988.
- [88] E. De Giorgi, M. Carriero, and A. Leaci. Existence theorem for a maximum problem with a free discontinuity set. *Archive for Rational Mechanics and Analysis*, 108:195–218, 1989.
- [89] F. Demengel and R. Temam. Convex functions of a measure and applications. *Indiana University Mathematics Journal*, 33:673–709, 1984.
- [90] G. Demoment. Image reconstruction and restoration: overview of common estimation structures and problems. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(12):2024–2036, 1989.
- [91] R. Deriche. Using Canny’s criteria to derive a recursively implemented optimal edge detector. *The International Journal of Computer Vision*, 1(2):167–187, May 1987.
- [92] R. Deriche, P. Kornprobst., and G. Aubert. Optical flow estimation while preserving its discontinuities: a variational approach. In *Proceedings of*

- the 2nd Asian Conference on Computer Vision*, volume 2/3, Singapore, December 1995.
- [93] B. Despas and F. Helt, editors. *La restauration numérique des films cinématographiques*. Commission Supérieure et Technique de l'Image et du Son, 1997.
  - [94] V. Devlaminck and J.P. Dubus. Estimation of compressible or incompressible deformable motions for density images. *International Conference on Image Processing*, I:125–128, 1996.
  - [95] R.A. DeVore, B. Jawerth, and B. Lucier. Image compression through wavelet transform coding. *IEEE Transactions on Information Theory, Special Issue Wavelet Transforms Multires. Anal.*, 38:719–746, 1992.
  - [96] R.A. DeVore and P. Popov. Interpolation of Bsov spaces. *Proceedings of the American Mathematical Society*, 305:397–414, 1988.
  - [97] E. DiBenedetto. *Degenerate parabolic equations*. Springer-Verlag, 1993.
  - [98] D. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41:613–627, 1995.
  - [99] D. Donoho. Nonlinear solution of linear inverse problems by wavelet-vaguelet decomposition. *Applied and Computational Harmonic Analysis*, 2:101–126, 1995.
  - [100] D. Donoho and I. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of American Statistical Association*, 90:1200–1224, 1995.
  - [101] I. Ekeland and R. Temam. *Convex analysis and variational problems. Translated by Minerva Translations, Ltd., London.*, volume 1 of *Studies in Mathematics and its Applications*. Amsterdam - Oxford: North-Holland Publishing Company; New York: American Elsevier Publishing Company, 1976.
  - [102] A.I. El-Fallah and G.E. Ford. On mean curvature diffusion in nonlinear image filtering. *Pattern Recognition Letters*, 19:433–437, 1998.
  - [103] W. Enkelmann. Investigation of multigrid algorithms for the estimation of optical flow fields in image sequences. *CVGIP*, 43:150–177, 1988.
  - [104] C.L. Epstein and Michael Gage. The curve shortening flow. In A.J. Chorin, A.J. Majda, and P.D. Lax, editors, *Wave motion: theory, modelling and computation*. Springer-Verlag, 1987.
  - [105] L.C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. Proceedings of the American Mathematical Society, 1998.
  - [106] L.C. Evans and R.F. Gariepy. *Measure theory and fine properties of functions*. CRC Press, 1992.
  - [107] C. Eveland, K. Konolige, and R.C. Bolles. Background modeling for segmentation of video-rate stereo sequences. *International Conference on Computer Vision and Pattern Recognition*, pages 266–272, 1998.
  - [108] J.M. Fitzpatrick. The existence of geometrical density-image transformations corresponding to object motion. *Computer Vision, Graphics, and Image Processing*, 44(2):155–174, 1988.
  - [109] G.B. Folland. *Real analysis. Modern techniques and their applications*. Pure and Applied Mathematics. A Wiley-Interscience Publication, 1984.

