Downloaded from orbit.dtu.dk on: Dec 18, 2017

Technical University of Denmark

Image reconstruction under non-Gaussian noise

Sciacchitano, Federica; Dong, Yiqiu; Hansen, Per Christian

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Sciacchitano, F., Dong, Y., & Hansen, P. C. (2017). Image reconstruction under non-Gaussian noise. Kgs. Lyngby: Technical University of Denmark (DTU). (DTU Compute PHD-2016; No. 426).

DTU Library
Technical Information Center of Denmark

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Ph.D. Thesis

DTU Compute
Department of Applied Mathematics and Computer Science

Image reconstruction under non-Gaussian noise

Federica Sciacchitano

Kgs. Lyngby, 2016

DTU Compute
Department of Applied Mathematics and Computer Science
Technical University of Denmark

Matematiktorvet
Building 303B
2800 Kongens Lyngby, Denmark
Phone +45 4525 3031
compute@compute.dtu.dk
www.compute.dtu.dk

PHD-2016-426
ISSN: 0909-3192

To my beloved grandfather Checco, who is highly missed

Preface

This thesis was prepared in partial fulfilment of the requirements for acquiring the PhD degree at the Technical University of Denmark (DTU). The work was carried out from September 2013 to August 2016 at the Scientific Computing section of the Department of Applied Mathematics and Computer Science (DTU Compute) under the supervision of Associate Professor Yiqiu Dong and Professor Per Christian Hansen.

From August 2014 to December 2014, the project was carried out at Department of Mathematics, Hong Kong Baptist University. This visit was hosted by Associate Professor Tieyong Zeng, and partially sponsored by Hong Kong Baptist University and Otto Mønsted foundation for which I am grateful.

The scholarship for this PhD project was funded by the Department of Applied Mathematics and Computer Science of the Technical University of Denmark.

Kongens Lyngby, August 31, 2016

Federica Sciacchitano

Abstract

During acquisition and transmission, images are often blurred and corrupted by noise. One of the fundamental tasks of image processing is to reconstruct the clean image from a degraded version. The process of recovering the original image from the data is an example of inverse problem. Due to the ill-posedness of the problem, the simple inversion of the degradation model does not give any good reconstructions. Therefore, to deal with the ill-posedness it is necessary to use some prior information on the solution or the model and the Bayesian approach.

Additive Gaussian noise has been extensively studied since it produces simple and tractable mathematical models. However, in the real applications, the noise is much more complicated and it cannot be well simulated by additive Gaussian noise, for instance, it may be signal dependent, very impulsive, multiplicative, mixed, etc. This PhD thesis intends to solve some of the many open questions for image restoration under non-Gaussian noise. The two main kinds of noise studied in this PhD project are the impulse noise and the Cauchy noise.

Impulse noise is due to for instance the malfunctioning pixel elements in the camera sensors, errors in analogue-to-digital conversion, faulty memory locations in hardware. Cauchy noise is characterized by a very impulsive behaviour and it is mainly used to simulate atmospheric and underwater acoustic noise, in radar and sonar applications, biomedical images and synthetic aperture radar images. For both noise models we introduce new variational models to recover the clean and sharp images from degraded images. Both methods are verified by using some simulated test problems. The experiments clearly show that the new methods outperform the former ones.

Furthermore, we have carried out a theoretical study on the two most known estimates: maximum a posteriori (MAP) estimate and conditional mean (CM) estimate for non-Gaussian noise. With only the convexity assumption on the data fidelity term, we introduce some cost functions for which the CM and MAP estimates are proper Bayes estimators and we also prove that the CM estimate outperforms the MAP estimate, when the error depends on Bregman distances.

This PhD project can have many applications in the modern society, in fact the reconstruction of high quality images with less noise and more details enhances the image processing operations, such as edge detection, segmentation, etc.

Resumé

Under transmission og dannelse af billeder opstår der ofte slørings- og støjforstyrrelser. En af de fundamentale opgaver indenfor billedbehandling er at rekonstruere det rene billede ud fra det forringede. Gendannelsen af det originale billede ud fra data er et eksempel på et inverst problem. Da problemet er "dårligt formuleret" (ill-posed) vil en simple inversion af degraderingsmodellen ikke give en god rekonstruktion. For at håndtere det dårligt formulerede problem er det nødvendigt at anvende forudgående (prior) information om løsningen eller modellen, og en Bayesiansk tilgang.

Additiv Gaussisk støj er blevet udførligt undersøgt da det producerer simple og medgørlige matematiske modeller. Men i forbindelse med faktisk anvendelse er støjen langt mere kompliceret og den kan ikke approksimeres ved hjælp af Gaussisk støj, for eksempel, kan den være signalafhængig, meget impulsiv, multiplikativ, en sammenblanding, osv. Intentionen med denne Ph.d.-afhandling er at besvare nogle af de mange åbne spørgsmål indenfor digital billed-restaurering i tilfælde af ikke-Gausisk støj. De to støjtyper der er omdrejningspunktet for dette ph.d.-projekt er impuls støj og Cauchy støj.

Impuls støj opstår blandt andet på grund af defekte pixelelementer i kamera-sensoren, fejl i analog-til-digital konvertering eller på grund af fejlbehæftet hukommelseslagring i hardwaren. Cauchy støj er karakteriseret ved at have en meget impulsiv opførsel og det bliver primært brugt til at simulere atmosfærisk- og akustisk undervands-støj, i radar og sonar applikationer, biomedicinske billeder og for syntetisk apertur radar billeder. For begge støj-modeller introducerer vi nye variationelle modeller for genvinde de rene og skarpe billeder ud fra de forringede. Vi bekræfter virkningen af begge metoder via simulerede testproblemer.

Derudover, har vi udført et teoretisk studie af de to bedst kendte estimater: maximum a posteriori (MAP) estimatet og betinget gennemsnit ("conditional mean" (CM)) estimatet for ikke-Gausisk støj. Vi har, kun baseret på en antagelse om konveksitet af data-overensstemmelsesledet, introduceret en udgiftsfunktion for hvilken CM og MAP estimaterne er egentlige ("proper") Bays etimatorer, og vi beviser også at CM estimatet udkonkurrerer MAP estimatet, når fejlen afhænger af Bregman afstande.

Dette ph.d.-projekt kan have mange forskellige anvendelsesmuligheder i det moderne samfund, faktisk leder rekonstruktion af højkvalitetsbilleder med mindre støj og flere detaljer til forbedrede billedbehandlings-operationer, som for eksempel kant-detektion, segmentering, osv.

Acknowledgments

I would like to express my gratitude to the people who contributed to this journey. First and foremost, I would like to thank my supervisors Yiqiu Dong and Per Christian Hansen for their invaluable guidance and support throughout this PhD project and for giving me the opportunity of working on this challenging and interesting topic.

A special thanks goes to all the members of the HD-Tomo and Hybrid Tomo projects; in particular, I would like to thank Martin S. Andersen, Jürgen Frikel, Henrik Garde, Jakob S. Jørgensen, Kim Knudsen, Rasmus D. Kongskov, Mikhail Romanov and Sara Soltani for interesting discussions and comments along the way. It has been a great honour being part of this group.

Furthermore, I would like to thank all PhDs and PostDocs of Scientific Computing section at DTU Compute. Most probably I will not miss the canteen, but I will definitely miss you; thank you so much for nice conversations, tasty cakes and great times. I'm also grateful to Mette E. Larsen, Dorthe Thøgersen, and the IT support at DTU Compute for their constant help.

During my PhD project, I visited several universities. I would like to thank Tieyong Zeng from Hong Kong Baptist University for being a great host during my external stay in Hong Kong. I also would like to thank the MIDA group, University of Genoa, and Fiorella Sgallari, University of Bologna, for their hospitality and useful discussions during my short external stays. I also would like to thank Martin Burger, Luca Calatroni and Adrian Martin for the stimulating collaborations.

I am grateful to Otto Mønsted Foundation for partially founding my external stay and trips to conferences.

Finally a huge and loving thanks goes to my family. My parents, Maria Luisa and Salvatore, for being always present in my life and for all the sacrifices they made on my behalf; my sister, Roberta, for her support and for cheering me up; my grandparents, Checco (to whom this thesis is dedicated) and Rosetta, for their love and for providing very good food during my stays in Italy; Luigi for being by my side and for making me laugh as none does.

List of symbols

Abbreviations

ACWM	Adaptive center-weighted median filter
AM	Adaptive median filter
CM	Conditional mean estimate
CT	Computed tomography
MAP	Maximum a posteriori estimate
MCMC	Markov chain Monte Carlo techniques
MRI	Magnetic resonance imaging
PDF	Probability density function
PET	Positron emission tomography
PSF	Point spread function
PSNR	Peak signal-to-noise ratio
ROF	Rudin-Osher-Fatemi model
SSIM	Structural similarity index
TV	Total variation regularization

Symbols

div	Divergence operator
$\hat{u}_C$	Bayes estimator
λ	Regularization parameter
$\langle \cdot \rangle$	Scalar product
$\|\cdot\|_p$	ℓ_p-norm
$\mathbb{E}$	Expected value
$\mathrm{TV}(u)$	Total variation regularization term
∇	Gradient operator
∇_x	Horizontal first order differences
∇_y	Vertical first order differences
Ω	Image domain
$\tilde{u}$	Original image
$\hat{u}_{\mathrm{CM}}$	Conditional mean (CM) estimate

$\hat{u}_{\text{MAP}}$	Maximum a posteriori (MAP) estimate
$BC_C(\hat{u})$	Bayes cost
$C(u, \hat{u})$	Cost function
$E(u; K, f)$	Data fidelity term
f	Degraded image
$J(u)$	Regularization term
K	Blurring operator
m	Number of pixels of the image
m_1	Number of rows of the image
m_2	Number of columns of the image
N	Number of samples
$u_{\max}$	Maximum intensity of the image
$u_{\min}$	Minimum intensity of the image
v	Random noise

List of figures

2.1 Additive noise models 8
2.2 Multiplicative noise, signal dependent and impulse noise 10
2.3 Different blur operators 11
2.4 Noisy image formation 12

3.1 Example of inverse problem 16
3.2 Influence of the regularization parameter 20

4.1 Sampling process 24
4.2 Quantization process 24
4.3 Gray image 25
4.4 Samples in Metropolis-Hastings algorithm 33
4.5 Acceptance ratio in Metropolis-Hastings algorithm 33

5.1 Salt-and-pepper noise and random-valued impulse noise 36
5.2 Filters for impulse noise removal 37
5.3 Variational methods for impulse noise removal 38
5.4 Our two-phase method in inpainting 39
5.5 Our two-phase method for super-resolution 40
5.6 Comparison alpha-stable PDFs 42
5.7 1D alpha-stable 42
5.8 2D alpha-stable 43
5.9 Differences between Gaussian, impulse and Cauchy noise 43

Contents

Preface iii

Abstract v

Resumé vii

Acknowledgments ix

List of symbols xi

List of figures xiii

1 Introduction and motivation 1
- 1.1 Contributions of the thesis . . . 2
- 1.2 Thesis structure . . . 3

2 Set the scene 5
- 2.1 What is an image? . . . 5
- 2.2 Noise . . . 6
  - 2.2.1 Additive noise . . . 6
  - 2.2.2 Multiplicative noise . . . 8
  - 2.2.3 Signal dependent noise . . . 8
  - 2.2.4 Impulse noise . . . 9
  - 2.2.5 Mixed noise . . . 9
- 2.3 Blur . . . 10
  - 2.3.1 Blurring models . . . 10
- 2.4 Quality of an image . . . 12

3 Inverse problems 15
- 3.1 Inverse problem . . . 15
- 3.2 The Bayesian approach . . . 17
  - 3.2.1 The data-fidelity term . . . 17
  - 3.2.2 The regularization term . . . 18
  - 3.2.3 The regularization parameter . . . 19
  - 3.2.4 MAP and CM estimates . . . 21

4 Numerical considerations **23**
4.1 What is a digital image? . . . 23
4.1.1 Sampling . . . 23
4.1.2 Quantization . . . 24
4.1.3 Notations in the discrete case . . . 25
4.2 Optimization . . . 26
4.2.1 Primal-dual algorithm . . . 28
4.2.2 Semismooth Newton's method . . . 30
4.3 Monte Carlo method-solving the integration . . . 31
4.3.1 Metropolis-Hastings sampling . . . 31
4.3.2 Gibbs sampling . . . 32

5 Contributions **35**
5.1 Impulse noise . . . 35
5.1.1 Overview . . . 35
5.1.2 Relevant paper . . . 38
5.1.3 Perspective and some future works . . . 39
5.2 Cauchy noise . . . 40
5.2.1 Overview . . . 41
5.2.2 Relevant paper . . . 44
5.2.3 Perspective and some future works . . . 44
5.3 Bregman cost for non-Gaussian noise . . . 44
5.3.1 Overview . . . 44
5.3.2 Relevant paper . . . 47
5.3.3 Perspective and some future works . . . 47

6 Summary/Conclusion **49**

Bibliography **51**

Journal Papers **59**

A Total variation based parameter-free model for impulse noise removal **61**

B Variational Approach for Restoring Blurred Images with Cauchy Noise **83**

C Bregman Cost for Non-Gaussian Noise **115**

CHAPTER 1

Introduction and motivation

In the last few decades digital images have reached a high importance in the world, they can be used for example, for medical purpose (such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET)), astronomic purpose, or simply in our daily life in our personal laptop and cellphone. Due to the popularity of digital cameras and smartphones, everyone knows how hard is to take a "'good" picture, in fact the picture could not be as clean as we would like mainly for two reasons: blur and noise. The presence of blur makes an image not sharp due to some motion defects or some focusing degradation. The noise does not allow to see all the details of the scene, it may be due to some issues during the acquisition and transmission for instance. Consequently, there is a huge demand for developing the imaging sciences.

Imaging sciences consist of three relatively independent components: image acquisition, image processing, and image interpretation. The first one studies the formation of the images by some physical mechanisms. The second discipline deals with the processing of images using mathematical operations, i.e. given an acquired image by using some mathematical model returns another image. The last one interprets the 3D world from images. Here, we just focus on the image processing. We underline that image processing is strongly connected with image acquisition, since the knowledge of an image acquisition process can be helpful in developing a good image processing model and algorithm.

Image processing is a very huge field that includes many disciplines, for instance: image denoising, image deblurring, image segmentation and image inpaiting. The first two deal with the restoration of the images, trying to recover the clean image from the degraded data. Image segmentation finds a visually meaningful partitioning of the image domain, see [63]. This process allows to simplify the acquired image into something that is more meaningful and easier to analyse, for instance for the biomedical images it can separate the bones from the soft tissues. Image inpainting [4] is the process of recover some hidden and lost parts of an image, for instance it may be the restoration of ancient paintings or the zooming and superresolution. In this thesis, we focus on the first two disciplines: image denoising and deblurring. These processes have a very important role in the imaging science, since the acquired images are always degrades due to either poor imaging conditions or problems during acquisition, storage and communication. Thus, image denoising and deblurring aim to recover the uncorrupted and sharp image knowing the blurred and noisy acquired image. The key idea is that some information of the clean and sharp image is hidden in the degraded image and it can be recovered if we know the

noise model and the blurring process. For this reason, sometimes image acquisition is combined with image processing, since the information of the formation of the data can be useful for getting a better reconstruction.

Unfortunately there is no hope to reconstruct the clean image exactly. In fact, during the the acquisition process there are some unavoidable errors, for instance the representation of the image with a limited number of digits and the noise puts a limit on the size of the details. Therefore, one of the main challenges of image reconstruction is the study a mathematical model that relates the degraded image to the unknown original image and to find some efficient algorithms that are able to recover the original image with much information and details as possible from the recorded image. According to [7], noise is an unwanted component of the image that inevitably occurs during the image acquisition and transmission. Over the years, additive white Gaussian noise has been extensively studied since it produces simple and tractable mathematical models. However, in many real applications, the noise is much more complicated, and cannot be well simulated by additive white Gaussian noise, since it can be signal dependent, multiplicative, impulsive, or even mixed. Thus, due to the complication and variety of the noise there are still a lot of open questions for image reconstruction under non-Gaussian noise. The goal of this thesis is the study of noise models and try to recover the ideal clean image. We will mainly focus on the denoising of images corrupted by non-Gaussian noise and for each of our projects we will try to extend our work also to the deblurring case. In some cases, the generalization to the combined deblurring and denoising case is quite straightforward, while some other times it can be more complicated. Later we will refer to image reconstruction when we will work on the image denoising and deblurring.

One of the main challenges of image reconstruction is to seek a good mathematical model that is able to denoise and deblur the image in order to get it as close as possible to the original one. Unfortunately, the way of solving the model is not straightforward, in fact the simply inversion of the blurring operator does not lead to any good result. This is due to the ill-posedness of the inverse problem that we need to solve.

Therefore, the use of the Bayesian inference is necessary. The Bayesian approach combines the known information about the image that we aim to recover to reconstruct it. In particular, by using some statistical properties, we can formulate the mathematical model for reconstructing the original image as a sum of two terms: one that is related with the noise model and the other one that describes some prior information about the expected behaviour of the original image. The tradeoff between these two terms, i.e. the good fit to the measurements and the smoothness of the solution is controlled by a parameter called the regularization parameter. Since the aim of the thesis is to study non-Gaussian noise, we will mainly focus on the term that controls the fit with the data.

1.1 Contributions of the thesis

This thesis focuses on the development of image reconstruction models and algorithms under non-Gaussian noise. Our main contributions regard three works:

- **Impulse noise**. This project deals with the restoration of blurred image corrupted by impulse noise. Impulse noise is characterized by having either uncorrupted pixels or noisy pixels which do not have any relation with the clean ones. By using the above property of this noise model, we introduce a new two-phase method. This method has the advantage of being parameter free. Furthermore,

in the denoising case we can prove that the proposed method improves the computational efficiency.

Relevant paper:

[A] F. Sciacchitano, Y. Dong, and M. S. Andersen, *Total variation based parameter-free model for impulse noise removal*, Numerical Mathematics: Theory, Methods and Applications; to appear.

- **Cauchy noise**. This project focuses on the restoration of blurred image corrupted by Cauchy noise. Cauchy noise together with Gaussian noise belongs to the alpha-stable distributions, but Cauchy noise is much more impulsive than the Gaussian noise. Here, based on the total variation (TV) we introduce a variational model for recovering images degraded by Cauchy noise and blurring. Due to the non-convexity of the variational model, we proposed to restore the image by using a convexification. Experimental results show the effectiveness of the proposed method for deblurring and denoising simultaneously images corrupted by Cauchy noise.

  Relevant paper:

  [B] F. Sciacchitano, Y. Dong, and T. Zeng, *Variational approach for restoring blurred images with cauchy noise*, SIAM Journal on Imaging Sciences 8.3 (2015): 1894-1922.

- **CM and MAP estimates**. This project provides a theoretical study of the maximum a posteriori (MAP) estimate and the conditional mean (CM) estimate for the non-Gaussian noise. In particular, we introduce some Bregman cost functions for which the CM and the MAP estimates are proper Bayes estimator. Moreover, we also prove that the CM estimate outperforms the MAP estimate, when the error depends on some Bregman distances.

  Relevant paper:

  [C] M. Burger, Y. Dong, and F. Sciacchitano, *Bregman Cost for Non-Gaussian Noise*, to be submitted. Available on arXiv http://arxiv.org/abs/1608.07483

1.2 Thesis structure

This PhD thesis is based on a collection of scientific articles published or to be submitted in peer-reviewed scientific journals. The rest of the thesis is organized as follows.

We first establish backgrounds and fundamentals of the image processing field and describe basic definitions in Chapters 2. A comparison between different noise models is provided as well. Chapter 3 provides the reader some background knowledge of inverse problems and image reconstruction. We particularly focus on the maximum a posteriori (MAP) estimate and the conditional mean (CM) estimate. Chapter 4 deals with the some numerical considerations, i.e. we introduce some algorithms that can be used for computing the MAP and the CM estimates. Chapter 5 is dedicated to the main contributions of this thesis. For each of our works, we give a broad introduction of the topic and we include some perspective. Finally, the conclusions and some open questions are given in Chapter 6.

Three papers are included in the attached appendices. It is recommended that enclosed papers are read before proceeding to the conclusions chapter.

CHAPTER 2

Set the scene

In this chapter we give an overview of the image processing field. Image processing includes many applications, for instance image deblurring, image denoising, image contrast enhancement, image inpainting, image segmentation and so on... In this thesis, we mainly focus on the first two topics, i.e. image deblurring and denoising, which deal with the reconstruction of the clean image from the degraded data. These disciplines are very important since the acquired image is inevitably degraded due to acquisition and transmission errors. Here, we start by giving a general introduction of the image, then we analyse the way how it can be corrupted, i.e. by noise and blur.

2.1 What is an image?

Nowadays images are omnipresent in our society, they are widely used in medicine, biology, astronomy, movies, forensic and video surveillance, 3-D reconstructions, robotics, games and special effects, and many others. The main reason of such popularity is because the images are a very powerful medium of communication and through them we can easily represent the physical world.

Mathematically speaking, we consider the following mapping

$$u : \Theta^d \to \mathbb{R}^s,$$

where $\Theta \subset \mathbb{R}$ is the domain acquisition, $d \geq 1$ and $s \geq 1$ is the dimensionality of the input and output space, respectively. When $d = 1$ and $s = 1$, u is a signal, when $d = 2$ and $s = 1$ it is a gray scale image, while if $d = 2$ and $s > 1$ color image, for instance if $d = 2$ and $s = 3$ is a color image with three channels (R for red, G green, B for blue). In this thesis, we only focus on the gray image, but since color image can be seen as the union of different gray images, most of our work can likely be extended also to the color image.

The images are analogue, which means that they are defined on a continuous domain and they take values that come from a continuum of possibilities. However, in order to turn images into a computer-readable digital format, we have to discretize them. The discretization process will be describe in Chapter 4, while in the first chapters we just consider analogue images.

Images are inevitable deteriorated during the formation, transmission and recording processes. The degradation is often due to two phenomena: blur and noise, which can be

caused by some possible defects of the imaging system during the image acquisition, for instance the image may look blurry due to the motion of the system or a miscalibrated lens in the system or the image may look noisy due to some dust in the lens. But they can also be random, for instance due to some issues during the image transmission or some motion defects due to the movement while the picture was taken.

Mathematically speaking, the forward model to acquire an image is described by

$$f = K\tilde{u} \odot v,$$

where f the degraded image defined on the acquisition domain $\Omega \subset \mathbb{R}^2$, K is the known linear and continuous blurring operator, $\tilde{u}$ is the desired clean image, v represents the noise component and $\odot$ indicates a mathematical operation that may be a sum or a multiplication or even a more complicated operation. Informally, the forward problem define how a sharp and clean image would look like if the camera was incorrectly focused and no noise occurred. In the next sections, we will first focus on the noise model, i.e. on the noise v and the operation $\odot$, then we will describe the blurring operator.

2.2 Noise

With the word noise we refer to any unwanted signal, variation of brightness or color information in images. Over the years additive Gaussian noise has very often been assumed due to its simplicity. Unfortunately, in real applications, the noise is much more complicated than just an additive Gaussian noise, for instance it can be signal dependent [38, 39, 48, 58], multiplicative [2, 35, 74, 82], more impulsive than the Gaussian noise [67, 78, 90], or even mixed [16, 57, 59, 88]. In order to get good restored images, it is very important to choose the noise model as close as possible to reality, thus the main aim of this section is to give an overview of the most well-known noise models with their main applications. In Chapter 3 we will show how to recover the noisy image according to the noise model. In this thesis, we have extensively studied two non-Gaussian noise models: impulse noise and Cauchy noise. More details about these two degradations will be given in Chapter 5.

2.2.1 Additive noise

We start with the most popular noise model: the additive noise. Given $\Omega \subset \mathbb{R}^2$ the acquisition domain, f the degraded image defined on Ω, in the additive case, we can decompose the corrupted image into the desired clean image $\tilde{u}$ and noise component v, i.e.

$$f = \tilde{u} + v.$$

We assume v to be white that means that every realization is independent from the others, which in other words, corresponds to say that it is spatially uncorrelated, i.e. the noise for each pixel is independent, [24, Def. 4.1]. Some examples of additive noise are Gaussian noise, uniform noise, Laplacian noise and Cauchy noise. Additive Gaussian noise has been extensively studied since it produces simple and tractable mathematical models, see [19, 24, 36, 73, 85]. In this case, the noise v follows the Gaussian distribution with zero mean and standard deviation $\sigma > 0$, thus, the probability distribution function

(also known as PDF) is defined as follows

$$\mathbb{P}(v(x) = \eta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\eta^2}{2\sigma^2}\right),\ x \in \Omega.$$

The Gaussian distribution is strongly related with the central limit theorem, which says that the arithmetic mean of a large number of independent and identically distributed random variables each with a finite expected value and variance is approximately Gaussian distributed, for more details the reader is referred to [37]. Although, in many applications the assumptions of the central limit theorem are almost, but not quite, true, [7], for instance the number of variables may not be large enough, or the terms may not be sufficiently independent. In order to get the best reconstruction, i.e. remove or diminish the noise, we have to choose the degradation model that is close enough to the real one. For this reason, the focus of this work is on the non-Gaussian models. We now introduce other additive noise models.

A less studied white additive noise is the uniform noise [30], which appears as a statistical model of quantization errors, thus is common during the digital acquisition [80, 86]. Here, the noise v follows the uniform distribution, which has the following probability density function

$$\mathbb{P}(v(x) = \eta) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \leq \eta \leq b \\ 0 & \text{otherwise,} \end{cases},\ x \in \Omega$$

with $a < b$. The additive noise models can be also more impulsive than the Gaussian noise: Laplacian noise and Cauchy are some examples of additive impulsive noise. These noise models are characterized by heavy tails, which means that for large values the density approaches 0 slower than the Gaussian distribution. The Laplace distribution is defined as follows,

$$\mathbb{P}\big(v(x) = \eta\big) = \frac{1}{2b} \exp\left(-\frac{|\eta|}{b}\right),\ x \in \Omega,$$

where $b > 0$ is the scale parameter and the variance is given by $2b^2$. Due to the heavy tailed behaviour Laplace noise usually models the impulse noise, see §2.2.4.

Cauchy noise is a degradation more impulsive than the Laplacian noise and it appears in atmospheric and underwater acoustic noises, radar and sonar applications, air turbulence, wireless communication system, biomedical images, synthetic aperture radar (SAR) images, see [B] and references therein. Cauchy noise together with the Gaussian noise belongs to the alpha-stable distribution, which is a family of distributions characterized by a bell-shaped distribution. The Cauchy distribution has the probability density function

$$\mathbb{P}(v(x) = \eta) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + (\eta - \delta)^2}, \tag{2.1}$$

where $\gamma > 0$ is the scale parameter and $\delta \in \mathbb{R}$ is called the localization parameter. The scale parameter determines the spread of the distribution around δ and plays a role similar to that of the variance in the Gaussian distribution. For a more detailed description of Cauchy noise, we refer the reader to Chapter 5.

In Fig. 2.1, we show an example of images corrupted by some additive noise models. As you can see Laplacian noise has a very impulsive behaviour compared with the Gaussian. Furthermore, a common characteristic of the additive noise is that the noise v is independent of the original image $\tilde{u}$.

Figure 2.1: From the left to the right: uncorrupted image, degraded image by Gaussian noise and by Laplacian noise. The standard deviation is 0.1 for both noisy images.

2.2.2 Multiplicative noise

Multiplicative noise is another kind of degradation where the noisy image is obtained by the multiplication between the noise v and the original image $\tilde{u}$. In the applications, it commonly appears in synthetic aperture radar (SAR), ultrasound imaging, microscope images, laser images, see for instance [2, 6, 35]. The degradation model for the multiplicative noise is given by

$$f = \tilde{u}v.$$

A very common assumption is that v follows Gamma distribution with mean 1, i.e.

$$\mathbb{P}(v(x) = \eta) = \frac{1}{\theta^K \Gamma(K)} \tilde{u}^{(k-1)} \exp\left(\frac{\tilde{u}}{\theta}\right), \ x \in \Omega,$$

with Γ as the Gamma-function, θ and K as the scale and the shape parameters respectively. Note that, since the mean is 1, we have $\theta K = 1$. In Fig. 2.2, we show an example of noisy image obtained by multiplicative noise. The main difference between the additive and multiplicative noise models is that in the multiplicative case the noisy image is darker than the original one. This is due to the fact that the noise v belongs to $[0, 1]$, thus when we multiply the original image by the noise, we reduce the intensity of the pixel values, therefore the noisy image is darker. While in the additive case, the noisy image is given by the sum of the original image and the zero-mean noise, thus the noisy image does not look obviously darker or lighter.

2.2.3 Signal dependent noise

There is also important situations in which neither the additive nor the multiplicative model fits the noise well: the signal dependent noise. Here, contrary to the additive and multiplicative case, the noise strongly depends on the original image. An example of signal dependent noise is the Poisson noise and the Rician noise. Poisson noise occurs whenever the device for image acquisition is a photon counter, for instance, Computed Tomography (CT), Positron Emission Tomography (PET), radiography, and so on, see [7, 38, 39, 58]. For each pixel, the observed image $f(x)$ is a realization of a Poisson

random value with the expected value $\tilde{u}(x)$, i.e.

$$\mathbb{P}(f(x) = \eta) = \frac{\tilde{u}(x)^{\eta} \exp(-\tilde{u}(x))}{\eta}, \ x \in \Omega.$$

The other very popular signal dependent noise model is the Rician noise which typically appears in Magnetic Resonance Imaging (MRI), [48, 81]. Here, the noisy image f is modelled as follows

$$f = \sqrt{(\tilde{u} + \eta_1)^2 + \eta_2^2},$$

with $\eta_1, \eta_2 \sim \mathcal{N}(0, \sigma^2)$ and $\sigma > 0$.

Figure 2.2 shows an image corrupted by Poisson noise. The noisy image is constructed by taking each pixel value in the original image $\tilde{u}$ and generating a Poisson random variable with parameter equals to the pixels value. As you can see, the white areas are noisier than the dark areas. Thus, the noise level increases with the intensity of the image while in the additive case the noise has almost the same power. Note that for a sufficiently large number of photon counts, the Poisson distribution can be approximated by Gaussian distribution with expected value and variance that depends on the intensity of the original image.

2.2.4 Impulse noise

Impulse noise is a common type of image degradation, which does not belong to any of the previous classes. Impulse noise is due to, e.g., malfunctioning pixel elements in the camera sensors, errors in analog-to-digital conversion, faulty memory locations in hardware, or transmission errors [67, 90]. A characteristic property of impulse noise is that only some pixels are corrupted while the rest are noise-free.

Impulse noise can be described as a stochastic degradation process of the form

$$f(x) = \begin{cases} v(x) & \text{with probability } r \\ \tilde{u}(x) & \text{with probability } 1 - r \end{cases}, \quad x \in \Omega$$

where $r > 0$ represents the corruption rate by impulse noise. The main characteristics of the impulse noise is that the noise v is independent from the original image $\tilde{u}$ and a part of the image remains unchanged. Two main types of impulse noise are the salt-and-pepper noise when, for each x, the noise $v(x)$ has values drawn from the set $\{u_{\min}, u_{\max}\}$, and the random-valued impulse noise when, for each x, the noise $v(x)$ is a uniformly distributed random variable with values in the gray-level range $[u_{\min}, u_{\max}]$. For salt-and-pepper noise corrupted pixels take the lowest or the highest pixel value (i.e., $u_{\min}$ or $u_{\max}$), whereas for random-valued impulse noise, the noisy pixels have values anywhere in the interval from $u_{\min}$ to $u_{\max}$. Therefore, the random-valued impulse noise is more general and more difficult to detect than the salt-and-pepper noise.

In Fig. 2.2, there is an example of image corrupted by salt-and-pepper noise (here, $r = 0.3$). As you can see, some pixels are uncorrupted, while the ones corrupted by noise carry no information about the original image and appear only white or black.

2.2.5 Mixed noise

In real world the noise often cannot be modelled only with one distribution, but is given by the combination of more different noise models. For instance the most common mixed

Figure 2.2: From the left to the right: image corrupted with multiplicative noise, with a signal dependent noise and with impulse noise. For these simulations, we used the same original image in Fig. 2.1.

noise models are given by Poisson and Gaussian or impulse and Gaussian, see for instance [16, 57, 59, 60, 88].

Knowing a priori the noise that is presented in the image helps to recover the clean image. However, due to the difficulty of the problem, the complete elimination of the noise is impossible, but, the aim is to eliminate as much noise as possible and without introducing any other artifacts in the restored image.

2.3 Blur

When we take a picture with a camera, the picture that we get is not a faithful representation of the scene that we see with our eyes, since it is inevitable blurred. The blurring often occurs inside the camera, it may arise because the camera lens may be out of focus or because the camera or the object is moving during the time the shutter was open, which happens very often in astronomical imaging because of turbulence in the atmosphere. These are only a few examples of situations that can give the inevitable result of blurred images.

Mathematically speaking, the blurred image f is given by

$$f(x) = K\tilde{u} := (k \star \tilde{u})(x) = \int_{\mathbb{R}^2} k(y)\tilde{u}(x-y)dy,$$

where $x \in \Omega$, $\tilde{u}$ is the original image and k is some suitable kernel function, which is called point spread function (PSF). In other words, the blurred image is given by the convolution of the original image with a two-dimensional point-spread function. For simplicity, here, we consider PSF spatially invariant, i.e. the image is blurred in exactly the same way at every spatial location.

2.3.1 Blurring models

Briefly, we give some examples of PSFs that occur often in real applications. We start with the uniform out-of-focus blur. This blur occurs every time that we take a 2-D picture from a 3-D scene, where some parts of the image are in focus, while some others

are not. If the aperture of the camera is circular, the image of any point source is a small disk that is called circle of confusion (COC). The PFS of the uniform out-of-focus blur with radius R is defined as follows,

$$k(x) = \begin{cases} \frac{1}{\pi R^2} & \text{if } \|x\|_2 \leq R^2 \\ 0 & \text{otherwise.} \end{cases}$$

Gaussian blur is another degradation that occurs in the atmospheric turbulence and it may depend on the temperature, wind speed, exposure time, and so on. The PSF is defined by the Gaussian function,

$$k(x) = C \exp\left(\frac{\|x\|_2^2}{2\sigma^2}\right),$$

where σ determines the spread of the blur and C is a normalization constant.

Finally, we consider the motion blur. Motion blur is due to the motion between the recording system and the scene or object. There exists many types of motion blur, for instance it can be due to a transition, a rotation, a change of scale or a combination of these things. Here we give the formula of the PSF only in the case of the transition. The PSF for the transition motion blur is defined by

$$k(x) = \begin{cases} \frac{1}{L} & \text{if } \|x\|_2 \leq \frac{L}{2},\ \frac{x_1}{x_2} = -\tan\Phi \\ 0 & \text{otherwise,} \end{cases}$$

where L defines the length of the motion and Φ is the angle in radiant of the motion.

Figure 2.3: From the left to the right: original image, blurred image by using the motion blur and the out-of-focus blur.

Although there is no hope to recover the original image exactly, our aim is to obtain the best estimation of the original sharp and uncorrupted image. One of our challenges is to study suitable mathematical model for non-Gaussian noise and to devise efficient and reliable algorithms for recovering as much information as possible from the given data. Usually, the characteristic of the degradation system (the operator K) and the noise are assumed to be known as a priori. If there are no information about the blurring operator, then the field is known as blind image deconvolution, [7]. Here, we will mainly focus on the denoising problem and then we generalize to the more general case, i.e. deblurring and denoising problem (always considering the blurred operator to be known). Note that, in some of our works the deblurring and denoising problem follows straightforward from the denoising problem only.

Figure 2.4: Formation of the noisy image. From left to the right: original image, original image degraded by blur, original image degraded by blur and noise.

In general, to simulate a degraded image, we will start with the original uncorrupted and sharp image and then we will degrade it with blur and noise. In Fig. 2.4, there is an example of how the degradation model works, i.e. the first image represents the clean image, the second one is given by corrupting the original image with blur and finally in the last one there is also the presence of the noise. In the real world, knowing the degraded image (the third one in Fig. 2.4), we need to solve an inverse problem, i.e., try to reconstruct the clean image (the first one in Fig. 2.4). As we will see in the next chapter, the inverse problem is not easy to solve due to the ill-posedness of the problem. Now, we give some objective quality assessment to evaluate the recovered image.

2.4 Quality of an image

The most straightforward way to evaluate the quality of the recovered image is through subjective evaluation. However, the evaluation from the human beings is usually too inconvenient (since it depends from person to person and it may be ambiguous), time-consuming and expensive. Over the years many quantitative measures have been introduced to automatically compute the image quality. In our studies, we often use the peak signal-to-noise ratio (PSNR) value [7] and the structural similarity (SSIM) [92].

The PSNR is widely used in image quality assessment and it is defined as follows,

$$\mathrm{PSNR} = 20 \log_{10} \frac{m_1 m_2 |u_{\max} - u_{\min}|}{\|u - \tilde{u}\|_2},$$

where u and $\tilde{u}$ are respectively the restored and the original image with values in the gray-level range $[u_{\min}, u_{\max}]$. We underline that, high value of PSNR implies a good restored image. It is a very useful tool, since it is able to measure quantitatively the quality of the reconstruct image compare to the original image. However the PSNR value is not very well matched to perceived visual quality. Recently, another measure has become very popular among the imaging community, the so-called SSIM measure. This measure compares local patterns of pixel intensities that have been normalized for luminance and contrast and it has been proved that it is more consistent with human eye perception than PSNR, we refer the reader to [92]. The SSIM measure is defined as

$$\mathrm{SSIM} = \frac{(2\mu_{\tilde{u}}\mu_u + c_1)(2\sigma_{\tilde{u}u} + c_2)}{(\mu_{\tilde{u}}^2 + \mu_u^2 + c_1)(\mu_{\tilde{u}}^2 + \mu_u^2 + c_2)},$$

where u and $\tilde{u}$ are the restored and original image, $\mu_{\tilde{u}}$ and μ_u denote the means of $\tilde{u}$ and u, $\sigma_{\tilde{u}}$ and σ_u denote the variances of $\tilde{u}$ and u, $\sigma_{\tilde{u}u}$ represents the covariance of $\tilde{u}$ and u and $c_1, c_2 > 0$ are constants. Also in this case, high value of SSIM means good reconstruction.

CHAPTER 3

Inverse problems

The process of recovering the original image from the data is an inverse problem. In most of the cases, the direct inversion of the degraded model does not give any good results, due to the ill-posedness of the problems. Thus a way to deal with ill-posedness is to use the Bayesian approach that is based on some a-priori information. Here, after a short introduction on the inverse problems, we study the two most common ways to estimate the image: the maximum-a-posteriori (MAP) and the conditional mean (CM) estimates. Furthermore, we show how the noise model influences the reconstruction problem.

3.1 Inverse problem

Solving an inverse problem is the opposite task of a direct (or forward) problem. Roughly speaking, in a direct problem we find an effect from a cause, while in an inverse problem we do the opposite, that corresponds to recover the cause knowing the effect. One of the main needs of inverse problem is the need to interpret indirect physical measurements of an unknown object of interest. Inverse problems arise everywhere, for instance in optics, radar, acoustics, communication theory, signal processing, medical imaging, computer vision, geophysics, oceanography, astronomy, remote sensing, machine learning, and many other fields. For instance, in computed tomography we are interested in looking inside the human body to see something that cannot be directly observed. By passing x-rays through the body, we get some projection images that correspond to the acquired data. The task of determining the data from the body and the knowledge of the physical model is called the forward model. Another example is the image reconstruction: here, the inverse problem consists of recovering the original sharp and noise-free image by knowing the degraded image, [50, 62]. In Fig. we 3.1, we show the forward and the inverse problem for image reconstruction.