- [110] G.B. Folland. *Fourier analysis and its applications*. The Wadsworth & Brooks/Cole Mathematics Series. Brooks/Cole Advanced Books & Software, 1992.
- [111] I. Fonseca and L. Tartar. The gradient theory of phase transitions for systems with two potential wells. *Proc. R. Soc. Edinb., Sect.,* 111(A)(1/2):89–102, 1989.
- [112] E. François and P. Bouthemy. The derivation of qualitative information in motion analysis. In O.D. Faugeras, editor, *Proceedings of the 1st European Conference on Computer Vision*, pages 226–230, Antibes, France, April 1990. Springer, Berlin, Heidelberg.
- [113] B. Galvin, B. McCane, K. Novins, D. Mason, and S. Mills. Recovering motion fields: an evaluation of eight optical flow algorithms. *British Machine Vision Conference*, pages 195–204, 1998.
- [114] D. Geman. *Random fields and inverse problems in imaging*, volume 1427 of *Lecture Notes in Mathematics, "Ecole d'été de Saint Flour"*. Springer Verlag, 1992.
- [115] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(3):367–383, 1993.
- [116] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.
- [117] D. Gilbarg and N.S. Trudinger. *Elliptic partial differential equations of second order*. 2nd ed., volume XIII of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, 1983.
- [118] G.L. Gimel’farb. *Image textures and Gibbs Random Fields*. Kluwer Academic Publishers, 1999.
- [119] E. De Giorgi and T. Franzoni. Su un tipo di convergenza variazionale. *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur.*, 68:842–850, 1975.
- [120] E. Giusti. *Minimal surfaces and functions of bounded variation*. Birkhäuser, 1984.
- [121] M. Gobbino. Finite difference approximation of the Mumford-Shah functional. *Communications on Pure Applied Mathematics*, 51(2):197–228, 1998.
- [122] E. Godlewski and P.A. Raviart. *Hyperbolic systems of conservation laws*, volume 3/4 of *Mathématiques et Applications*. SMAI, 1991.
- [123] C. Goffman and J. Serrin. Sublinear functions of measures and variational integrals. *Duke Mathematical Journal*, 31:159–178, 1964.
- [124] J. Gomes and O. Faugeras. Representing and evolving smooth manifolds of arbitrary dimension embedded in  $R^n$  as the intersection of  $n$  hypersurfaces: the vector distance functions. Technical Report 4012, INRIA, 2000.
- [125] J. Gomes and O.D. Faugeras. Reconciling distance functions and level sets. *Journal of Visual Communication and Image Representation*, 11:209–223, 2000.

- [126] W.E.L. Grimson, C. Stauffer, R. Romano, and L. Lee. Using adaptive tracking to classify and monitor activities in a site. *International Conference on Computer Vision and Pattern Recognition*, pages 22–31, 1998.
- [127] P. Grisvard. *Elliptic problems in nonsmooth domains*, volume XIV of *Monographs and Studies in Mathematics*. Pitman Advanced Publishing Program. Boston-London-Melbourne: Pitman Publishing Inc., 1985.
- [128] F. Guichard. Multiscale analysis of movies. *Proceedings of the eighth workshop on image and multidimensional signal processing, IEEE*, pages 236–237, 1993.
- [129] F. Guichard. A morphological, affine, and galilean invariant scale-space for movies. *IEEE Transactions on Image Processing*, 7(3):444–456, 1998.
- [130] F. Guichard and J.M. Morel. Image iterative smoothing and P.D.E.’s - Questions Mathématiques en Traitement du Signal et de l’Image. Technical Report 12, Institut Henri Poincaré, 1998.
- [131] F. Guichard and L. Rudin. Accurate estimation of discontinuous optical flow by minimizing divergence related functionals. *International Conference on Image Processing*, I:497–500, 1996.
- [132] S.N. Gupta and J.L. Prince. On div-curl regularization for motion estimation in 3-d volumetric imaging. *International Conference on Image Processing*, pages 929–932, 1996.
- [133] G.D. Hager and P.N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1025–1039, 1998.
- [134] R. Haralick. Digital step edges from zero crossing of second directional derivatives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(1):58–68, January 1984.
- [135] H.W. Haussecker and D.J. Fleet. Computing optical flow with physical models of brightness variations. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 760–767, Hilton Head Island, South Carolina, June 2000. IEEE Computer Society.
- [136] C. Hirsch. *Numerical computation of internal and external flows. Volume 1: Fundamentals of numerical discretization*. Wiley Series in Numerical Methods in Engineering; Wiley-Interscience Publication, 1988.
- [137] B.K. Horn. *Robot vision*. MIT Press, 1986.
- [138] B.K. Horn and B.G. Schunck. Determining Optical Flow. *Artificial Intelligence*, 17:185–203, 1981.
- [139] M. Irani and S. Peleg. Motion analysis for image enhancement: resolution, occlusion, and transparency. *Journal on Visual Communications and Image Representation*, 4(4):324–335, 1993.
- [140] Anil K. Jain. *Fundamentals of digital image processing*. Prentice-Hall International Editions, 1989.
- [141] G. Kanizsa. *Grammatica del vedere*. Bologna: Il Mulino, 1980.

- [142] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. In *First International Conference on Computer Vision*, pages 259–268, London, June 1987.
- [143] S. Kichenassamy. The Perona-Malik paradox. *SIAM Journal of Applied Mathematics*, 57(5):1328–1342, October 1997.
- [144] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi. Conformal curvature flows: from phase transitions to active vision. *Archive for Rational Mechanics and Analysis*, 134:275–301, 1996.
- [145] R. Kimmel, R. Malladi, and N. Sochen. Images as embedded maps and minimal surfaces: movies, color, texture, and volumetric medical images. *International Journal of Computer Vision*, 39(2):111–129, September 2000.
- [146] D. Kinderlehrer and G. Stampacchia. *An introduction to variational inequalities and their applications*. Academic Press, 1980.
- [147] A. Kirsch. *An introduction to the mathematical theory of inverse problems*, volume 120 of *Applied Mathematical Sciences*. Springer, 1996.
- [148] J.J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.
- [149] A. Kokaram. *Motion Picture Restoration : Digital Algorithms for Artefact Suppression in Degraded Motion Picture Film and Video*. Springer, 1998.
- [150] A. C. Kokaram. *Motion Picture Restoration*. PhD thesis, Cambridge University, England, May 1993.
- [151] P. Kornprobst. *Contributions à la restauration d'images et à l'analyse de séquences : Approches Variationnelles et Equations aux Dérivées Partielles*. PhD thesis, Université de Nice-Sophia Antipolis, 1998.
- [152] P. Kornprobst, R. Deriche, and G. Aubert. Image coupling, restoration and enhancement via PDE's. In *International Conference on Image Processing*, volume IIII, pages 458–461, Santa-Barbara, California, October 1997.
- [153] P. Kornprobst, R. Deriche, and G. Aubert. Nonlinear operators in image restoration. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pages 325–331, Puerto-Rico, June 1997. IEEE Computer Society, IEEE.
- [154] P. Kornprobst, R. Deriche, and G. Aubert. Image Sequence Analysis via Partial Differential Equations. *Journal of Mathematical Imaging and Vision*, 11(1):5–26, October 1999.
- [155] A. Kumar, A. Tannenbaum, and G. Balas. Optical flow : a curve evolution approach. *IEEE Transactions on Image Processing*, 5:598–611, 1996.
- [156] O.A. Ladyzhenskaya, V.A. Solonnikov, and N.N. Ural'ceva. *Linear and quasilinear equations of parabolic type*. Proceedings of the American Mathematical Society, Providence, RI, 1968.
- [157] R.J. LeVeque. *Numerical methods for conservation laws*. Birkhauser, Basel, 1992.
- [158] S.Z. Li. *Markov Random Field modeling in computer vision*. Springer-Verlag, 1995.
- [159] J.L. Lions and E. Magenes. *Problèmes aux limites non homogènes et applications*. Paris, Dunod, 1968. Volumes 1, 2, 3.

- [160] P.L. Lions. *Generalized solutions of Hamilton-Jacobi equations*. Pitman, 1982.
- [161] M.M. Lipschutz. *Differential geometry, theory and problems*. Schaum's Outline Series. Mc Graw-Hill, 1969.
- [162] W. Lohmiller and J.J. Slotine. Global convergence rates of nonlinear diffusion for time-varying images. In Mads Nielsen, P. Johansen, O.F. Olsen, and J. Weickert, editors, *Scale-Space Theories in Computer Vision*, volume 1682 of *Lecture Notes in Computer Science*, pages 525–529. Springer Verlag, 1999.
- [163] B. Lucas. *Generalized image matching by method of differences*. PhD thesis, Carnegie-Mellon University, 1984.
- [164] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence*, pages 674–679, 1981.
- [165] R. Malladi, J.A. Sethian, and B.C. Vemuri. Evolutionary fronts for topology-independent shape modeling and recovery. In J-O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision*, volume 800 of *Lecture Notes in Computer Science*, pages 3–13, Stockholm, Sweden, May 1994. Springer-Verlag.
- [166] R. Malladi, J.A. Sethian, and B.C. Vemuri. Shape modeling with front propagation: A level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2):158–175, February 1995.
- [167] S. Mallat. *A wavelet tour of signal processing*. Academic Press, 1998.
- [168] R. March. Visual reconstructions with discontinuities using variational methods. *Image and Vision Computing*, 10:30–38, 1992.
- [169] D. Marr. *Vision*. W.H. Freeman and Co., 1982.
- [170] D. Marr and E. Hildreth. Theory of edge detection. *Proceedings of the Royal Society London, B*, 207:187–217, 1980.
- [171] S. Masnou. *Filtrage et désocclusion d'images par méthodes d'ensembles de niveau (Image filtering and disocclusion using level sets)*. PhD thesis, CEREMADE, Université Paris-Dauphine, 1998.
- [172] S. Masnou and J.M. Morel. Level lines based disocclusion. *International Conference on Image Processing*, III:259–263, 1998.
- [173] U. Massari and I. Tamanini. Regularity properties of optimal segmentations. *J. Reine Angew. Math.*, 420:61–84, 1991.
- [174] G. Matheron. *Random Sets and Integral Geometry*. John Wiley & Sons, 1975.
- [175] M. Mattavelli and A. Nicoulin. Motion estimation relaxing the constant brightness constraint. *International Conference on Image Processing*, II:770–774, November 1994.
- [176] E. Mémin and P. Pérez. A multigrid approach for hierarchical motion estimation. In *Proceedings of the 6th International Conference on Computer Vision*, pages 933–938. IEEE Computer Society Press, Bombay, India, January 1998.