Mathematically speaking, in image processing, solving the inverse problem means recovering the original image $\tilde{u}$ based on

$$f = K\tilde{u} \odot v, \tag{3.1}$$

knowing the corrupted image f and the blurring operator K. For more details about the forward model in (3.1), we refer the reader to the previous chapter. Unfortunately, computing $\tilde{u}$ in (3.1) by using the simple inversion does not lead to any suitable solution.

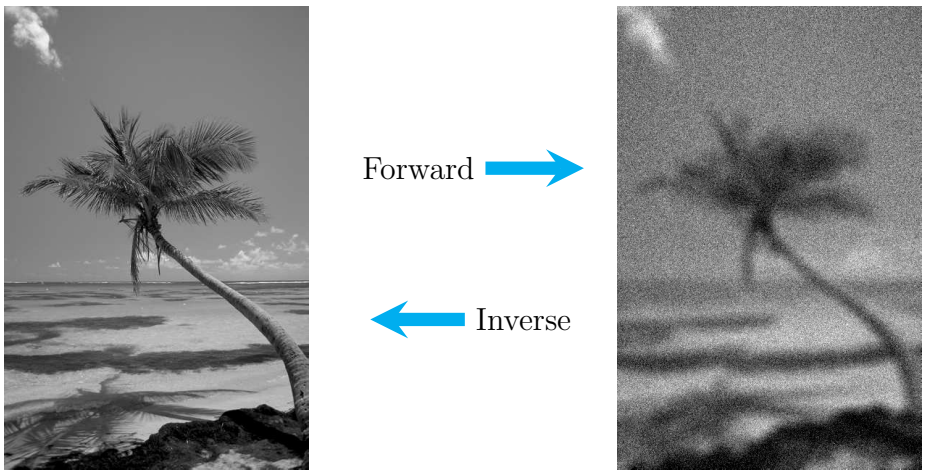


Figure 3.1: Example of forward and inverse problem in image processing. Left: original image. Right: corrupted image during the acquisition and transmission.

This is due to the fact that inverse problems very often are ill-posed, and thus solving them may be very challenging. The term ill-posed problem was introduced the first time in the early 20th century by Hadamard. A problem is well-posed if the following conditions are satisfied:

- **Existence:** The problem must have a solution.
- **Uniqueness:** The solution must be unique.
- **Stability:** The solution must depend continuously on the data.

If a problem does not satisfy one or more of these conditions, it is called ill-posed. Therefore, the forward model in (3.1) is well-posed if K is well-defined, single-valued, and continuous function. But, very often in image processing, K^{-1} does not exist or is not continuous, thus the problem is ill-posed.

At first glance, one way to handle the first condition, i.e. the existence of the solution, is to compute the least-squared solution,

$$\hat{u} = \underset{u}{\operatorname{argmin}} \|Ku - f\|_2^2.$$

If the solution is unique solution, then it is given by

$$(K^\star K)\hat{u} = K^\star f.$$

But computing the least-squared solution does not guarantee the uniqueness of the solution. For this reason, a way to handle the ill-posed problem is to introduce some further regularity to the problem, adding a regularization term to the minimization problem, for instance it can be done as follows,

$$\hat{u} = \underset{u}{\operatorname{argmin}} \|Ku - f\|_2^2 + \lambda \|Lu\|_2^2, \tag{3.2}$$

where $\lambda > 0$ and L can be for instance the identity operator or the gradient operator. The stability condition can be handled by tuning the regularization parameter λ in

(3.2). In the next section, we will see that the minimization problem in (3.2) can be also formulated from the statistical point of view, also known as Bayesian approach, [2, 17, 55].

3.2 The Bayesian approach

To reconstruct the clean image one can employ the Bayesian approach that combines prior information on $\tilde{u}$ and the forward model in (3.1), see [2, 17, 55]. For simplicity, we will use the lower-case letters for indicating both random variables and instances. From Bayes' rule, [44] we have

$$p(u|f) = \frac{p(f|u)p(u)}{p(f)}, \tag{3.3}$$

where $p(u|f)$ is called posterior and represents the conditional probability density of u given the degraded image f, $p(f|u)$ is called likelihood function and it encodes the likelihood that the data f is due to the image u, $p(u)$ is called prior and describes the properties of the image that we want to recover and $p(f)$ is a normalization factor (independent of u). To point out the connection to the previous section and the problem in (3.2), we give an example of priori and likelihood. When the image is corrupted by blur and additive Gaussian noise, the likelihood density is given by

$$p(f|u) \propto \exp\Big(-E(u; K, f)\Big),$$

where $\propto$ means "proportional to" and $E(u; K, f) = \frac{1}{2}\|f - Ku\|_2^2$. In image processing, the most common prior use are the so-called Gibbs distributions, [40],

$$p(u) \propto \exp(-\lambda J(u)),$$

where $J(u)$ is a convex functional and $\lambda > 0$ is a parameter. Thus, the posterior can be written as follows

$$p(u|f) \propto \exp\Big(-E(u; K, f) - \lambda J(u)\Big). \tag{3.4}$$

The term $E(u; K, f)$ is often called data fidelity term, $J(u)$ is known as regularization term and $\lambda > 0$ is the regularization parameter. We now analyse these two terms and the regularization parameter. Finally in the end of the chapter, we show how to compute the clean image by knowing the posterior distribution.

3.2.1 The data-fidelity term

The data fidelity term strongly depends on the forward model, which means that different noise models lead to different data fidelity terms. In Chapter 2, we give some examples of noise models. Here, we give their corresponding data-fidelity terms for the noise models introduced in the previous chapter.

We start with the additive noise model, i.e. when the noisy image is given by

$$f = K\tilde{u} + v.$$

We have already seen that in the Gaussian case, the data-fidelity term is given by the L^2-norm, that means

$$E(u; K, f) = \|Ku - f\|_2^2.$$

As suggested in [30], the data-fidelity for recovering an image corrupted by the uniform noise is given by the norm infinity, i.e.

$$E(u; K, f) = \|Ku - f\|_\infty.$$

In one of our contributions, we propose the following data fidelity term for the additive Cauchy noise removal

$$E(u; K, f) = \int_\Omega \log\Big(\gamma^2 + (Ku - f)^2\Big)\mathrm{d}x.$$

For more details about Cauchy noise we refer the reader to Chapter 5.

The data fidelity term for the multiplicative noise was introduced the first time in [2] and it reads

$$E(u; K, f) = \int_\Omega \Big(\log(Ku) + \frac{f}{Ku}\Big)\mathrm{d}x.$$

Regarding the signal dependent noise model, the data fidelity for the Poisson noise has been introduced in [58] only for the denoising case. Later, it has been generalize for the deblurring and denoising see for instance [38, 79] and it is defined by

$$E(u; K, f) = \int_\Omega \Big(Ku - f\log(Ku)\Big)\mathrm{d}x, \quad \text{with} \quad Ku \geq 0.$$

The other well-known signal-dependent noise model is the Rician noise which is common found in the magnetic resonance imaging, see [39, 42, 48]. In this case the data-fidelity term is given by

$$E(u; K, f) = \frac{1}{2\sigma^2}\int_\Omega (Ku)^2\mathrm{d}x - \int_\Omega \log I_0\Big(\frac{f(Ku)}{\sigma^2}\Big)\mathrm{d}x,$$

where $\sigma > 0$ is the noise level and I_0 corresponds to the modified Bessel functions of the first kind with order zero. As in the Cauchy case, also here the noise level is required in the data-fidelity term. Thus, in these two cases, in order to reconstruct the image we need to know (or to estimate) the corruption rate.

Finally, for impulse noise it is used the following term, [66, 67],

$$E(u; K, f) = \|Ku - f\|_1.$$

Since our contribution mainly deals with Cauchy noise and impulse noise, in the following chapter we will extensively study these two noise models and their variational methods.

3.2.2 The regularization term

The regularization term $J(u)$ is based on some prior information about the expected behaviour of the original image $\tilde{u}$. One of the most straightforward choices of regularization term is given by Tikhonov regularization [84], that corresponds to

$$J(u) = \int_\Omega |\nabla u|^2\mathrm{d}x,$$

where ∇u indicates the gradient of u. The main idea is to recover u that best fits the data and has lowest gradient. In this setting, the functional space for which the data fidelity term and the Tikhonov regularization term are defined is

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

with Ω a bounded open subset of $\mathbb{R}^2$. The reason why Tikhonov regularization was so attractive is because it leads to a quadratic optimization problem for which the solution has a closed-form,

$$K^*(Ku - f) - \lambda \Delta u = 0,$$

where Δ indicates the Laplacian operator, and there exist fast algorithms to solve it, for instance we refer the reader to [49]. However, due to the L^2 norm of the gradient the reconstructions of natural images given by Tikhonov regularization are too smooth, since we penalize too much on the high values of the gradients which corresponds to edges. For this reason, in [73] the total variation (TV) regularization, which corresponds to the L^1 norm of the gradient of u, was proposed and it is defined as follows

$$J(u) = \int_\Omega |Du| := \sup\left\{ \int_\Omega u(x)\mathrm{div}(\xi(x))\mathrm{d}x; \xi \in C_c^1(\Omega; \mathbb{R}^2), |\xi(x)| \leq 1 \quad \forall x \in \Omega \right\}, \quad (3.5)$$

with $u \in L^1(\Omega)$, $\mathrm{div}(\xi(x)) = \frac{\partial \xi_1}{\partial x_1}(x) + \frac{\partial \xi_2}{\partial x_2}(x)$ and $C_c^1(\Omega; \mathbb{R}^2)$ is the space of continuously differentiable functions with compact support in Ω.

If $u \in C^1(\Omega)$, then $\int_\Omega u(x)\mathrm{div}(\xi(x))\mathrm{d}x = -\int_\Omega \nabla u \cdot \xi dx$ and therefore $J(u) = \int_\Omega |\nabla u(x)|_2 dx$.

Furthermore, we define the space of the functions of bounded variation $BV(\Omega)$ as

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

The space $BV(\Omega)$ has the lower semicontinuity property, i.e. let $u_j \in BV(\Omega)$ and $u_j \to u$ in $L^1(\Omega)$, then $\int_\Omega |Du| \leq \liminf_{j \to +\infty} \int_\Omega |Du_j|$. The space $BV(\Omega)$ endowed with the norm $\|u\|_{BV} = \|u\|_{L^1} + \int_\Omega |Du|$ is a Banach space. Total variation regularization has the good property of preserving the edges, but it has some drawbacks, for instance the staircasing effect. To overcome this issue other modern regularization terms have been proposed, such as nonlocal TV [43, 89], high order TV [22], total generalized variation (TGV) [9].

3.2.3 The regularization parameter

Last but not least we consider the choice of the regularization parameter $\lambda > 0$. The regularization parameter represents the tradeoff between a good fit to the data and the smoothness due to the regularization term, thus it is very important to find the right regularization parameter in order to get the best solution. In particular, from (3.4) we can see that if λ is too big then the reconstruction will be too blurred, because of the predominance of the regularization term, on the other hand if the regularization parameter is too small the noise will not be removed from the image, and the reconstruction will still look very noisy.

In Fig. 3.2, we show the influence of the regularization parameter for the reconstruction. The noisy image has been created with 10% of salt-and-pepper noise. The

quality of the restored images is computed quantitatively using the peak signal noise ratio (PSNR) value, see section 2.4. From the plot of the regularization parameter λ versus the PSNR, we can see how the recovered image depends on the tuning of the parameter λ. Visually, we can see the difference between the reconstructions by using different λ-values. In fact, when $\lambda = 2$ the image looks too blurry, because during the reconstruction process the image is over-regularized, on the other hand when $\lambda = 0.3$ there is still some noise in the image. Thus, the reconstruction strongly depends on the regularization parameter, therefore finding a good way to compute the regularization parameter is extremely important.

The problem of finding the best regularization parameter has been extensively studied over the years. But at this time there is no a general automatic parameter-choice method that produce the best reconstruction. On the other hand there exist many methods that, under certain conditions, work well, but all of them can fail to produce good results.

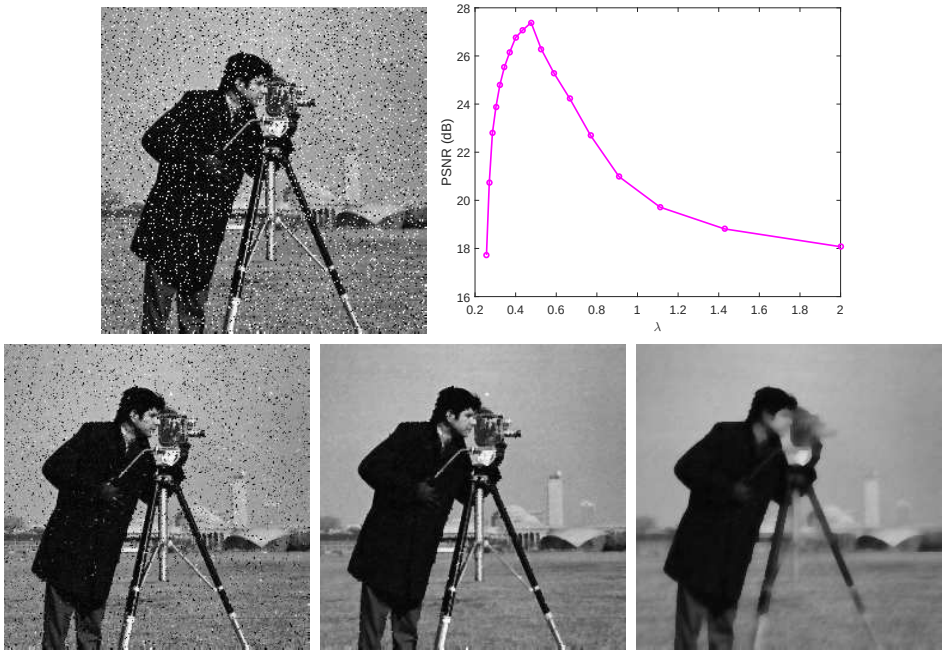


Figure 3.2: Influence of the regularization parameter for recovering the image. First row: degraded image with 10% of salt-and-pepper noise; plot regularization parameter λ versus the PSNR(dB). Second row: reconstructed images by using $\lambda = 0.3, 0.5, 2$.

One of the most intuitive and straightforward choice is based on visual criterion [7]. If there is a considerable prior knowledge on the scene, it may be reasonable to choose the regularization parameter through a simple visual consideration of the reconstructed images. For instance the way to find the best parameter corresponds in finding the image that visually looks better, i.e. it is not over-regularized and the noise is fully removed.

In the literature, there exist many possible strategies to choose the regularization parameter: the discrepancy principle [61], the L-curve [49, 51], the generalized cross-validation [47], just to mention some of them. The discrepancy principle is a method

that finds the best regularization parameter if a priori is known the error level. The L-curve is based on a log-log scale plot of the regularization term $J(\hat{u}(\lambda))$ versus the data fidelity term $D(\hat{u}(\lambda), f)$, where $\hat{u}(\lambda)$ represents the recovered image using the regularization parameter λ. With the L-curve method the best parameter is computed using the corner of the L-shaped plot. The generalized cross-validation determines the regularization parameter without the knowledge of the noise properties. The behaviour of these methods is very problem-dependent, for instance the discrepancy principle relies on a good estimate of the error norm, which may be not known, the L-curve often fail when the solution is too smooth, the generalized cross-validation is often a robust method, but sometimes fails leading to under-regularization.

Recently, many other techniques to automatically select the regularization parameter are introduced, for an overview we refer the reader to [32, 33] and the references therein.

One of the contribution of the thesis, is given by introducing a parameter-free method to remove impulse noise from blurred and noisy images. For more details, we refer the reader to 5 and to our paper [77].

We now focus on the process of extracting information from the posterior distribution. The process of finding some information from the posterior is called Bayesian estimation. In the next section, we introduce the Bayesian formalism by giving two ways to reconstruct the image.

3.2.4 MAP and CM estimates

Probably the most common choice to extract information from the posterior is given by computing the maximum-a-posteriori estimate (MAP),

$$\hat{u}_{\text{MAP}} := \operatorname*{argmax}_u p(u|f). \tag{3.6}$$

In practise, since the joint log-likelihood of a set of independent variables is the sum of the log-likelihood of individual variables, it is more convenient to minimize the opposite of the log-likelihood instead of maximizing the likelihood,

$$\hat{u}_{\text{MAP}} := \operatorname*{argmin}_u \log(p(F|U)) + \log(p(U)).$$

Let's note that in the Gaussian case, the above equality corresponds of solving the following minimization problem

$$\hat{u}_{\text{MAP}} := \operatorname*{argmin}_u \frac{1}{2}\|f - Ku\|_2^2 + \lambda J(u),$$

which exactly corresponds to the minimization problem in (3.6). Thus, the MAP estimation is directly related to the variational methods and regularization techniques.

Another way to compute an estimate is given by the conditional mean estimate (CM), which is defined as follows

$$\hat{u}_{\text{CM}} := \mathbb{E}[u|f] = \int up(u|f)\mathrm{d}u \tag{3.7}$$

Therefore, the difference between the two estimates is that the MAP estimate corresponds to find the mode of the posterior, while the CM estimate corresponds to compute

the expected value of the posterior. In order to get the best reconstruction we need to understand which of these two estimates is best. Unfortunately, there is no a general answer to the question, but there are some considerations that we should take into account. The conditional mean looks like the most intuitive choice, since it represents the average of samples. But, from the numerical point of view, the MAP estimate requires to solve an optimization problem which can be usually computed easily, while the CM estimate requires to solve an integration problem which in high-dimension scenarios can cause some trouble. In the next chapter, we will show some ways to compute these two estimates numerically. In Chapter 5, we will analyse the two estimates from the theoretical point of view.

CHAPTER 4

Numerical considerations

In this chapter we show how to compute the restored image numerically by knowing the noisy and blurred image. First of all, we start with the digital images and some notation in the discrete case. For simplicity we will mainly use the same notation as in the continuous case. After that, we will show how the maximum-a-posterior (MAP) and the conditional mean (CM) estimate can be numerically computed. By definitions, the first one requires to solve an optimization problem while the second one requires to solve an integral problem. Therefore, we will briefly give some general notions about optimization, by mainly focusing on the algorithms that we used in our contributions, i.e. the primal-dual algorithm and the semismooth Newton's method. To calculate the CM estimate, the quadrature rules cannot be used due to the very large dimension of the image, thus we will apply the Monte Carlo methods. Since in general, it is not possible to draw samples from the posterior distribution we will need the Markov chain Monte Carlo techniques. Examples of how apply this algorithm in image restoration are shown as well.

4.1 What is a digital image?

As we have seen in Chapter 3, the images are analogue, which means that they are defined on a continuous domain and that they take values that come from a continuum of possibilities. However, in order to have an image which has computer-readable digital format, we need to introduce the digital images. A digital image is defined on a discrete domain and it takes values from a discrete set of possibilities. We briefly describe the process of analogue-to-digital conversion, which consists of two processes: the sampling and the quantization

4.1.1 Sampling

Sampling is the process of converting a continuous signal into a discrete-space signal. During this process it is very important to not loose information of the image by limiting the presence of artifacts and noise. Since some information is lost, it may not be possible to reconstruct the original continuous signal from the samples. A sampled image is an array that is arranged in a row-column format. Each element of the array is called

picture element or shortly pixels. In Figure 4.1, we have an example of different sampling densities (256×256, 128×128 and 32×32) for the same image. As you can see, the details of the image substantially change when we undersample the image. In the 128-by-128 image, the shape of the boats can be still recognized, but some details are missing, while in the 32-by-32 image the subject of the scene is completely lost.

Figure 4.1: Sampling process. From left to right: 256-by-256 image, 128-by-128 image and 32-by-32 image.

4.1.2 Quantization

The other process involved during the conversion from analogue-to-digital is the quantization. This process consists of converting the continuous intensities into a discrete values. When the gray level of an image pixel is quantized the finite set of numbers is called gray level range. In Fig. 4.2, we have an example of images at various levels of gray level resolution (4, 2, and 1 bit representation). The 8-bit image can be found in Fig. 4.1. In the 4-bit image we can still recognize the subject of the image, but some of the details are lost compared to the 8-bit image. In the other representations (2 bit and binary) most of the information disappear, making difficult to recognize the scene represented.

Figure 4.2: Quantization process. From left to right: 4-bit image, $[0, 15]$; 2-bit image, $[0, 3]$; and binary image, $\{0, 1\}$. The corresponding 8-bit image, $[0, 255]$, is represented on the left in Fig. 4.1.

From now on in this chapter, we will just focus on digital images.

4.1.3 Notations in the discrete case

Let $u \in \mathbb{R}^{m_1 \times m_2}$ a two-dimensional image with size $m_1 \times m_2$ ($m_1 \in \mathbb{N}$ and $m_2 \in \mathbb{N}$ indicate the number of pixels in the horizontal and vertical direction, respectively) and $m = m_1 m_2$ represents the total number of pixels in the image. For $i = 1, \ldots, m_1$ and $j = 1, \ldots, m_2$, the intensity value of the pixel (i, j) is denoted with $u_{i,j}$. Along this thesis, we will often use the following notation $u_k = u_{i,j}$, where $k = (i-1)m_2 + j$ for $i = 1, \ldots, m_1$ and $j = 1, \ldots, m_2$, to identify the pixel at position (i, j) with the kth element of u, and we define a set $\Omega = \{1, 2, \ldots, m\}$ that contains all pixel indices. Often, it will be convenient to consider a two-dimensional pixel-array by concatenation in the usual columnwise fashion. Given $u_{\min}$ and $u_{\max}$ the minimum and maximum intensity of the image, respectively, the intensity of the pixels $u_{i,j} \in [u_{\min}, u_{\max}]$ with $i = 1, \ldots, m_1$ and $j = 1, \ldots, m_2$.

For instance the following matrix

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0.5 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0.5 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0.5 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

corresponds to the image in Fig. 4.3. Using the previous notations, we have $m_1 = 6$, $m_2 = 7$, $u_{\min} = 0$ and $u_{\max} = 1$. The intensity value 0 corresponds to the black pixels while the 1 produces a white pixels and the values between 0 and 1 create a gray color.

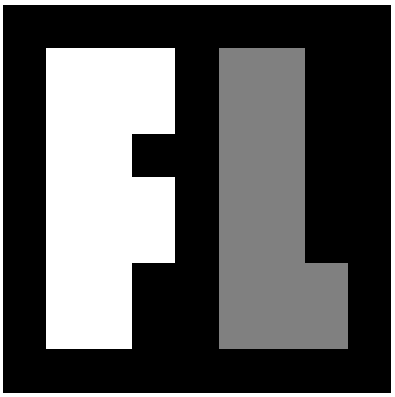


Figure 4.3: Example of 6-by-7 gray level image.

The most popular choices for the range of the image u is $[0, 1]$ or $[0, 255]$; in our work, we mainly considered the first one. Each pixel of the image has an intensity, which characterize the color of a small rectangular segment of the scene. For example, if an image is 256-by-256, then it contains 65536 pixels, while a high-resolution image can be much more than 5 million of pixels.

In this context we need to define also the discrete total variation, [3], which reads

$$\mathrm{TV}(u) := \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} |(\nabla u)_{i,j}|_2, \text{ with } |(\nabla u)_{i,j}|_2 = \sqrt{|(\nabla_x u)_{i,j}|^2 + |(\nabla_y u)_{i,j}|^2}.$$

The discrete gradient operator $\nabla \in \mathbb{R}^{2m\times m}$ is given by

$$(\nabla u) = \begin{pmatrix} \nabla_x u \\ \nabla_y u \end{pmatrix} \in \mathbb{R}^{2m},$$

and $\nabla_x u \in \mathbb{R}^m$ and $\nabla_y u \in \mathbb{R}^m$ denote the horizontal and vertical first order differences, i.e., by using the symmetric boundary conditions, we have

$$(\nabla_x u)_{i,j} = \begin{cases} u_{i+1,j} - u_{i,j} & \text{if } i < m_1, \\ 0 & \text{if } i = m_1, \end{cases} \quad \text{and} \quad (\nabla_y u)_{i,j} = \begin{cases} u_{i,j+1} - u_{i,j} & \text{if } j < m_2, \\ 0 & \text{if } j = m_2, \end{cases}$$

for $i = 1, \dots, m_1$ and $j = 1, \dots, m_2$.

In order to solve the optimization problems introduced in Chapter 3, we need to introduce the divergence. The divergence is defined as $\operatorname{div} = -\nabla^\top$, where $\nabla^\top$ is the transpose of the gradient operator. Hence, the explicit formula of the divergence can be found by using the definition of transpose

$$\langle -\operatorname{div} p, v\rangle_{\mathbb{R}^m} = \langle p, \nabla v\rangle_{\mathbb{R}^{2m}},$$

for every $p \in \mathbb{R}^{2m}$ and $v \in \mathbb{R}^m$, where $\langle\cdot,\cdot\rangle_{\mathbb{R}^m}$ and $\langle\cdot,\cdot\rangle_{\mathbb{R}^{2m}}$ denote the standard scalar products in $\mathbb{R}^m$ and $\mathbb{R}^{2m}$, respectively. Thus,

$$(\operatorname{div} p)_{i,j} = (\operatorname{div} p)^1_{i,j} + (\operatorname{div} p)^2_{i,j},$$

where

$$(\operatorname{div} p)^1_{i,j} = \begin{cases} p^1_{i,j} - p^1_{i-1,j} & \text{if} \quad 1 < i < m_1 \\ p^1_{i,j} & \text{if} \quad i = 1 \\ -p^1_{i-1,j} & \text{if} \quad i = m_1 \end{cases} \quad \text{and}$$

$$(\operatorname{div} p)^2_{i,j} = \begin{cases} p^2_{i,j} - p^2_{i,j-1} & \text{if} \quad 1 < j < m_2 \\ p^2_{i,j} & \text{if} \quad j = 1 \\ -p^2_{i,j-1} & \text{if} \quad j = m_2 \end{cases}$$

for every $p = (p^1, p^2) \in \mathbb{R}^{2m}$.

In the next chapters, we will describe how to numerically denoise and deblur the degraded image.

4.2 Optimization

A way to numerically solve an ill-posed inverse problem for reconstructing an image is given by using the variational methods. Variational methods constitute the computation of a reconstructed image u based on the observed image (or more generally data) f as a minimizer of a functional, which is called as a variational model.

Computing a solution of the variational method requires to solve an optimization problem. We now give an overview of how solving the variational models. Variational models for image reconstruction usually has such a format:

$$\min_{u\in U} E(u; K, f) + \lambda J(u),$$

here, U indicates the feasible set, and, as in the previous section, $J(u)$ represents the regularization term, $\lambda > 0$ is the regularization parameter, and $E(u; K, f)$ is the data fidelity term. As we have already discussed in §3.2.3, the problem of finding the best regularization parameter can be very tricky. For this reason, sometimes is more convenient to consider a constrained minimization problem instead, i.e.

$$\min_{u \in U} J(u) \quad \text{such that} \quad E(u; K, f) \leq \tau,$$

where $\tau > 0$ is a given constant that can be related with the noise level, see for instance [64, 83]. If the noise level is known or it can be estimated, then the constrained minimization problem is easier to solve than the unconstrained one, since τ is known while λ has to be tuned.

Optimization problems can be classified as follows: convex and non-convex problems, see [72]. Convex problems have the advantages that the global optimum can be computed independently of the initialization. Thus, the quality of the solution depends only on the accuracy of the denoising and deblurring model. Non-convex problems are able to model more precisely the process behind the image acquisition, but on the other hand the solution depends on the initialization and the algorithm used to solve the optimization problem.

The convexity of the optimization problem depends on the objective function and on the feasible set. For instance, the data-fidelity terms for denoising an image corrupted by Gaussian noise, impulse noise or Poisson noise are convex, while the original one for Cauchy noise is non-convex. In one of our contribution, we add a quadratic penalty term to guarantee the convexity of the data-fidelity term, see [B] for more details. Regarding the regularization term, total variation plays a very important role in the image processing field, since it allows discontinuities in the recovered image, thus it guarantees the presence of edges which are very important for a wide range of applications. However, due to the non-smoothness of total variation, see §3.2.2, computing the solution of the optimization problem with TV is not so straightforward. From now on, we will focus only on convex problems.

The first classification that we gave, i.e. constrained and unconstrained problems is very important in our paper on impulse noise [A]; the second one, i.e. convex and nonconvex, will be useful for our paper on Cauchy noise [B].

We now introduce some numerical algorithms that are often used in image processing. In literature there exist a lot of optimization algorithms available, but as usual it is not clear which one is better in general. In fact, up to our knowledge, there is no an algorithm that produces the most accurate solutions in the shortest time with less memory requirement. For this reason, the choice of the numerical algorithm strictly depends on the application. To compute the solution for the convex problems, there exist a huge selection of algorithms, for instance the primal dual algorithm [19, 23, 31], the alternating direction method with multipliers (ADMM) [8], the split-Bregman algorithm [45] and the Chambolle-Pock algorithm [20]. Since the convergence of the Chambolle-Pock algorithm is well studied, see [20], in our contributions we have mainly employed it to solve our minimization problem. Furthermore, since in our paper on impulse noise, A, we need to compare the computational time with a second-order method, here we also introduce the Newton's method.

4.2.1 Primal-dual algorithm

Due to the simplicity and the convergence theorem of the Chambolle-Pock algorithm, in most of our works we used this algorithm to recover the clean image. Now, we present the Chambolle-Pock algorithm for a general convex functions and later, we will show how to apply to the image denoising.

In the general continuous case, given X, Y two finite-dimensional real vector space, $L : X \to Y$ a continuous operator with induced norm

$$\|L\| = \max\{\|Lx\| : x \in X \text{ with } \|x\| \leq 1\},$$

the primal problem is defined as follows

$$\min_{x \in X} F(Lx) + G(x), \tag{4.1}$$

where $F : Y \to [0, +\infty)$, $G : X \to [0, +\infty)$ are proper, convex and lower-semicontinuous functions. The dual problem is then given by

$$\max_{y \in Y} -\Big(G^*(-L^*y) + F^*(y)\Big),$$

where F^* (resp. G^*) indicates the convex conjugate of a convex lower-semicontinuous function F (resp. G). Then, the primal-dual formulation is the saddle-point problem and it reads

$$\min_{x \in X} \max_{y \in Y} \langle Lx, y\rangle_Y + G(x) - F^*(y),$$

with $\langle \cdot, \cdot \rangle_Y$ indicates the inner product in Y. If there exists at least one solution $(\hat{x}, \hat{y}) \in X \times Y$ we have

$$L\hat{x} \in \partial F^*(\hat{y}) \text{ and } -(L^*\hat{y}) \in \partial G(\hat{x}),$$

where ∂ denotes the subdifferential[1] of a function, see [72] for more details.

Now, we can present the primal-dual algorithm in [20] which is also known as Chambolle-Pock algorithm. Initialize $(x^0, y^0) \in X \times Y$ and choose the parameters $\sigma, \tau > 0$ and $\theta \in [0, 1]$. Until the convergence is reached repeat the following three steps ($n \geq 0$)

1. Update y^{n+1}:
$$y^{n+1} = (I + \sigma \partial F^*)^{-1}(y^n + \sigma L\bar{x}^n).$$
2. Update x^{n+1}:
$$x^{n+1} = (I + \tau \partial G)^{-1}(x^n - \tau L^* y^{n+1}).$$
3. Update $\bar{x}^{n+1}$:
$$\bar{x}^{n+1} = x^{n+1} + \theta(x^{n+1} - x^n).$$

If $\theta = 1$ and $\tau\sigma\|L\|^2 < 1$, in [20], it has been proved that the algorithm converges to the saddle-point (x^*, y^*).

Now, we show two applications of the Chambolle-Pock algorithm to image processing.

[1] In general, we recall that $z \in \partial G(x)$ means

$$G(w) \geq F(x) + \langle w - x, z\rangle_X \text{ for all } w \text{ in } X.$$

Example 4.2.1. The ROF model. The variational method to recover an image corrupted by Gaussian noise is given by the well known ROF model, see [73], which, in the discrete setting, reads as follows

$$\min_{u\in X} \mathrm{TV}(u) + \frac{\lambda}{2}\|u-f\|_2^2,$$

with $u \in \mathbb{R}^m$ the unknown image, $f \in \mathbb{R}^m$ the noisy image and $\lambda > 0$ the regularization parameter, see Chapter 3. Using the notations in the primal problem (4.1), we have $F(\nabla u) = \mathrm{TV}(u)$ and $G(u) = \frac{\lambda}{2}\|u-f\|_2^2$. Thus, given the dual variable $p \in \mathbb{R}^{2m}$, based on the definition of TV, the primal-dual problem of the ROF model is defined as follows

$$\min_{u\in\mathbb{R}^m}\max_{p\in\mathbb{R}^{2m}} -\langle u, -\mathrm{div} p\rangle + \frac{\lambda}{2}\|u-f\|_2^2 - \delta_P(p),$$

where δ_P is the indicator function of the set P, i.e.

$$\delta_P(p) = \begin{cases} 0 & \text{if } p \in P \\ +\infty & \text{otherwise} \end{cases}$$

and P is the convex set given by

$$P = \{p \in \mathbb{R}^{2m} : \||p|_2\|_\infty \le 1\}.$$

The discrete maximum norm $\| |p|_2 \|_\infty$ is defined as

$$\| |p|_2 \|_\infty = \max_{i,j} |p_{i,j}|_2, \quad \text{and} \quad |p_{i,j}|_2 = \sqrt{(p^1_{i,j})^2 + (p^2_{i,j})^2}.$$

Now, we can apply the Chambolle-Pock algorithm to the saddle-point problem with $F^*(p) = \delta_P(p)$ and $G(u) = \frac{\lambda}{2}\|u-f\|_2^2$. Since F^* is the indicator function of a convex set, the resolvent operator is the pointwise Euclidean projectors onto the ℓ^2 unit ball, i.e. $p = (I + \sigma\partial F^*)^{-1}(\tilde{p})$ is equivalent to

$$p_{i,j} = \frac{\tilde{p}_{i,j}}{\max(1, +|\tilde{p}_{i,j}|_2)}.$$

The second step of the algorithm, is even simpler since it leads to a pointwise quadratic problem, i.e. $u = (I + \tau\partial G)^{-1}(\tilde{u})$ is equivalent to

$$u = \frac{\tilde{u} + \tau\lambda f}{1 + \tau\lambda}.$$

In [19], it has been computed the bound of the norm of the linear operator ∇. Using the discrete gradient and divergence operator, we have

$$\|\nabla\|^2 = \|\mathrm{div}\|^2 \le 8.$$

Thus, in order to have a convergent algorithm we just need to set the two parameter τ and σ such that $\tau\sigma < 8$.

Example 4.2.2. The ℓ^1-TV model. In case the image is corrupted by impulse noise, the ℓ^1-TV model works quite well,

$$\min_{u\in\mathbb{R}^m} \mathrm{TV}(u) + \lambda\|u-f\|_1,$$

where the main difference is given by the ℓ^1-norm in the data fidelity term. Using the notations in the primal problem (4.1), we have $F(\nabla u) = \mathrm{TV}(u)$ and $G(u) = \|u - f\|_1$. Thus, using the same notations in the previous example, the primal-dual problem of the ROF model is defined as follows

$$\min_{u\in\mathbb{R}^m} \max_{p\in\mathbb{R}^{2m}} -\langle u, -\mathrm{div} p\rangle + \lambda\|u - f\|_1 - \delta_P(p).$$

Therefore, the only difference with the ROF model is the resolvent operator with respect to G, which in this case is given by the pointwise shrinkage operator, i.e. $u = (I + \tau\partial G)^{-1}(\tilde{u})$ corresponds to

$$u_{i,j} = \begin{cases} \tilde{u}_{i,j} - \tau\lambda & \text{if } \tilde{u}_{i,j} - g_{i,j} > \tau\lambda \\ \tilde{u}_{i,j} + \tau\lambda & \text{if } \tilde{u}_{i,j} - g_{i,j} < -\tau\lambda \\ g_{i,j} & \text{otherwise.} \end{cases}$$

4.2.2 Semismooth Newton's method

In this section we introduce the second order method that has been used in our paper [A]: the semismooth Newton's method. Given a nonlinear equation

$$F(x) = 0,$$

where $F : \mathbb{R}^s \to \mathbb{R}^t$ is a continuously differentiable function, one of the classical way to solve it is by employing the Newton's method. Newton's method reads

$$x^{k+1} = x^k - \big(F'(x^k)\big)^{-1} F(x^k),$$

where F' indicates the first derivative of F and x^k represents the kth iterate. This problem is quite easy to solve, however, in some cases the optimality system of the optimization problem is not smooth. For instance, in our contribution [A] we have to deal with a max-operator which is nonsmooth. In this case, the semismooth Newton's method can be employed, see [52]. We can define the semismooth Newton's method as the generalized version of Newton's method for semismooth maps [70]. Then, if $F : \mathbb{R}^s \to \mathbb{R}^t$ is semismooth, then the semismooth Newton's method is defined as follows

$$x^{k+1} = x^k - \big(G_F(x^k)\big)^{-1} F(x^k),$$

where G_F denotes the generalized derivative of F. The mapping F is called generalized differentiable in an open set $U \subset \mathbb{R}^s$ if there exists $G_F : \mathbb{R}^s \to \mathbb{R}^{s\times t}$ such that

$$\lim_{\|h\|\to 0} \frac{1}{\|h\|}\|F(x + h) - F(x) - G_F(x + h)h\| = 0,$$

for every $x \in U$; see e.g. [52]. For an example on how to apply the semismooth Newton's method for solving a system we refer to our paper [A].

4.3 Monte Carlo method-solving the integration

While the maximum a posteriori estimate leads to a minimization problem, the conditional mean requires the integration over $\mathbb{R}^m$, where m represents the dimension of the image u. Usually, the dimension of the image, that we want recover is very large, therefore the numerical quadrature rules fail due to the exceeding the computational capacity of the computers. Another drawback of the quadrature rules is that they required the knowledge of the support of the probability distribution, which is in most of the cases unknown (since it is part of the information that we want to recover). For this reason, for computing the conditional mean estimate we will use the Monte Carlo integration. Given a sequence of N samples u_i, for $i = 1, \ldots, N$, distributed as the posterior distribution, from the law of large numbers we have

$$\mathbb{E}[u] = \int up(u|f) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} u^i, \tag{4.2}$$

for any measurable f almost certainly, [55].

A problem that can arise is the difficulty of drawing samples from the posterior, since it is known only up to a normalization factor and it usually does not belong to a class of distributions for which independent sampling schemes are known. To overcome this problem, one can employ the strong ergodic theorem, and we have that (4.2) holds if the sequence is dependent but generated from an ergodic Markov chain that has $p(u|f)$ as its equilibrium distribution. For the ergodic theory we refer the reader to [54]. Systematic ways of generating a sample ensemble such that (4.2) holds is called the Markov chain Monte Carlo techniques (MCMC for short). The most common procedures to generate a Markov chain are the Metropolis-Hastings algorithm and the Gibbs algorithm, [55]. For simplicity, we will denote the posterior by using $p(x)$, $x \in \mathbb{R}^n$.

4.3.1 Metropolis-Hastings sampling

Pick the initial value $x^1 \in \mathbb{R}^n$. For $k = 1, \ldots, N$, where N indicates the number of samples, repeat the following procedure

1. Draw $y \in \mathbb{R}^n$ from a proposal distribution $q(x^k, y)$.
2. Compute the acceptance ratio

$$r(x^k, y) = \min\left(1, \frac{p(y)q(y, x^k)}{p(x^k)q(x^k, y)}\right).$$

3. Draw $t \in [0, 1]$ from a uniform probability density.
4. If $r(x_k, y) \geq t$, set $x^{k+1} = y$ else $x^{k+1} = x^k$.