- [177] E. Mémin and P. Pérez. Optical flow estimation and object-based segmentation with robust techniques. *IEEE Trans. on Image Processing*, 7(5):703–719, May 1998.
- [178] A. Mitiche and P. Bouthemy. Computation and analysis of image motion: a synopsis of current problems and methods. *The International Journal of Computer Vision*, 19(1):29–55, July 1996.
- [179] L. Modica. The gradient theory of phase transitions and the minimal interface criterion. *Archive for Rational Mechanics and Analysis*, 98:123–142, 1987.
- [180] L. Moisan. Analyse multiéchelle de films pour la reconstruction du relief. *Comptes Rendus de l'Académie des Sciences*, 320-I:279–284, 1995.
- [181] L. Moisan. Multiscale analysis of movies for depth recovery. *International Conference on Image Processing*, 3:25–28, 1995.
- [182] L. Moisan. Perspective invariant multiscale analysis of movies. *Proceedings of the International Society for Optical Engineering*, 2567:84–94, 1995.
- [183] J.M. Morel and S. Solimini. Segmentation of images by variational methods: A constructive approach. *Rev. Math. Univ. Complut. Madrid*, 1:169–182, 1988.
- [184] J.M. Morel and S. Solimini. Segmentation d'images par methode variationnelle: Une preuve constructive d'existence. *Comptes Rendus de l'Académie des Sciences*, 308(15):465–470, 1989.
- [185] J.M. Morel and S. Solimini. *Variational methods in image segmentation*. Progress in Nonlinear Differential Equations and Their Applications. Birkhauser, Basel, 1995.
- [186] D. Mumford and J. Shah. Boundary detection by minimizing functionals. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, June 1985. IEEE.
- [187] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42:577–684, 1989.
- [188] H.H. Nagel. Constraints for the estimation of displacement vector fields from image sequences. In *International Joint Conference on Artificial Intelligence*, pages 156–160, 1983.
- [189] H.H. Nagel. Displacement vectors derived from second order intensity variations in image sequences. *Computer Vision, Graphics, and Image Processing*, 21:85–117, 1983.
- [190] S.K. Nayar and S.G. Narasimhan. Vision in bad weather. In *Proceedings of the 6th International Conference on Computer Vision*, pages 820–827, Bombay, India, January 1998. IEEE Computer Society, IEEE Computer Society Press.
- [191] S. Negahdaripour. Revised definition of optical flow: integration of radiometric and geometric cues for dynamic scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(9):961–979, 1998.
- [192] S. Negahdaripour and C.H. Yu. A generalized brightness change model for computing optical flow. *International Conference on Computer Vision*, pages 2–11, 1993.

- [193] P. Nési. Variational approach to optical flow estimation managing discontinuities. *Image and Vision Computing*, 11(7):419–439, September 1993.
- [194] M. Nitzberg and T. Shiota. Nonlinear image filtering with edge and corner enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:826–833, 1992.
- [195] J.A. Noble. *Description of image surfaces*. PhD thesis, Robotics research group, Department of Engineering Science, Oxford University, 1989.
- [196] J.M. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *Journal of Visual Communication and Image Representation*, 6(4):348–365, December 1995.
- [197] M. Orkisz and P. Clarysse. Estimation du flot optique en présence de discontinuités: une revue. *Traitements du Signal*, 13(5):489–513, 1996.
- [198] S. Osher and R.P. Fedkiw. Level set methods. Technical Report 00-08, UCLA CAM Report, 2000.
- [199] S. Osher and L.I. Rudin. Feature-oriented image enhancement using shock filters. *SIAM Journal of Numerical Analysis*, 27(4):919–940, August 1990.
- [200] S. Osher and J. Sethian. Fronts propagating with curvature dependent speed : algorithms based on the Hamilton-Jacobi formulation. *Journal of Computational Physics*, 79:12–49, 1988.
- [201] M. Otte and H.H Nagel. Extraction of line drawings from gray value images by non-local analysis of edge element structures. In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision*, pages 687–695, Santa Margherita, Italy, May 1992. Springer-Verlag.
- [202] M. Otte and H.H. Nagel. Optical flow estimation: Advances and comparisons. In Jan-Olof Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision*, volume 800 of *Lecture Notes in Computer Science*, pages 51–70. Springer-Verlag, 1994.
- [203] N. Paragios and R. Deriche. A PDE-based level set approach for detection and tracking of moving objects. In *Proceedings of the 6th International Conference on Computer Vision*, pages 1139–1145, Bombay, India, January 1998. IEEE Computer Society, IEEE Computer Society Press.
- [204] N. Paragios and R. Deriche. Geodesic Active Regions for Supervised Texture Segmentation. In *Proceedings of the 7th International Conference on Computer Vision*, Kerkyra, Greece, September 1999. IEEE Computer Society, IEEE Computer Society Press.
- [205] N. Paragios and R. Deriche. Geodesic active contours and level sets for the detection and tracking of moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:266–280, March 2000.
- [206] N. Paragios and R. Deriche. Geodesic active regions: a new paradigm to deal with frame partition problems in computer vision. *International Journal of Visual Communication and Image Representation, Special Issue on Partial Differential Equations in Image Processing, Computer Vision and Computer Graphics*, 2001. Accepted for Publication. To appear in 2001.

- [207] T. Pavlidis and Y.T. Liow. Integrating region growing and edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:225–233, 1990.
- [208] D. Peng, B. Merriman, S. Osher, H. Zhao, and M. Kang. A PDE-based fast local level set method. *Journal on Computational Physics*, 155(2):410–438, 1999.
- [209] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, July 1990.
- [210] J.M.S. Prewitt. Object enhancement and extraction. In Lipkin, B.S. and Rosenfeld, A., editor, *Picture Processing and Psychopictorics*, pages 75–149. New York: Academic, 1970.
- [211] L.G. Roberts. Machine perception of three-dimensional solids. In Tippett, J. and Berkowitz, D. and Clapp, L. and Koester, C. and Vanderburgh, A., editor, *Optical and Electrooptical Information processing*, pages 159–197. MIT Press, 1965.
- [212] M. Rosati. Asymptotic behavior of a Geman-Mac Lure discrete model. Technical Report 8, Istituto per le Applicazioni del Calcolo Mauro Picone, 1997.
- [213] A. Rosenfeld and A.C. Kak. *Digital picture processing*, volume 1. Academic Press, New York, 1982. Second Edition.
- [214] L. Rudin and S. Osher. Total variation based image restoration with free local constraints. In *International Conference on Image Processing*, volume I, pages 31–35, November 1994.
- [215] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
- [216] W. Rudin. *Real and complex analysis*. McGraw-Hill, 1966.
- [217] W. Rudin. *Principles of mathematical analysis*. 3rd ed. International Series in Pure and Applied Mathematics. McGraw-Hill, 1976.
- [218] W. Rudin. *Functional analysis*. 2nd ed. International Series in Pure and Applied Mathematics. McGraw-Hill, 1991.
- [219] S.J. Ruuth, B. Merriman, and S. Osher. A fixed grid method for capturing the motion of self-intersecting interfaces and related pdes. Technical Report 99-22, UCLA Computational and Applied Mathematics Reports, July 1999.
- [220] C. Samson. *Contribution à la classification d'images satellitaires par approche variationnelle et équations aux dérivées partielles*. PhD thesis, Université de Nice Sophia-Antipolis, 2000.
- [221] C. Samson, L. Blanc-Féraud, G. Aubert, and J. Zerubia. A level-set model in image classification. *Second International Conference on Scale Space, Lecture Notes in Computer Sciences*, 1682:306–317, 1999.
- [222] C. Samson, L. Blanc-Féraud, G. Aubert, and J. Zérubia. A variational model for image classification and restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(5):460–472, May 2000.
- [223] G. Sapiro. *Geometric partial differential equations and image analysis*. Cambridge University Press, 2001.