Note that, to compute the acceptance ratio we just need to know $p(x)$ up to a scaling factor. Furthermore, if the proposal distribution is symmetric, i.e. $q(x, y) = q(y, x)$ with $x, y \in \mathbb{R}^n$, then the acceptance ratio can be simplified as following

$$r(x^k, y) = \min\left(1, \frac{p(y)}{p(x^k)}\right).$$

The main difficulty of this algorithm is finding a good proposal distribution. In fact, the algorithm works only if the proposal distribution $q(x, y)$ leads to a chain that moves fast in the sampling space, in this way the region would be explored reasonably fast and the consecutive samples would be uncorrelated as possibles, see [55]. A very common way of choosing the proposal distribution is given by

$$q(x, y) \propto \exp\left(-\frac{1}{2\gamma^2}\|x - y\|^2\right),$$

therefore, in step 1, y is set as follows

$$y = x + w, \qquad \text{where} \qquad w \sim \mathcal{N}(0, \gamma^2 I).$$

4.3.2 Gibbs sampling

However, there may exist some situations in which the direct sampling of a m-dim multivariate distribution is not possible or computational too expensive. In this case the Gibbs sampling can be employed. The basic idea behind the Gibbs algorithm is that the samples are drawn not from the multivariate distribution, but from some conditioned versions of the multivariate distribution. Before describing the Gibbs sampling we need to introduce some notations. Let $I = \{1, 2, \ldots, n\}$ be the index set of $\mathbb{R}^n$ and $I = \cup_{j=1}^{s} I_j$ a partitioning of the index set into disjoint nonempty subsets. Denoting κ_j the number of the elements in I_j, we can partition $\mathbb{R}^n = \mathbb{R}^{\kappa_1} \times \cdots \times \mathbb{R}^{\kappa_m}$ and

$$x = [x_{I_1}; \ldots; \ldots x_{I_m}] \in \mathbb{R}^n, \text{ with } x_{I_j} \in \mathbb{R}^{\kappa_j},$$

where the components of x are rearranged so that $x_i \in \mathbb{R}$ is a component of the vector x_{I_j} if and only if $i \in I_j$. In the following we will use the notation with the hat to indicates that the corresponding elements are not included in the vector, i.e.

$$\begin{aligned} x_{-I_j} &= [x_{I_1}; \ldots; \widehat{x_{I_j}}; \ldots; x_{I_m}] \\ &= [x_{I_1}, \ldots; x_{I_{j-1}}, x_{I_{j+1}}; \ldots; x_{I_m}]. \end{aligned}$$

In the Gibbs sampling algorithm the samples are generated as follows: Pick the initial value $x^1 \in \mathbb{R}^n$. For $k = 1, \ldots, N$, where N indicates the number of samples, repeat the following procedure

1. Set $x = x^k$.
2. For $1 \leq j \leq n$, draw $y_{I_j} \in \mathbb{R}^{k_j}$ from the k_j-dimensional distribution $p(y_{I_j}|y_{I_1}, \ldots, y_{I_{j-1}}, x_{I_{j+1}}, \ldots, x_{I_n})$.
3. Set $x^{k+1} = y$.

Although the Gibbs sampler does not require any proposal distribution, it is often seen as a special case of the Metropolis-Hastings algorithm, where the main difference is that in the Gibbs sampler the proposal is always accepted. On the other hand, if the proposal distribution q is simple to handle, the drawing process in the Metropolis-Hastings algorithm is easier and less time consuming than the Gibbs sampler.

The convergences of the previous sampling methods strongly depends on the size of the sample. In fact, it is not simple to decide when the sample in the MCMC is large

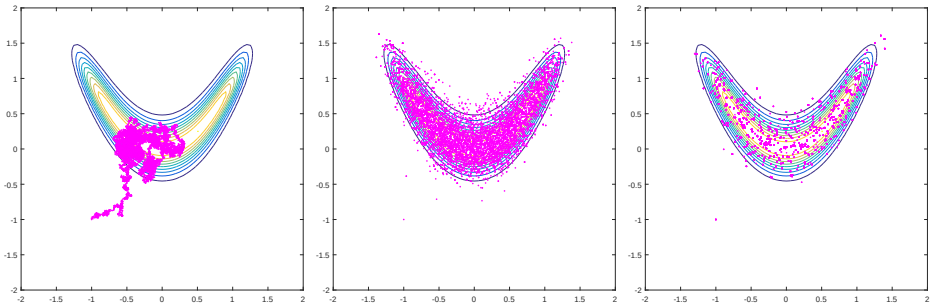


Figure 4.4: From left to the right: samples ($N = 10000$) from the random walk Metropolis-Hastings algorithm with different step sizes $\gamma = 0.01, 0.3, 2$.

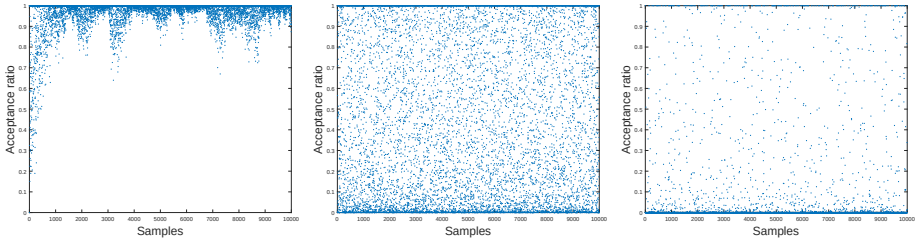


Figure 4.5: From left to the right: acceptance ratio for $N = 10000$ samples with $\gamma = 0.01, 0.3, 2$, respectively.

enough to cover the probability distribution p. For more details about the convergence see [55, Sect. 3.6.4].

Here we include two examples that show how the Metropolis-Hasting algorithm works.

Example 4.3.1. In this example, we show how the Metropolis-Hastings algorithm works with the density p in $\mathbb{R}^2$, defined as follows

$$p(x) = \exp\left(-10(x[1]^2 - x[2])^2 - \left(x[2] - \frac{1}{4}\right)^4\right), \quad \text{with} \quad x = (x[1], x[2]) \in \mathbb{R}^2.$$

The contour plot of the density p is shown in Fig. 4.4. For this simulation we draw $N = 10000$ samples and we initialize the algorithm to the point $x^1 = (-1, -1) \in \mathbb{R}^2$. For the proposal distribution q we use the normal distribution. In Fig. 4.4, we show three random walks obtained with different step sizes γ, i.e. $\gamma = 0.01, 0.3, 2$. From the figure, it is clear that when the step size is too small ($\gamma = 0.01$) the density is not full explored and it also takes a while before moving from the initial point x_1 and drawing right sample from the distribution. On the other hand, when the step size is too big ($\gamma = 2$) the random walk move too fast and only a few samples are accepted. To further point out the influence of the step size γ, in Fig. 4.5 we present the acceptance ratio. It clearly shows that when the step size γ is too big the proposals are very often not accepted, while when it is too small they are usually accepted.

Example 4.3.2. The conditional mean estimate Given a nosy image $f \in \mathbb{R}^m$ a way to compute the condition mean (CM) estimate is to use the Metropolis-Hastings

algorithm. In fact, by using the Metropolis-Hastings algorithm, we can draw N samples from the posterior $p(u|f)$. Then, for N big enough, the conditional mean estimate can be calculated by computing the weighted sum of the samples, i.e.

$$\hat{u}_{\mathrm{CM}} \approx \frac{1}{N} \sum_{i=1}^{N} u^i,$$

where u^i for $i = 1, \ldots, N$ are the samples drawn from the posterior density. This method seems quite simple to solve, the main issues are the how to find the proposal distribution $q(x, y)$ in the Metropolis-Hastings algorithm and the value of N.

CHAPTER 5

Contributions

This thesis is aimed at studied non-Gaussian noise models in the image processing field. The main contributions regard the introduction of a parameter-free two phase method for impulse noise removal, the study of a variational method for Cauchy noise removal and the study of the maximum-a-posterior (MAP) and the conditional mean (CM) estimates for non-Gaussian noise models. In this chapter, for each of our works we give an overview of the topic, we highlight the most significant results and we conclude with some possible future works on that topic. For a more detail description of our works we refer the reader to the papers enclosed at the end of this thesis.

5.1 Impulse noise

We focus on degraded image with impulse noise. We will start giving an overview of the noise model and the variational model for reconstructing images corrupted by impulse noise. Then we will show our contribution in this field and we will conclude given some ideas for future works.

5.1.1 Overview

Impulse noise is a common type of image degradation due to, e.g., malfunctioning pixel elements in the camera sensors, errors in analogue-to-digital conversion, faulty memory locations in hardware, or transmission errors [7]. A characteristic property of impulse noise is that a certain number of pixels are uncorrupted and the noise-corrupted pixels contain no information about the true pixel value. Since our paper on impulse noise is only in the discrete case, here we focus directly on the discrete.

Impulse noise can be described as a stochastic degradation process of the form

$$f_k = \begin{cases} \eta_k & \text{with probability } r \\ \tilde{u}_k & \text{with probability } 1-r \end{cases}, \quad k \in \Omega$$

where $\tilde{u} \in \mathbb{R}^m$ is the original image, $f \in \mathbb{R}^m$ is the corrupted image, and $\eta \in \mathbb{R}^m$ is the noise, which is independent from the original image $\tilde{u}$. We refer to the parameter $r \in [0, 1]$ as the noise level since it can be interpreted as the probability that a pixel is

Figure 5.1: From the left to the right: original image, noisy image corrupted with salt-and-pepper noise $r = 0.3$, noisy image corrupted with random-valued impulse noise $r = 0.3$.

corrupted. Notice that some pixels remain unchanged, and the pixels that are corrupted by noise carry no information about the noise-free image.

The two most common kinds of impulse noise degradation are given by the salt-and-pepper noise and the random-valued impulse noise. Given $u_{\min} = \min_k \tilde{u}_k$ and $u_{\max} = \max_k \tilde{u}_k$, then for the salt-and-pepper noise

$$\eta_k = \begin{cases} u_{\min} & \text{with probability } \frac{1}{2} \\ u_{\max} & \text{with probability } \frac{1}{2}, \end{cases} \quad k \in \Omega,$$

while for the random-valued noise

$$\eta_k \sim \mathcal{U}([u_{\min}, u_{\max}]), \quad k \in \Omega,$$

where $\mathcal{U}([u_{\min}, u_{\max}])$ represents the uniform distribution in $[u_{\min}, u_{\max}]$. In Fig. 5.1, we show an example of image corrupted by salt-and-pepper and by random-valued impulse noise with $r = 0.3$. Although we use the same noise level, the image corrupted by salt-and-pepper noise looks more noisy that the other one, this is because, noisy pixels corrupted by salt-and-pepper noise take the lowest or the highest pixel value (i.e., $u_{\min}$ or $u_{\max}$), whereas the degraded pixels with random-valued impulse noise have values anywhere in the interval from $u_{\min}$ to $u_{\max}$. Thus, the image with salt-and-pepper noise looks more corrupted, even if it has the same rate of corruption as the random-valued impulse noise. On the other hand, since the random-valued impulse noise may assume any value in the gray-level range, it is more difficult to detect than the salt-and-pepper noise, which assumes only the maximum and the minimum of the range.

Over the years, many nonlinear digital filters methods have been proposed, see [1]. The most common filters used to remove impulse noise are the median-type filters: median filter [71], weighted median filter [11], adaptive median filter (AM) [53], multistate median filter [28], center weighted median filter [56] and adaptive center-weighted median filter (ACWM) [29]. Basically, these filters are based on a sliding window and the pixels are reconstructed by using the values of the neighbourhood. In Fig. 5.2, we give an example of reconstruction by using the median filter and the adaptive median filter (if the image is degraded with salt-and-pepper noise) and the adaptive center-weighted median filter (if the image is degraded with random-valued impulse noise). Here, it is clear that the images corrupted by salt-and-pepper are easier to reconstruct than the

Figure 5.2: From the left to the right: reconstruction by using the median filter, the adaptive median filter and the adaptive center-weighted median filter. The clean and noisy images are represented in Fig. 5.1. For the first two reconstructions we use the image corrupted with salt-and-pepper noise while for the last we use the image degraded with random-valued impulse noise.

one with random-valued impulse noise. Although these filters are efficient and easy to implement, they cannot achieve good results in general, in particular they are not able to restore blurred images and they do not preserve the image edges well.

In order to preserve the edges, in 2004, Nikolova [67] proposed a variational model which combines an ℓ^1-data fidelity term with total variation (TV), which has been in shown in [66, 67] to work better than the classical ℓ^2-term, [73].

The resulting reconstruction problem is convex and takes the following form

$$\min_{u\in\mathbb{R}^m} \|u-f\|_1 + \lambda\,\mathrm{TV}(u), \tag{5.1}$$

where $\|u-f\|_1 = \sum_{k\in\Omega}|u_k - f_k|$ is the data-fidelity term, $\mathrm{TV}(u)$ is a regularization term, and $\lambda > 0$ is a regularization parameter.

The ℓ^1-TV model has some nice properties, such as contrast preservation, multiscale decomposition and morphological invariance [21, 66, 91]. However, the main disadvantage of this approach is that we have to reconstruct all the pixels of the image, including the ones that are noise-free. Furthermore, including the noisy pixels in the data-fidelity term introduces errors since the noise-corrupted pixels contain no information about the true image. To address this issue, Chan et al. [27] studied a two-phase method (the CDH method) in which they first detect the noisy pixels (phase 1) and then exclude these pixels from the data-fidelity term when computing a reconstruction (phase 2). Thus, in the first phase, they use a detector (an AM filter for salt-and-pepper and an ACWM filter for random-valued impulse noise) to split the domain Ω into two sets: $\mathcal{N}$ that includes all indices of the corrupted pixels and $\mathcal{U}$ that includes the indices of the noise-free pixels. We will henceforth assume that there are $|\mathcal{N}| = n$ noisy pixels and $|\mathcal{U}| = m-n$ noise-free pixels. In the second phase, for the denoising case (here, instead of the blurring operator we consider the identity), they reconstruct the image based on the following model

$$\min_{u\in\mathbb{R}^m} \sum_{k\in\mathcal{U}} |u_k - f_k| + \lambda\,\mathrm{TV}(u). \tag{5.2}$$

The main advantage of the CDH method is that the noise detector improves the data-fidelity term in ℓ^1-TV model (5.1), and this often yields a great improvement in terms of

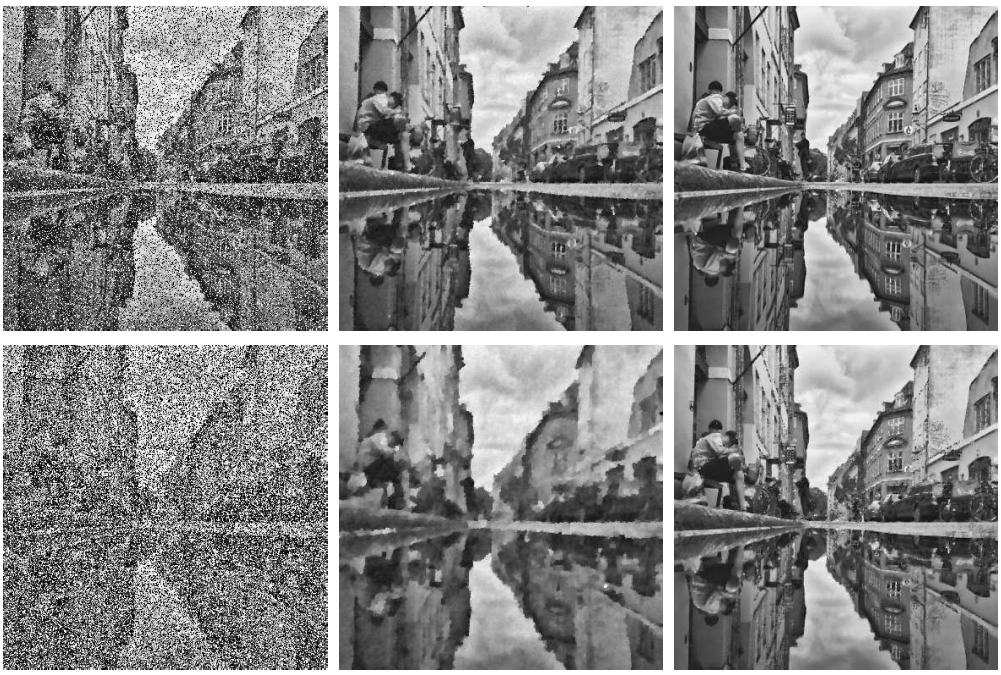


Figure 5.3: From the left to the right: noisy image with salt-and-pepper noise (first row: $r = 0.3$ and second row: $r = 0.6$); reconstruction given by the ℓ^1-TV; reconstruction given by the two phase method. The clean image is represented in Fig. 5.1.

restoration capabilities. Furthermore, the ℓ^1-norm in the data-fidelity term allows many noisy-free pixels to maintain their exact values.

In Fig. 5.3, we show the reconstruction for ℓ^1-TV and the two-phase method for two different noise levels, $r = 0.3, 0.6$. It is clear, especially when the noise level is high, that the two-phase method outperforms the ℓ^1-TV. In particular, when $r = 0.6$, in the recovered image given by the two-phase method we can still recognize the windows in the buildings, the cars and even the scratches in the wall, while the one given by the ℓ^1-TV looks very blurry and we can barely recognize the two guys sitting on the left of the image.

Even if the two-phase method seems to work quite well, some drawbacks are still present. In fact, the presence of a regularization parameter in the model necessitates multiple reconstructions or tests in order to find a good choice for the parameter. Moreover, the problem in (5.2) includes all pixels of the image as variables, including the ones that are assumed to be free of noise.

5.1.2 Relevant paper

In this framework our contribution regards the study of a new two phase method for deblurring and denoising impulse noise. In particular, to overcome the disadvantages of the CDH method, we propose to alter the second phase of the method such that the noise-free pixels are required to be equal to their known values. The main characteristic

Figure 5.4: First column: given images. Second column: reconstructions given by [41] (Cameraman: PSNR=28.78 and Parrot: PSNR=27.73). Third column: our reconstructions (Cameraman: PSNR=28.86 and Parrot: PSNR=27.75).

of our method is that it is regularization parameter-free and that in the denoising case is faster than the other two-phase methods.

For more details we refer the reader to the following paper:

[A] F. Sciacchitano, Y. Dong, and M. S. Andersen, *Total variation based parameter-free model for impulse noise removal*, Numerical Mathematics: Theory, Methods and Applications; to appear.

5.1.3 Perspective and some future works

The method presented in [A] can be applied also for the problem of missing data. The missing data can occur when some pixel elements in the camera sensors do not work well, during the transmission, the analogue-to-digital conversion etc, see [7] for more details. The missing regions can have arbitrary size, but, usually, the location is known. Therefore, one could think to apply the second phase of model in [A]. We remark that, since the location of the regions is known, we do not need the detection in the first phase. Some examples of missing data are image inpainting and super-resolution reconstruction.

Image inpainting corresponds to the problem to fill in missing regions in an image, see for instance [25, 26, 41]. The missing region, which need to be inpaintend, is usually called mask and it is denoted with $\mathcal{M}$. In general, the variational model for image inpainting in the discrete case can be seen as

$$\min_u \sum_{k\in\Omega\setminus\mathcal{M}} |u_k - f_k|^2 + \lambda TV(u).$$

Figure 5.5: From left to right: given image; double size image filled with zeros; reconstruction given by the Matlab function *imresize* (PSNR=23.76); our reconstruction (PSNR=24.93).

In Fig. 5.4, we show an example of image inpainting problem. Here, we compare the results given by our method with the one in [41]. As we can see from the pictures and the PSNR values, both methods yield good reconstruction.

Super-resolution reconstruction corresponds to generate a higher resolution image from lower resolution images, in other words, it represents the procedure to fill in gray-level values of the missing pixels after up-sampling (for more details we refer to [23, 65], and their reference therein). The high resolution can be applied in different fields, for instance for the diagnosis of the medical imaging, for surveillance, forensic and satellite imaging applications.

In Fig. 5.5, we show an example of super-resolution reconstruction. The given image is a 128-by-128 gray level image of the Parrot. The second image in the figure represents a 256-by-256 gray level image, where the missing pixels are filled with zero values. In the last two images we show the super-resolution reconstructions given by the Matlab function imresize and our approach. Note that, as for the inpainting problem, since we know the location of the missing data, we do not need the first phase of our method.

The proposed two-phase method can be likely also applied for ring artifacts removal in computed tomography (CT). In CT, an image of an object is reconstructed from projections obtained by measuring the attenuation of the x-rays passed through the object, see [15]. During the tomographic acquisition process some artifacts are often generated in the CT image. In particular, due to the malfunctioning and miscalibration of the detector pixels, impurities and dust on the scintillator screens, and the non-linear response of individual detector elements, stripe artifacts occur in the sinogram. During the reconstruction process, these artifacts generate concentric ring artifacts in the reconstructed CT images. The interesting thing is that the stripes in the sinogram can be seen as a missing data, therefore one could try to detect these artifacts and then try to reconstruct the object and remove the stripes simultaneously, by using a modified version of the proposed two-phase method.

5.2 Cauchy noise

As in the previous chapter, first we describe the Cauchy degradation and then we give an overview of some open questions.

5.2.1 Overview

Many studies in image and signal processing rely on the fundamental assumption that the noise follows a Gaussian distribution. This hypothesis is justified due to the existence of the central limit theorem, see [44]. Unfortunately, most of the real world problems cannot be modeled by Gaussian distribution, since the noise is much more impulsive than the one that is modeled by additive Gaussian noise. Examples of these applications can be found in the radar and sonar applications, where there are atmospheric and underwater acoustic noises, in biomedical images, in SAR images and so on. These types of noise follow the so called alpha-stable distributions [68, 69, 75].

The alpha-stable distributions are closed under additions, i.e. the sum of two alpha-stable random variables is still an alpha-stable random variable. Moreover, the alpha-stable random variables obey to the Generalized Central Limit Theorem [69]. But, this class of random variables has no close formula for densities and distribution functions (apart from Gaussian, Cauchy and Lévy distributions). The easiest and most common way to define these distributions is through the characteristic function or Fourier transform, see [75].

Generally speaking, an alpha-stable distribution is characterized by four parameters: an index of stability $\alpha \in (0, 2]$, a skewness parameter $\beta \in [-1, 1]$, a scale parameter $\gamma > 0$ and a location parameter $\delta \in \mathbb{R}$. When $\alpha \in (1, 2]$, the location parameter δ corresponds to the mean of the distribution, otherwise δ corresponds to its median. The scale parameter determines the spread of the distribution around δ and it plays a similar rule as the variance in the Gaussian distribution. If $\gamma = 1$ and $\delta = 0$ we say that the distribution is standardized, furthermore if $\beta = 0$ the distribution is symmetric around zero, in this case we call it symmetric alpha-stable distributions. Since Cauchy distribution belongs to the symmetric alpha-stable distribution, from now on we will focus only on symmetric alpha-stable distributions.

The distributions of this class are all bell-shaped, with increasing density on the left and decreasing on the right. The heaviness of the distribution tails is controlled by the parameter $\alpha \in (0, 2]$, i.e., the tails grow thicker as α becomes smaller.

In Figure 5.6, we show the probability density functions (PDFs) of alpha-stable distributions with different values of α and γ. The distribution with $\alpha = 2$ corresponds to the well-known Gaussian distribution with variance $\sigma^2 = 2\gamma^2$ and the one with $\alpha = 1$ corresponds to the Cauchy distribution. Comparing the PDFs, we see that the tails of the bells become heavier as α decreases. In fact, the Cauchy bell ($\alpha = 1$) has thicker tail than the Gaussian distribution ($\alpha = 2$). Thus, the rare events have more probability of occurring in the Cauchy bell curve than in the Gaussian bell curve and for this reason, the noise generated from the Cauchy distribution is more impulsive than the Gaussian one.

To further illustrate the difference between the symmetric α-stable distribution, in Figure 5.7 we show the realizations for different α values: $\alpha = 2, 1.5, 1, 0.5$ with fixed $\gamma = 1$ and $\delta = 0$. For having a better comparison, we use the same vertical scale for all the images. From the figures, one can see that the spikes increases when α decreases. Thus, Cauchy noise is more impulsive than the Gaussian noise.

Finally, we now describe how Cauchy noise influences the clean image. Given the original image $\tilde{u} : \Omega \to \mathbb{R}$, with $\Omega \subset \mathbb{R}^2$ be a bounded, open and connected set with compact Lipschitz boundary, the noisy image $f : \Omega \to \mathbb{R}$ is given by,

$$f = \tilde{u} + v,$$

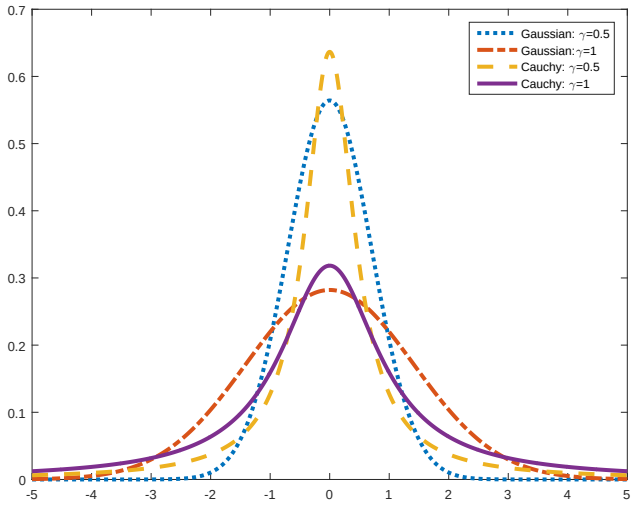


Figure 5.6: Comparison of the PDFs of symmetric alpha-stable distributions with $\delta = 0$, $\alpha = 2$ (Gaussian) and $\alpha = 1$ (Cauchy) and $\gamma = 0.5, 1$.

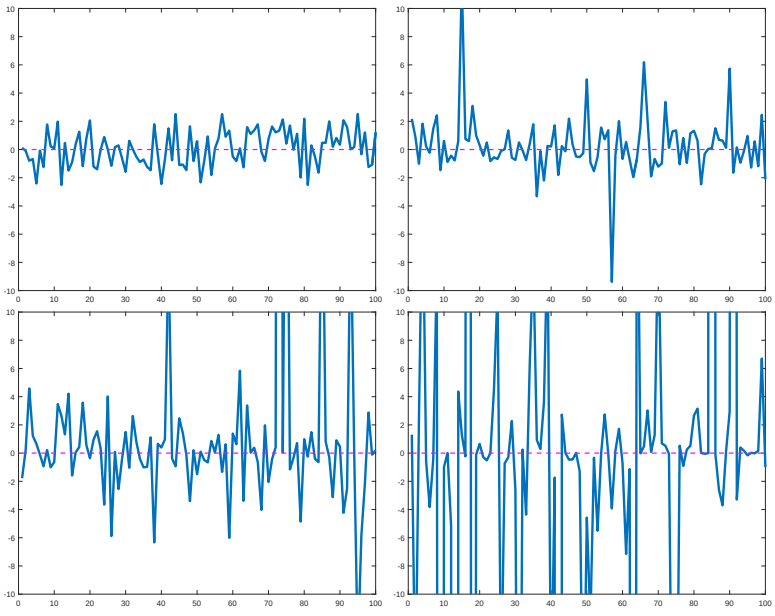


Figure 5.7: Symmetric alpha-stable realizations with $\delta = 0$ and $\gamma = 1$. From left to the right: $\alpha = 2$ (Gaussian), $\alpha = 1.5$, $\alpha = 1$ (Cauchy) and $\alpha = 0.5$.

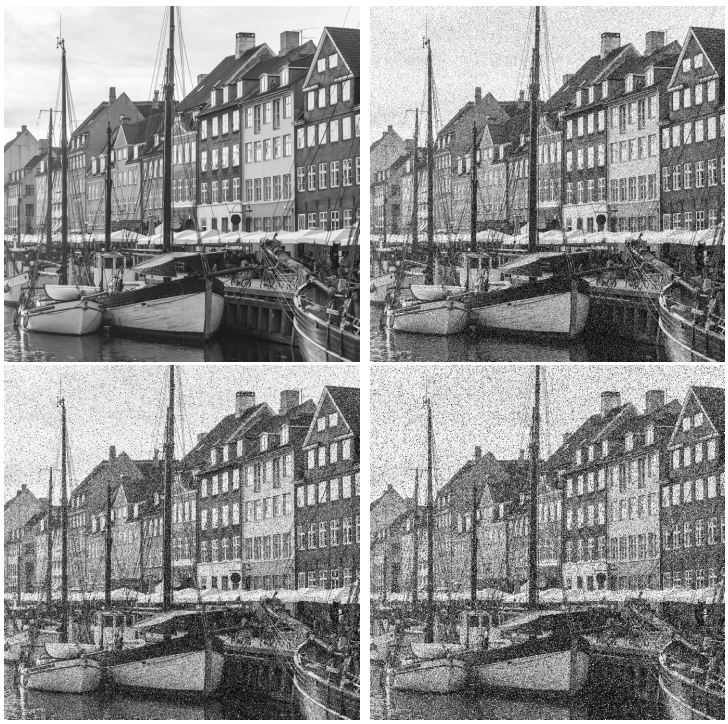


Figure 5.8: From left to the right: original image, noisy images corrupted by symmetric alpha-stable realizations with $\delta = 0$, $\gamma = 0.1$ and $\alpha = 2$ (Gaussian), $\alpha = 1.5$, and $\alpha = 1$ (Cauchy).

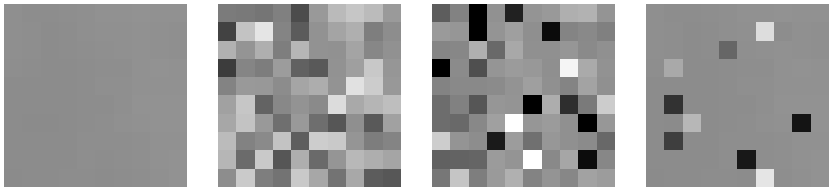


Figure 5.9: Comparison of different 10-by-10 noisy images. From left to right: clean image; degraded image by an additive Gaussian noise; degraded image by an additive Cauchy noise; degraded image by random-valued impulse noise.

where v represents the random noise that models a Cauchy distribution. A random variable V follows the Cauchy distribution, $V \sim \text{Cauchy}(\gamma, \delta)$, if it has density as in (2.1). Without loss of generality, from now on, in our analysis we consider $\delta = 0$.

In Fig. 5.8, we show an example of image corrupted by symmetric alpha stable distributions with $\gamma = 0.1$ and different α values. Although we use the same noise level γ, the image corrupted by Cauchy noise is much more impulsive than the other ones.

Finally, to further show the impulsive behaviour of Cauchy noise, in Fig. 5.9, we compare it with Gaussian noise and impulse noise. In the impulse noise and in the Cauchy noisy there are some degraded pixels that assume values either white or black, while the image corrupted by Gaussian noise is uniformly modified and white and black

pixels are very rare. On the other hand, for Cauchy noise and Gaussian noise, all the pixels of the image are corrupted by noise, while in the impulse noise some pixels are noise free.

5.2.2 Relevant paper

Here, our contribution regards the introduction of a variational method to restore blurred images corrupted with Cauchy noise. Since the model is non-convex, we suggest a convexified version of it by adding a quadratic penalty term .

For more details we refer the reader to the following paper:

[B] F. Sciacchitano, Y. Dong, and T. Zeng, *Variational approach for restoring blurred images with Cauchy noise*, SIAM Journal on Imaging Sciences 8.3 (2015): 1894-1922.

5.2.3 Perspective and some future works

There are some open questions about Cauchy noise.

A deeper study of the scale parameter γ will be very interesting. In fact, since the scale parameter appears explicitly in the data fidelity term, in order to reconstruct an image an estimation of γ is required.

Another important topic would be the study of the regularization parameter for the proposed variational model. For instance, by using the similar argument in [5, 83], a discrepancy principle for the Cauchy model may be formulated. Furthermore, since for some pixels the image corrupted by Cauchy noise is highly corrupted, an extension to the spatially dependent regularization parameter would be interesting. A future work regards the estimation of the regularization parameter and the scale parameter by using a learning approach, [32].

Finally, since Gaussian noise and Cauchy noise belongs to the alpha-stable distributions, a generalization of the proposed variational model to the symmetric alpha-stable distribution would be interesting. Here, the main difficulty is the lack of close formula of the alpha-stable distributions.

5.3 Bregman cost for non-Gaussian noise

Here, we focus on the Bayes formalism. We start by giving the definition of Bayes cost and Bayes estimate, then we show our contribution.

5.3.1 Overview

One of the tasks of the Bayesian inverse problem is to find a good estimate based on the posterior probability density. The most common point estimators are the conditional mean (CM) and maximum a posteriori (MAP) estimates, which correspond to the mean and the mode of the posterior, respectively. As we saw in the previous chapters, from the numerical point of view, the MAP estimate can be computed rather easily, while the CM estimate requires to solve a much harder and more time-consuming integration problem. Although there are numerical challenges to calculate the CM estimate, it has

many theoretical benefits. Before introducing the advantages of the CM estimate, we need some theoretical remark about the Bayes cost formalism.

The Bayesian estimation of u from the given noisy image f, relies on the minimization of a Bayes cost, which is defined as follows

$$\begin{aligned} BC_C(\hat{u}) :=& \mathbb{E}[C(u,\hat{u})] \\ &= \int\int C(u,\hat{u})p(u,f)\mathrm{d}u\mathrm{d}f \\ &= \int\int C(u,\hat{u})p(u|f)\mathrm{d}up(f)\mathrm{d}f, \end{aligned}$$

where $C : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a mapping that measures the cost of estimating $\hat{u}$ by f instead of the true image. $C(u,\hat{u})$ is called the cost function or loss function, since it shows the loss when the true value is $\tilde{u}$ but we estimate it by f. In Bayesian estimation framework, a Bayes estimator $\hat{u}_C$ is the estimator which minimizes the Bayes cost function $BC_C(\hat{u})$, i.e.

$$\hat{u}_C := \operatorname*{argmin}_{\hat{u}} BC_C(\hat{u}).$$

Since $\hat{u}$ only depends on the noisy image f and $p(f)$ is non-negative, the Bayesian estimator can be computed as follows

$$\hat{u}_C := \operatorname*{argmin}_{\hat{u}} \int C(u,\hat{u})p(u|f)\mathrm{d}u = \operatorname*{argmin}_{\hat{u}} \mathbb{E}[C(u,\hat{u})].$$

Therefore, in order to compute the Bayesian estimator we need to find a suitable cost function. One of the most common choice for the cost function is given by the mean squared error, i.e.

$$C(u,\hat{u}) = \|u-\hat{u}\|_2^2. \tag{5.3}$$

The conditional mean estimate $\hat{u}_{\mathrm{CM}}$ is the Bayes estimator for the mean squared error, in fact

$$\begin{aligned} \int \|u-\hat{u}\|_2^2 p(u|f)\mathrm{d}u =& \int \|u-\hat{u}_{\mathrm{CM}}\|_2^2 p(u|f)\mathrm{d}u + \int \|\hat{u}-\hat{u}_{\mathrm{CM}}\|_2^2 p(u|f)\mathrm{d}u \\ &- 2\int \langle u-\hat{u}_{\mathrm{CM}}, \hat{u}-\hat{u}_{\mathrm{CM}}\rangle p(u|f)\mathrm{d}u \\ =& \int \|u-\hat{u}_{\mathrm{CM}}\|_2^2 p(u|f)\mathrm{d}u + \int \|\hat{u}-\hat{u}_{\mathrm{CM}}\|_2^2 p(u|f)\mathrm{d}u \end{aligned}$$

where the last equality holds by definition of CM estimate. Thus, since the first term is independent of $\hat{u}$, we have that the minimizer is given by $\hat{u} = \hat{u}_{\mathrm{CM}}$.

Another popular choice for the cost function is the uniform cost, defined as follows

$$C(u,\hat{u}) = \begin{cases} 0 & |u_k - \hat{u}_k| < \epsilon \quad \text{for all } k, 1 \le k \le n \\ 1 & \text{otherwise,} \end{cases}$$

where $\epsilon > 0$ is a small constant. It turns out that the MAP estimate is asymptotically a solution to the Bayes cost optimization with the uniform cost function, in fact

$$\int_{|u_k-\hat{u}_k|>\epsilon} p(u|f)\mathrm{d}u = 1 - \prod_{k=1}^{n} \int_{\hat{u}_k-\epsilon}^{\hat{u}_k+\epsilon} p(u|f)\mathrm{d}u_k \approx 1 - (2\epsilon)^n p(\hat{u}|f),$$

where the last approximation holds by the mean value theorem. Thus, minimizing the Bayes cost is equivalent to computing the MAP estimate.

Recently, in [12] it has been proved that the MAP estimate is a proper Bayes estimator for

$$C(u, \hat{u}) = \|K(\hat{u} - u)\|_2^2 + 2\lambda D_J^q(\hat{u}, u), \tag{5.4}$$

where K is the blurring operator, λ is the regularization parameter and D_J represents the Bregman distance for the regularization term J, which is defined

$$D_J^q(\hat{u}, u) = J(\hat{u}) - J(u) - \langle q, \hat{u} - u \rangle,$$

for the subgradient $q \in \partial J(u)$. We will often refer to the Bregman distance by omitting q, i.e. $D_J(\hat{u}, u)$. If J is Fréchet differentiable in u, than the subgradient q corresponds to the standard Fréchet derivative J'. The Bregman distance has been introduced in [10] and it is a very useful tool in image processing, see for instance [13, 14, 46]. Although it is called distance it is not a distance in the mathematical sense, since it is not symmetric and the triangle inequality does not hold. On the other hand, some nice properties hold: $D_J(\hat{u}, u) \geq 0$; if J is strictly convex, $D_J(\hat{u}, u) = 0$ implies $\hat{u} = u$; $D_J(\hat{u}, u)$ is convex in $\hat{u}$. We underline that in [12], it has only been proven that the MAP estimate is a proper Bayes estimator in the additive Gaussian noise case, i.e. when the data fidelity term is the L^2 norm.

Moreover, in [12] the authors show that the MAP estimate outperforms the CM estimate when the error is measured using the Bregman distance $D_J(\hat{u}, u)$ and that the CM estimate outperforms the MAP estimate when the error is measured using the quadratic distance $\|K(\hat{u} - u)\|_2^2$. In particular, we have

$$\begin{aligned} \mathbb{E}\Big[\|K(\hat{u}_{\mathrm{CM}} - u)\|_2^2\Big] &\leq \mathbb{E}\Big[\|K(\hat{u}_{\mathrm{MAP}} - u)\|_2^2\Big] \\ \mathbb{E}\Big[D_J(\hat{u}_{\mathrm{MAP}}, u)\Big] &\leq \mathbb{E}\Big[D_J(\hat{u}_{\mathrm{CM}}, u)\Big]. \end{aligned}$$

Since the CM estimate is a Bayes estimator for the mean squared error (see (5.3)), the first inequality is straightforward. Regarding the second inequality, by using the fact that the MAP estimate is a Bayes estimator for the cost function in (5.4), we have

$$\begin{aligned} \int \Big(\|K(\hat{u}_{\mathrm{MAP}} - u)\|_2^2 + 2\alpha D_J^q(\hat{u}_{\mathrm{MAP}}, u) \Big) p(u|f) \mathrm{d}u \\ \leq \int \Big(\|K(\hat{u}_{\mathrm{CM}} - u)\|_2^2 + 2\alpha D_J^q(\hat{u}_{\mathrm{CM}}, u) \Big) p(u|f) \mathrm{d}u \\ \leq \int \Big(\|K(\hat{u}_{\mathrm{MAP}} - u)\|_2^2 + 2\alpha D_J^q(\hat{u}_{\mathrm{CM}}, u) \Big) p(u|f) \mathrm{d}u, \end{aligned}$$

where the last inequality follows from the fact that the mean squared error is a cost function for the CM estimate.

The results in [12] give some new points of view for the MAP and the CM estimate, but all the results are achieved under the assumption that the image is corrupted by Gaussian noise. In our contribution we extend some results of [12] to a more general noise model.

5.3.2 Relevant paper

With only the convexity assumption on the data fidelity term, we introduce some Bregman cost functions for which the CM and MAP estimates are proper Bayes estimators. Moreover, we also prove that the CM estimate outperforms the MAP estimate, when the error depends on some Bregman distance.

For more details we refer the reader to the following paper:

[C] M. Burger, Y. Dong, and F. Sciacchitano, *Bregman Cost for Non-Gaussian Noise*, to be submitted. Available on arXiv http://arxiv.org/abs/1608.07483

5.3.3 Perspective and some future works

This study has been done only from the theoretical point of view, a numerical validation of the results would be interesting. Furthermore, another very interesting point would be finding other costs functions for both MAP and CM estimates and try to see if there exist other relations between them.

CHAPTER 6

Summary/Conclusion

This PhD study investigates new methods in the image processing field for achieving sharp and noise-free images from the given data. Due to the popularity of the Gaussian noise, the main focus of the thesis is study of the non-Gaussian noise models. This project can be split into two parts, in the first one we propose variational methods to recover blurred images corrupted with impulsive noise models: the impulse noise and Cauchy noise; the second one is more theoretical and it deals with the Bayes estimators for reconstructing blurred images corrupted by non-Gaussian noise with convex data fidelity term.

For the impulse noise removal, we proposed a parameter-free two-phase method for restoring degraded images. In the denoising case, by separating the noisy pixels and the noise-free ones, we can recover the original image by simply solving an unconstrained problem with only the noisy pixels as variables. Therefore, especially for low noise-levels, the size of the problem is considerably reduced. Our method does not require any tuning of the regularization parameter and for the denoising case, it provides competitive results in less time when compared to the other two-phase methods. A very interesting future work would be the application of this method to the ring artifacts removal in computed tomography.

Concerning the Cauchy noise, we propose a convex variational model for deblurring and denoising of degraded images. Due to the convexity of the minimization problem a primal-dual algorithm has been employed to recover the original image. In future works, it would be very interesting to estimate the scale parameter γ of the Cauchy distribution. Moreover, a generalization of our variational model to the alpha-stable distribution could be useful. Furthermore, a study of a discrepancy principle for the Cauchy noise removal would be also very interesting.

The last project regards the study of the two typical point estimators for the posterior probability density: the conditional mean (CM) estimate and the maximum-a-posterior (MAP) estimate. Here, based on the Bregman distance, we propose new cost functions for which the CM and the MAP estimates are Bayes estimators. Furthermore, we give a new comparison result on these two estimates. An ongoing work deals with testing the theoretical results with some numerical simulations.

Bibliography

[1] J. Astola and P. Kuosmanen, *Fundamentals of Nonlinear Digital Filtering*, vol. 8, CRC, Boca Raton, FL, 1997.

[2] G. Aubert and J. Aujol, *A variational approach to removing multiplicative noise*, SIAM J. Appl. Math., 68, pp. 925–946, 2008.

[3] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing. Partial Differential Equations and the Calculus of Variations*, Appl. Math. Sci. 147, Springer, New York, 2006.

[4] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, *Image inpainting* Proceedings of the 27th annual conference on Computer graphics and interactive techniques. ACM Press/Addison-Wesley Publishing Co., 2000.

[5] M. Bertero, P. Boccacci, G. Talenti, R. Zanella, and L. Zanni, *A discrepancy principle for Poisson data*, Inverse problems, 26.10, 2010.

[6] J. Bioucas-Dias and M. Figueiredo, *Multiplicative noise removal using variable splitting and constrained optimization*, IEEE Transactions on Image Processing 19.7, pp. 1720-1730, 2008.

[7] A. Bovik, *Handbook of Image and Video Processing*, Academic Press, New York, 2000.

[8] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. and Trends Mach. Learning, 3, pp. 1–122, 2010.

[9] K. Bredies, K. Kunisch, and T. Pock, *Total generalized variation*, SIAM Journal on Imaging Sciences 3.3, pp. 492-526, 2010.

[10] L.M. Bregman, *The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming*, USSR computational mathematics and mathematical physics 7.3, pp. 200-217, 1967.

[11] D. Brownrigg, *The weighted median filter*, Comm. ACM, 27, pp. 807—818, 1984.

[12] M. Burger and F. Lucka, *Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper Bayes estimators*, Inverse Problems 30.11, 114004, 2014.

[13] M. Burger and S. Osher, *Convergence rates of convex variational regularization* Inverse problems 20.5, pp. 1411-1421, 2004.

[14] M. Burger, E. Resmerita, and L. He, *Error estimation for Bregman iterations and inverse scale space methods in image restoration*, Computing 81.2-3, pp. 109-135, 2007.

[15] T. Buzug, *Computed tomography: from photon statistics to modern cone-beam CT*, Springer Science and Business Media, 2008.

[16] J. Cai, R. Chan, and M. Nikolova, *Two-phase approach for deblurring images corrupted by impulse plus Gaussian noise*, Inverse Probl. Imaging, 2, pp. 187–204, 2008.

[17] D. Calvetti and E. Somersalo, *An Introduction to Bayesian Scientific Computing: Ten Lectures on Subjective Computing*, Vol. 2. Springer Science and Business Media, 2007.

[18] A. Chambolle, *An algorithm for mean curvature motion*, Interface Free Bound., 4, pp. 1–24, 2004.

[19] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vis., 20, pp. 89–97, 2004.

[20] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vis., 40, pp. 120–145, 2011.

[21] T. Chan and S. Esedoglu, *Aspects of total variation regularized L^1 function approximation*, SIAM J. Appl. Math., 65, pp. 1817–1837, 2005.

[22] T. Chan, A. Marquina, and P. Mulet, *High-order total variation-based image restoration*, SIAM J. Scient. Comput., 22.2, pp. 503-516, 2000.

[23] T. Chan, M. Ng, A. Yau, and A. Yip *Superresolution image reconstruction using fast inpainting algorithms*, Applied and Computational Harmonic Analysis 23.1, pp. 3-24, 2007.

[24] T. Chan and J. Shen, *Image Processing And Analysis: Variational, PDE, Wavelet, And Stochastic Methods*, SIAM Publisher, 2005.

[25] T. Chan and J. Shen, *Non-Texture Inpainting by Curvature-Driven Diffusions (CDD)*, Journal of Visual Communication and Image Representation 12.4, pp. 436-449, 2001.

[26] T. Chan, J. Shen, H. Zhou,*Total variation wavelet inpainting* Journal of Mathematical imaging and Vision, 25.1, pp. 107-125, 2006.

[27] R. Chan, Y. Dong, and M. Hintermüller, *An efficient two-phase L1-TV method for restoring blurred images with impulse noise*, IEEE Trans. Image Process., 19, pp. 1731–1739, 2010.

[28] T. Chen and H. Wu, *Space variant median filters for the restoration of impulse noise corrupted images*, IEEE Trans. Circuits Syst. II, 48, pp. 784-789, 2001.

[29] T. Chen and H. Wu, *Adaptive impulse detection using center-weighted median filters*, IEEE Signal Process. Lett., 8, pp. 1–3, 2001.

[30] C. Clason, *L^∞ fitting for inverse problems with uniform noise*, Inverse Problems, 28.10, 104007, 2012.

[31] L. Condat, *A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms*, J. Optim. Theory and Applic., 158, pp. 460–479, 2013.

[32] J. De los Reyes, L. Calatroni and C. Schönlieb, *Image denoising: Learning the noise model via nonsmooth PDE-constrained optimization*, Inverse Probl. Imaging 7.4, pp. 1139-1155, 2013.

[33] Y. Dong , M. Hintermüller, and M. M. Rincon-Camacho, *Automated regularization parameter selection in multi-scale total variation models for image restoration*, J. Mathem. Imaging and Vision 40.1, pp. 82-104, 2011.

[34] Y. Dong and S. Xu, *A new directional weighted median filter for removal of random-valued impulse noise*, IEEE Signal Proc. Lett., 14, pp. 193–196, 2007.

[35] Y. Dong and Tieyong Zeng, *A convex variational model for restoring blurred images with multiplicative noise*, SIAM Journal on Imaging Sciences 6.3, pp. 1598-1625, 2013.

[36] M. Elad and M. Aharon *Image denoising via sparse and redundant representations over learned dictionaries*, IEEE Tran. Image Process., 15, pp. 3736–3745, 2006.

[37] W. Feller *An introduction to probability theory and its applications*, Wiley, New York, 1968.

[38] B. Figueiredo and J. Bioucas-Dias, *Restoration of Poissonian images using alternating direction optimization*, IEEE Trans. Image Process., 19, pp. 3133-3145, 2010.

[39] A. Foi, *Noise estimation and removal in MR imaging: the variance-stabilization approach*, IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 1809–1814, 2011.

[40] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, no. 6, pp. 721–741, 1984.

[41] P. Getreuer, *Total Variation Inpainting using Split Bregman*, Image Processing On Line, 2, pp. 147–157, 2012.

[42] P. Getreuer, M. Tong, and L. Vese, *A variational model for the restoration of MR images corrupted by blur and Rician noise*, International Symposium on Visual Computing. Springer Berlin Heidelberg, 2011.

[43] G. Gilboa and S. Osher, *Nonlocal operators with applications to image processing*, Multiscale Model. Simul., 7, pp. 1005–1028, 2009.

[44] G. Grimmett and D. Welsh, *Probability: an Introduction*, Oxford Science Publications, London, 1986.

[45] T. Goldstein and S. Osher, *The split Bregman algorithm for L1-regularized problems*, SIAM J. Imaging Sci., 2, pp. 323–343, 2009.

[46] T. Goldstein and S. Osher, *The split Bregman method for L1-regularized problems* SIAM journal on imaging sciences 2.2, pp. 323-343, 2009.

[47] G. Golub, M. Heath, and G. Wahba, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics 21.2, pp. 215-223, 1979.

[48] H. Gudbjartsson and S. Patz, *The Rician distribution of noisy MRI data*, J.Magn. Reson. Med. 34.6, pp. 910–914, 1995.

[49] P.C. Hansen, *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*, Vol. 4, SIAM, 1998.

[50] P.C. Hansen, *Discrete inverse problems: insight and algorithms*, Vol. 7. Siam, 2010.

[51] P.C. Hansen and D.P. O'Leary *The use of the L-curve in the regularization of discrete ill-posed problems*, SIAM Journal on Scientific Computing 14.6, pp. 1487-1503, 1993.

[52] M. Hintermüller, K. Ito and K. Kunisch *The primal-dual active set strategy as a semismooth Newton method*, SIAM Journal on Optimization 13.3, pp. 865-888, 2002.

[53] H. Hwang and R. Haddad, *Adaptive median filters: new algorithms and results*, IEEE Trans. Image Process., 4, pp. 499-–502, 1995.

[54] A. Klenke, *Probability theory: a comprehensive course*, Springer Science and Business Media, 2013.

[55] J. Kaipio and E. Somersalo, *Statistical and computational inverse problems*, Vol, 160. Springer Science and Business Media, 2006.

[56] S. Ko and Y. Lee, *Center weighted median filters and their applications to image enhancement*, IEEE Trans. Circuits Syst., 38, pp. 984-993, 1991.

[57] A. Lanza, S. Morigi, F. Sgallari, and Y. W. Wen, *Image restoration with Poisson–Gaussian mixed noise*. Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization, 2.1, pp. 12-24, 2014.

[58] T. Le, T. Chartrand, and T. Asaki, *A variational approach to reconstructing images corrupted by Poisson noise*, J.Math. Imaging Vis., 27, pp. 257–263, 2007.

[59] F. Luisier, T. Blu, and M. Unser, *Image denoising in mixed Poisson-Gaussian noise* Image Processing, IEEE Transactions on 20.3, pp. 696-708, 2011.

[60] M. Mäkitalo, and A.Foi, *Noise parameter mismatch in variance stabilization, with an application to Poisson–Gaussian noise estimation*, IEEE Transactions on Image Processing 23.12, pp. 5348-5359, 2014.

[61] V. Morozov,*Methods for solving incorrectly posed problems*, Springer Science and Business Media, 2012.

[62] J. Mueller and S. Siltanen,*Linear and nonlinear inverse problems with practical applications* Vol. 10. Siam, 2012.

[63] D. Mumford and J. Shah, *Optimal approximations by piecewise smooth functions and associated variational problems*, Communications on pure and applied mathematics 42.5, pp. 577-685, 1989.

[64] M. Ng, P. Weiss, and X. Yuan,*Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods*, SIAM journal on Scientific Computing 32.5, pp. 2710-2736, 2010.

[65] N. Nguyen, P. Milanfar, and G. Golub, *A computationally efficient superresolution image reconstruction algorithm*, Image Processing, IEEE Transactions on 10.4, pp. 573-583, 2001.

[66] M. Nikolova, *Minimizers of cost functions involving nonsmooth data fidelity terms. Application to the processing of outliers*, SIAM J. Numer. Anal., 40, pp. 965–994, 2002.

[67] M. Nikolova, *A variational approach to remove outliers and impulse noise*, J. Math. Imaging Vis., 20, pp. 90–120, 2004.

[68] J. Nolan, *Numerical calculation of stable densities and distribution functions*, Comm. Statist. Stochastics Models, 13, pp. 759–774, 1997.

[69] J. Nolan, *Stable Distributions - Models for Heavy Tailed Data*, Birkhäuser Boston, 2015 (in progress: Chapter 1 online at academic2.american.edu/∼jpnolan).

[70] L. Qi and J. Sun, *A nonsmooth version of Newton's method* Mathematical programming 58.1-3, pp. 353-367, 1993.

[71] W. Pratt, *Median Filtering*, Technical report, Image Processing Institute, University of Southern California, Los Angeles, CA, 1975.

[72] R. Rockafellar, *Convex analysis*, Princeton university press, 2015.

[73] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60, pp. 259–268, 1992.

[74] L. Rudin, P. Lions, and S. Osher, *Multiplicative denoising and deblurring: theory and algorithms*, S. Osher and N. Paragios, eds, Geometric Level Sets in Imaging, Vision and Graphics, pp. 103–119, Springer, 2003.

[75] G. Samorodnitsky and M. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Chapman & Hall, New York, 1994.

[76] O. Scherzer, *Handbook of mathematical methods in imaging* Springer Science and Business Media, 2010.

[77] F. Sciacchitano, Y. Dong, and M. S. Andersen, *Total variation based parameter-free model for impulse noise removal*, Numerical Mathematics: Theory, Methods and Applications; to appear.

[78] F. Sciacchitano, Y. Dong, and T. Zeng, *Variational approach for restoring blurred images with Cauchy noise*, SIAM Journal on Imaging Sciences 8.3, pp. 1894-1922, 2015.

[79] S. Setzer and G. Steidl and T. Teuber, *Deblurring Poissonian images by split Bregman techniques*, J. Visual Commun. and Image Represent. 21, pp. 193–199, 2010.

[80] M. Shykula, and O. Seleznjev, *Stochastic structure of asymptotic quantization errors*, Statistics and probability letters 76.5, pp. 453-464, 2006.

[81] J. Sijbers, A. Den Dekker, J., Van Audekerke, M. Verhoye, and D. Van Dyck, *Estimation of the noise in magnitude MR images*, Magnetic Resonance Imaging, 16.1, pp. 87-90, 1998.

[82] G. Steidl and T. Teuber , *Removing multiplicative noise by Douglas-Rachford splitting methods*, Journal of Mathematical Imaging and Vision 36.2, pp. 168-184, 2010.

[83] T. Teuber, G. Steidl, and R. Chan , *Minimization and parameter estimation for seminorm regularization models with I-divergence constraints* Inverse Problems 29.3, 035007, 2013.

[84] A. Tikhonov and V. Arsenin,*Solutions of ill-posed problems*, Winston and Sons, Washington, DC, 1977.

[85] P. Weiss, L. Blanc-Féraud, and G. Aubert, *Efficient schemes for total variation minimization under constraints in image processing*, SIAM J. Sci. Comput., 31, pp. 2047–2080, 2009.

[86] B. Widrow, and I. Kollár, *Quantization Noise: Roundoff Error in Digital Computation*, Signal Processing, Control, and Communications, pp. 485-528, 2008.

[87] C. Wu and X. Tai, *Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models*, SIAM J. Imaging Sci., 3, pp. 300—339, 2010.

[88] Y. Xiao, T. Zeng, J. Yu, and M. Ng, *Restoration of images corrupted by mixed Gaussian-impulse noise via l 1–l 0 minimization*, Pattern Recognition, 44.8, pp. 1708-1720, 2011.

[89] Z. Yang and M. Jacob, *Nonlocal regularization of inverse problems: a unified variational framework*, IEEE Trans. Image Process., 22, pp. 3192-3203, 2013.

[90] J. Yang, Y. Zhang, and W. Yin, *An efficient TVL1 algorithm for deblurring multichannel images corrupted by impulsive noise*, SIAM J. Sci. Comput., 31, pp. 2842–2865, 2009.

[91] W. Yin, D. Goldfarb and S. Osher, *The total variation regularized L^1 model for multiscale decomposition*, Multiscale Model. Simul., 6, pp. 190–211, 2007.

[92] W. Zhou, A. Bovik, H. Sheikh, and E. Simoncelli, *Image Qualifty Assessment: From Error Visibility to Structural Similarity*, IEEE Trans. Image Process., 13, pp. 600-–612, 2004.

List of Papers

PAPER A

Total variation based parameter-free model for impulse noise removal

Authors:

Federica Sciacchitano, Yiqiu Dong and Martin S. Andersen

To appear in:

Numerical Mathematics: Theory, Methods and Applications.

NUMERICAL MATHEMATICS: Theory, Methods and Applications
Numer. Math. Theor. Meth. Appl., Vol. **xx**, No. x, pp. 1-19 (200x)

Total variation based parameter-free model for impulse noise removal

Federica Sciacchitano*, Yiqiu Dong and Martin S. Andersen

Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark.

Abstract. We propose a new two-phase method for reconstruction of blurred images corrupted by impulse noise. In the first phase, we use a noise detector to identify the pixels that are contaminated by noise, and then, in the second phase, we reconstruct the noisy pixels by solving an equality constrained total variation minimization problem that preserves the exact values of the noise-free pixels. For images that are only corrupted by impulse noise (i.e., not blurred) we apply the semismooth Newton's method to a reduced problem, and if the images are also blurred, we solve the equality constrained reconstruction problem using a first-order primal-dual algorithm. The proposed model improves the computational efficiency (in the denoising case) and has the advantage of being regularization parameter-free. Our numerical results suggest that the method is competitive in terms of its restoration capabilities with respect to the other two-phase methods.

AMS subject classifications: 68U10, 94A08, 49J40, 52A41, 65K10, 90C47, 49M15

Key words: Image deblurring, image denoising, impulse noise, noise detector, primal-dual first-order algorithm, semismooth Newton method, total variation regularization.

1. Introduction

During the image acquisition and transmission, observed images are inevitably degraded by blur and noise. In the literature, many kinds of noise have been widely considered, Gaussian noise [14, 20, 36], impulse noise [7, 11, 28, 29, 31], multiplicative noise [3, 19, 35], Poisson noise [21, 26, 37] or mixed noise [8, 27, 38]. In this paper, we focus on blurred image with impulse noise, which is a common type of image degradation due to, e.g., malfunctioning pixel elements in the camera sensors, errors in analog-to-digital conversion, faulty memory locations in hardware, or transmission errors [5]. A characteristic property of impulse noise is that a certain number of pixels are uncorrupted and the noise-corrupted pixels contain no information about the true pixel value.

*Corresponding author. *Email addresses:* feds@dtu.dk (F. Sciacchitano), yido@dtu.dk (Y. Dong), mskan@dtu.dk (M. S. Andersen)

 ©200x Global-Science Press

Over the years, many nonlinear digital filters methods have been proposed, see [2]. The most common filters used to remove impulse noise are the median-type filters: median filter [34], weighted median filter [6], adaptive median filter [24], multistate median filter [15], center weighted median filter [25] and adaptive center-weighted median filter [16]. Although these filters are efficient and easy to implement, they cannot achieve good results in general, in particular they are not able to restore a blurred image and they do not preserve the image edges well.

In order to preserve the edges, in 2004, Nikolova [31] proposed a variational model which combines an ℓ^1-data fidelity term with total variation (TV), which has been in shown in [30, 31] to work better than the classical ℓ^2-term, [36].

Later, other approaches based on the ℓ^1-TV have been proposed to handle the deblurring problem and the non-differentiability of the ℓ^1-norm, for instance: Bar et al. [4] introduce a model using the Mumford-Shah regularizer and the ℓ^1-data fidelity term; Yang et al. [39] suggested an efficient algorithm to solve the ℓ^1-TV model; Dong et al. [17] solved the ℓ^1-TV model using a primal-dual approach.

However, since the ℓ^1-TV minimization method negatively affects the noisy-free pixels, in 2005, Chan, Ho, and Nikolova [12] proposed the so-called two-phase method. The basic idea behind this method, which we will refer to as the CHN method, is to separate noise detection and image reconstruction. In the first phase, the method uses a noise detector to identify which pixels are corrupted, and in the second phase, it reconstructs only the noisy pixels based on an objective function with an ℓ^1-data fidelity term and with TV as a regularization term. The two-phase model has also been studied for other applications, for instance in [8], the authors apply the two-phase method to restore blurred images with impulse and Gaussian noise; in [23], a two-phase method is used for recovering images corrupted by multiplicative noise; in [7] and [11], a two-phase method is used to simultaneously deblur and denoise an image with impulse noise. Different from [12], in the second phase of [7] and [11] the authors reconstruct the image based on a modified ℓ^1-TV model where only noise-free pixels are kept in the ℓ^1-data fidelity term, due to no useful information contained in impulse noise. We will focus only on the method in [11] (the CDH method in short), since it outperforms the one in [7] and [12] with respect to both image restoration capability and computational efficiency.

While the CDH method has been shown to perform well on many test problems, the inclusion of noise-free pixels in the data-fidelity term is somewhat at odds with the assumption that their true values are known. If the pixels are indeed noise-free, then they can either be treated as constants or eliminated from the problem. In this work, we investigate such an approach and propose a modified two-phase method. In particular, as suggested in [11], in the first phase we distinguish noisy pixels from the noise-free pixels by the adaptive median (AM) filter [24] for detecting salt-and-pepper noise, and the adaptive center-weighted median (ACWM) filter [16] for random-valued impulse noise. The detector for salt-and-pepper noise is able to detect almost all noisy pixels even for noise level around 90%, while for random-valued impulse noise, the ACWM works quite well until a noise level of 40%, since in this case the noisy pixels

can be confused as clean ones and vice-versa. In case of higher noise for random-valued impulse noise, we suggest to use other filters, e.g. [1] and [18], which have the capability to detect noisy pixels in images with noise percentage up to 60%. The study of a good detector for random-valued impulse nose is outside the scope of this paper.

In the second phase, we compute the reconstruction by solving a reduced TV minimization problem that involves only the corrupted pixels. For solving the problem, we employ the semismooth Newton method, [32]. Furthermore, if the noise level is below 30%, we speed up the process by preprocessing the independent noisy pixels. The main difference between our method and the CDH method is that we reconstruct only the pixels that are identified as corrupted by noise instead of all pixels. For this reason, our method often leads to computational savings, and it can be viewed as an exact method in the sense that the reconstruction model matches the information about the noise-free pixels. Moreover, a notable advantage of our model is that the reconstruction results do not rely on the adjustment of any regularization parameters.

In addition to the impulse noise denoising problem, we also consider simultaneous deblurring and denoising. Instead of including the noise-free pixels in a data-fidelity term, we propose a regularization parameter-free model based on a constrained minimization problem. We solve this problem numerically using a primal-dual first-order algorithm [10].

This study focuses on the second phase of the two-phase method, and the main contributions regard the introduction of a regularization parameter-free model to reconstruct the corrupted image and, for the denoising case, the computational efficiency compared to the latest two-phase method, especially given by the preprocessing part.

The paper is organized as follows. In Section 2, we review the impulse noise model and propose a two-phase methods for denoising and also deblurring images. In Section 3, we present the implementation details, and in Section 4, we show some numerical results. We conclude the paper in Section 5.

2. The regularization parameter-free two-phase models

We start this section by introducing two impulse noise models, namely salt-and-pepper noise and random-valued impulse noise. Then, inspired by the existing denoising models in the literature [11], we propose (i) a two-phase method for denoising, and (ii) a two-phase method for simultaneous denoising and deblurring.

Given a discrete image of size $m_1 \times m_2$, we define a vector $\boldsymbol{u} \in \mathbb{R}^m$ with the $m = m_1 m_2$ pixels. We shall use the following notation $u_k = u_{i,j}$, where $k = (i-1)m_2 + j$ for $i = 1, \ldots, m_1$ and $j = 1, \ldots, m_2$, to identify the pixel at position (i, j) with the kth element of $\boldsymbol{u}$, and we define a set $\Omega = \{1, 2, \ldots, m\}$ that contains all pixel indices.

2.1. Impulse Noise Models

Impulse noise can be described as a stochastic degradation process of the form

$$z_k = N_r(\tilde{u}_k) = \begin{cases} \eta_k & \text{with probability } r \\ \tilde{u}_k & \text{with probability } 1-r \end{cases}, \quad k \in \Omega$$

where $\tilde{\boldsymbol{u}} \in \mathbb{R}^m$ is the original image, $\boldsymbol{z} \in \mathbb{R}^m$ is the corrupted image, and $\boldsymbol{\eta} \in \mathbb{R}^m$ is the noise, which is independent from the original image $\tilde{\boldsymbol{u}}$. Both images are assumed to be obtained from a two-dimensional pixel-array by means of columnwise concatenation. We refer to the parameter $r \in [0,1]$ as the noise level since it can be interpreted as the probability that a pixel is corrupted. Notice that some pixels remain unchanged, and the pixels that are corrupted by noise carry no information about the noise-free image.

Two main types of impulse noise are the salt-and-pepper noise when, for each k, the noise η_k is a discrete random variable with values drawn from the set $\{d_{\min}, d_{\max}\}$ with equal probability (with $d_{\min} = \min_k \tilde{u}_k$ and $d_{\max} = \max_k \tilde{u}_k$), and the random-valued impulse noise when, for each k, the noise η_k is a uniformly distributed random variable with values in the gray-level range $[d_{\min}, d_{\max}]$. For salt-and-pepper noise corrupted pixels take the lowest or the highest pixel value (i.e., $d_{\min}$ or $d_{\max}$), whereas for random-valued impulse noise, the noisy pixels have values anywhere in the interval from $d_{\min}$ to $d_{\max}$. Then, the random-valued impulse noise is more general and more difficult to detect than the salt-and-pepper noise.

2.2. Denoising Models

The ℓ^1-TV model for impulse noise denoising proposed by Nikolova [30, 31] combines the TV regularization term with an ℓ^1 data-fidelity term. The resulting reconstruction problem is convex and takes the following form

$$\min_{\boldsymbol{u} \in \mathbb{R}^m} \|\boldsymbol{u} - \boldsymbol{z}\|_1 + \alpha \, \mathrm{TV}(\boldsymbol{u}), \tag{2.1}$$

where $\|\boldsymbol{u}-\boldsymbol{z}\|_1 = \sum_{k\in\Omega} |u_k - z_k|$ is the data-fidelity term, $\mathrm{TV}(\boldsymbol{u})$ is a regularization term, and $\alpha > 0$ is a regularization parameter. The (discrete) TV is defined as

$$\mathrm{TV}(\boldsymbol{u}) := \sum_{k\in\Omega} |(\nabla \boldsymbol{u})_k|_2 = \sum_{k\in\Omega} \sqrt{|(\nabla_x \boldsymbol{u})_k|^2 + |(\nabla_y \boldsymbol{u})_k|^2},$$

where the discrete gradient operator $\nabla \in \mathbb{R}^{2m\times m}$ is given by

$$(\nabla \boldsymbol{u})_k = \begin{pmatrix} (\nabla_x \boldsymbol{u})_k \\ (\nabla_y \boldsymbol{u})_k \end{pmatrix},$$

and $\nabla_x \boldsymbol{u}$ and $\nabla_y \boldsymbol{u}$ denote the horizontal and vertical first order differences, i.e., using the symmetric boundary conditions, we have

$$(\nabla_x \boldsymbol{u})_k = \begin{cases} u_{i+1,j} - u_{i,j} & \text{if } i < m_1, \\ 0 & \text{if } i = m_1, \end{cases} \quad \text{and} \quad (\nabla_y \boldsymbol{u})_k = \begin{cases} u_{i,j+1} - u_{i,j} & \text{if } j < m_2, \\ 0 & \text{if } j = m_2, \end{cases}$$

for $k=(i-1)m_2+j$ with $i=1,\dots,m_1$ and $j=1,\dots,m_2$.

The ℓ^1-TV model has some nice properties, such as contrast preservation, multiscale decomposition and morphological invariance [13,30,40]. However, as mentioned in the introduction, the main disadvantage of this approach is that we have to reconstruct all the pixels of the image, including the ones that are noise-free. Furthermore, including the noisy pixels in the data-fidelity term introduces errors since the noise-corrupted pixels contain no information about the true image. To address this issue, Chan et al. [11] studied a two-phase method (the CDH method) in which they first detect the noisy pixels (phase 1) and then exclude these pixels from the data-fidelity term when computing a reconstruction (phase 2). Thus, in the first phase, they use a detector (an AM filter for salt-and-pepper and an ACWM filter for random-valued impulse noise) to split the domain Ω into two sets: $\mathcal{N}$ that includes all indices of the corrupted pixels and $\mathcal{U}$ that includes the indices of the noise-free pixels. We will henceforth assume that there are $|\mathcal{N}|=n$ noisy pixels and $|\mathcal{U}|=m-n$ noise-free pixels. In the second phase, for the denoising case (here, instead of the blurring operator we consider the identity), they reconstruct the image based on the following model

$$\min_{\boldsymbol{u}\in\mathbb{R}^m} \sum_{k\in\mathcal{U}} |u_k - z_k| + \alpha\,\mathrm{TV}(\boldsymbol{u}). \tag{2.2}$$

The main advantage of the CDH method is that the noise detector improves the data-fidelity term in ℓ^1-TV model (2.1), and this often yields a great improvement in terms of restoration capabilities. Furthermore, the ℓ^1-norm in the data-fidelity term allows many noisy-free pixels to maintain their exact values. However, the presence of a regularization parameter in the model necessitates multiple reconstructions or tests in order to find a good choice for the parameter. Moreover, the problem (2.2) includes all pixels of the image as variables, including the ones that are assumed to be free of noise. To overcome these disadvantages of the CDH method, we propose to alter the second phase of the method such that the noise-free pixels are required to be equal to their known values, i.e., we consider the following constrained optimization problem

$$\begin{aligned} &\min_{\boldsymbol{u}\in\mathbb{R}^m} \mathrm{TV}(\boldsymbol{u}) \\ &\text{s.t.} \quad u_k = z_k \quad k\in\mathcal{U}. \end{aligned} \tag{2.3}$$

In other words, instead of looking at the unconstrained minimization problem in (2.2), we are considering the constrained version of it. The equality constraints in this model reflect the exact prior that some pixels are known, assuming that all pixels were correctly identified as either noise-free or corrupted in the first phase. For this reason, the reconstruction model (2.2) can be seen as an approximation model since it allows the noise-free pixels to deviate from their known value.

Although the model (2.3) does not require the determination of any regularization parameters, it can be shown to be equivalent to (2.2) if the regularization parameter α is chosen sufficiently small. Specifically, we refer the reader to [32, Thm. 17.3] (note that, this theorem holds in the continuous case).

Unlike the model (2.2), which includes all pixels as variables, the constrained problem (2.3) allows us to eliminate the variables that correspond to noise-free pixels from the problem formulation. To write (2.3) as an unconstrained optimization problem, we define a vector $\boldsymbol{u}_{\mathcal{N}} \in \mathbb{R}^n$ that corresponds to the corrupted pixels. With this notation, $\boldsymbol{u}$ can be expressed as

$$\boldsymbol{u} = \Lambda_{\mathcal{N}} \boldsymbol{u}_{\mathcal{N}} + \Lambda_{\mathcal{U}} \boldsymbol{u}, \tag{2.4}$$

where $\Lambda_{\mathcal{U}} \in \mathbb{R}^{m\times m}$ is a diagonal matrix defined as

$$(\Lambda_{\mathcal{U}})_{i,i} = \begin{cases} 1 & \text{if } i \in \mathcal{U} \\ 0 & \text{if } i \in \mathcal{N} \end{cases}$$

and $\Lambda_{\mathcal{N}} \in \mathbb{R}^{m\times n}$ is a matrix with n unit vectors $\boldsymbol{e}_j \in \mathbb{R}^m$, $j \in \mathcal{N}$, as columns. Note that $\Lambda_{\mathcal{U}}\Lambda_{\mathcal{N}} = 0$ by construction. Since $\Lambda_{\mathcal{U}}\boldsymbol{u}$ represents the intensity of the noise-free pixels, based on the constraint in (2.3), we can substitute $\Lambda_{\mathcal{U}}\boldsymbol{z}$ into (2.4) and express the image $\boldsymbol{u}$ as follows

$$\boldsymbol{u} = \Lambda_{\mathcal{N}} \boldsymbol{u}_{\mathcal{N}} + \Lambda_{\mathcal{U}} \boldsymbol{z}. \tag{2.5}$$

The problem (2.3) can therefore be expressed in terms of $\boldsymbol{u}_{\mathcal{N}}$ as follows

$$\min_{\boldsymbol{u}_{\mathcal{N}}\in\mathbb{R}^n} \mathrm{TV}(\Lambda_{\mathcal{N}} \boldsymbol{u}_{\mathcal{N}} + \Lambda_{\mathcal{U}} \boldsymbol{z}). \tag{2.6}$$

Comparing this model with (2.2), we see that both are unconstrained minimization problems, but (2.6) has some advantages. Firstly, the minimization problem involves only n variables instead of m variables, and if the noise level is relatively low (i.e. $n \ll m$) the reduction in the number of variables is quite substantial. Secondly, it does not require the determination of the regularization parameter.

Before we discuss how to solve (2.6), we first consider an extension of our denoising approach to simultaneous deblurring and denoising.

2.3. Deblurring and Denoising Models

Suppose the observed image $\boldsymbol{z}$ is not only corrupted by impulse noise but also blurred by a known linear blur operator $K \in \mathbb{R}^{m\times m}$, i.e., we define $\boldsymbol{z} = N_r(K\boldsymbol{u})$. To solve the deblurring and denoising problem, we consider the two-phase method in [11], which extends the CHN method for the general case. In the first phase, the authors of [11] identify the corrupted pixels, and then, in the second phase, they compute a reconstruction by solving the following problem

$$\min_{\boldsymbol{u}\in\mathbb{R}^m} \sum_{k\in\mathcal{U}} |(K\boldsymbol{u})_k - z_k| + \alpha\,\mathrm{TV}(\boldsymbol{u}). \tag{2.7}$$

Note that only the noise-free pixels are included in the data-fidelity term. As in the denoising problem (2.3), we can formulate a constrained minimization problem

$$\begin{aligned} &\min_{\boldsymbol{u}\in\mathbb{R}^m} \mathrm{TV}(\boldsymbol{u}) \\ &\text{s.t.} \quad (K\boldsymbol{u})_k = z_k \quad k \in \mathcal{U} \end{aligned} \tag{2.8}$$

which implies that the value of noise-free pixels in the blurred image are treated as known constants. However, unlike in the denoising case, the blur operator K introduces coupling and makes it difficult to eliminate the equality constraints. In the next section, we address how to solve the problem numerically.

3. The algorithms

We now present methods for solving the denoising problem (2.6) as well as the deblurring and denoising problem (2.8).

3.1. Solving the Denoising Problem

Since the objective function in (2.6) is not differentiable everywhere, we introduce the following smooth approximation of TV,

$$\mathrm{TV}^\gamma(\boldsymbol{u}) = \sum_{k\in\Omega} \Phi^\gamma(|(\nabla \boldsymbol{u})_k|_2),$$

where the function Φ^γ is the Huber function which is defined as

$$\Phi^\gamma(t) = \begin{cases} |t| - \frac{\gamma}{2} & \text{if } |t| \geq \gamma \\ \frac{1}{2\gamma}|t|^2 & \text{else,} \end{cases}$$

with parameter $\gamma > 0$. Other smooth approximations may be used instead, such as e.g. $\sqrt{t^2+\gamma^2}$. The gradient of $\mathrm{TV}^\gamma(\Lambda_{\mathcal{N}}\boldsymbol{u}_{\mathcal{N}} + \Lambda_{\mathcal{U}}\boldsymbol{z})$ can be expressed as

$$F^\gamma(\boldsymbol{u}_{\mathcal{N}}) = -\Lambda_{\mathcal{N}}^\top \mathrm{div}\, \mathscr{D}^\gamma(\boldsymbol{u}_{\mathcal{N}})^{-1} \nabla(\Lambda_{\mathcal{N}}\boldsymbol{u}_{\mathcal{N}} + \Lambda_{\mathcal{U}}\boldsymbol{z}),$$

where $\mathrm{div} \in \mathbb{R}^{m\times 2m}$ represents the divergence, $\mathscr{D}^\gamma(\boldsymbol{u}_{\mathcal{N}}) \in \mathbb{R}^{2m\times 2m}$ is defined as

$$\mathscr{D}^\gamma(\boldsymbol{u}_{\mathcal{N}}) = \begin{pmatrix} N^\gamma(\boldsymbol{u}_{\mathcal{N}}) & 0 \\ 0 & N^\gamma(\boldsymbol{u}_{\mathcal{N}}) \end{pmatrix},$$

and $N^\gamma(\boldsymbol{u}_{\mathcal{N}}) \in \mathbb{R}^{m\times m}$ is a diagonal matrix with diagonal $\max\big(|\nabla(\Lambda_{\mathcal{N}}\boldsymbol{u}_{\mathcal{N}} + \Lambda_{\mathcal{U}}\boldsymbol{z})|_2, \gamma\big)$. The divergence satisfies the equation $\mathrm{div} = -\nabla^\top$, where $\nabla^\top$ is the transpose of the gradient operator. Hence, the explicit formula of the divergence can be found using the definition of transpose

$$\langle -\mathrm{div}\, \boldsymbol{p}, \boldsymbol{v}\rangle_{\mathbb{R}^m} = \langle \boldsymbol{p}, \nabla \boldsymbol{v}\rangle_{\mathbb{R}^{2m}},$$

for every $\boldsymbol{p} \in \mathbb{R}^{2m}$ and $\boldsymbol{v} \in \mathbb{R}^m$, where $\langle\cdot,\cdot\rangle_{\mathbb{R}^m}$ and $\langle\cdot,\cdot\rangle_{\mathbb{R}^{2m}}$ denote the standard scalar products in $\mathbb{R}^m$ and $\mathbb{R}^{2m}$, respectively.