- [224] C. Schnörr. Determining optical flow for irregular domains by minimizing quadratic functionals of a certain class. *The International Journal of Computer Vision*, 6(1):25–38, 1991.
- [225] B.G. Schunck. The motion constraint equation for optical flow. *International Conference on Pattern Recognition*, pages 20–22, 1984.
- [226] B.G. Schunck. Image flow continuity equations for motion and density. *Workshop motion: representation and analysis*, pages 89–94, 1986.
- [227] J.A. Sethian. Recent numerical algorithms for hypersurfaces moving with curvature-dependent speed: Hamilton-Jacobi equations and conservation laws. *Journal Differential Geometry*, 31:131–136, 1990.
- [228] J.A. Sethian. A fast marching sevel set method for monotonically advancing fronts. In *Proceedings of the National Academy of Sciences*, volume 93, pages 1591–1694, 1996.
- [229] J.A. Sethian. *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials sciences*. Cambridge Monograph on Applied and Computational Mathematics. Cambridge University Press, 1999.
- [230] J. Shen and S. Castan. An optimal linear operator for step edge detection. *CVGIP: Graphics Models and Image Processing*, 54(2):112–133, March 1992.
- [231] I. Sobel. An isotropic 3x3 image gradient operator. *Machine Vision for Three-Dimensional Scenes*, H. Freeman editor, pages 376–379, 1990.
- [232] N. Sochen, R. Kimmel, and R. Malladi. From high energy physics to low level vision. Technical Report 39243, LBNL report, UC Berkeley, 1996.
- [233] N. Sochen, R. Kimmel, and R. Malladi. A geometrical framework for low level vision. *IEEE Trans. on Image Processing, Special Issue on PDE based Image Processing*, 7(3):310–318, 1998.
- [234] P. Sternberg. The effect of a singular perturbation on nonconvex variational problems. *Archive for Rational Mechanics and Analysis*, 101(3):209–260, 1988.
- [235] P. Sternberg. Vector-valued local minimizers of nonconvex variational problems. *Rocky Mt. J. Math.*, 21(2):799–807, 1991.
- [236] D. Suter. Motion estimation and vector splines. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pages 939–942, Seattle, WA, June 1994. IEEE.
- [237] S. Teboul, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Variational approach for edge-preserving regularization using coupled pde's. *IEEE Transactions on Image Processing*, 7(3):387–397, 1998.
- [238] J.W. Thomas. *Numerical partial differential equations: finite difference methods*. Springer-Verlag, 1995.
- [239] A.N. Tikhonov and V.Y. Arsenin. *Solutions of ill-posed problems*. Winston and Sons, Washington, D.C., 1977.
- [240] M. Tistarelli. Computation of coherent optical flow by using multiple constraints. *International Conference on Computer Vision*, pages 263–268, 1995.

- [241] J.N. Tsitsiklis. Efficient algorithms for globally optimal trajectories. *IEEE Transactions on Automatic Control*, 40(9):1528–1538, September 1995.
- [242] A. Verri, F. Girosi, and V. Torre. Differential techniques for optical flow. *Journal of the Optical Society of America A*, 7:912–922, 1990.
- [243] A. Verri and T. Poggio. Against quantitative optical flow. In *Proceedings First International Conference on Computer Vision*, pages 171–180. IEEE Computer Society, 1987.
- [244] A. Verri and T. Poggio. Motion field and optical flow: qualitative properties. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(5):490–498, 1989.
- [245] L. Vese. *Problèmes variationnels et EDP pour l'analyse d'images et l'évolution de courbes*. PhD thesis, Université de Nice Sophia-Antipolis, November 1996.
- [246] L. Vese. A study in the BV space of a denoising-deblurring variational problem. *Applied Mathematics and Optimization (an International Journal with Applications to Stochastics)*, 2001. (in press).
- [247] A.I. Vol'pert. The spaces BV and quasilinear equations. *Math. USSR-Sbornik*, 2(2):225–267, 1967.
- [248] J. Weickert. Theoretical foundations of anisotropic diffusion in image processing. *Computing Supplement*, 11:221–236, 1996.
- [249] J. Weickert. *Anisotropic diffusion in image processing*. Teubner-Verlag, Stuttgart, 1998.
- [250] J. Weickert. Coherence-enhancing diffusion filtering. *The International Journal of Computer Vision*, 31(2/3):111–127, April 1999.
- [251] J. Weickert. Linear scale space has first been proposed in Japan. *Journal of Mathematical Imaging and Vision*, 10(3):237–252, May 1999.
- [252] R.P. Wildes, M.J. Amabile, A.M. Lanzillotto, and T.S. Leu. Physically based fluid flow recovery from image sequences. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pages 969–975, San Juan, Puerto Rico, June 1997. IEEE Computer Society, IEEE.
- [253] A.P. Witkin. Scale-space filtering. In *International Joint Conference on Artificial Intelligence*, pages 1019–1021, 1983.
- [254] C.R. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfnder: real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):780–785, July 1997.
- [255] K. Yosida. *Functional analysis. 6th ed.*, volume XII of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, 1980.
- [256] H.K. Zhao, T. Chan, B. Merriman, and S. Osher. A variational level set approach to multiphase motion. *Journal of Computational Physics*, 127(0167):179–195, 1996.
- [257] L. Zhou, C. Kambhamettu, and D. Goldgof. Fluid structure and motion analysis from multi-spectrum 2d cloud image sequences. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 744–751, Hilton Head Island, South Carolina, June 2000. IEEE Computer Society.