It follows from the first-order optimality condition associated with (2.6) that the solution to the smooth approximation should satisfy the following equation

$$\Lambda_{\mathcal{N}}^\top \mathrm{div}\, \mathscr{D}^\gamma(\boldsymbol{u}_{\mathcal{N}})^{-1} \nabla(\Lambda_{\mathcal{N}}\boldsymbol{u}_{\mathcal{N}} + \Lambda_{\mathcal{U}}\boldsymbol{z}) = 0. \tag{3.1}$$

The nonlinear equation in (3.1) can be solved by the semismooth Newton method. Before describing the algorithm, we give the definition of generalized differentiability of a mapping $F : \mathbb{R}^s \to \mathbb{R}^t$ with $s, t \in \mathbb{N}$. The mapping F is called generalized differentiable in an open set $\mathcal{V} \subset \mathbb{R}^s$ if there exists $G_F : \mathbb{R}^s \to \mathbb{R}^{s\times t}$ such that

$$\lim_{\|\delta_{\boldsymbol{v}}\|\to 0} \frac{1}{\|\delta_{\boldsymbol{v}}\|}\|F(\boldsymbol{v}+\delta_{\boldsymbol{v}})-F(\boldsymbol{v})-G_F(\boldsymbol{v}+\delta_{\boldsymbol{v}})\delta_{\boldsymbol{v}}\| = 0,$$

for every $\boldsymbol{v} \in \mathcal{V}$; see e.g. [22]. This definition is equivalent to the semismoothness of locally Lipschitz maps F in [33]. Thus, we can define the semismooth Newton's method as the generalized version of Newton's method for semismooth maps. In particular, given the current iterate $\boldsymbol{u}^l_{\mathcal{N}}$, the semismooth Newton iteration is

$$G_F(\boldsymbol{u}^l_{\mathcal{N}})\delta_{\boldsymbol{u},l} = -F^\gamma(\boldsymbol{u}^l_{\mathcal{N}}), \tag{3.2}$$

with $\delta_{\boldsymbol{u},l} = \boldsymbol{u}^{l+1}_{\mathcal{N}} - \boldsymbol{u}^l_{\mathcal{N}}$.

The generalized derivative of F^γ is given by

$$G_F(\boldsymbol{u}_{\mathcal{N}}) = \Lambda^\top_{\mathcal{N}}\mathrm{div}(\mathscr{D}^\gamma(\boldsymbol{u}_{\mathcal{N}}))^{-1}\nabla\Lambda_{\mathcal{N}} + \frac{1}{2}h(\boldsymbol{u}_{\mathcal{N}})\boldsymbol{w}^\top + \frac{1}{2}\boldsymbol{w}(h(\boldsymbol{u}_{\mathcal{N}}))^\top,$$

where $h(\boldsymbol{u}_{\mathcal{N}}) = \Lambda^\top_{\mathcal{N}}\mathrm{div}\nabla(\Lambda_{\mathcal{N}}\boldsymbol{u}_{\mathcal{N}}+\Lambda_{\mathscr{U}}\boldsymbol{z})$ and $\boldsymbol{w} \in \partial((\mathscr{D}^\gamma(\boldsymbol{u}_{\mathcal{N}}))^{-1})$, with $\partial((\mathscr{D}^\gamma(\boldsymbol{u}_{\mathcal{N}}))^{-1})$ indicates the generalized derivative of $(\mathscr{D}^\gamma(\boldsymbol{u}_{\mathcal{N}}))^{-1}$, [22]. Thus, $\boldsymbol{w}$ is given by $\boldsymbol{w} = \Lambda^\top_{\mathcal{N}}\tilde{\boldsymbol{w}}$ where

$$\tilde{\boldsymbol{w}} = \begin{cases} -\mathrm{div}(\mathscr{D}^\gamma(\boldsymbol{u}_{\mathcal{N}}))^{-3}\nabla(\Lambda_{\mathcal{N}}\boldsymbol{u}_{\mathcal{N}} + \Lambda_{\mathscr{U}}\boldsymbol{z}) & \text{if } \boldsymbol{m}(\boldsymbol{u}_{\mathcal{N}}) > \gamma \\ 0 & \text{otherwise} \end{cases}$$

and $\boldsymbol{m}(\boldsymbol{u}_{\mathcal{N}}) = |\nabla(\Lambda_{\mathcal{N}}\boldsymbol{u}_{\mathcal{N}} + \Lambda_{\mathscr{U}}\boldsymbol{z})|_2$.

A solution $\delta_{\boldsymbol{u},l}$ in (3.2) may not exist or may not be unique, since it is not ensured that G_F is positive definite. For this reason, we add a small multiple of the identity matrix to G_F,

$$G^\varepsilon_F = G_F + \varepsilon I, \tag{3.3}$$

where ε is a small positive constant. Thus, substituting G_F with G^ε_F in (3.2), the semismooth Newton iteration is given by

$$G^\varepsilon_F(\boldsymbol{u}^l_{\mathcal{N}})\delta_{\boldsymbol{u},l} = -F^\gamma(\boldsymbol{u}^l_{\mathcal{N}}), \tag{3.4}$$

with $\delta_{\boldsymbol{u},l} = \boldsymbol{u}^{l+1}_{\mathcal{N}} - \boldsymbol{u}^l_{\mathcal{N}}$. Since the regularized matrix is positive definite and symmetric, we can solve (3.4) using the conjugate gradient method, [32].

In our numerical experiments, we have tested our implementation of the algorithm with different values of $\varepsilon \in [10^{-3}, 1]$, and we always obtained good reconstructions which suggests that our method is robust with respect to the choice of ε. In the numerical experiments reported in Section 4, we fix $\varepsilon = 0.1$.

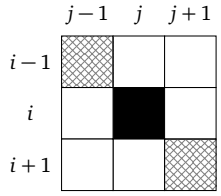


Figure 1: Neighborhood of a noisy pixel $u_{i,j}$ (black) containing six noise-free pixels (white).

3.1.1. Preprocessing

The problem (2.6) is very structured and partially separable. It follows from the definition of the discrete gradient operator ∇ that a noisy pixel with a noise-free neighborhood as shown in Fig. 1 is completely independent of other noisy pixels and hence can be computed independently. Specifically, if we regard $\boldsymbol{u}$ as a matrix instead of a vector and consider a noisy pixel $u_{i,j}$ with six noise-free neighboring pixels

$$\{u_{i,j-1}, u_{i+1,j-1}, u_{i-1,j}, u_{i+1,j}, u_{i-1,j+1}, u_{i,j+1}\},$$

then it follows from (2.6) that $u_{i,j}$ can be obtained by solving an unconstrained univariate optimization problem

$$u_{i,j} = \arg\min_{v} \; H_{i,j}(v) \tag{3.5}$$

where

$$H_{i,j}(v) \;=\; \left\| \begin{pmatrix} u_{i+1,j} - v \\ u_{i,j+1} - v \end{pmatrix} \right\|_2 + \left\| \begin{pmatrix} u_{i+1,j-1} - u_{i,j-1} \\ v - u_{i,j-1} \end{pmatrix} \right\|_2 + \left\| \begin{pmatrix} v - u_{i-1,j} \\ u_{i-1,j+1} - u_{i-1,j} \end{pmatrix} \right\|_2 .$$

The function $H_{i,j}(v)$ is clearly convex, and hence the minimization problem (3.5) can be solved using e.g. the golden section search method. To this end, it is worth noticing that each of the three terms of $H_{i,j}(v)$ is coercive, and hence we can easily derive a lower bound and an upper bound on $u_{i,j}$. Indeed, the three terms of $H_{i,j}(v)$ have minimizers $(u_{i+1,j} + u_{i,j+1})/2$, $u_{i,j-1}$, and $u_{i-1,j}$, so the interval $[a, b]$ with

$$a = \min\{(u_{i+1,j} + u_{i,j+1})/2, u_{i,j-1}, u_{i-1,j}\}$$

and

$$b = \max\{(u_{i+1,j} + u_{i,j+1})/2, u_{i,j-1}, u_{i-1,j}\}$$

must contain a solution.

When the noise-level is high, there may not be many pixels with a noise-free neighborhood as shown in Fig. 1. However, it may still be possible to separate the problem (2.6) into a number of independent subproblems. Extraction of such subproblems can easily be automated using morphological image processing and image analysis, but in this paper, we will only consider corrupted pixels with a neighborhood as in Fig. 1 and solve for these independently.

3.1.2. Algorithm

Our denoising algorithm is summarized in Algorithm 3.1.

Algorithm 3.1 Impulse noise denoising

1: Detect the noise-free pixels using a noise detector.
2: If the estimated noise level $r \leq 30\%$, do preprocessing:

- find pixels with noise-free neighborhood;
- solve (3.5) to find optimal pixel value.

3: Initialize $\boldsymbol{u}_{\mathcal{N}}^0 \in \mathbb{R}^n$ as the value obtained from the previous steps and set $l = 0$.
4: Compute $\delta_{\boldsymbol{u},l}$, by solving the equation (3.4), and update $\boldsymbol{u}_{\mathcal{N}}^{l+1} = \boldsymbol{u}_{\mathcal{N}}^l + \delta_{\boldsymbol{u},l}$. (Note that, with preprocessing, a part of noisy pixels have been restored in step 2. In this case, the size of $\boldsymbol{u}_{\mathcal{N}}$ in (3.4) is further reduced.)
5: Stop if stopping criteria are satisfied; otherwise set $l = l + 1$ and go to step 4.

3.2. Solving the Deblurring and Denoising Problem

A saddle-point formulation of the problem (2.8) is given by

$$\max_{\boldsymbol{b}\in\mathscr{B}} \min_{\boldsymbol{u}\in\mathscr{C}} \ \boldsymbol{b}^\top \nabla \boldsymbol{u}, \tag{3.6}$$

where $\boldsymbol{b} \in \mathbb{R}^{2m}$ is a dual variable, and the sets $\mathscr{C}$ and $\mathscr{B}$ are defined as

$$\mathscr{C} = \{\boldsymbol{u} \in \mathbb{R}^m \,|\, \Lambda_{\mathscr{Y}} K \boldsymbol{u} = \Lambda_{\mathscr{Y}} \boldsymbol{z}\}$$

and

$$\mathscr{B} = \{\boldsymbol{b} \in \mathbb{R}^{2m} \,|\, \|\boldsymbol{b}\|_\infty \leq 1\}.$$

The norm $\|\boldsymbol{b}\|_\infty$ denotes the discrete maximum norm, defined as

$$\|\boldsymbol{b}\|_\infty = \max_k |b_k|_2 = \max_k \sqrt{|(\boldsymbol{b}_x)_k|^2 + |(\boldsymbol{b}_y)_k|^2},$$

with

$$\boldsymbol{b} = \begin{pmatrix} \boldsymbol{b}_x \\ \boldsymbol{b}_y \end{pmatrix} \text{ and } \boldsymbol{b}_x, \boldsymbol{b}_y \in \mathbb{R}^m.$$

The Chambolle–Pock algorithm [10] for solving the convex–concave saddle-point problem (3.6) is summarized in Algorithm 3.2.

The projection operator $P_{\mathscr{C}}$ in (3.7) can be evaluated by solving the following least-norm problem

$$\begin{aligned} &\min_{\boldsymbol{u}} \|\boldsymbol{u} - \boldsymbol{w}^l\|_2^2 \\ &\text{s.t. } \Lambda_{\mathscr{Y}} K \boldsymbol{u} = \Lambda_{\mathscr{Y}} \boldsymbol{z}, \end{aligned} \tag{3.10}$$

Algorithm 3.2 Chambolle–Pock algorithm for deblurring and denoising

1: Detect the noise-free pixels using a noise detector.
2: Initialize $\boldsymbol{u}^0$ with image from step 1 and $\boldsymbol{b}^0 = \mathbf{0}$ and set $l = 0$.
3: Set $\theta \in [0,1]$ and $\tau,\ \sigma > 0$ such that $\tau\sigma\|\nabla\|_2^2 < 1$.
4: Compute $\boldsymbol{u}^{l+1}$:

$$\boldsymbol{u}^{l+1} = P_{\mathscr{C}}(\boldsymbol{u}^l + \tau \mathrm{div}\boldsymbol{b}^l). \tag{3.7}$$

5: Compute $\bar{\boldsymbol{u}}^{l+1}$:

$$\bar{\boldsymbol{u}}^{l+1} = \boldsymbol{u}^{l+1} + \theta(\boldsymbol{u}^{l+1} - \boldsymbol{u}^l). \tag{3.8}$$

6: Compute $\boldsymbol{b}^{l+1}$:

$$\boldsymbol{b}^{l+1} = P_{\mathscr{B}}(\boldsymbol{b}^l + \sigma\nabla\bar{\boldsymbol{u}}^{l+1}). \tag{3.9}$$

7: Stop if stopping criteria are satisfied; otherwise set $l = l+1$ and go to step 4.

where $\boldsymbol{w}^l = \boldsymbol{u}^l + \tau \mathrm{div}\boldsymbol{b}^l$. This problem has the closed-form solution

$$\boldsymbol{u}^{l+1} = \boldsymbol{w}^l - K^\top \Lambda_{\mathscr{U}}^\top (\Lambda_{\mathscr{U}} K K^\top \Lambda_{\mathscr{U}}^\top)^{-1} \Lambda_{\mathscr{U}} (K\boldsymbol{w}^l - \boldsymbol{z}).$$

In order to compute the above quantity, we use the conjugate gradient method, [32]. Due to the singularity of the matrix $\Lambda_{\mathscr{U}}$, to guarantee the stability of the conjugate gradient algorithm, we add a small multiple of the identity matrix to $\Lambda_{\mathscr{U}} K K^\top \Lambda_{\mathscr{U}}^\top$.

Furthermore, the projection $P_{\mathscr{B}}$ in (3.9) is a pointwise Euclidean projection onto L^2 balls, i.e.

$$\boldsymbol{b}^{l+1} = \frac{\boldsymbol{b}^l + \sigma\nabla\bar{\boldsymbol{u}}^{l+1}}{\max(1, |\boldsymbol{b}^l + \sigma\nabla\bar{\boldsymbol{u}}^{l+1}|_2)}.$$

The Chambolle-Pock primal-dual algorithm ensures convergence if $\theta = 1$ and

$$\tau\sigma\|\nabla\|_2^2 < 1. \tag{3.11}$$

For more details we refer the reader to [10]. From [9], we know that the bound on the norm of the linear operator ∇ is

$$\|\nabla\|_2^2 = \|\mathrm{div}\|_2^2 < 8,$$

and hence the algorithm converges if $8\tau\sigma < 1$. In our numerical experiments, we use

$$\tau = \frac{\beta}{3} \text{ and } \sigma = \frac{1}{3\beta},$$

with $\beta > 0$. In this way, the convergence of the algorithm is ensured. In our experiments described in the next section, we fixed $\beta = 0.01$ which worked well. Thus, the proposed model for deblurring and denoising is also regularization parameter-free.

Table 1: PSNR values for restored images with different levels of salt-and-pepper noise given by different approaches.

	20% noise			40% noise			60% noise			80% noise		
	AM	**CDH**	**Ours**	**AM**	**CDH**	**Ours**	**AM**	**CDH**	**Ours**	**AM**	**CDH**	**Ours**
Boat	30.27	34.81	**34.82**	26.85	30.25	**30.32**	24.14	27.07	**27.18**	20.87	24.03	**24.19**
Bridge	28.41	32.01	**32.06**	25.11	28.17	**28.25**	22.44	25.09	**25.18**	19.47	22.11	**22.26**
Cameraman	28.66	**33.60**	33.59	25.27	**29.14**	29.13	22.59	26.23	**26.29**	19.97	23.07	**23.14**
Goldhill	29.91	34.07	**34.12**	27.46	30.64	**30.70**	24.80	27.76	**27.84**	22.15	25.06	**25.22**
Lena	32.05	36.71	**36.72**	28.13	32.15	**32.16**	24.82	**28.96**	28.95	21.10	25.07	**25.21**
Parrot	29.03	34.43	**34.44**	25.23	29.52	**29.53**	22.03	25.90	**26.01**	18.73	21.20	**21.29**
Peppers	25.47	25.58	**25.60**	24.44	25.26	**25.27**	23.02	24.67	**24.68**	20.25	22.91	**22.96**

Figure 2: Original images: "Parrot", "Cameraman" and "'Boat".

4. Numerical results

In this section, we show some reconstructions obtained by applying the proposed methods to sharp and blurred images corrupted by impulse noise. For the illustrations, we use the 256×256 gray-level images: "Boat","Cameraman" and "Parrot"; see Fig. 2. The quality of the images is compared in terms of the peak signal to noise ratio (PSNR) which is defined as

$$\mathrm{PSNR}(\boldsymbol{u}^\star) = 20\log_{10}\frac{m|\tilde{\boldsymbol{u}}_{\max} - \tilde{\boldsymbol{u}}_{\min}|}{\|\tilde{\boldsymbol{u}} - \boldsymbol{u}^\star\|_2},$$

where $\tilde{\boldsymbol{u}}$ and $\boldsymbol{u}^\star$ represent respectively the original image and the reconstructed image with values in the gray-level range $[\tilde{\boldsymbol{u}}_{\min}, \tilde{\boldsymbol{u}}_{\max}]$. Our reconstructions are compared with the ones given by the detector (AM or ACWM) and the latest two-phase method, i.e. the one proposed in [11]. As suggested in [11], we set the parameters $\lambda = 0.0005$ and $\gamma = 0.01$ and we tune α to get the highest value of PSNR (for more details about the parameters we refer the reader to [11]). Concerning our reconstruction, based on numerical experiment, we set $\varepsilon = 0.1$ and $\gamma = 0.01$, as in the CDH method. In our simulations, we stop our algorithm as soon as there are not big changes in the iterations, i.e.,

$$\frac{\|\boldsymbol{u}^l - \boldsymbol{u}^{l-1}\|_2}{\|\boldsymbol{u}^l\|_2} < 10^{-5}.$$

In our first experiment, we consider denoising without blurring. Recall that in the

first phase of our algorithm, we detect the noisy pixels using the adaptive median (AM) filter [24] for salt-and-pepper noise and the adaptive center-weighted median (ACWM) filter [16] for random-valued impulse noise. In the second phase, we solve the minimization problem (2.6) to denoise the corrupted pixels.

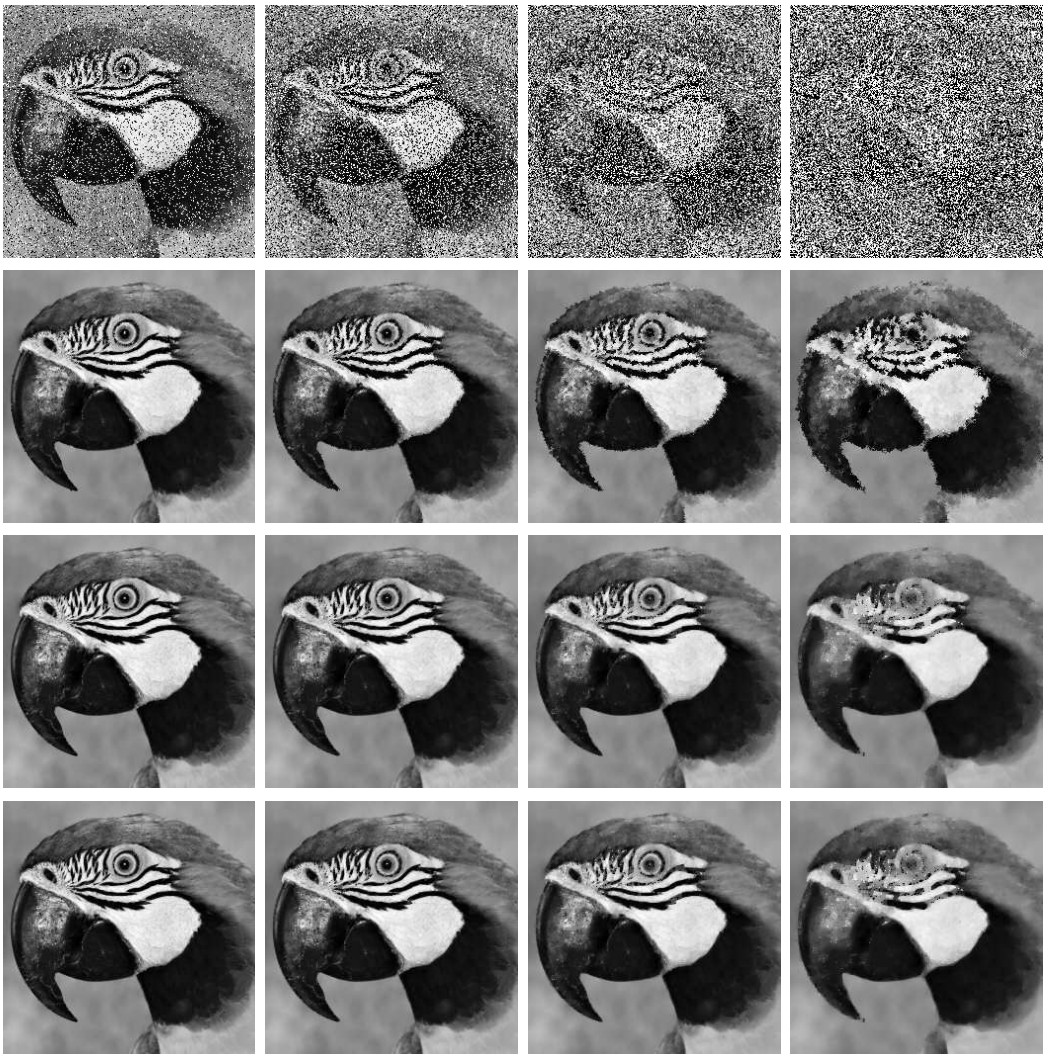


Figure 3: First row: noisy images "Parrot" with salt-and-pepper noise with noise level 20%, 40%, 60%, and 80% (left to right). Second row: results obtained with AM filter. Third row: results obtained with CDH method ($\alpha = 0.16$, $\alpha = 0.18$, $\alpha = 0.2$ and $\alpha = 0.25$). Fourth row: results obtained with proposed method.

In Fig. 3, we show the restored images from salt-and-pepper noise using (i) the adaptive median filter [24], (ii) the CDH method [11], and (iii) our method. For the CDH method, we adjusted the regularization parameter through numerical tests and show the best results. Comparing the PSNRs listed in Table 1, we see that the pro-

Table 2: Comparison of CPU time (in seconds) for "Pirate" corrupted by salt-and-pepper noise.

Noise	**CDH method**	**Our method**	
		No preprocessing	**Preprocessing**
10%	177.9	54.7	33.7
20%	202.1	83.4	69.6
30%	199.9	110.3	101.3
40%	284.0	142.8	-
50%	293.9	195.2	-
60%	353.3	245.1	-
70%	359.3	313.0	-
80%	425.8	375.5	-

posed method outperforms the AM filter and is competitive when compared to the CDH method. Taking into account that the proposed model is regularization parameter-free and only restores the noisy pixels, it is more practical and more efficient than the CDH method.

To compare the computational cost, we list the CPU time of the CDH method and the proposed method in Table 2. All of the numerical experiments were carried out in MATLAB R2014a on a PC equipped with a 3.20GHz CPU and 8GB memory. The results are based on the 1024×1024 test image "Pirate" and represent the average computation time based on ten noise realizations. Note that since the first phase of the CDH method and our method is same and the main computational load is in the second phase, we only give the CPU time associated with the second phase in Table 2. To show the effect of the preprocessing step in our method, we report the CPU times for restoration with and without the preprocessing when the noise level is at most 30%. Based on the results in Table 2, we find that the CDH method is slower than our method, especially when the noise-level is low. Moreover, the results verify that preprocessing is beneficial when the noise-level is low. Furthermore, the computation times for the CDH method do not include the overhead of tuning the regularization parameter, and hence the results do not reflect the added advantage of our model being regularization parameter-free.

In Fig. 4, we show the results when restoring the image "Parrot" corrupted by 30% random-valued impulse noise. In the first phase, we use the ACWM filter [16] as noise detector. In case of higher noise level, others filter (see for instance [1] and [18]), which have the capability to detect noisy pixels up to 60% noise level, might be employed. From the Table 3, it is clear that our method still provides results similar to those obtained with the CDH method, and both methods outperform the ACWM filter. As for salt-and-pepper case, the main advantages are that the proposed model does not require the adjustment of the regularization parameter and, since it reconstructs only the noisy pixels, it is faster than the CDH method.

As a final experiment, we compare the CDH method [11] and the method proposed in Section 3.2 for restoring blurred images with salt-and-pepper noise. In our simula-

Figure 4: From left to right: noisy image "Parrot" corrupted by 30% random-valued impulse noise, image restored with ACWM filter, image restored with CDH method ($\alpha = 0.15$), and image restored with proposed method.

Table 3: PSNR values for restored images with different levels of random valued impulse noise given by different approaches.

	30% noise			40% noise		
	ACWM	**CDH**	**Ours**	**ACWM**	**CDH**	**Ours**
Boat	25.85	26.08	**26.09**	24.27	24.74	**24.75**
Bridge	24.17	**24.61**	**24.61**	22.57	23.18	**23.19**
Cameraman	24.17	24.19	**24.20**	22.75	**23.15**	23.14
Goldhill	26.68	**27.04**	27.02	25.02	25.73	**25.75**
Lena	27.13	27.01	**27.03**	25.01	**26.16**	**26.16**
Parrot	23.62	**23.83**	23.82	21.72	**22.05**	22.04
Peppers	26.33	**26.88**	26.83	24.37	**25.26**	25.25

tion, we consider Gaussian blur with window size 7×7 and standard deviation 5. Fig. 5 shows the degraded images and the restored images obtained with the AM filter, the CDH method and our method. In addition, in Table 4 we list the PSNRs for different images restored with different methods. Moreover, to see how the amount of blur affects the quality of the reconstructions we test our model with Gaussian blur with window size 11×11 and standard deviation 7. The results for the image "'Boat" are shown in Fig. 6. From Fig. 5 and 6, we see that the results given by the AM filter are still blurred, since there is no deblurring step in the filter. Comparing the results of the two TV-based methods, i.e. CDH and ours, we have that both methods yield good restorations. However, like the proposed denoising model, our denoising and deblurring model does not require any regularization parameters adjustment.

In Fig. 7, we compare the PSNR values for the CDH method, using different regularization parameters α in (2.1), and our method. For this purpose, we use the blurred image "Cameraman" with 40% and 60% salt-and-pepper noise (see Fig. 5 second and third column, respectively). Note that, only in the continuous case the constrained minimization problem in (2.8) is equivalent to the unconstrained minimization problem in (2.7), if the regularization parameter is sufficiently small (see for more details [32, Thm. 17.3]). While, in the discrete case, we can see that due to numerical issues the theorem does not hold. From the figure, it is clear that the performance of the CDH method

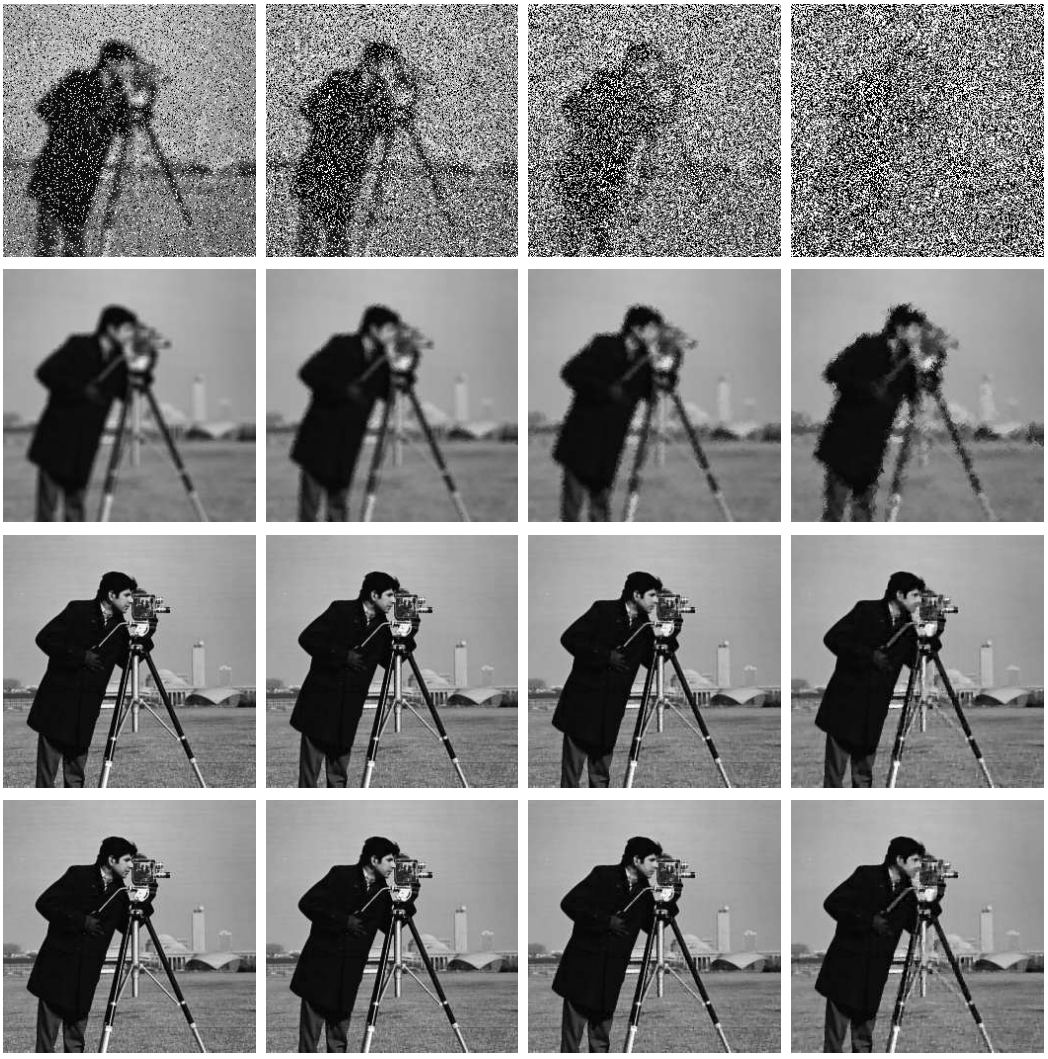


Figure 5: First row: Corrupted image "Cameraman" with salt-and-pepper noise with noise level 20%, 40%, 60%, and 80% and Gaussian blur with window size 7×7 and standard deviation 5. Second row: results obtained with the AM filter. Third row: results obtained with the CDH method ($\alpha = 0.0008$, $\alpha = 0.0011$, $\alpha = 0.0018$ and $\alpha = 0.0033$). Fourth row: results obtained with our method.

strongly depends on the selection of α. With a value of α that differs only slightly from the best choice, the restoration results can be much worse. Hence, being regularization parameter-free, our method is much more practical and always provides comparable restored images with the CDH method.

5. Conclusion

We have introduced a total-variation based regularization parameter-free model for restoring images corrupted by impulse noise. Since impulse noise only partly cor-

Table 4: PSNR values for restored blurred images with different levels of salt-and-pepper noise given by different approaches.

	20% noise			40% noise			60% noise			80% noise		
	AM	**CDH**	**Ours**	**AM**	**CDH**	**Ours**	**AM**	**CDH**	**Ours**	**AM**	**CDH**	**Ours**
Boat	23.00	38.85	**39.00**	22.90	35.78	**36.02**	22.68	**32.13**	32.09	21.88	27.74	**27.79**
Bridge	21.55	**33.74**	33.62	21.46	**31.09**	31.08	21.19	**28.34**	28.32	20.49	**25.47**	25.46
Cameraman	21.87	38.09	**38.16**	21.79	34.63	**35.05**	21.55	31.07	**31.33**	20.77	26.62	**26.65**
Goldhill	24.31	**35.73**	35.56	24.21	33.31	**33.36**	23.94	**30.70**	30.69	23.20	**27.94**	27.85
Lena	24.03	38.15	**38.19**	23.80	35.38	**35.60**	23.32	32.17	**32.38**	22.00	**29.16**	29.14
Parrot	21.02	37.16	**37.24**	20.88	33.95	**34.08**	20.55	30.18	**30.43**	19.41	26.47	**26.48**
Peppers	22.96	**41.04**	40.83	22.81	**38.74**	38.57	22.38	35.09	**35.32**	21.24	**30.31**	30.27

Figure 6: From the left to the right: corrupted image "Boat" with salt-and-pepper noise with noise level 40% and Gaussian blur with window size 11×11 and standard deviation 7; recovered image by using the AM filter (PSNR=21.55); recovered image by using the CDH method with $\alpha = 0.0013$ (PSNR=33.76); recovered image by using our method (PSNR=33.92).

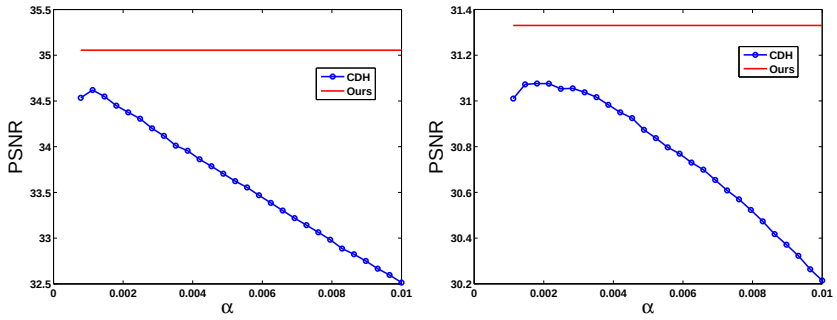


Figure 7: Comparison of the performance for the CDH method, with different values of α, and for our method for the blurred "Cameraman" with 40% and 60% salt-and-pepper noise (see Fig. 5 second and third column, respectively).

rupts images, we start with a constrained minimization problem for which the CDH model [11] can be viewed as an ℓ^1-approximation of our model. In the denoising case, by separating the noisy pixels and the noise-free ones, our formulation yields an unconstrained problem with only the noisy pixels as variables. This reduces the size of the problem, especially for low noise-levels. We also extend our method to the simultaneous deblurring and denoising case. The main advantage of our model is that it does not require the tuning of the regularization parameter, and we have demonstrated numerically that, for the denoising case, our method provides competitive results in less time

when compared to the CDH method.

Acknowledgments

The work of Y. Dong is supported by Advanced Grant No. 291405 from the European Research Council.

References

[1] S. Akkoul, R. Ledee, R. Leconge, and R. Harba, *A new adaptive switching median filter*, IEEE Signal Process. Lett, 17 (2010), pp. 587–590.
[2] J. Astola and P. Kuosmanen, *Fundamentals of Nonlinear Digital Filtering*, vol. 8, CRC, Boca Raton, FL, 1997.
[3] G. Aubert and J. Aujol, *A variational approach to removing multiplicative noise*, SIAM J. Appl. Math., 68 (2008), pp. 925–946.
[4] L. Bar, N. Kiryati and N. Sochen, *Image deblurring in the presence of impulsive noise*, International Journal of Computer Vision, 70 (2006), pp. 279–298.
[5] A. Bovik, *Handbook of Image and Video Processing*, New York: Academic, 2010.
[6] D. Brownrigg, *The weighted median filter*, Comm. ACM, 27 (1984), pp. 807âĂŞ818.
[7] J. Cai, R. Chan, and M. Nikolova, *Fast two-phase image deblurring under impulse noise*, J. Math. Imaging Vision, 36 (2010), pp. 46–53.
[8] J. Cai, R. Chan, and M. Nikolova, *Two-phase approach for deblurring images corrupted by impulse plus Gaussian noise*, Inverse Probl. Imaging, 2 (2008), pp. 187–204.
[9] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imag. Vis., 20 (2004), pp. 89–97.
[10] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imag. Vis., 40 (2011), pp. 120–145.
[11] R. Chan, Y. Dong and M. Hintermüller, *An efficient two-phase L^1- TV method for restoring blurred images with impulse noise*, IEEE Trans. Image Process., 19 (2010), pp. 1731–1739.
[12] R. Chan, C. Ho and M. Nikolova. *Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization*, IEEE Trans. Image Process., 14 (2005), pp. 1479–1485.
[13] T. Chan and S. Esedoglu, *Aspects of total variation regularized L^1 function approximation*, SIAM J. Appl. Math., 65 (2005), pp. 1817–1837.
[14] T. Chan and J. Shen, *Image processing and analysis: variational, PDE, wavelet, and stochastic methods*, SIAM, 2005.
[15] T. Chen and H. Wu, *Space variant median filters for the restoration of impulse noise corrupted images*, IEEE Trans. Circuits Syst. II, 48 (2001), pp. 784-789.
[16] T. Chen and H. Wu, *Adaptive impulse detection using center-weighted median filters*, IEEE Signal Process. Lett., 8 (2001), pp. 1–3.
[17] Y. Dong, M. Hintermüller and M. Neri *An efficient primal dual method for L^1- TV image restoration*, SIAM J. Imag. Sci., 2 (2009), pp. 1168–1189.
[18] Y. Dong, R. Chan, and S. Xu, *A detection statistic for random-valued impulse noise*, IEEE Trans. Image Process., 16 (2007), pp. 1112-1120.
[19] Y. Dong and T. Zeng, *A convex variational model for restoring blurred images with multiplicative noise*, SIAM J. Imaging Sci., 6 (2013), pp. 1598–1625.

[20] M. Elad and M. Aharon, *Image denoising via sparse and redundant representations over learned dictionaries*, IEEE Trans. Image Process., 15 (2006), pp. 3736–3745.
[21] B. Figueiredo and J. Bioucas-Dias, *Restoration of Poissonian images using alternating direction optimization*, IEEE Trans. Image Process., 19 (2010), pp. 3133-3145.
[22] M. Hintermüller, K. Ito and K. Kunisch, *The primal-dual active set strategy as a semismooth Newton method*, SIAM J. Opt., 13 (2002), pp. 865–888.
[23] Y. Huang, D. Lu and T. Zeng, *A Two-Step Approach for the Restoration of Images Corrupted by Multiplicative*, SIAM J. Sci. Comput., 35 (2013), pp. A2856–A2873.
[24] H. Hwang and R. Haddad *Adaptive median filters: new algorithms and results*, IEEE Trans. Image Process., 4 (1995), pp. 499–502.
[25] S. Ko and Y. Lee, *Center weighted median filters and their applications to image enhancement*, IEEE Trans. Circuits Syst., 38 (1991), pp. 984-993.
[26] T. Le, T. Chartrand, and T. Asaki, *A variational approach to reconstructing images corrupted by Poisson noise*, J. Math. Imaging Vis., 27 (2007), pp. 257–263.
[27] Y. Li, L. Shen, D. Dai and B. Suter, *Framelet algorithms for de-blurring images corrupted by impulse plus Gaussian noise*, IEEE Trans. Image Process., 20 (2011), pp. 1822-1837.
[28] L. Ma, J. Yu, and T. Zeng, *Sparse Representation Prior and Total Variation–Based Image Deblurring Under Impulse Noise*, SIAM J. Imag Sci, 6 (2013), pp. 2258-2284.
[29] L. Ma, M. Ng, J. Yu, and T. Zeng, *Efficient box-constrained TV-type-l^1 Algorithms for Restoring Images with Impulse Noise*, J. Comp. Math., 31 (2013), pp. 249–270.
[30] M. Nikolova, *Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers*, SIAM J. Numer. Anal., 40 (2002), pp. 965–994.
[31] M. Nikolova, *A variational approach to remove outliers and impulse noise*, J. Math. Imag. Vis., 20 (2004), pp. 99–120, .
[32] J. Nocedal and S. Wright, *Numerical Optimization*, New York: Springer, Second edition, 2006.
[33] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Programm., 58 (1993), pp 353–367.
[34] W. Pratt, *Median Filtering*, Technical report, Image Processing Institute, University of Southern California, Los Angeles, CA, 1975.
[35] L. Rudin, P. Lions, and S. Osher, *Multiplicative denoising and deblurring: theory and algorithms*, Geometric Level Sets in Imaging, Vision and Graphics, S. Osher and N. Paragios, Eds. New York: Springer, pp. 103–119, 2003.
[36] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
[37] S. Setzer and G. Steidl and T. Teuber, *Deblurring Poissonian images by split Bregman techniques*, J. Visual Commun. and Image Represent., 21 (2010), pp. 193–199.
[38] Y. Xiao, T. Zeng, J. Yu and M. Ng, *Restoration of Images Corrupted by Mixed Gaussian-Impulse Noise via l_1-l_0 Minimization*, Pattern Recogn., 44 (2011), pp. 1708–1728.
[39] J. Yang, Y. Zhang and W. Yin, *An efficient TVL1 algorithm for deblurring multichannel images corrupted by impulsive noise*, SIAM J. Sci. Comput., 31 (2009), pp. 2842–2865.
[40] W. Yin, D. Goldfarb and S. Osher, *The total variation regularized L^1 model for multiscale decomposition*, Multiscale Model. Simul., 6 (2007), pp. 190–211.