- [258] S. Zhu and A. Yuille. Region competition: unifying snakes, region growing, and bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):884–900, September 1996.

# Index

- Active contours, 161
- Approximate
  - derivative, 49, 78
  - limit, 50, 209
- Axiom
  - locality, 106
  - recursivity, 106
  - regularity, 106
- Bernstein method, 178
- Bhattacharyya distance, 224
- Camera model, 193
- Characteristics
  - curve, 129
  - equations, 130
  - lines, 250
  - method of, 128, 253
- Classification, 218
  - regular, 220
  - supervised, 219
- Coarea formula, 64, 222, 236
- Coercive, 34, 159, 199
- Compactness
  - weak sequential, 34
- Condition
  - CFL, 252
- entropy, 257, 259
- Rankine-Hugoniot, 256
- Conjecture
  - Mumford-Shah, 141, 145, 153
  - Osher-Rudin conjecture, 134
- Convergence
  - BV-weak\*, 48, 52
  - Gamma-convergence, 42, 78, 154, 228, 234
  - strong, 32
  - weak, 32
  - weak\*, 33
- Crack tip, 141, 150
- Curvature
  - mean curvature motion, 167, 268
  - radius, 59
  - tensor, 59
- Cut-off function, 208
- Derivative
  - approximate, 49, 78
  - Gâteaux, 38, 186, 221
  - length, 147, 166
  - of domain integral, 147
  - Radon Nikodym derivative, 45
- Diffusion
  - nonlinear, 96

- operator, 73
- tensor, 113
- Direct method of the calculus of variations, 34, 142, 199, 207
- Discrete image, 3, 126, 156, 259
- Domain of dependence, 252
- Dual variable, 78, 80, 200, 210, 231, 263
- Edge, 120, 168
- Elliptic
  - approximation by elliptic functionals, 154
  - degenerate, 54
  - equation, 73
- Equation
  - backward heat, 121
  - Burgers, 253
  - eikonal, 52
  - Euler-Lagrange equation, 38, 71, 72, 78, 80, 86, 117, 162
  - Hamilton Jacobi equations, 54, 172
  - heat, 93, 259
  - mean curvature motion, 167, 268
  - transport, 250
- Equicontinuous
  - uniformly, 65
- Eulerian formulation, 170, 172
- Fast marching algorithm, 270
- Filter
  - Gaussian, 93
  - high-pass, 120
  - low-pass, 94, 120
  - Osher and Rudin's shock-filters, 127, 133
- Finite
  - differences, 238
    - backward, 251
    - centered, 251, 267
    - derivatives, 240, 261
    - forward, 251
  - elements, 238
- Fixed-point
  - Schauder's fixed point Theorem, 125
- Formula
  - chain rule, 209, 210
  - coarea, 64, 222, 236
- integration by part, 63
- Functional
  - coercive, 34, 159, 199
  - convex, 35
  - equicoercive, 42
  - lower semi-continuous, 34, 52, 75, 207
  - Mumford-Shah, 141
  - nonconvex, 88, 207
  - relaxed, 39, 75, 101, 201
- Gamma-convergence, 42, 78, 154, 228, 234
- Gaussian kernel, 94
- Generalized solution, 126
- Half-quadratic minimization, 78, 80, 200, 210, 231, 263
- Hamiltonian, 54
  - degenerate elliptic, 54
- Ill-posed, 66, 70, 88, 122, 133
- Inequality
  - Cauchy, 61
  - Cauchy-Schwarz, 61, 166, 179
  - Gronwall, 61, 124, 181
  - Jensen, 62
  - Minkowski, 62
  - Poincaré, 62
  - Poincaré-Wirtinger, 62, 76, 88
  - Young, 63
- Invariance, 85
  - gray-level, 86, 96, 106
  - gray-scale, 111
  - isometry, 86, 96, 110
  - projection, 111
  - scale, 96, 111
  - translation, 86, 96, 106
- Isophotes, 72
- Lebesgue
  - Lebesgue decomposition Theorem, 46
  - point, 50
- Level sets method, 170, 220, 268
- Limit
  - approximate lower/upper limit, 50
- Lower semi-continuous
  - functional, 34, 52, 75, 207