PAPER B

Variational Approach for Restoring Blurred Images with Cauchy Noise

Authors:

Federica Sciacchitano, Yiqiu Dong and Tieyong Zeng

Published in:

SIAM Journal on Imaging Sciences, vol. **8**(3), pp. 1894-1922, 2015.

SIAM J. IMAGING SCIENCES
Vol. 8, No. 3, pp. 1894–1922
© 2015 Society for Industrial and Applied Mathematics

Variational Approach for Restoring Blurred Images with Cauchy Noise*

Federica Sciacchitano†, Yiqiu Dong†, and Tieyong Zeng‡

Abstract. The restoration of images degraded by blurring and noise is one of the most important tasks in image processing. In this paper, based on the total variation (TV) we propose a new variational method for recovering images degraded by Cauchy noise and blurring. In order to obtain a strictly convex model, we add a quadratic penalty term, which guarantees the uniqueness of the solution. Due to the convexity of our model, the primal dual algorithm is employed to solve the minimization problem. Experimental results show the effectiveness of the proposed method for simultaneously deblurring and denoising images corrupted by Cauchy noise. Comparison with other existing and well-known methods is provided as well.

Key words. Cauchy noise, image deblurring, image denoising, primal dual algorithm, total variation regularization, variational model

AMS subject classifications. 68U10, 94A08, 49J40, 52A41, 65K10, 90C47, 35B45

DOI. 10.1137/140997816

1. Introduction.

Image deblurring and image denoising are fundamental problems in the applied mathematics community; see, for instance, [4, 5]. Most of the literature deals with the restoration of images corrupted by additive Gaussian noise [10, 15, 22, 54, 60]. Unfortunately, in many engineering applications the noise has a very impulsive character, and thus it cannot be modeled by this kind of noise. The most common example of impulsive noise is given by the impulse noise [12, 45, 46], which can be caused, for instance, by analogue-to-digital converter errors, by malfunctioning pixel elements in the camera sensors, and so on. Another impulsive degradation is given by Cauchy noise, which appears frequently in atmospheric and underwater acoustic noises, radar and sonar applications, air turbulence, wireless communication systems, biomedical images, and synthetic aperture radar (SAR) images. For an overview we refer the reader to [36, 39, 40, 49, 50, 52] and references therein.

Mathematically speaking, the degraded image f in the presence of blurring and Cauchy noise is given by $f = Ku + v$, where u is the original image defined on the image domain $\Omega \subset \mathbb{R}^2$, K is the blurring operator, and v is some Cauchy noise. A random variable V follows the Cauchy distribution if it has density

$$g(v) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + (v - \delta)^2}, \tag{1.1}$$

*Received by the editors December 1, 2014; accepted for publication (in revised form) July 6, 2015; published electronically September 17, 2015.
http://www.siam.org/journals/siims/8-3/99781.html
†Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark (feds@dtu.dk, yido@dtu.dk). The work of the second author was supported by Advanced Grant 291405 from the European Research Council.
‡Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong (zeng@hkbu.edu.hk). The work of this author was partially supported by NSFC 11271049, RGC 211911, 12302714, and RFGs of HKBU.

where $\gamma > 0$ is the scale parameter and $\delta \in \mathbb{R}$ is called the localization parameter. The scale parameter determines the spread of the distribution around δ and plays a role similar to that of the variance in the Gaussian distribution, while the localization parameter corresponds to the median of the distribution.

Recently, several approaches have been proposed to deal with Cauchy noise; for instance, Chang et al. [16] used recursive Markov random field models for reconstructing images under Cauchy noise. Achim and Kuruoğlu [1] proposed a method for denoising a image degraded by Cauchy noise in the complex wavelet domain. Wan, Canagarajah, and Achim [59] studied a segmentation technique for noisy color images corrupted by Cauchy noise. As far as we know, in the literature, no one has ever studied a variational model for removing Cauchy noise. Hence, our contribution is to propose a variational model for deblurring and denoising degraded images with Cauchy noise.

One of the most famous variational models is the ROF model [54]. This approach was introduced in 1992 by Rudin, Osher, and Fatemi, and it is defined as follows:

$$\inf_{u\in BV(\Omega)} J(u) + \frac{\lambda}{2}\int_{\Omega}(f-u)^2 \mathrm{d}x, \tag{1.2}$$

where $J(u) = \int_{\Omega} |Du|$ is the total variation (TV) regularization term, BV is the space of the functions of bounded variation (see [4] or below), the last term is the data fidelity term, and $\lambda > 0$ is the regularization parameter, which represents the trade-off between a good fit of f and a smoothness due to the TV regularization term. Due to its capability of preserving sharp edges, it is a very successful and popular approach for denoising images corrupted by additive Gaussian noise.

Over the years, many variational models based on TV have been introduced for removing other noises, such as multiplicative noise [3, 21, 53], impulse noise [12, 46, 62], Poisson noise [19, 24, 41, 51, 56], etc. In our work, inspired by the above studies, we introduce a variational model, based on TV as the regularization term, for restoring images with blur and Cauchy noise. In particular, we propose the following problem for removing Cauchy noise:

$$\inf_{u\in BV(\Omega)} J(u) + \frac{\lambda}{2}\int_{\Omega}\log\Big(\gamma^2 + (u-f)^2\Big)\mathrm{d}x, \tag{1.3}$$

where $\gamma > 0$ is the scale parameter; see (1.1). As one can see, we keep the same regularization term as in the ROF model, but we adapt the data fidelity term to the Cauchy noise, introducing one that is suitable for such a type of noise. We emphasize that TV regularization is a very useful tool for preserving edges but is not so good for texture recovery; thus, clearly, the proposed model can be extended to other modern regularization terms such as nonlocal TV [26, 58, 63], high order TV [61], dictionary learning [22, 33], or a tight-frame approach [8, 38].

Unfortunately, since the data fidelity term is not convex, the restored results depend on the initialization and the numerical scheme. Hence, to overcome this problem we use a quadratic penalty function technique; in particular, we introduce the following minimization problem:

$$\inf_{u\in BV(\Omega)} J(u) + \frac{\lambda}{2}\left(\int_{\Omega}\log\Big(\gamma^2 + (u-f)^2\Big)\mathrm{d}x + \mu\|u-u_0\|_2^2\right), \tag{1.4}$$

where u_0 is the image obtained by applying the median filter [5] to the noisy image. Under some assumptions on μ, we are able to prove that (1.4) is convex and there exists a unique solution of (1.4). We employ the median filter in the quadratic penalty term, since it has been shown that it works well for removing impulse noise [12], and the Cauchy degradation has some similarities with the impulse degradation.

Readily, we can also generalize our model for restoring a blurred image corrupted by Cauchy noise. Given a linear blurring operator K, we consider the following convex model for deblurring and denoising simultaneously:

$$\inf_{u\in BV(\Omega)} J(u) + \frac{\lambda}{2}\left(\int_\Omega \log\left(\gamma^2 + (Ku-f)^2\right)\mathrm{d}x + \mu\|Ku-u_0\|_2^2\right). \tag{1.5}$$

The minimization problem in (1.5) could be solved by the primal dual algorithm proposed by Chambolle and Pock in [11] or other efficient optimization algorithms [14, 17, 26, 61, 62].

Numerical results show the potential and the effectiveness of the proposed method for restoring blurred images degraded by Cauchy noise. Furthermore, we compare the reconstructed images obtained by our method with those given by the ROF model [54], the median filter [25], the myriad filter [29], the BM3D [18], the SURE-LET [43], the wavelet shrinkage [5], and the L^1-TV model [46].

The rest of the paper is organized as follows. In section 2, we describe the alpha-stable distribution, focusing on the Cauchy distribution. Using the MAP estimator, in section 3 we derive our model for simultaneously deblurring and denoising an image and we analyze some theoretical properties of this model. Adding a quadratic penalty term, which depends on the median filter, in section 4, we propose a convex model to restore blurred and degraded images by Cauchy noise and we prove the existence and uniqueness of the solution. In section 5, using the primal dual algorithm, we show some numerical results and we compare them with the reconstructions obtained with other existing approaches. Finally, in section 6, we draw some conclusions.

2. Cauchy noise modeling.

Many studies in image and signal processing rely on the fundamental assumption that the noise follows a Gaussian distribution. This hypothesis is justified due to the existence of the central limit theorem; see [31]. Unfortunately, most of the real world problems cannot be modeled by Gaussian distribution, since the noise is much more impulsive than the one that is modeled by additive Gaussian noise. Examples of these applications can be found in the radar and sonar applications, where there are atmospheric and underwater acoustic noises, in biomedical images, in SAR images, and so on. These types of noise follow the so called alpha-stable distributions [47, 48, 55].

The alpha-stable distributions are closed under additions; i.e., the sum of two alpha-stable random variables is still an alpha-stable random variable. Moreover, the alpha-stable random variables obey to the generalized central limit theorem [48]. But, this class of random variables has no close formula for densities and distribution functions (apart from Gaussian, Cauchy, and Lévy distributions). The distributions of this class are all bell-shaped, with increasing density on the left and decreasing density on the right. The heaviness of the distribution tails is controlled by the parameter $\alpha \in (0, 2]$; i.e., the tails grow thicker as α becomes smaller.

In Figure 1, we show the probability density functions (PDFs) of alpha-stable distributions with different values of α. The distribution with $\alpha = 2$ corresponds to the well-known Gaussian

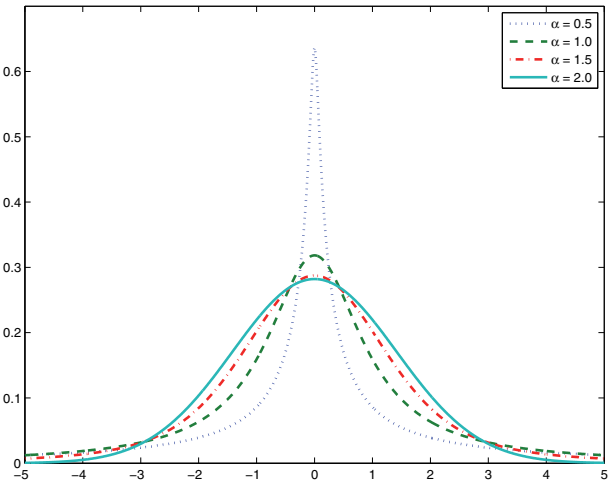
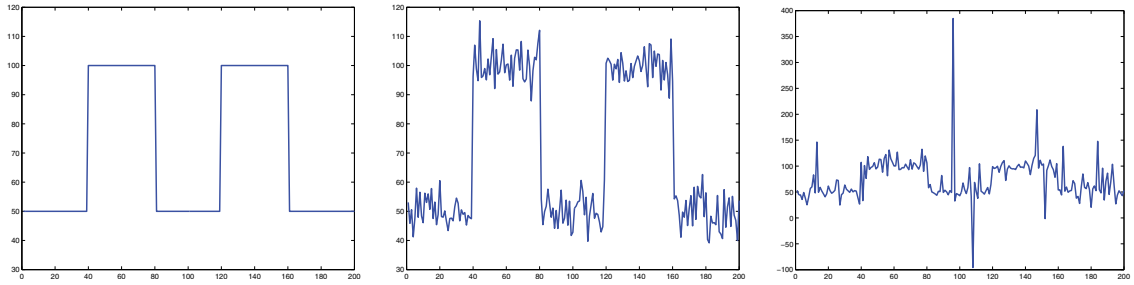


Figure 1. *Comparison of the PDFs of alpha-stable distributions with* $\alpha = 0.5$, $\alpha = 1$, $\alpha = 1.5$, *and* $\alpha = 2$.



(a) Noise free signal. (b) Degraded by Gaussian noise. (c) Degraded by Cauchy noise.

Figure 2. *Alpha-stable noise in one dimension: notice that the y-axis has different scales (scale between* 30 *and* 120 *on* (a) *and* (b) *and* −100 *and* 400 *on* (c)*).* (a) *one-dimensional noise-free signal;* (b) *signal degraded by an additive Gaussian noise;* (c) *signal degraded by an additive Cauchy noise. The Cauchy noise is more impulsive than the Gaussian noise.*

distribution, and the one with $\alpha = 1$ corresponds to the Cauchy distribution. Comparing the PDFs, we see that the tails of the bells become heavier as α decreases. In fact, the Cauchy bell ($\alpha = 1$) has a thicker tail than the Gaussian distribution ($\alpha = 2$). Thus, the rare events have more probability of occurring in the Cauchy bell curve than in the Gaussian bell curve, and for this reason, the noise generated from the Cauchy distribution is more impulsive than the Gaussian one. For instance, the Cauchy noise can contain powerful noise spikes that can be more than a hundred times the magnitude of the humbler Gaussian noise spikes.

In order to illustrate the difference between the Gaussian noise and the Cauchy noise, in Figure 2 we show a one-dimensional noise-free signal and the corresponding degraded signal by the Gaussian noise and the Cauchy noise. The noisy signal corrupted by the Gaussian noise has been obtained simply by adding random values from a Gaussian distribution. From [47, 48], we know that the Cauchy noise can be obtained from the ratio of two independent Gaussian variables. Hence, to create the noisy signal with the Cauchy noise, first we generate two vectors

containing random values from the Gaussian distribution, and then we add the ratio between these two vectors to the original signal. From the figures, one can see that the Cauchy noise is much more impulsive than the Gaussian noise, since the rare events have more probability to occur. Note that the vertical scale goes from 30 to 120 for the original signal and the one degraded by the Gaussian noise, while it goes from -100 to 400 for the signal degraded by the Cauchy noise.

Finally, we now describe how Cauchy noise influences the clean image. Given the original image $u : \Omega \to \mathbb{R}$, with $\Omega \subset \mathbb{R}^2$ being a bounded, open, and connected set with compact Lipschitz boundary, the noisy image $f : \Omega \to \mathbb{R}$ is given by

$$f = u + v,$$

where v represents the random noise that models a Cauchy distribution. A random variable V follows the Cauchy distribution, $V \sim \text{Cauchy}(\gamma, \delta)$, if it has density as in (1.1). Without loss of generality, from now on, in our analysis we consider $\delta = 0$.

3. Variational model.

In this section we analyze a variational model for deblurring and denoising images corrupted by Cauchy noise. In the first part, we focus only on the denoising case and using the maximum a posteriori (MAP) estimator (see [31]) to derive a variational model. Then, we study some properties of the restoration model, i.e., the existence of a minimizer and the minimum maximum principle. Later, we incorporate a blurring operator K in our variational model for simultaneously deblurring and denoising an image corrupted by Cauchy noise.

3.1. Variational model via MAP estimator.

Our goal is to find a variational model to restore an image corrupted by Cauchy noise; in particular, we want to recover the original image u, given the noisy image $f = u + v$, where v follows the Cauchy noise. Based on [3], we derive our model using the Bayes rule and the MAP estimator; see [31]. In the following, we denote the random variables with the uppercase letters F, U, and V, the respective instances with the lowercase letters f, u, and v, and the respective density functions with g_F, g_U, and g_V.

As already said in the previous section, we assume that v follows a "zero-centered" Cauchy law, and thus its density function is defined as follows:

$$g_V(v) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + v^2}.$$

Given the noisy image F, for restoring the original image U, we have to maximize the conditional probability $P(U|F)$. From Bayes's rule [31], we know that

$$P(U|F) = \frac{P(F|U)P(U)}{P(F)}. \tag{3.1}$$

Based on (3.1), we can equivalently minimize

$$-\log(P(U|F)) = -\log(P(F|U)) - \log(P(U)) + \log(P(F)). \tag{3.2}$$

Since the quantity $P(F)$ is constant with respect to the variable U, we just need to minimize $-\log(P(F|U)) - \log(P(U))$.

The pixels of our image are corrupted by Cauchy noise; thus for $x \in \Omega$, with Ω the set of the pixels of the image, we have

$$P(f(x)|u(x)) = P_{u(x)}(f(x)) = \frac{\gamma}{\pi\big(\gamma^2 + (u(x) - f(x))^2\big)}.$$

Inspired by [3], we assume that U follows a Gibbs prior:

$$g_U(u) = \frac{1}{Z}\exp(-\beta J(u)),$$

where Z is the normalization factor, $\beta > 0$, and J is a nonnegative given function such as $J(u) = \int_\Omega |Du|$ (the notation will be explained in the next section).

Now, since the pixels $x \in \Omega$ are mutually independent and identically distributed (i.i.d.), we have $P(U) = \prod_{x\in\Omega} P(U(x))$, where $U(x)$ is the instance of the random variable U at the pixels x. Hence, minimizing (3.2) is equivalent to minimizing

$$-\log(P(F|U)) = -\int_\Omega \Big(\log\big(P(F(x)|U(x))\big) + \log\big(P(U(x))\big)\Big). \tag{3.3}$$

Substituting the explicit expressions of $\log P(F(x)|U(x))$ and $\log P(U(x))$ in (3.3), we can easily write (3.3) as follows:

$$-\log(P(F|U)) = \int_\Omega \Big(\log\big(\gamma^2 + (U(x) - F(x))^2\big) + \beta J(U(x)) + \log\pi + \log Z - \log\gamma\Big). \tag{3.4}$$

Since the last three terms are constants, our model for restoring images corrupted with Cauchy noise is given by

$$\inf_{u\in BV(\Omega)} E(u) := \int_\Omega |Du| + \frac{\lambda}{2}\int_\Omega \log\Big(\gamma^2 + (u - f)^2\Big)\mathrm{d}x, \tag{3.5}$$

where $\lambda = \frac{2}{\beta}$ is a strictly positive parameter and we assume $f \in L^\infty(\Omega)$.

As in [3, 21, 54], in our work we consider the recovered image u in the space of the functions of bounded variation (BV). In particular, $u \in BV(\Omega)$ iff $u \in L^1(\Omega)$ and the seminorm in the space $BV(\Omega)$ is finite, where the BV-seminorm is defined as follows:

$$\int_\Omega |Du| := \sup\left\{\int_\Omega u \cdot \operatorname{div}(\xi(x))\mathrm{d}x \,\middle|\, \xi \in C_0^\infty(\Omega, \mathbb{R}^2), \|\xi\|_{L^\infty(\Omega,\mathbb{R}^2)} \le 1\right\}. \tag{3.6}$$

The space $BV(\Omega)$ endowed with the norm $\|u\|_{BV} = \|u\|_{L^1} + \int_\Omega |Du|$ is a Banach space. If $u \in BV(\Omega)$, (3.6) corresponds to the TV. From the compactness of the space $BV(\Omega)$, we have the following embedding, $BV(\Omega) \hookrightarrow L^p(\Omega)$, with $1 \le p \le 2$, and for $p < 2$ it is compact (see [2, 3] for more explanations).

In the following section, we give some theoretical results on the existence of the minimizer and we enunciate the minimum-maximum principle.

3.2. Properties of the model (3.5). We start this section with proving that there exists at least one solution for the minimization problem (3.5).

Theorem 3.1. *Let f be in $L^\infty(\Omega)$; then the problem (3.5) has at least one solution in $BV(\Omega)$ satisfying*

$$\inf_\Omega f \le u \le \sup_\Omega f.$$

Proof. Let us denote $a = \inf f$ and $b = \sup f$, and consider a minimizing sequence $\{u_n\} \in BV(\Omega)$ for (3.5). First of all, we show that we can assume $a \le u_n \le b$ without loss of generality, and so the sequence $\{u_n\}$ is bounded in $L^1(\Omega)$. Fixing $x \in \Omega$ and denoting the data fidelity term with $h : \mathbb{R} \to \mathbb{R}$, where $h(t) := \log(\gamma^2 + (t - f(x))^2)$, we have

$$h'(t) = \frac{2\big(t - f(x)\big)}{\gamma^2 + \big(t - f(x)\big)^2}.$$

Thus, the function h is decreasing if $t < f(x)$ and increasing if $t > f(x)$. For every $M \ge f(x)$, we have

$$h(\min(t, M)) \le h(t).$$

Hence, if $M = b$, we have

$$\int_\Omega \log\Big(\gamma^2 + \big(\inf(t, b) - f(x)\big)^2\Big)\mathrm{d}x \le \int_\Omega \log\Big(\gamma^2 + \big(t - f(x)\big)^2\Big)\mathrm{d}x.$$

Furthermore, from [34], we know that $\int |D \inf(u, b)| \le \int |Du|$. By definition of our functional E, we can conclude that $E(\inf(u, b)) \le E(u)$. In the same way, we can prove that $E(\sup(u, a)) \le E(u)$, with $a = \inf f$. Hence, since $a \le u_n \le b$, the sequence $\{u_n\}$ is bounded in $L^1(\Omega)$.

Now, applying our functional E in (3.5) to the sequence $\{u_n\}$, we have that $E(u_n)$ is bounded. In particular, there exists a constant $C > 0$ such that $E(u_n) \le C$. The data fidelity term has minimum value $2 \log \gamma$ when $u = f$ and $E(u_n)$ is bounded, and hence the regularization term $\int |D(u_n)|$ is also bounded. Thus, the sequence $\{u_n\}$ is bounded in $BV(\Omega)$ and there exists $u \in BV(\Omega)$ such that up to a subsequence, we have $u_n \to u$ in $BV(\Omega)$-weak and $u_n \to u$ in $L^1(\Omega)$-strong. Furthermore, using $a \le u \le b$, the lower semicontinuity of the TV, and Fatou's lemma, we have that u is a minimizer of the problem (3.5). Remark that if $\gamma \ge 1$, we can directly apply Fatou's lemma, since the logarithm does not take negative values; if $\gamma < 1$, we can still use Fatou's lemma, but we need some considerations. In fact, letting $u = \gamma v$ and $f = \gamma f_0$ (with $v \in BV(\Omega)$ and $f_0 \in L^\infty(\Omega)$), the minimization problem in (3.5) can be rewritten as follows:

$$\inf_{v \in BV(\Omega)} E(v) := \gamma \int_\Omega |Dv| + \frac{\lambda}{2} \int_\Omega \log\Big(1 + (v - f_0)^2\Big)\mathrm{d}x.$$

Hence, also in the case $\gamma < 1$, the logarithm does not take negative values, and then the use of Fatou's lemma is ensured. ∎

Now we are able to prove, under some hypothesis, that there exists a unique solution for our minimization problem (3.5).

Proposition 3.2. *Let f be in $L^\infty(\Omega)$; then the problem (3.5) has only one solution u such that $f - \gamma < u < f + \gamma$.*

Proof. Using the same notation as before and fixing $x \in \Omega$, we have

$$h''(t) = \frac{2\big(\gamma^2 - (t - f(x))^2\big)}{\big(\gamma^2 + (t - f(x))^2\big)^2},$$

where $t \in \mathbb{R}$. If $f - \gamma < t < f + \gamma$, the function h is strictly convex, and hence there exists a unique minimizer for the problem defined in (3.5). ∎

In the following proposition we enunciate the minimum-maximum principle.

Proposition 3.3. *Let f_1 and f_2 be in $L^\infty(\Omega)$, with $a_1 = \inf_\Omega f_1$, $a_2 = \inf_\Omega f_2$, $b_1 = \sup_\Omega f_1$, and $b_2 = \sup_\Omega f_2$. Let us assume that $f_1 < f_2$. Then, denoting with u_1 (resp., u_2) a solution of (3.5) for $f = f_1$ (resp., $f = f_2$), we have $u_1 \le u_2$ if $b_2 < \gamma + a_1$.*

Proof. From Theorem 3.1, we know that problem (3.5) admits solutions. Thus, by definition of u_1 and u_2 we have

$$J(u_1 \wedge u_2) + \frac{\lambda}{2}\int_\Omega \log(\gamma^2 + (u_1 \wedge u_2 - f_1)^2)\mathrm{d}x \ge J(u_1) + \frac{\lambda}{2}\int_\Omega \log(\gamma^2 + (u_1 - f_1)^2)\mathrm{d}x$$

and

$$J(u_1 \vee u_2) + \frac{\lambda}{2}\int_\Omega \log(\gamma^2 + (u_1 \vee u_2 - f_2)^2)\mathrm{d}x \ge J(u_2) + \frac{\lambda}{2}\int_\Omega \log(\gamma^2 + (u_2 - f_2)^2)\mathrm{d}x,$$

where $u_1 \wedge u_2 = \inf(u_1, u_2)$ and $u_1 \vee u_2 = \sup(u_1, u_2)$. From [9, 27], we know that $J(u_1 \wedge u_2) + J(u_1 \vee u_2) \le J(u_1) + J(u_2)$; thus, adding the two inequalities above we have

$$\int_\Omega \Big(\log(\gamma^2 + (u_1 \wedge u_2 - f_1)^2) - \log(\gamma^2 + (u_1 - f_1)^2) + \log(\gamma^2 + (u_1 \vee u_2 - f_2)^2) - \log(\gamma^2 + (u_2 - f_2)^2)\Big)\mathrm{d}x \ge 0.$$

We now split the domain Ω into two parts $\Omega = \{u_1 > u_2\} \cup \{u_1 \le u_2\}$ and deduce that

(3.7)
$$\int_{\{u_1 > u_2\}} \Big(\log(\gamma^2 + (u_2 - f_1)^2) - \log(\gamma^2 + (u_1 - f_1)^2) + \log(\gamma^2 + (u_1 - f_2)^2) - \log(\gamma^2 + (u_2 - f_2)^2)\Big)\mathrm{d}x \ge 0.$$

With the hypothesis $b_2 < \gamma + a_1$, one can prove that the integrand of the above integral is strictly negative (for further details we refer the reader to the appendix. Hence, we have that $\{u_1 > u_2\}$ has a zero Lesbegue measure, and thus we have proved that $u_1 \le u_2$ a.e. in the domain Ω. ∎

Although we proved under some conditions that there exists a unique solution for (3.5), the model is not convex. Due to the nonconvexity of (3.5), the restored results from (3.5) strongly depend on the initialization and the numerical schemes. To overcome this problem, in section 4 we introduce a convex model by adding a quadratic penalty term. Before that we first extend (3.5) to the deblurring case.

3.3. Deblurring and denoising case. Since in the real applications the observed image f is usually not only corrupted by noise but also blurred, we extend the minimization model in (3.5) to the deblurring and denoising case. In particular, the blurred and noisy image is given by $f = Ku+v$, where $K \in \mathcal{L}(L^1(\Omega), L^2(\Omega))$ is a known linear and continuous blurring operator and $v \in L^2(\Omega)$, as above, represents the Cauchy noise. In the deblurring and denoising case, the minimization problem becomes

$$\inf_{u\in BV(\Omega)} \int_\Omega |Du| + \frac{\lambda}{2}\int_\Omega \log\Big(\gamma^2 + (Ku-f)^2\Big)\mathrm{d}x. \tag{3.8}$$

As in the denoising case, (3.8) is nonconvex; in the next section, to overcome this problem we introduce a convex model.

4. Convex variational model. In this section we introduce a convex variation model for deblurring and denoising an image corrupted by Cauchy noise. At the beginning, we focus only on the denoising case, and then we generalize the model for the deblurring case. Drawing inspiration from the nonconvex model defined in (3.5), we introduce a new model by adding a quadratic penalty term that is based on the image given by applying the median filter to the noisy image f. The reason why we choose to use the median filter will be explained in subsection 4.1.

In particular, introducing a quadratic penalty term into the previous nonconvex model (3.5), we have

$$\inf_{u\in BV(\Omega)} \int_\Omega |Du| + \frac{\lambda}{2}\left(\int_\Omega \log\Big(\gamma^2 + (u-f)^2\Big)\mathrm{d}x + \mu\|u-u_0\|_2^2\right), \tag{4.1}$$

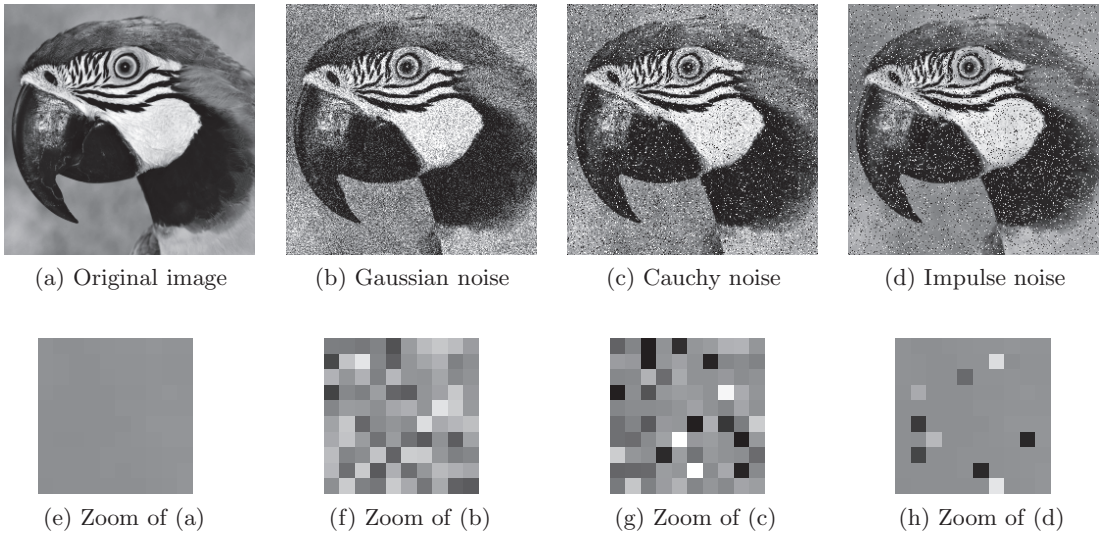
where u_0 is the image obtained by applying the median filter to the noisy image f and $\lambda > 0$ and $\mu > 0$ are the regularization parameters. In this way, we will prove that the model, under some conditions, is strictly convex.

4.1. Median filter. In this part we explain the reason why we choose the median filter [25] as a quadratic penalty term, focusing on the analogies between the Cauchy noise and impulse noise. Due to its simplicity and its capability of preserving image edges, in past decades, the median filter has attracted much attention in image processing [7, 35, 42, 57], especially for denoising images corrupted by impulse noise; see [12, 20]. Given the original image u, the noisy image f corrupted by impulse noise is defined as follows:

$$f(x) = \begin{cases} u(x) & \text{with probability } 1-\sigma, \\ \eta & \text{with probability } \sigma, \end{cases} \quad \text{with} \quad x \in \Omega,$$

where η is a uniformly distributed random variable with values in $[\min u, \max u]$ and $\sigma > 0$ is the noise level.

Figure 3(a) shows the original Parrot image, and Figures 3(b), 3(c), and 3(d), respectively, represent the images corrupted by additive Gaussian noise, impulse noise, and Cauchy noise. In Figures 3(e)–3(h) we show the zooms of the top left corners of 3(a)–3(d). One can see that the image degraded by Gaussian noise looks slightly different from the images corrupted by Cauchy noise and impulse noise, while in some way Cauchy noise and impulse noise are



(a) Original image (b) Gaussian noise (c) Cauchy noise (d) Impulse noise

(e) Zoom of (a) (f) Zoom of (b) (g) Zoom of (c) (h) Zoom of (d)

Figure 3. *Comparison of different noisy images.* (a) *Original image* u_0*;* (b) u *corrupted by an additive Gaussian noise;* (c) u *corrupted by an additive Cauchy noise;* (d) u *corrupted by an impulse noise;* (e)–(h) *zooms of the top left corners of images* (a)–(d)*, respectively. The Cauchy noise and impulse noise are more impulsive than the Gaussian noise.*

quite close to each other. For instance, with the impulse noise and the Cauchy noisy there are some pixels degraded to white or black, while the image corrupted by the Gaussian noise is uniformly modified and white and black pixels are very rare. Although the Cauchy noise has some analogies with the impulse noise, there are also some very important differences; for example, in the impulse noise some pixels are noise-free (see Figure 3(h)), while in the Cauchy noise all the pixels are corrupted by noise (see Figure 3(g)). Thus, due to the impulsive character of the Cauchy noise and to its analogies with the impulse noise, we decide to employ the median filter in our minimization problem (4.1).

In the literature, there also exist some filters created for removing noise in impulsive environments, for instance the myriad filter [29, 30]. The myriad filter theory is based on the definition of the sample myriad as the maximum likelihood location estimator of the alpha-stable distribution. It is a very robust filter for suppressing impulsive noise, in particular alpha-stable noise, but comparing with the median filter it is much more sensitive with respect to the parameter selection and much more time-consuming. In section 5, we show the results obtained by applying the median filter and the myriad filter to the noisy images, and we can see that the myriad filter slightly outperforms the median filter. Furthermore, there is not any significant improvement if, in our model, we use the myriad filter instead of the median filter. Thus, for simplicity, we keep the median filter result as u_0 in our model (4.1).

4.2. Existence and uniqueness of a solution. We now prove that under certain conditions, there exists a unique solution for the minimization problem defined in (4.1). To do this, first of all we show that under certain conditions the objective function of (4.1) is strictly convex.

Proposition 4.1. *If* $8\mu\gamma^2 \geq 1$*, the model defined in* (4.1) *is strictly convex.*

Proof. We start to prove that the data fidelity term in (4.1) is strictly convex. Fixed $x \in \Omega$, we define a function $h : \mathbb{R} \to \mathbb{R}$ as

$$h(t) := \log\Big(\gamma^2 + \big(t - f(x)\big)^2\Big) + \mu\Big(t - u_0(x)\Big)^2, \tag{4.2}$$

and we prove that it is strictly convex. Easily, we can compute the first and the second order derivatives of h, and we have

$$h'(t) = 2\frac{t - f(x)}{\gamma^2 + (t - f(x))^2} + 2\mu(t - u_0(x)) \quad \text{and} \quad h''(t) = 2\frac{\gamma^2 - (t - f(x))^2}{\big(\gamma^2 + (t - f(x))^2\big)^2} + 2\mu.$$

A direct computation shows that h is strictly convex for $8\mu\gamma^2 \geq 1$. Since TV regularization is convex, we can also conclude that the objective function in (4.1) is strictly convex, for $8\mu\gamma^2 \geq 1$, and hence we have the thesis. ∎

We now prove the existence and uniqueness of a solution to (4.1).

Theorem 4.2. *Let f be in $L^\infty(\Omega)$; then the model (4.1) has a unique solution $u \in BV(\Omega)$ satisfying*

$$\min\left\{\inf_\Omega f, \inf_\Omega u_0\right\} \leq u \leq \max\left\{\sup_\Omega f, \sup_\Omega u_0\right\}.$$

Proof. The proof of the existence of a solution to (4.1) is similar to that for Theorem 3.1. We would like to mention that in this case the function defined in (4.2) is decreasing if $t < \min\{\inf f, \inf u_0\}$ and is increasing if $t > \max\{\sup f, \sup u_0\}$.

The uniqueness of the solution follows directly from the strict convexity of our model. ∎

As in section 3, we enunciate the minimum-maximum principle for the convex minimization problem. The proof of this proposition follows the same arguments as in Proposition 3.3.

Proposition 4.3. *Let f_1 and f_2 be in $L^\infty(\Omega)$ with $a_1 = \inf_\Omega f_1$ and $a_2 = \inf_\Omega f_2$, and we denote $b_1 = \sup_\Omega f_1$ and $b_2 = \sup_\Omega f_2$. Let us assume that $f_1 < f_2$. Then, denoting with u_1 (resp., u_2) a solution of (4.1) for $f = f_1$ (resp., $f = f_2$), we have $u_1 \leq u_2$ if $b_2 < a_1 + \gamma$.*

4.3. Deblurring and denoising case. We now modify our model to include a linear and continuous blurring operator $K \in \mathcal{L}(L^1(\Omega), L^2(\Omega))$. To restore a blurred image corrupted by Cauchy noise, we introduce the following optimization problem:

$$\inf_{u \in BV(\Omega)} \int_\Omega |Du| + \frac{\lambda}{2}\left(\int_\Omega \log\Big(\gamma^2 + (Ku - f)^2\Big)\mathrm{d}x + \mu\|Ku - u_0\|_2^2\right), \tag{4.3}$$

where u_0 is the image obtained by applying the median filter to the blurred and noisy image f.

Since the blurring operator K is nonnegative and it is linear, we can conclude that the model in (4.3) is convex when $8\mu\gamma^2 \geq 1$. In the following theorem we give the existence and uniqueness results to (4.3).

Theorem 4.4. *Let f be in $L^\infty(\Omega)$, let $u_0 \in L^2(\Omega)$, and let $K \in \mathcal{L}(L^1(\Omega), L^2(\Omega))$ be a nonnegative linear operator, and let assume that K does not annihilate constant functions, i.e., $KI \neq 0$. Then the model (4.3) admits a solution. If $8\mu\gamma^2 \geq 1$ and K is injective, there exists a unique solution.*

Proof. As in the proof of Theorem 3.1, one can prove that the objective function of the minimization problem in (4.3) is bounded from below. Consider a minimizing sequence

$\{u_n\} \in \mathrm{BV}(\Omega)$ for (4.3). From the boundedness of the objective function of our model, we have that $\{\int_\Omega |Du_n|\}$ is bounded. We define $m_\Omega(u_n) = \frac{1}{|\Omega|}\int_\Omega u_n dx$ with $|\Omega|$ as the measure of Ω. Based on the Poincaré inequality [23] we have

$$\|u_n - m_\Omega(u_n)\|_2 \leq C \int_\Omega |D(u_n - m_\Omega(u_n))| = C \int_\Omega |Du_n|,$$

where C is a positive constant. Recalling that Ω is bounded, we have that $\|u_n - m_\Omega(u_n)\|_2$ and $\|u_n - m_\Omega(u_n)\|_1$ are bounded for each n. Due to the continuity of the blurring operator $K \in \mathcal{L}(L^1(\Omega), L^2(\Omega))$, the sequence $\{K(u_n - m_\Omega(u_n))\}$ is bounded in $L^2(\Omega)$ and in $L^1(\Omega)$. Furthermore, since the objective function of (4.3) is bounded, we also have that $(Ku_n - u_0)^2$ is bounded in $L^1(\Omega)$ for each n; thus $\|Ku_n - u_0\|_1$ is bounded and hence $\|Ku_n\|_1$ is bounded. We are now ready to prove that $|m_\Omega(u_n)|\|KI\|_1$ is bounded from

$$|m_\Omega(u_n)|\|KI\|_1 = \|K(u_n - m_\Omega(u_n)) - Ku_n\|_1 \leq \|K(u_n - m_\Omega)\|_1 + \|Ku_n\|_1.$$

Since $KI \neq 0$, we have that $m_\Omega(u_n)$ is uniformly bounded. Thus, now we can conclude that the sequence $\{u_n\}$ is bounded in $L^2(\Omega)$ and hence in $L^1(\Omega)$. Thus, since $BV(\Omega)$ is closed and convex, $\{u_n\}$ is also bounded in $BV(\Omega)$. Thus, there exists a subsequence $\{u_{n_k}\}$ which converges strongly in $L^1(\Omega)$ to some $u^\star \in BV(\Omega)$, and $\{Du_{n_k}\}$ converges weakly as a measure to $Du^\star$. Since K is a continuous linear operator, $\{Ku_{n_k}\}$ converges strongly to $Ku^\star$ in $L^2(\Omega)$. Moreover, up to a subsequence, $\{Ku_{n_k}\}$ converges almost everywhere to this $Ku^\star$. Based on the lower semicontinuity of TV and Fatou's lemma, we have that $u^\star$ is a solution of (4.3).

The uniqueness of the solution follows directly from the injectivity of the operator K and the assumption of $8\mu\gamma^2 \geq 1$, since in this case the model is strictly convex. ■

Note that the assumption that $K \in \mathcal{L}(L^1(\Omega), L^2(\Omega))$ is classical; see [15]. Basically, the above proof also works when K =Id. We leave the details to the interested reader.

4.4. Numerical method. In this part we show how to compute numerically the minimizer of (4.3). We focus directly on the general case, since the denoising case can be seen as a special case of the deblurring and denoising one, when K is the identity operator. First of all, we derive the discrete version of our minimization problem (4.3), and then we study how to solve it numerically. For the sake of simplicity we keep the notation from the continuous contest. Let $f \in \mathbb{R}^{mn}$ be the noisy image obtained from a two-dimensional pixel-array, with dimension $m \times n$, by concatenation in the usual columnwise fashion, and let $K \in \mathbb{R}^{mn \times mn}$ be the discretization of the continuous blurring operator K. Due to the convexity of (4.3), there exist many algorithms to solve the proposed model, for instance the primal dual algorithm [10, 14, 17], the alternating direction method with multipliers (ADMM) [6], the split-Bregman algorithm [28], and the Chambolle–Pock algorithm [11]. Since, under some hypothesis, the convergence of the Chambolle–Pock algorithm is guaranteed (see [11]), we decide to employ it to solve our minimization problem (3.8).

In order to compute numerically the solution of our minimization problem, we introduce the discrete version of (4.3),

$$\min_u \|\nabla u\|_1 + \frac{\lambda}{2} G(Ku), \tag{4.4}$$

where $G : \mathbb{R}^{mn} \to \mathbb{R}$ represents the data fidelity term and it is defined as follows:

$$G(u) := \sum_i \log\left(\gamma^2 + (u_i - f_i)^2\right) + \mu\|u - u_0\|_2^2.$$

The first term of (4.4) represents the discrete TV of the image u, and it is defined as follows:

$$\|\nabla u\|_1 = \sum_i \sqrt{(\nabla_x u)_i^2 + (\nabla_y u)_i^2}.$$

The discrete gradient $\nabla \in \mathbb{R}^{2mn \times mn}$ is given by

$$\nabla u = \begin{pmatrix} \nabla_x u \\ \nabla_y u \end{pmatrix},$$

where the discrete derivative operators in the x-direction and y-direction, respectively, ∇_x and ∇_y, are obtained using the finite difference approximations to the derivatives with symmetric boundary conditions,

$$(\nabla_x u)_{l,j} = \begin{cases} u_{l+1,j} - u_{l,j} & \text{if } l < n, \\ 0 & \text{if } l = n \end{cases} \quad \text{and} \quad (\nabla_y u)_{l,j} = \begin{cases} u_{l,j+1} - u_{l,j} & \text{if } j < m, \\ 0 & \text{if } j = m. \end{cases}$$

As in [13], for using the primal dual algorithm, we introduce two new variables $v \in \mathbb{R}^{2mn}$ and $w \in \mathbb{R}^{mn}$, and, instead of consider the unconstrained problem, we look at the following constrained optimization problem:

$$\min_{u,v,w} \|v\|_1 + \frac{\lambda}{2} G(w) \text{ subject to } v = (v_x, v_y)^\top = \nabla u \text{ and } w = Ku. \tag{4.5}$$

To apply the Chambolle–Pock algorithm, we study the following optimization problem:

$$\min_{u,v,w \in \mathbb{R}^{mn}} \max_{p,q \in Y} \|v\|_1 + \frac{\lambda}{2} G(w) + \langle v - \nabla u, p\rangle + \langle w - Ku, q\rangle, \tag{4.6}$$

where $p \in \mathbb{R}^{2mn}$ and $q \in \mathbb{R}^{mn}$ are the dual variables, and $Y = \{q \in \mathbb{R}^{2mn} : \|q\|_\infty \le 1\}$, where $\|q\|_\infty$ is the ℓ^∞-vector norm and it is defined as follows:

$$\|q\|_\infty = \max_{i \in \{1,\dots,mn\}} \sqrt{q_i^2 + q_{i+mn}^2}.$$

Then the Chambolle–Pock algorithm for solving (4.6) is described in Algorithm 1. The main calculation is carried out in (4.7)–(4.11). In the following, we give the details on how to solve them.

Algorithm 1. Solving (4.6) by using the Chambolle–Pock algorithm.

1: Fixed $\sigma > 0$ and $\tau > 0$. Initialize: $p^0 = 0$, $q^0 = 0$, $u^0 = \bar{u}^0 = f$, $v^0 = \bar{v}^0 = \nabla u^0$, and $w^0 = \bar{w}^0 = Ku^0$.

2: Calculate p^{k+1}, q^{k+1}, u^{k+1}, v^{k+1}, w^{k+1}, $\bar{u}^{k+1}$, $\bar{v}^{k+1}$, and $\bar{v}^{k+1}$ using the following equations:

$$p^{k+1} = \arg\max_p \langle \bar{v}^k - \nabla \bar{u}^k, p\rangle - \frac{1}{2\sigma}\|p - p^k\|_2^2, \tag{4.7}$$

$$q^{k+1} = \arg\min_q \langle \bar{w}^k - K\bar{u}^k, q\rangle - \frac{1}{2\sigma}\|q - q^k\|_2^2, \tag{4.8}$$

$$u^{k+1} = \arg\min_u -\langle \nabla u, p^{k+1}\rangle - \langle Ku, q^{k+1}\rangle + \frac{1}{2\tau}\|u - u^k\|_2^2, \tag{4.9}$$

$$v^{k+1} = \arg\min_v \|v\|_1 + \langle v, p^{k+1}\rangle + \frac{1}{2\tau}\|v - v^k\|_2^2, \tag{4.10}$$

$$w^{k+1} = \arg\min_w \frac{\lambda}{2}G(w) + \langle w, q^{k+1}\rangle + \frac{1}{2\tau}\|w - w^k\|_2^2, \tag{4.11}$$

$$\bar{u}^{k+1} = 2u^{k+1} - u^k, \tag{4.12}$$

$$\bar{v}^{k+1} = 2v^{k+1} - v^k, \tag{4.13}$$

$$\bar{w}^{k+1} = 2w^{k+1} - w^k. \tag{4.14}$$

3: Stop or set $k := k + 1$ and go back to step 2.

The objective functions (4.7)–(4.9) are quadratics, and thus the update of p, q, and u is given by

$$\begin{aligned} p^{k+1} &= \sigma(\bar{v}^k - \nabla \bar{u}^k) + p^k, \\ q^{k+1} &= \sigma(\bar{w}^k - K\bar{u}^k) + q^k, \\ u^{k+1} &= u^k + \tau(K^\top q^{k+1} - \mathrm{div} p^{k+1}). \end{aligned} \tag{4.15}$$

The equation in (4.10) can be rewritten in the following way:

$$v^{k+1} = \arg\min_v \|v\|_1 + \frac{1}{2\tau}\|v - t^k\|_2^2,$$

where $t^k = v^k - \tau p^{k+1}$. Thus, the update of v is easily given by applying the soft shrinkage operator,

$$v_i^{k+1} = \frac{t_i^k}{|t_i^k|}\max\{|t_i^k| - \tau, 0\} \text{ and } v_{i+mn}^{k+1} = \frac{t_{i+mn}^k}{|t_i^k|}\max\{|t_i^k| - \tau, 0\} \text{ for } i = 1, \ldots, mn,$$

with $|t_i^k| = \sqrt{(t_i^k)^2 + (t_{i+mn}^k)^2}$.

The optimality condition for (4.11) is given by

$$\lambda\frac{w - f}{\gamma^2 + (w - f)^2} + \mu\lambda(w - u_0) + q^{k+1} + \frac{1}{\tau}(w - w^k) = 0, \tag{4.16}$$

where, as usual, the division and the exponentiation have to be considered pointwise.

Multiplying both sides of the above equation for $\tau(\gamma^2+(w-f)^2)$, collecting the term with the same factors, one can see that (4.16) is equivalent to the following cubic equation:

$$aw^3+bw^2+cw+d=0,$$

with

$$\begin{aligned}
a &= \mu\lambda\tau+1;\\
b &= -(\mu\lambda\tau(2f+u_0)-\tau q^{k+1}+2f+w^k);\\
c &= \tau\lambda+\mu\lambda\tau(\gamma^2+f^2+2u_0f)-2\tau q^{k+1}f+\gamma^2+f^2+2w^kf;\\
d &= -\tau\lambda f-\mu\lambda\tau u_0(\gamma^2+f^2)+\tau q^{k+1}(\gamma^2+f^2)-w^k(\gamma^2+f^2).
\end{aligned}$$

From Cardano's formula, we can find the explicit expression for the solutions of a cubic equation; see the following proposition. For more details, we refer the reader to [37].

Proposition 4.5. *A generic cubic equation with real coefficients*

$$ax^3+bx^2+cx+d=0, \text{ with } a\neq 0, \tag{4.17}$$

has at least one solution among the real numbers. Let

$$q=\frac{3ac-b^2}{9a^2} \quad \text{and} \quad r=\frac{9abc-27a^2d-2b^3}{54a^3};$$

if there exists a unique real solution of (4.17)*, the discriminant,* $\Delta=q^3+r^2$*, has to be positive. Furthermore, if* $\Delta\geq 0$*, the only real root of* (4.17) *is given by*

$$x=\sqrt[3]{r+\sqrt{\Delta}}+\sqrt[3]{r-\sqrt{\Delta}}-\frac{b}{3a}. \tag{4.18}$$

Due to the strict convexity of our problem, we know that there exists a unique real solution for (4.16) and, from the above proposition, it can be computed explicitly using (4.18). Otherwise, since the objective function in (4.11) has the second derivative, one can also determine the solution in an efficient way using the Newton method followed by one projection step, in order to guarantee the nonnegativity of u; see [21, 34]. In our simulations, we decide to compute the explicit expression of unique real solution by using Cardano's formula.

We remark that if K is the identity operator, i.e., the degraded image f is not blurred but it is only corrupted by noise, there is no need to introduce the primal variable w and the dual variable q, and the algorithm can be simplified accordingly.

In the last part of this section, we study the existence of the solution and the convergence of the algorithm. First of all, we reformulate (4.6) in the following way:

$$\min_x \max_y \; H(x)+\langle Ax,y\rangle, \tag{4.19}$$

with $H(x)=\|v\|_1+\frac{\lambda}{2}G(w)$ and

$$A=\begin{pmatrix}-\nabla & I & 0\\ -K & 0 & I\end{pmatrix}, \quad x=\begin{pmatrix}u\\ v\\ w\end{pmatrix}, \quad \bar{x}=\begin{pmatrix}\bar{u}\\ \bar{v}\\ \bar{w}\end{pmatrix}, \quad y=\begin{pmatrix}p\\ q\end{pmatrix}.$$

Proposition 4.6. *The saddle-point set of* (4.19) *is nonempty.*

For the proof, we refer the reader to Proposition 2 in [44].

The following proposition shows the convergence of the algorithm described in Algorithm 1.

Proposition 4.7. *The iterates* (x^k, y^k) *defined in Algorithm 1 converge to a saddle point of the primal dual problem defined in* (4.19) *if* $\sigma\tau\|A\|_2^2 < 1$, *where* $\|A\|_2$ *denotes the operator* 2*-norm of* A.

This proposition can be seen as a special case of the theorem proved by Chambolle and Pock [11, Theorem 1].

In order to use the inequality given in the above proposition, we need to give an estimate of $\|A\|_2$. By using the property of the norm, one can find that

$$\|Ax\|_2 \leq \sqrt{\|\nabla\|_2^2 + \|K\|_2^2}\|u\|_2 + \left\|\begin{pmatrix} v \\ w \end{pmatrix}\right\|_2.$$

If $\|x\|_2 = 1$, by definition of x, we have that $\|u\|_2^2 + \left\|\binom{v}{w}\right\|_2^2 = 1$; therefore, from the Cauchy inequality, we obtain

$$\|Ax\|_2 \leq \sqrt{\|\nabla\|_2^2 + \|K\|_2^2 + 1}.$$

Hence, we have $\|A\|_2 \leq \sqrt{\|\nabla\|_2^2 + \|K\|_2^2 + 1}$.

From [10], we know that $\|\nabla\|_2^2 \leq 8$, and from [44], we have that $\|K\|_2 \leq 1$, and thus $\|A\|_2 \leq \sqrt{10}$. Therefore, in order to ensure the convergence of our algorithm we just need that $\sigma\tau < 0.1$. In our numerical simulations we set $\sigma = \tau = 0.3$, which ensures the convergence of the algorithm.

5. Numerical simulations. In this section, we show some numerical reconstructions obtained by applying our proposed model to blurred images corrupted by Cauchy noise. First of all, we focus only on the denoising case, and then we consider also the deblurring case. In order to show the potentiality of our method, we compare our reconstructions with images obtained by employing other well-known methods, such as the ROF model [54], the median filter [25], the myriad filter [29], and the L^1-TV model. The L^1-TV model was introduced by Nikolova in [45, 46] for restoring images corrupted by impulse noise; in particular, in this model, the TV regularization is combined with an L^1 data fidelity term. Motivated by the impulsive character of the Cauchy noise, we decide to compare our reconstructions also with the L^1-TV model. For the ROF model and the L^1-TV model, we employ the primal dual algorithm proposed in [11] to solve the minimization problem. Furthermore, in Figure 9, we also compare our method with other methods, such as the wavelet shrinkage [5], the SURE-LET [43], and the BM3D [18].

For illustrations, we use the 256-by-256 gray level images Peppers, Parrot, and Cameraman; the original images are presented in Figure 4. The quality of the restored images is compared quantitatively using the peak signal noise ratio (PSNR) value [5] and the measure of structural similarity (SSIM) [64]. The PSNR is a measure widely used in image quality assessment, and it is defined as follows:

$$\mathrm{PSNR} = 20\log_{10}\frac{mn|u_{\max} - u_{\min}|}{\|u^\star - u\|_2},$$

(a) (b) (c)

Figure 4. *Original images.* (a) *Peppers;* (b) *Parrot;* (c) *Cameraman.*

where $u^\star$ and u are, respectively, the restored and the original images with values in the gray-level range $[u_{\min}, u_{\max}]$. It is a very useful tool, since it is able to measure quantitatively the quality of the reconstructed image compared to the original image. Recently, another measure has become very popular in the imaging community, the so-called SSIM measure. This measure compares local patterns of pixel intensities that have been normalized for luminance and contrast, and it has been proved that it is more consistent with human eye perception than PSNR [64].

In our simulations, we stop our algorithm as soon as there are not big changes in the objective function, i.e.,

$$\frac{E(u^k) - E(u^{k-1})}{E(u^k)} < 5 \cdot 10^{-5},$$

where E denotes the objective function of the proposed minimization problem. In our method, we choose the regularization parameter λ to give a good balance between a good fit to f and the smoothness from TV. Since γ depends on the noise level, we use the same value of γ for all test images under the same noise level. Based on our numerical experiments, our method is robust with respect to μ; so we choose it such that the convexity condition is just satisfied, i.e., $8\mu\gamma^2 = 1$. The development of an automatic procedure for choosing these parameters is outside the scope of this paper. In addition, all the simulations are run in MATLAB R2014a.

5.1. Image denoising. In this section we focus only on the denoising case. Our aim is to recover the original image u, knowing the corrupted image f. Since the ratio of two independent standard normal variables gives a standard Cauchy random variable, we generate the noisy image f by using the following equation:

$$f = u + v = u + \xi\frac{\eta_1}{\eta_2},$$

where the random variable v follows the Cauchy distribution, $\xi > 0$ gives the noise level, and η_1 and η_2 follow the Gaussian distribution with mean 0 and variance 1.

In the following, we compare our reconstructions with those obtained by applying the ROF model, the median filter (MD), the myriad filter (MR), and the L^1-TV model. By tuning the regularization parameter λ, in the ROF model and the L^1-TV model, we use the results with the best PSNRs to compare with our method.

In Figures 5 and 7, we give the results for denoising the corrupted images Peppers, Parrot, and Cameraman for different noise levels, $\xi = 0.02$ and $\xi = 0.04$. In order to make evident

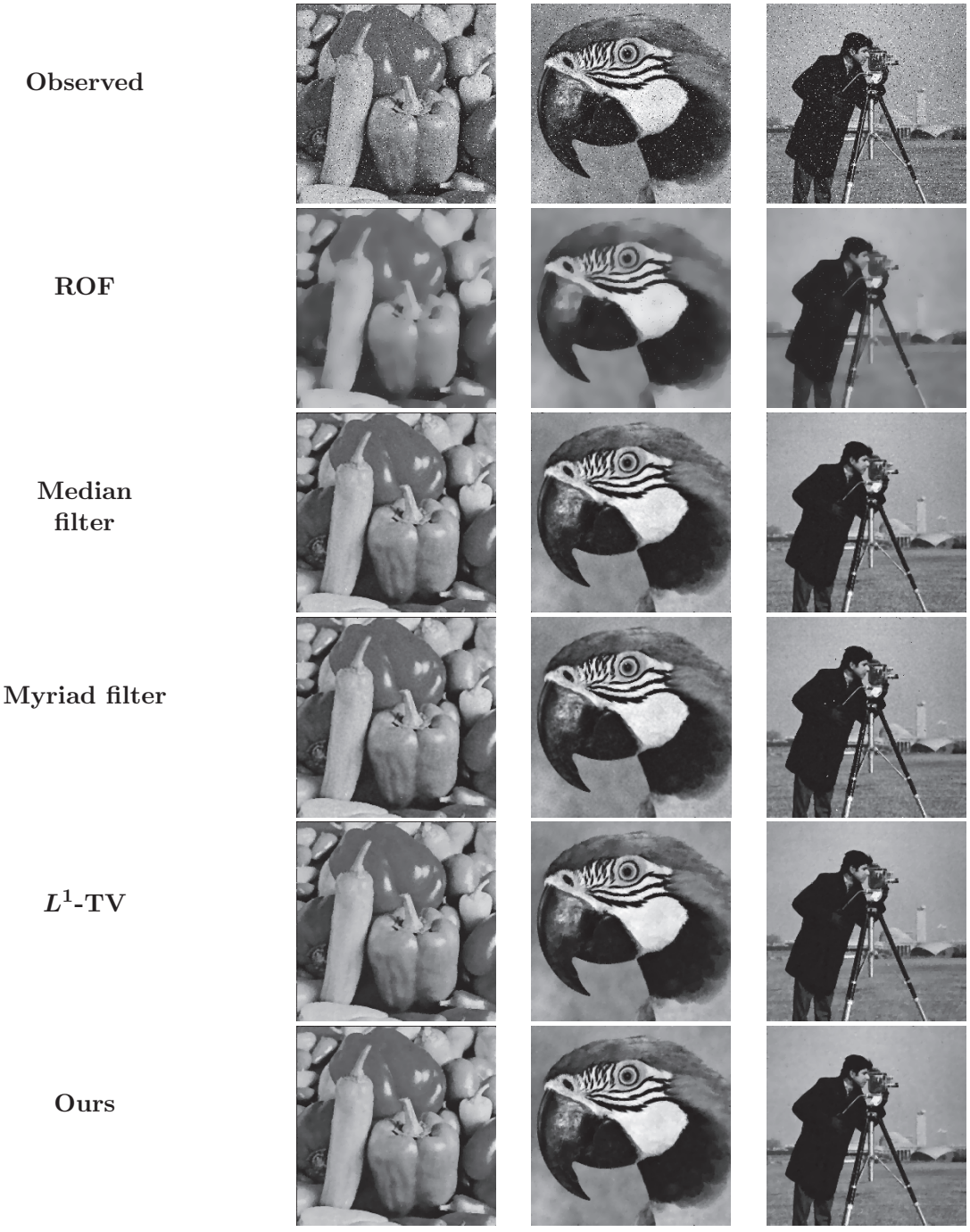


Figure 5. *Comparison of the recovered images from different methods for removing Cauchy noise in Peppers ("Pe" for short), Parrot ("Pa"), and Cameraman ("C"). First row: noisy images f ($\xi = 0.02$); second row: restored images by the ROF approach ($\lambda = 5$ ("Pe"); 6 ("Pa"); 5.8 ("C")); third row: restored images by the median filter (MD); fourth row: restored images by the myriad filter (MR); fifth row: restored images by the L^1-TV approach ($\lambda = 1.5$ ("Pe"); 1.5 ("Pa"); 1.6 ("C")); sixth row: restored images by our approach ($\lambda = 0.7$ ("Pe"); 0.8 ("Pa"); 0.7 ("C"), $\mu = 6.25$, and $\gamma = \frac{\sqrt{2}}{10}$).*

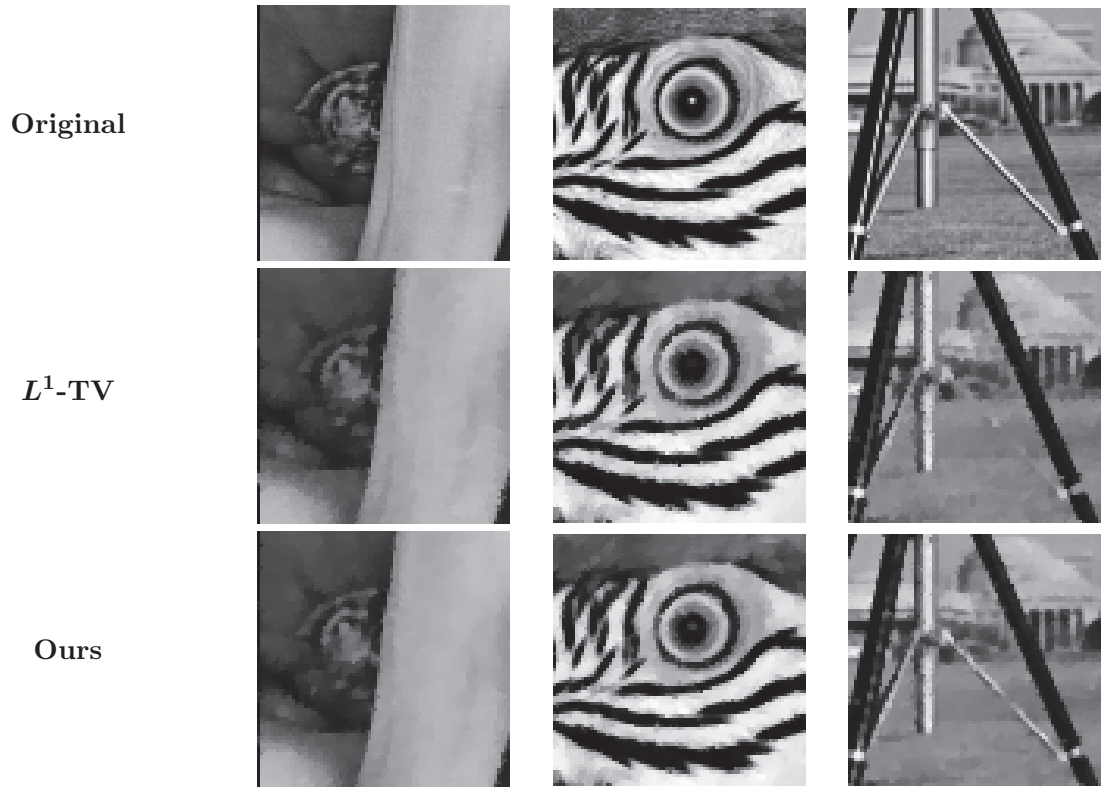


Figure 6. *The zoomed-in regions of the recovered images in Figure 5. First row: details of original images; second row: details of restored images by the L^1-TV approach; third row: details of restored images by our approach.*

the differences between the L^1-TV approach and ours, in Figure 6 we present some details of Figure 5 (here, we also include the original images in the first row). It can be seen that the L^1-TV model outperforms the ROF model, the median filter, and the myriad filter, but our method gives even better visual quality. The reason why our method and the L^1-TV approach perform better is because Cauchy noise is very impulsive and in some way it is very similar to impulse noise; see subsection 4.1. Since the ROF model was introduced for removing Gaussian noise, in order to remove highly impulsive Cauchy noise, it has to oversmooth the image. For example, in Cameraman many details are missing and the contrast of the image is reduced. The median filter and the myriad filter work quite well if the noise level is low; otherwise they are not able to eliminate all the noise and at the same time to preserve most details. From the details in Figure 6, we can see that our reconstructions preserve better the details of the image, for example the stalk of the Peppers, the eye and the stripes of the Parrot, and the tripod and the column of the building in Cameraman.

For the comparison of the performance quantitatively, in Tables 1–4 we list the values of the PSNR and SSIM for the noisy and recovered images. Here, we also provide the values of PSNR and SSIM for other popular test images in image processing, such as Lena, Baboon,

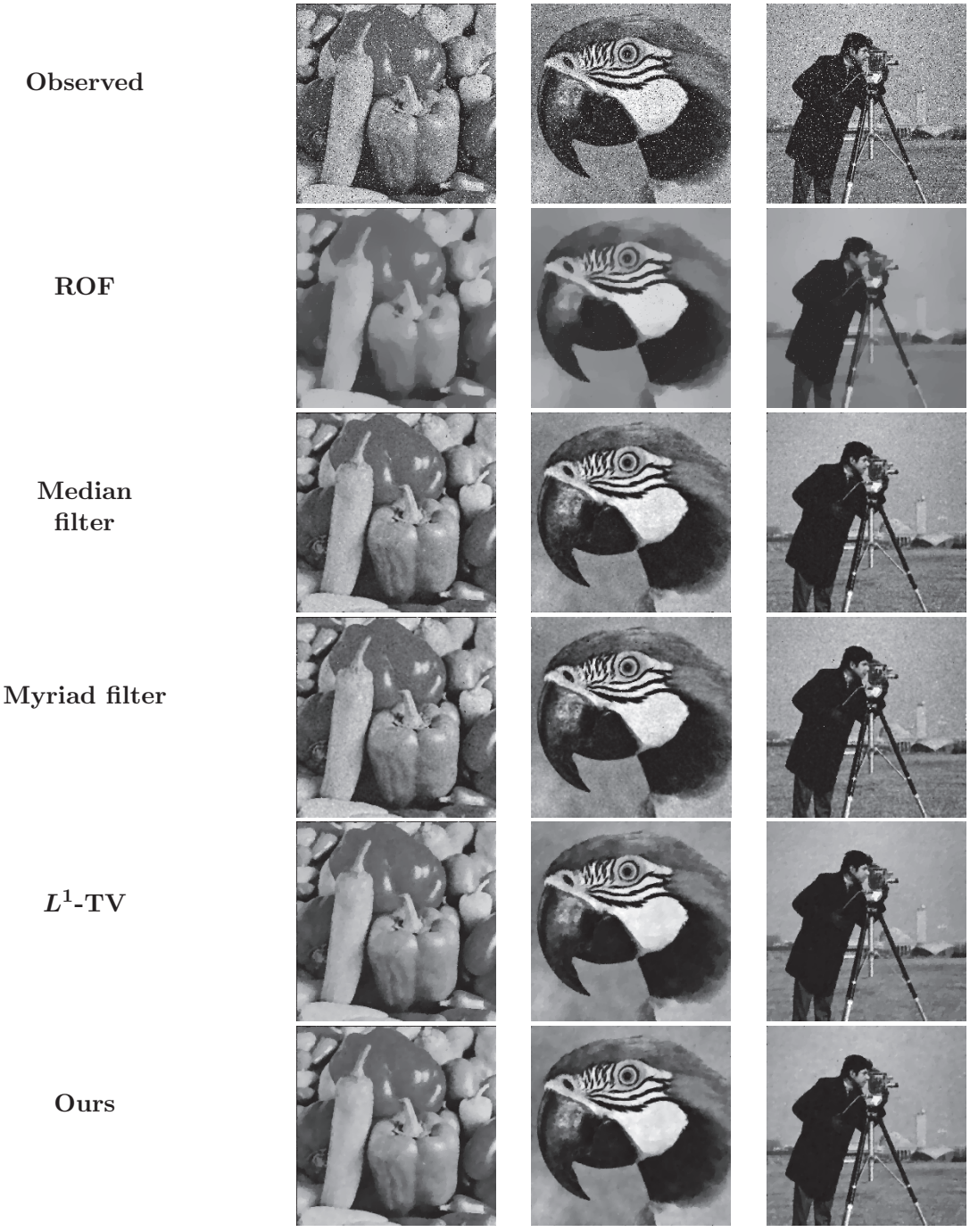


Figure 7. *Comparison of the recovered images from different methods for removing Cauchy noise. First row: noisy images f ($\xi = 0.04$); second row: restored images by the ROF approach ($\lambda = 4.5$ ("Pe"); 4.7 ("Pa"); 5 ("C")); third row: restored images by the median filter (MD); fourth row: restored images by the myriad filter (MR); fifth row: restored images by the L^1-TV approach ($\lambda = 1.3$ ("Pe"); 1.3 ("Pa"); 1.5 ("C")); sixth row: restored images by our approach ($\lambda = 0.6$ ("Pe"); 0.8 ("Pa"); 0.9 ("C"), $\mu = 3.125$, and $\gamma = 0.2$).*

Table 1

PSNR values for noisy images and recovered images given by different methods ($\xi = 0.02$). In the last line of the table, we compute the average of the values.

	Noisy	ROF	MD	MR	L^1-TV	**Ours**
Peppers	19.15	25.03	29.64	29.85	30.34	**30.94**
Parrot	19.13	23.88	27.05	27.13	28.02	**28.98**
Cameraman	19.07	24.00	26.14	26.57	27.21	**27.91**
Lena	19.06	24.58	28.94	28.98	29.84	**30.36**
Baboon	19.17	21.16	21.38	21.64	24.24	**24.96**
Goldhill	18.99	24.40	26.80	27.12	28.23	**28.80**
Boat	19.03	24.21	27.27	27.49	28.70	**29.20**
Average	19.09	23.89	26.75	26.97	28.08	**28.74**

Table 2

SSIM measures for noisy images and recovered images given by different methods ($\xi = 0.02$). In the last line of the table, we compute the average of the values.

	Noisy	ROF	MD	MR	L^1-TV	**Ours**
Peppers	0.3243	0.4820	0.6743	0.6790	0.7163	**0.7168**
Parrot	0.3179	0.4083	0.5894	0.5697	0.6571	**0.6641**
Cameraman	0.2743	0.2314	0.4115	0.4180	0.4715	**0.4707**
Lena	0.3240	0.4022	0.6488	0.6370	0.6950	**0.6880**
Baboon	0.5174	0.2115	0.4231	0.4175	0.6980	**0.6950**
Goldhill	0.3744	0.3191	0.5875	0.5952	0.6692	**0.6811**
Boat	0.3566	0.3474	0.6437	0.6387	0.6908	**0.6931**
Average	0.3556	0.3431	0.5683	0.5650	0.6568	**0.6584**

Table 3

PSNR values for noisy images and recovered images given by different methods ($\xi = 0.04$). In the last line of the table, we compute the average of the values.

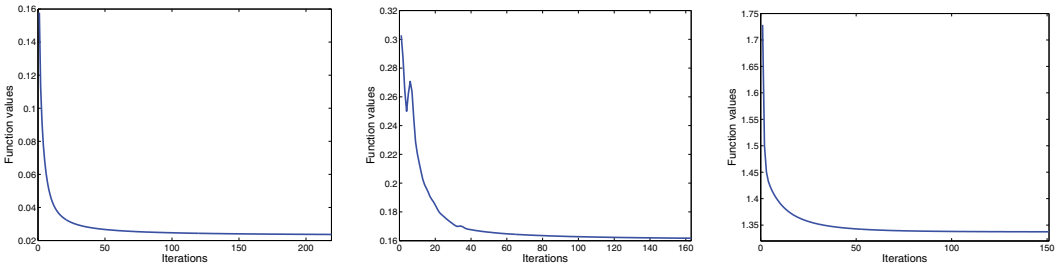
	Noisy	ROF	MD	MR	L^1-TV	**Ours**
Peppers	16.25	23.95	27.25	27.50	28.29	**28.80**
Parrot	16.27	22.75	25.50	25.85	26.55	**27.16**
Cameraman	16.08	23.17	24.87	25.19	25.99	**26.66**
Lena	16.21	24.29	26.88	27.03	28.79	**29.30**
Baboon	16.16	20.67	20.90	21.55	22.50	**23.05**
Goldhill	16.21	23.72	25.48	25.85	26.49	**27.00**
Boat	16.28	23.55	25.67	25.16	26.67	**27.18**
Average	16.21	23.16	25.22	25.45	26.47	**27.02**

Goldhill, and Boat. From Tables 1–4, we can see that, on average, with our method we can increase the PSNRs of the recovered images of 0.66 dB for $\xi = 0.02$ and 0.55 dB for $\xi = 0.04$

Table 4

SSIM measures for noisy images and recovered images given by different methods ($\xi = 0.04$). In the last line of the table, we compute the average of the values.

	Noisy	ROF	MD	MR	L^1-TV	**Ours**
Peppers	0.2246	0.4294	0.5605	0.5734	0.6347	**0.6411**
Parrot	0.2334	0.3289	0.4706	0.4800	0.5529	**0.5676**
Cameraman	0.1989	0.2081	0.3379	0.3452	0.3857	**0.3920**
Lena	0.2220	0.4025	0.5394	0.5500	0.5993	**0.6170**
Baboon	0.3651	0.1588	0.3681	0.3795	0.5525	**0.5650**
Goldhill	0.2426	0.2786	0.5108	0.5256	0.5205	**0.5684**
Boat	0.2479	0.3266	0.5429	0.5498	0.5843	**0.5930**
Average	0.2478	0.3047	0.4757	0.4862	0.5471	**0.5634**



(a) (b) (c)

Figure 8. *Plots of the objective function values versus iterations of the three TV-based methods corresponding to the experiments in the first line of Figure 5.* (a) *ROF model;* (b) *L^1-TV model;* (c) *our model.*

and also obtain largest SSIM values.

The convergence of the algorithm of the three TV-based methods is presented in Figure 8, where we plot the objective function values versus the number of iterations (we use the image of the Parrot when $\xi = 0.02$). Here, we see that the objective function values for the ROF model and ours are monotonically decreasing.

Finally, we compare our method with some well-known techniques in image denoising. Here, we use the noisy image Peppers with PSNR=19.15 and we compare with the wavelet shrinkage [5], the SURE-LET [43], and the BM3D [18]. From Figure 9, we can clearly see that our method outperforms all of them. Visually, there is still some noise left in the results from the other three methods, which is due to the impulsive behavior of Cauchy noise.

5.2. Image deblurring and denoising. In this section, we consider restoring blurred images corrupted by Cauchy noise. In our simulation, we use the Gaussian blur with a window size 9×9 and standard deviation of 1. After the blurring operation, we corrupt the images by adding Cauchy noise with $\xi = 0.02$. As in the previous section we compare our reconstructions with those obtained by employing the ROF model, the median filter, the myriad filter, and the L^1-TV model; see Figures 10 and 11. In Tables 5 and 6, we list the values of the PSNR and SSIM for different images and different variational methods.

(a) Wavelet: 23.13 (b) SURE-LET: 23.22 (c) BM3D: 28.74 (d) Ours: 30.94

Figure 9. *Recovered images (with PSNR(dB)) of different approaches for removing Cauchy noise using as the noisy image that in the first column of Figure 5.* (a) *Wavelet shrinkage;* (b) *SURE-LET;* (c) *BM3D;* (d) *our model.*

Comparing the results of the three TV-based methods, i.e., the ROF, the L^1-TV, and ours, one can see that our method performs best visually. The images given by the ROF model are too smooth, and the details are missed. The L^1-TV model preserves more details than the ROF model, but still some features are lost or not well recovered as in our model, such as the eye of the Parrot and the columns of the building in the Cameramen; see Figure 11. In the third and fourth rows of Figure 10, the reconstructions given by the median filter and the myriad filter are shown. We can see that the images are still blurred, because there are no deblurring steps in both filters. Comparing the values of the PSNR and SSIM, we can clearly see that our method outperforms the others even in presence of blur.

6. Conclusion. In this paper, we introduce a variational method for deblurring and denoising of blurred images corrupted by Cauchy noise. Inspired by the ROF model we combine a TV regularization term with a data fidelity term suitable for the Cauchy noise. In order to obtain a convex minimization problem, we add a quadratic penalty term based on the median filter. Due to the strict convexity of our problem, we are able to prove the existence and the uniqueness of a solution to our proposed model. Then, we introduce the primal dual algorithm to solve our convex minimization problem and the convergence is ensured. Numerical results show that our method outperforms other existing and well-known methods.

Appendix. Proof of Proposition 3.3. In this section we give more details on the proof of Proposition 3.3; in particular we show that the integrand of (3.7) is strictly negative.