- Measure  
 absolutely continuous, 46, 200  
 bounded, 45  
 convex function of, 51, 206  
 Hausdorff, 49  
 mutually singular, 46  
 positive, 45  
 Radon, 45  
 signed, 45  
 vector-valued, 45
- Minimizing sequence, 34
- Minmod function, 267
- Model  
 Alvarez-Guichard-Lions-Morel  
 scale space theory, 105  
 Caselles-Kimmel-Sapiro, 164  
 Catté et al, 122  
 degradation, 68  
 Horn and Schunck, 195  
 Kass-Witkin-Terzopoulos, 161  
 Nitzberg-Shiota, 126  
 Osher-Rudin, 127  
 Perona-Malik, 97, 120, 121, 264  
 smoothing-enhancing, 98  
 variational model for restoration,  
     72, 263  
 Weickert, 112
- Mollifier, 64
- Motion  
 2D motion field, 192  
 apparent, 192
- Movies, 111
- Narrow band, 223, 272
- Negligible set, 45
- Operator  
 divergence, 263  
 domain, 99  
 maximal monotone, 99–101  
 minmod, 267  
 monotone, 99  
 morphological, 111  
 multivalued, 99  
 range, 99  
 surjective, 99
- Optical flow, 192, 203, 263  
 constraint (OFC), 194, 201
- Partition condition, 220
- Partitioning process, 219
- PDEs  
 backward parabolic, 92, 98  
 enhancing PDEs, 127  
 forward parabolic, 92, 98  
 hyperbolic, 92, 250  
 smoothing enhancing PDEs, 120  
 smoothing PDEs, 93, 94  
 smoothing-enhancing, 98  
 system of coupled, 223
- Perimeter, 48, 228  
 minimal, 219
- Phases transitions, 227
- Polar function, 40, 156
- Precise representation, 51, 201
- Principle  
 comparison, 106  
 extremum, 113  
 maximum, 58, 95, 180
- Process  
 dynamical, 224  
 restoration, 224
- Regularity  
 Bonnet results, 153  
 heat equation, 96  
 of the edge set, 150
- Regularization  
 edge preserving, 73, 88  
 of the Perona and Malik model, 122  
 Tikhonov-Arsenin regularization,  
     71
- Relaxed functional, 39, 75, 101, 201
- Scale  
 Alvarez-Guichard-Lions-Morel  
 scale space theory, 105  
 parameter, 94  
 space, 88  
 variable, 92
- Scheme  
 conservative, 257, 259  
 consistent, 244, 258, 259  
 convergent, 242  
 Crank-Nicholson, 249  
 explicit, 240  
 implicit, 249  
 monotone, 259

- order of accuracy, 244
- order of convergence, 242
- pointwise consistent, 243
- pointwise convergent, 241
- stable, 245, 251
  - (un)conditionally, 246
- symbol, 248
- truncature error, 244
- upwind, 252
- Seed initialization, 224
- Semi-group, 96, 99
- Sequence, 190
  - restoration, 212
  - sampling, 190
  - segmentation, 203, 263
- Snakes, 162
- Space
  - Banach, 32
  - of bounded variation, 47, 75, 143, 198, 207
  - of special functions of bounded variation, 144
  - reflexive, 33, 75
  - separable, 33
- Sub-jet, 57
- Subdifferential, 100
  - of a convex function, 100
- Super-jet, 56
- Theorem
  - Arzelà-Ascoli, 65, 180
  - Dominated convergence, 65
  - Fubini, 96
  - Gauss-Green, 63
  - Green, 63, 87, 102, 148
  - Lax-Wendroff, 258
  - Lebesgue decomposition Theorem, 46
  - Schauder, 125
- Topology
  - BV-weak\*, 48, 52
  - strong, 32
  - weak, 32
  - weak\*, 33
- Total variation, 27, 45, 51, 71, 78, 118, 134, 196
- Triple junction, 141, 150
- Vector distance function, 187