Using the properties of the logarithm, we can rewrite the last inequality in the proof as follows:

$$\int_{\{u_1>u_2\}} \log \frac{(\gamma^2+(u_2-f_1)^2)(\gamma^2+(u_1-f_2)^2)}{(\gamma^2+(u_1-f_1)^2)(\gamma^2+(u_2-f_2)^2)} \mathrm{d}x \geq 0. \tag{A.1}$$

Now we show that, under our assumptions, the integrand of (A.1) is strictly negative. Since the argument of the logarithm in (A.1) is strictly positive, we just need to show that

$$\frac{(\gamma^2+(u_2-f_1)^2)(\gamma^2+(u_1-f_2)^2)}{(\gamma^2+(u_1-f_1)^2)(\gamma^2+(u_2-f_2)^2)} < 1. \tag{A.2}$$

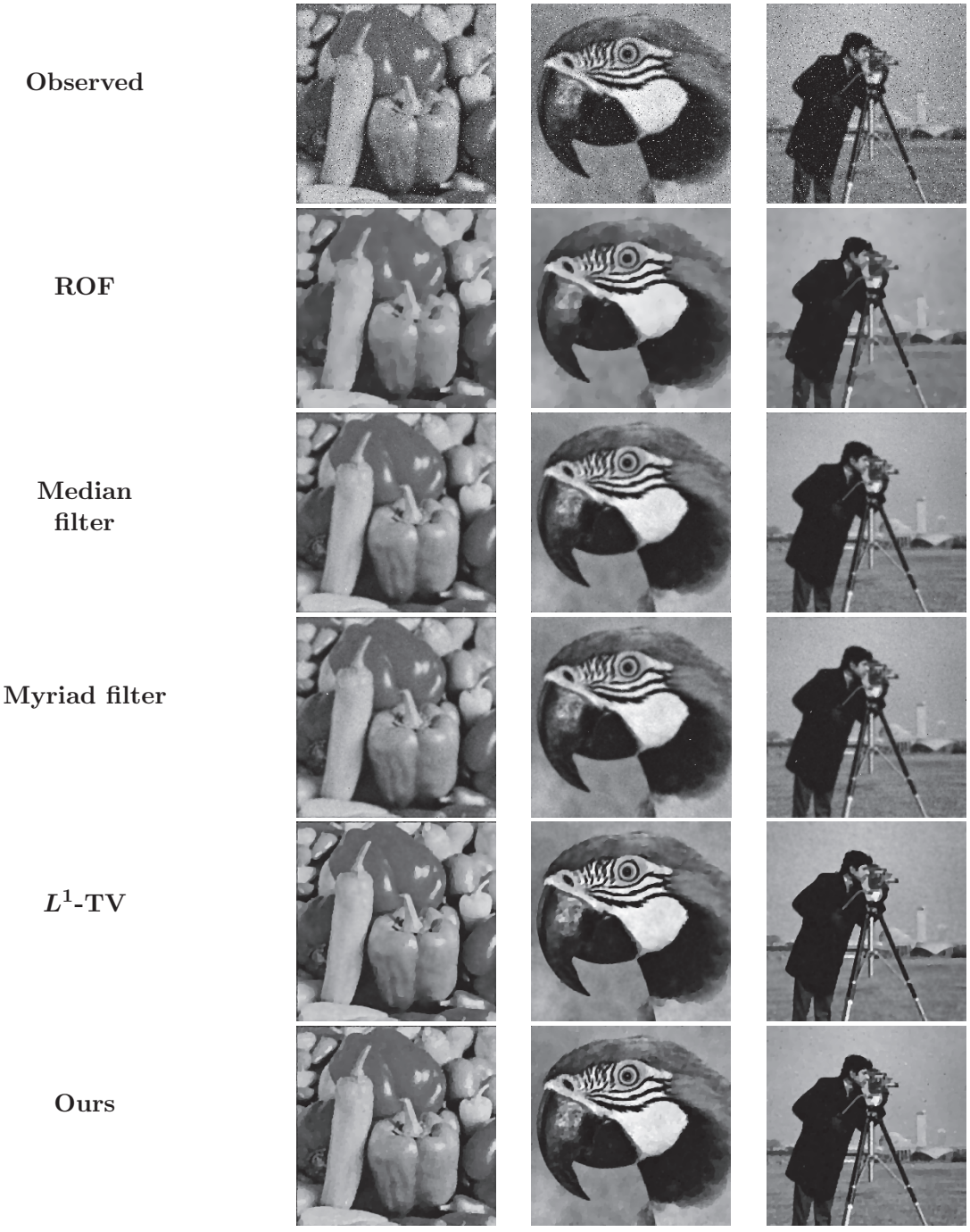


Figure 10. *Comparison of the recovered images from different methods for deblurring and denoising an image blurred and corrupted by Cauchy noise. First row: blurred and noisy images f ($\xi = 0.02$); second row: restored images by the ROF approach ($\lambda = 15$ ("Pe"); 16 ("Pa"); 16 ("C")); third row: restored images by the median filter (MD); fourth row: restored images by the myriad filter (MR); fifth row: restored images by the L^1-TV approach ($\lambda = 3$ ("Pe"); 3.4 ("Pa"); 3.5 ("C")); sixth row: restored images by our approach ($\lambda = 2$ ("Pe"); 2.1 ("Pa"); 2.1 ("C"), $\mu = 6.25$, and $\gamma = \frac{\sqrt{2}}{10}$).*

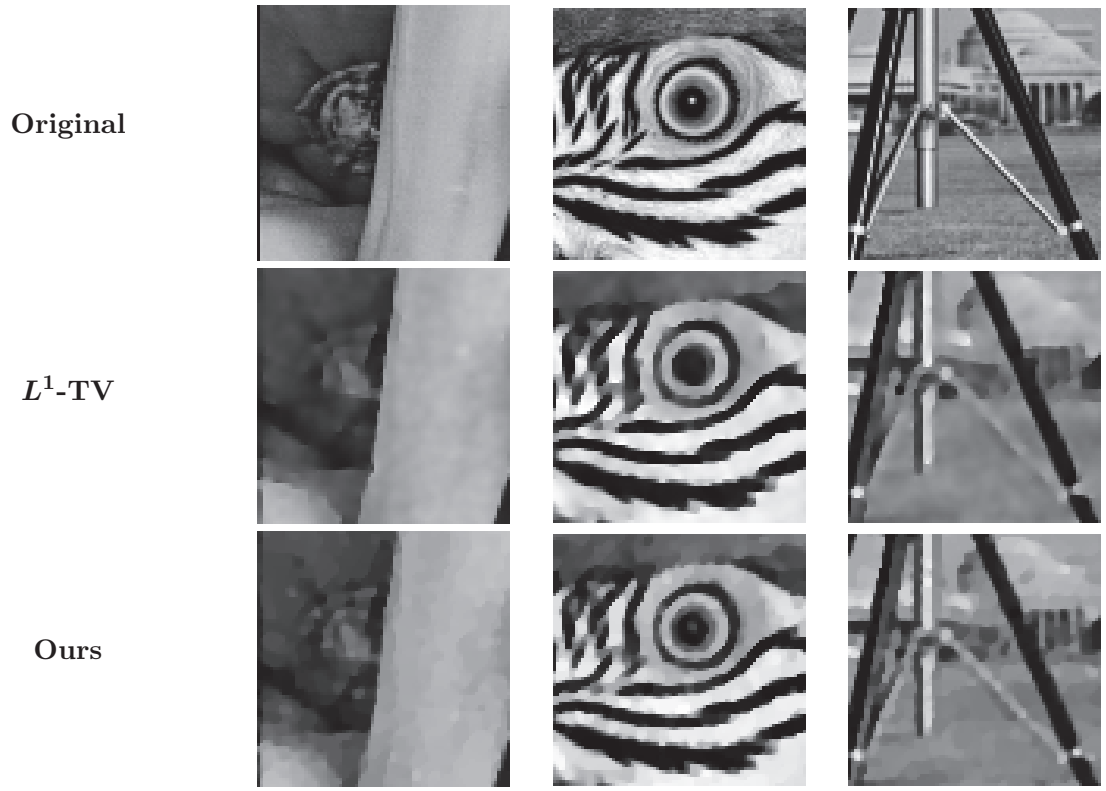


Figure 11. *The zoomed-in regions of the recovered images in Figure 10. First row: details of original images; second row: details of restored images by L^1-TV approach; third row: details of restored images by our approach.*

Table 5

PSNR values for noisy images and recovered images given by different methods ($\xi = 0.02$). In the last line of the table, we compute the average of the values.

	Noisy	ROF	MD	MR	L^1-TV	**Ours**
Peppers	18.31	24.21	25.19	25.01	26.70	**27.46**
Parrot	18.23	24.06	24.48	24.57	25.75	**26.79**
Cameraman	18.29	23.98	24.43	24.39	25.49	**26.27**
Lena	18.64	25.74	26.70	26.72	27.26	**28.14**
Baboon	17.42	20.84	21.54	21.49	21.36	**21.81**
Goldhill	18.47	24.84	25.88	25.85	26.17	**26.76**
Boat	18.48	24.36	25.42	25.43	26.18	**26.69**
Average	18.28	24.00	24.81	24.78	25.56	**26.31**

Now, collecting the term with the same factor γ^2, we have

$$\gamma^2\Big((u_2-f_1)^2+(u_1-f_2)^2-(u_1-f_1)^2-(u_2-f_2)^2\Big)+(u_2-f_1)^2(u_1-f_2)^2-(u_1-f_1)^2(u_2-f_2)^2<0.$$

Table 6

SSIM measures for noisy images and recovered images given by different methods ($\xi = 0.02$). In the last line of the table, we compute the average of the values.

	Noisy	ROF	MD	MR	L^1-TV	**Ours**
Peppers	0.2413	0.4974	0.5909	0.5762	0.6086	**0.6297**
Parrot	0.2316	0.4439	0.5145	0.4991	0.5278	**0.5655**
Cameraman	0.1753	0.2609	0.3433	0.3296	0.3516	**0.3880**
Lena	0.2487	0.4748	0.5764	0.5631	0.5712	**0.6071**
Baboon	0.1955	0.2167	0.3573	0.3502	0.3208	**0.3905**
Goldhill	0.2262	0.3678	0.5070	0.4949	0.4911	**0.5390**
Boat	0.2410	0.4059	0.5313	0.5243	0.5478	**0.5721**
Average	0.2228	0.3811	0.4887	0.4767	0.4884	**0.5274**

The above inequality can be simply rewritten as follows:

$$2\gamma^2(u_1f_1+u_2f_2-u_2f_1-u_1f_2)+\Big(f_1^2u_1^2-2u_1^2u_2f_1+u_2^2f_2^2-2u_2f_1f_2^2-2u_1u_2^2f_2-2u_1f_1^2f_2\Big)\\ -\Big(f_1^2u_2^2-2u_1u_2^2f_1+u_1^2f_2^2-2u_1f_1f_2^2-2u_1^2u_2f_2-2u_2f_1^2f_2\Big)<0,$$

and collecting some terms together we need to prove that

$$2\gamma^2(f_1-f_2)(u_1-u_2)+f_1^2(u_1^2-u_2^2)+f_2^2(u_2^2-u_1^2)\\ +2\Big(-u_1^2u_2f_1-u_2f_1f_2^2-u_1u_2^2f_2-u_1f_1^2f_2+u_1u_2^2f_1+u_1f_1f_2^2+u_1^2u_2f_2+u_2f_1^2f_2\Big)<0.$$

Thus,

$$2\gamma^2(f_1-f_2)(u_1-u_2)+(u_1-u_2)(f_1-f_2)(u_1+u_2)(f_1+f_2)\\ +2\Big(u_1^2u_2(f_2-f_1)+f_1f_2^2(u_1-u_2)+u_1u_2^2(f_1-f_2)+f_1^2f_2(u_2-u_1)\Big)<0$$

and hence

$$2\gamma^2(f_1-f_2)(u_1-u_2)+(u_1-u_2)(f_1-f_2)(u_1+u_2)(f_1+f_2)\\ +2\Big(u_1u_2(u_1-u_2)(f_2-f_1)+f_1f_2(f_2-f_1)(u_1-u_2)\Big)<0.$$

Finally, collecting $(f_1-f_2)(u_1-u_2)$ we need to prove that

$$(f_1-f_2)(u_1-u_2)\Big(2\gamma^2+(u_1+u_2)(f_1+f_2)-2(u_1u_2+f_1f_2)\Big)<0.$$

Hence, since $f_1<f_2$ and $u_1>u_2$, we just need to show that

$$2\gamma^2+(f_1+f_2)(u_1+u_2)-2(f_1f_2+u_1u_2)>0. \tag{A.3}$$

Using the Cauchy inequality, we can easily find that (A.3) is consequence of the following:

$$(\sqrt{u_1 u_2} - f_1)(\sqrt{u_1 u_2} - f_2) < \gamma^2. \tag{A.4}$$

From Theorem 3.1, we know that $a_1 \leq u_1 \leq b_1$ and $a_2 \leq u_2 \leq b_2$; thus, by the hypothesis that $f_1 < f_2$, we have $a_1 < \sqrt{u_1 u_2} < b_2$. Furthermore, $|\sqrt{u_1 u_2} - f_1| < b_2 - a_1$ and $|\sqrt{u_1 u_2} - f_2| < b_2 - a_1$. Hence, the inequality in (A.4) always holds if $(b_2 - a_1)^2 < \gamma^2$, and thus $b_2 < \gamma + a_1$.

Acknowledgment. The authors would like to thank the reviewers for their careful reading of the manuscript and the insightful and constructive comments.

REFERENCES

[1] A. Achim and E. Kuruoğlu, *Image denoising using bivariate K-stable distributions in the complex wavelet domain*, IEEE Signal Process. Lett., 12 (2005), pp. 17–20.
[2] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problem*, Oxford University Press, London, 2000.
[3] G. Aubert and J.-F. Aujol, *A variational approach to removing multiplicative noise*, SIAM J. Appl. Math., 68 (2008), pp. 925–946.
[4] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, Appl. Math. Sci. 147, Springer, New York, 2006.
[5] A. Bovik, *Handbook of Image and Video Processing*, Academic Press, New York, 2000.
[6] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2010), pp. 1–122.
[7] D. Brownrigg, *The weighted median filter*, Commun. Ass. Comput. Mach., 27 (1984), pp. 807–818.
[8] J. Cai, R. Chan, and Z. Shen, *A framelet-based image inpainting algorithm*, Appl. Comput. Harmon. Anal., 24 (2008), pp. 131–149.
[9] A. Chambolle, *An algorithm for mean curvature motion*, Interfaces Free Bound., 4 (2004), pp. 1–24.
[10] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision, 20 (2004), pp. 89–97.
[11] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145.
[12] R. Chan, Y. Dong, and M. Hintermüller, *An efficient two-phase L1-TV method for restoring blurred images with impulse noise*, IEEE Trans. Image Process., 19 (2010), pp. 1731–1739.
[13] R. Chan, H. Yang, and T. Zeng, *A two-stage image segmentation method for blurry images with Poisson or multiplicative gamma noise*, SIAM J. Imaging Sci., 7 (2014), pp. 98–127.
[14] T. F. Chan, G. H. Golub, and P. Mulet, *A nonlinear primal-dual method for total variation-based image restoration*, SIAM J. Sci. Comput., 20 (1999), pp. 1964–1977.
[15] T. F. Chan and J. Shen, *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*, SIAM, Philadelphia, 2005.
[16] Y. Chang, S. Kadaba, P. Doerschuk, and S. Gelfand, *Image restoration using recursive Markov random field models driven by Cauchy distributed noise*, IEEE Signal Process. Lett., 8 (2001), pp. 65–66.
[17] L. Condat, *A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms*, J. Optim. Theory Appl., 158 (2013), pp. 460–479.
[18] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, *Image denoising by sparse 3D transform-domain collaborative filtering*, IEEE Trans. Image Process., 16 (2007), pp. 2080–2095.
[19] N. Dey, L. Blanc-Féraud, C. Zimmer, P. Roux, Z. Kam, J. Olivo-Marin, and J. Zerubia, *Richardson-Lucy algorithm with total variation regularization for 3D confocal microscope deconvolution*, Microsc. Res. Tech., 69 (2006), pp. 260–266.
[20] Y. Dong and S. Xu, *A new directional weighted median filter for removal of random-valued impulse noise*, IEEE Signal Process. Lett., 14 (2007), pp. 193–196.

[21] Y. Dong and T. Zeng, *A convex variational model for restoring blurred images with multiplicative noise*, SIAM J. Imaging Sci., 6 (2013), pp. 1598–1625.
[22] M. Elad and M. Aharon, *Image denoising via sparse and redundant representations over learned dictionaries*, IEEE Trans. Image Process., 15 (2006), pp. 3736–3745.
[23] L. Evans, *Partial Differential Equations*, American Mathematical Society, Providence, RI, 2010.
[24] B. Figueiredo and J. Bioucas-Dias, *Restoration of Poissonian images using alternating direction optimization*, IEEE Trans. Image Process., 19 (2010), pp. 3133–3145.
[25] B. Frieden, *A new restoring algorithm for the preferential enhancement of edge gradients*, J. Opt. Soc. Amer., 66 (1976), pp. 116–123.
[26] G. Gilboa and S. Osher, *Nonlocal operators with applications to image processing*, Multiscale Model. Simul., 7 (2008), pp. 1005–1028.
[27] E. Giusti, *Minimal Surfaces and Functions of Bounded Variation*, Birkhäuser Boston, Cambridge, MA, 1984.
[28] T. Goldstein and S. Osher, *The split Bregman method for L1-regularized problems*, SIAM J. Imaging Sci., 2 (2009), pp. 323–343.
[29] J. Gonzalez and G. Arce, *Optimality of the myriad filter in practical impulsive noise environments*, IEEE Trans. Signal Process., 49 (2001), pp. 438–441.
[30] J. Gonzalez and G. Arce, *Weighted myriad filters: A robust filtering framework derived from alpha-table distributions*, in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vol. 5, Atlanta, GA, 1996, pp. 2833–2836.
[31] G. Grimmett and D. Welsh, *Probability: An Introduction*, Oxford Science Publications, London, 1986.
[32] M. El Hassouni and H. Cherifi, *Alpha-stable noise reduction in video sequences*, in Image Analysis and Recognition, A. C. Campilho and M. S. Kamel, eds., Springer, Berlin, 2004, pp. 580–587.
[33] Y. Huang, L. Moisan, M. Ng, and T. Zeng, *Multiplicative noise removal via a learned dictionary*, IEEE Trans. Image Process., 21 (2012), pp. 4534–4543.
[34] Y.-M. Huang, M. K. Ng, and Y.-W. Wen, *A new total variation method for multiplicative noise removal*, SIAM J. Imaging Sci., 2 (2009), pp. 20–40.
[35] H. Hwang and R. Haddad, *Adaptive median filters: New algorithms and results*, IEEE Trans. Image Process., 4 (1995), pp. 499–502.
[36] M. Idan and J. Speyer, *Cauchy estimation for linear scalar systems*, IEEE Trans. Automat. Control, 55 (2010), pp. 1329–1342.
[37] N. Jacobson, *Basic Algebra*, Freeman, New York, 1974.
[38] L. Jacques, L. Duval, C. Chaux, and G. Peyré, *A panorama on multiscale geometric representations, intertwining spatial, directional and frequency selectivity*, Signal Process., 91 (2011), pp. 2699–2730.
[39] B. Kosko, *Noise*, Viking Adult, New York, 2006.
[40] E. Kuruoglu, W. Fitzgerald, and P. Rayner, *Near optimal detection of signals in impulsive noise modeled with asymmetric alpha-stable distribution*, IEEE Commun. Lett., 2 (1998), pp. 282–284.
[41] T. Le, T. Chartrand, and T. Asaki, *A variational approach to reconstructing images corrupted by Poisson noise*, J. Math. Imaging Vision, 27 (2007), pp. 257–263.
[42] H. Lin and A. N. Willson, *Median filter with adaptive length*, IEEE Trans. Circuits Syst., 35 (1988), pp. 675–690.
[43] F. Luisier, T. Blu, and M. Unser, *A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding*, IEEE Trans. Image Process., 16 (2007), pp. 593–606.
[44] L. Ma, M. Ng, J. Yu, and T. Zeng, *Efficient box-constrained TV-type-l^1 algorithms for restoring images with impulse noise*, J. Comput. Math., 31 (2013), pp. 249–270.
[45] M. Nikolova, *Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers*, SIAM J. Numer. Anal., 40 (2002), pp. 965–994.
[46] M. Nikolova, *A variational approach to remove outliers and impulse noise*, J. Math. Imaging Vision, 20 (2004), pp. 90–120.
[47] J. Nolan, *Numerical calculation of stable densities and distribution functions*, Comm. Statist. Stochastics Models, 13 (1997), pp. 759–774.
[48] J. Nolan, *Stable Distributions - Models for Heavy Tailed Data*, Birkhäuser Boston, Cambridge, MA, to appear (Chapter 1 available online from http://academic2.american.edu/~jpnolan).
[49] T. Pander, *New polynomial approach to myriad filter computation*, Signal Process., 90 (2010), pp. 1991–

2001.

[50] Y. Peng, J. Chen, X. Xu, and F. Pu, *SAR images statistical modeling and classification based on the mixture of alpha-stable distributions*, Remote Sens., 5 (2013), pp. 2145–2163.

[51] N. Pustelnik, C. Chaux, and J. Pesquet, *Parallel proximal algorithm for image restoration using hybrid regularization*, IEEE Trans. Image Process., 20 (2011), pp. 2450–2462.

[52] P. Reeves, *A Non-Gaussian Turbulence Simulation*, Technical Report AFFDL-TR-69-67, Air Force Flight Dynamics Laboratory, Wright-Patterson Air Force Base, OH, 1969.

[53] L. Rudin, P. Lions, and S. Osher, *Multiplicative denoising and deblurring: Theory and algorithms*, in Geometric Level Sets in Imaging, Vision and Graphics, S. Osher and N. Paragios, eds., Springer, New York, 2003, pp. 103–119.

[54] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.

[55] G. Samorodnitsky and M. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Chapman & Hall, New York, 1994.

[56] S. Setzer, G. Steidl, and T. Teuber, *Deblurring Poissonian images by split Bregman techniques*, J. Visual Commun. Image Represent., 21 (2010), pp. 193–199.

[57] T. Sun and Y. Neuvo, *Detail-preserving median based filters in image processing*, Pattern Recogn. Lett., 15 (1994), pp. 341–347.

[58] C. Sutour, C. Deledalle, and J. Aujol, *Adaptive regularization of the NL-means: Application to image and video denoising*, IEEE Trans. Image Process., 23 (2014), pp. 3506–3521.

[59] T. Wan, N. Canagarajah, and A. Achim, *Segmentation of noisy colour images using Cauchy distribution in the complex wavelet domain*, IET Image Process., 5 (2011), pp. 159–170.

[60] P. Weiss, L. Blanc-Féraud, and G. Aubert, *Efficient schemes for total variation minimization under constraints in image processing*, SIAM J. Sci. Comput., 31 (2009), pp. 2047–2080.

[61] C. Wu and X.-C. Tai, *Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models*, SIAM J. Imaging Sci., 3 (2010), pp. 300–339.

[62] J. Yang, Y. Zhang, and W. Yin, *An efficient TVL1 algorithm for deblurring multichannel images corrupted by impulsive noise*, SIAM J. Sci. Comput., 31 (2009), pp. 2842–2865.

[63] Z. Yang and M. Jacob, *Nonlocal regularization of inverse problems: A unified variational framework*, IEEE Trans. Image Process., 22 (2013), pp. 3192–3203.

[64] W. Zhou, A. Bovik, H. Sheikh, and E. Simoncelli, *Image quality assessment: From error visibility to structural similarity*, IEEE Trans. Image Process., 13 (2004), pp. 600–612.

PAPER C

Bregman Cost for Non-Gaussian Noise

Authors:

Martin Burger, Yiqiu Dong and Federica Sciacchitano

To be submitted

Available on arXiv http://arxiv.org/abs/1608.07483

Bregman Cost for Non-Gaussian Noise

Martin Burger * Yiqiu Dong † Federica Sciacchitano ‡

Abstract One of the tasks of the Bayesian inverse problem is to find a good estimate based on the posterior probability density. The most common point estimators are the conditional mean (CM) and maximum a posteriori (MAP) estimates, which correspond to the mean and the mode of the posterior, respectively. From a theoretical point of view it has been argued that the MAP estimate is only in an asymptotic sense a Bayes estimator for the uniform cost function, while the CM estimate is a Bayes estimator for the means squared cost function. Recently, it has been proven that the MAP estimate is a proper Bayes estimator for the Bregman cost if the image is corrupted by Gaussian noise. In this work we extend this result to other noise models with log-concave likelihood density, by introducing two related Bregman cost functions for which the CM and the MAP estimates are proper Bayes estimators. Moreover, we also prove that the CM estimate outperforms the MAP estimate, when the error is measured in a certain Bregman distance, a result previously unknown also in the case of additive Gaussian noise.

1 Introduction

Bayesian estimation theory deals with the determination of the best estimate of an unknown vector from a related observation. In particular, here, we are interested in recovering the original unknown $u \in \mathbb{R}^n$ from the knowledge of an indirect measurement $f \in \mathbb{R}^m$ and the following forward degradation model

$$f = Ku \odot \eta, \tag{1.1}$$

with $K \in \mathbb{R}^{m\times n}$ being the linear forward operator and $\eta \in \mathbb{R}^m$ represents the noise, where $\odot$ might be a multiplication or a sum or even a more complicated operation. In literature, for simplicity most of the work focused on the additive white Gaussian noise, i.e. when η follows the Gaussian distribution, $\mathcal{N}(0, \Sigma)$, with mean 0 and covariance matrix Σ (often a multiple of the identity), see for instance [24, 32, 30]. However, in many real applications, the noise can be much more complicated than additive white Gaussian noise. For example, it might be impulsive [21, 26, 31], signal dependent [11, 12, 15, 18], multiplicative [2, 10, 23], or even mixed [8, 20, 17]. In this paper, we consider a general case, where the noise is just an unknown realization of a known random noise process.

*Institut für Numerische und Angewandte Mathematik, Westfälische Wilhelms-Universität Münster, Einsteinstr. 62, 48149 Münster, Germany (martin.burger@wwu.de). The work of this author was supported by by ERC via Grant EU FP 7ERC Consolidator Grant 615216 LifeInverse

†Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark (yido@dtu.dk). The work of this author was supported by Advanced Grant 291405 from the European Research Council.

‡Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark (feds@dtu.dk)

Due to the ill-posedness of the inverse problem defined in (1.1), the simple matrix inversion does not lead to any meaningful solution. Thus, to recover the original image we can employ the Bayesian approach that combines prior information on u and the forward model in (1.1) to reconstruct the image. By the Bayes' rule [14], we have

$$p(u|f) = \frac{p(f|u)p(u)}{p(f)},$$

where $p(u|f)$ is called posterior density and represents the conditional probability density of u given the noisy image f, $p(f|u)$ is the likelihood density and it encodes the likelihood that the data f is due to the image u, $p(u)$ is the prior density and it describes the properties of the image that we would like to recover, and $p(f)$ is a normalization factor independent of u. The Bayesian inference deals with extracting the useful information from the posterior to recover the sharp and noise-free image. In particular, it intends to find the best estimate by using the probability of the unknown u given the known observations f and to quantify the uncertainty of the estimate.

A common prior density is the so-called log-concave Gibbs distribution [3] (using $\propto$ for equality up to a normalization factor)

$$p(u) \propto \exp(-\alpha R(u)),$$

where $R : \mathbb{R}^n \to \mathbb{R}$ is a convex functional and $\alpha > 0$ is a scaling parameter, which is often called as regularization parameter. Two popular examples of Gibbs distribution are the Tikhonov regularization and the total variation (TV), see [27] and the references therein.

The likelihood density depends on the forward degradation model and the noise model. The most common way to write is

$$p(f|u) \propto \exp(-E(u; K, f)),$$

where E is called data fidelity term and it represents the good fit with the data based on the forward model. In this paper, we assume that E is convex with respect to u. For instance, if the original image is corrupted by Poisson noise [11, 12, 18], based on Poisson distribution E can be written as

$$E(u; K, f) = \sum_{i=1}^{n} \left[(Ku)_i - f_i \log(Ku)_i + \log(f_i!)\right] \tag{1.2}$$

with $Ku \geq 0$.

Thus, based on the Bayes' rule the posterior can be rewritten as

$$p(u|f) \propto \exp(-E(u; K, f) - \alpha R(u)), \tag{1.3}$$

up to some terms independent of u. Now, the question is how to obtain a suitable estimate of the unknown by using the information in the posterior (1.3). In the Bayesian inversion approach (cf. [22, 29]), there are two popular estimates: the maximum a posteriori (MAP) estimate, $\hat{u}_{\text{MAP}}$, and the conditional mean (CM) estimate, $\hat{u}_{\text{CM}}$. They are defined as (cf. [9, 16] for infinite-dimensional versions)

$$\begin{aligned} \hat{u}_{\text{MAP}} &= \operatorname*{argmax}_{u \in \mathbb{R}^n} p(u|f) = \operatorname*{argmin}_{u \in \mathbb{R}^n} E(u; K, f) + \alpha R(u), \\ \hat{u}_{\text{CM}} &= \mathbb{E}[u|f] = \int u\, p(u|f)\, \mathrm{d}u. \end{aligned} \tag{1.4}$$

According to their definitions, the MAP estimate corresponds to find the mode of the posterior, while the CM estimate corresponds to compute the expected value of the posterior. Of course the quality of the different estimates as representations of the posterior distribution is an important question.

Computing the MAP estimate leads to solving a high-dimensional optimization problem and the CM estimate leads to solving a high-dimensional integration problem. From the numerical point of view, the MAP estimate can be computed rather efficiently, see for instance [27], while the CM estimate requires to solve a much harder and more time-consuming integration problem. To calculate the CM estimate the classical techniques of numerical quadrature seem prohibitive in high-dimension, hence Monte Carlo methods or special sparsity techniques [28] have to be employed. Further, drawing samples from the posterior is often not straight-forward, so the Markov chain Monte Carlo (MCMC) techniques need be used, for an overview see [22]. Although there are computational challenges to calculate the CM estimate, it has many theoretical benefits. Comparing with the MAP estimate, the CM estimate is a more intuitive choice, since it represents the average of the samples. Moreover, from the theoretical point of view, the CM estimate is the Bayes estimator for the mean squared error cost, while the MAP estimate is only asymptotically the Bayes estimator for the uniform cost function. Recently, in [5], in the Gaussian noise case, it has been shown that the MAP estimate is a Bayes estimator for the cost function given by an L^2 term and a Bregman distance. More details will be given in Section 2. In [16], the results in [5] are extended to the infinite-dimensional setting.

The main novelty of this paper is to study from the Bayesian cost point of view the CM and MAP estimator in non-Gaussian noise cases. In this paper, we provide several cost functions for which the MAP and CM estimate are Bayes estimator not only under the Gaussian noise model, but under more general noise models. The only assumption that we need is the convexity and Lipschitz-continuity of E. In addition, we will show that under some assumptions the CM estimate outperforms the MAP estimate in an appropriate error measure.

The remainder of the paper is organized as follows. In Section 2 we analyse the difference between the MAP estimate and the CM estimate, and give an overview of the Bayesian approach. In Section 3 we provide a cost function for which the MAP estimate is Bayes estimator. In Section 4 we study the optimality condition for the CM estimate, and give some suitable cost functions for the CM estimate. In Section 5, we compare the two estimates by proving that the CM estimate outperforms the MAP estimate when the error is measured using a cost function that depends on the Bregman distance. The conclusion are drawn in Section 6.

2 Review of Bayes Cost Formalism

One main focus of Bayesian technique is the determination of the best estimate of an unknown data. For instance, in the inverse problem defined in (1.1) it would be to find the best estimation of the original image u, which is corrupted by blur and noise. The Bayesian estimation of u from the given noisy image f relies on the minimization of a Bayes cost,

which is defined as follows

$$\begin{aligned} BC_C(\hat{u}) :=& \mathbb{E}[C(u,\hat{u})] \\ =& \int\int C(u,\hat{u})p(u,f)\ \mathrm{d}u\ \mathrm{d}f \\ =& \int\int C(u,\hat{u})p(u|f)\ \mathrm{d}u\ p(f)\ \mathrm{d}f, \end{aligned}$$

where $C : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a cost functional measuring the distance between u and $\hat{u}$. A Bayes estimator $\hat{u}_C$ is the estimator minimizing the Bayes cost function $BC_C(\hat{u})$, that is

$$\hat{u}_C := \operatorname*{argmin}_{\hat{u}} BC_C(\hat{u}). \tag{2.1}$$

Since $\hat{u}$ only depends on f and the marginal density $p(f)$ is non-negative, the Bayes estimator can be also computed by

$$\hat{u}_C := \operatorname*{argmin}_{\hat{u}} \int C(u,\hat{u}(f))p(u|f)\ \mathrm{d}u.$$

Therefore, the Bayes estimator is always corresponding to certain cost functions, and it's very important to find a suitable cost function.

One of the most common choice for the cost function is the mean squared error, i.e.

$$C(u,\hat{u}) = \|u-\hat{u}\|_2^2, \tag{2.2}$$

and the conditional mean estimate $\hat{u}_{\mathrm{CM}}$ is the corresponding Bayes estimator. Another popular choice for the cost function is the uniform cost, i.e.

$$C(u,\hat{u}) = \begin{cases} 0, & |u_k - \hat{u}_k| < \epsilon \quad \text{for } 1 \le k \le n, \\ 1, & \text{otherwise}, \end{cases}$$

where $\epsilon > 0$ is a small constant. It turns out that the MAP estimate $\hat{u}_{\mathrm{MAP}}$ is an asymptotic Bayes estimator for this cost function.

Although it seems intuitively optimal to use the squared Euclidean norm, i.e. variance, as a Bayes cost functional, there is no real justification in high-dimensional version with a non-Gaussian prior. Assume e.g. that R is some power of a norm different from the Euclidean one (e.g. the popular ℓ^1- or total variation norm), then effectively R induces the relevant Banach space geometry on $\mathbb{R}^n$. For increasing n this geometry is very different from the Euclidean one for large n and in a limit $n \to \infty$ one might even converge to a Banach space setting where no equivalent of the Euclidean norm exists, hence the standard variance becomes questionable. Hence, in such a setting different cost functionals better adapted to the structure of induced by R shall be benefitial, both for characterizing the MAP and CM estimate.

Recently, in [5] it has been shown that the MAP estimate is a Bayes estimator for

$$C(u,\hat{u}) = \|K(\hat{u}-u)\|_2^2 + 2\alpha D_R^q(\hat{u},u),$$

where K is the blurring operator, $\alpha > 0$ is a regularization parameter, and $D_R^q(\hat{u},u)$ represents the Bregman distance between $\hat{u}$ and u for a convex regularization functional R and a subgradient $q \in \partial R(u)$, which is defined as

$$D_R^q(\hat{u},u) = R(\hat{u}) - R(u) - \langle q, \hat{u}-u\rangle.$$

If R is Fréchet differentiable in u, then the subgradient q corresponds to the standard Fréchet derivative R'. In this paper, we refer to the Bregman distance by omitting q, i.e. $D_R(\hat{u}, u)$. The Bregman distance has been introduced in [4] and it is a very useful tool in image processing, see for instance [6, 7, 13]. Since it is not symmetric and the triangle inequality does not hold, it is not a distance in the mathematical sense. But some nice properties hold, such as

- $D_R(\hat{u}, u) \geq 0$.
- If R is strictly convex, $D_R(\hat{u}, u) = 0$ implies $\hat{u} = u$.
- $D_R(\hat{u}, u)$ is convex in $\hat{u}$.

In [5], all the comparison of the MAP and CM esitmates are under the additive Gaussian noise model, i.e., $E(u; K, f) = \|Ku - f\|_2^2$, and it has been proven that the MAP estimate is a proper Bayes estimator in this case. In this paper, we will discuss and compare the MAP and CM esitmates under more general noise models. The main assumption we need is that the considered noise model leads to a convex data fitting term $E(u; K, f)$. More precisely we shall assume that the functionals $u \mapsto E(u; K, f)$ and R are nonnegative, convex and Lipschitz-continuous on $\mathbb{R}^n$ without further notice. Moreover, we assume that the posterior is well specified by (1.3), i.e.

$$\int_{\mathbb{R}^n} \exp(-E(u; K, f) - \alpha R(u))\, du < \infty.$$

3 Cost Function for the MAP Estimate

To propose a cost function for the MAP estimate, we first show that the posterior distribution in (1.3) can be rewritten in a MAP-centred form by using the optimality condition of the MAP estimate.

Since the MAP estimate $\hat{u}_{\text{MAP}} \in \mathbb{R}^n$ is a maximizer of the posterior defined in (1.3), it satisfies the optimality condition

$$K^\top \hat{q}_{\text{MAP}} + \alpha \hat{p}_{\text{MAP}} = 0, \tag{3.1}$$

where $\hat{q}_{\text{MAP}} \in \partial_u E(\hat{u}_{\text{MAP}}; K, f)$ and $\hat{p}_{\text{MAP}} \in \partial R(\hat{u}_{\text{MAP}})$. Then, we can obtain the following result.

Lemma 3.1. *The posterior in* (1.3) *can be rewritten in a MAP-centred form*

$$p(u|f) \propto \exp(-D_E^{\hat{q}_{MAP}}(u, \hat{u}_{MAP}) - \alpha D_R^{\hat{p}_{MAP}}(u, \hat{u}_{MAP})), \tag{3.2}$$

where $D_E^{\hat{q}_{MAP}}(u, \hat{u}_{MAP})$ (resp. $D_R^{\hat{p}_{MAP}}(u, \hat{u}_{MAP})$) indicates the Bregman distance between u and $\hat{u}_{MAP}$.

Proof. First, we would like to point out that we are allowed to ignore the terms independent of u, since the minimization problem in (1.4) is only on u. Then, based on the definition of

Bregman distance, we have

$$\begin{aligned}
&\exp(-D_E(u,\hat{u}_{\mathrm{MAP}}) - \alpha D_R(u,\hat{u}_{\mathrm{MAP}})) \\
&\quad \propto \exp(-E(u;K,f) + \langle \hat{q}_{\mathrm{MAP}}, Ku - K\hat{u}_{\mathrm{MAP}}\rangle - \alpha R(u) + \alpha\langle \hat{p}_{\mathrm{MAP}}, u - \hat{u}_{\mathrm{MAP}}\rangle) \\
&\quad = \exp(-E(u;K,f) - \alpha R(u) + \langle K^\top \hat{q}_{\mathrm{MAP}} + \alpha\hat{p}_{\mathrm{MAP}}, u - \hat{u}_{\mathrm{MAP}}\rangle) \\
&\quad = \exp(-E(u;K,f) - \alpha R(u)) \\
&\quad \propto p(u|f),
\end{aligned} \tag{3.3}$$

where $\hat{q}_{\mathrm{MAP}} \in \partial E(\hat{u}_{\mathrm{MAP}};K,f)$, $\hat{p}_{\mathrm{MAP}} \in \partial R(\hat{u}_{\mathrm{MAP}})$, and the last equality follows directly from the optimality condition in (3.1). □

Now, we suggest a cost function for which the MAP estimate is a Bayes estimator. For simplicity of notation we omit writing the subgradient as a superscript in the Bregman distance, since due to the Lipschitz-continuity of the involved functionals the subgradient contains a single element almost everywhere.

Theorem 3.2. *Under the decay assumption*

$$\lim_{r\to\infty} \int_{\partial\mathcal{B}_r(0)} p(u|f)\,ds = 0, \tag{3.4}$$

for the posterior defined by (1.3)*, the MAP estimate minimizes the Bayes cost with cost functional*

$$C(\hat{u},u) = D_E(\hat{u},u) + \alpha D_R(\hat{u},u). \tag{3.5}$$

Proof. Based on the definition in (2.1), $\hat{u}_{\mathrm{MAP}}$ is a Bayes estimator for the cost function $C(\hat{u},u)$, if it satisfies

$$\hat{u}_{\mathrm{MAP}} \in \arg\min_{\hat{u}} \int C(\hat{u},u)p(u|f)\ \mathrm{d}u.$$

Using the elementary identity for the Bregman distance

$$D_R(\hat{u},u) = D_R(\hat{u},\hat{u}_{\mathrm{MAP}}) + D_R(\hat{u}_{\mathrm{MAP}},u) + \langle \hat{p}_{\mathrm{MAP}} - p, \hat{u} - \hat{u}_{\mathrm{MAP}}\rangle$$

with $\hat{p}_{\mathrm{MAP}} \in \partial R(\hat{u}_{\mathrm{MAP}})$, $p \in \partial R(u)$ and ignoring the terms independent on $\hat{u}$, we have

$$\begin{aligned}
C(\hat{u},u) &= D_E(\hat{u},\hat{u}_{\mathrm{MAP}}) + \alpha D_R(\hat{u},\hat{u}_{\mathrm{MAP}}) \\
&\quad + \langle \hat{q}_{\mathrm{MAP}}, K\hat{u} - K\hat{u}_{\mathrm{MAP}}\rangle + \alpha\langle \hat{p}_{\mathrm{MAP}}, \hat{u} - \hat{u}_{\mathrm{MAP}}\rangle \\
&\quad - \langle q, K\hat{u} - K\hat{u}_{\mathrm{MAP}}\rangle - \alpha\langle p, \hat{u} - \hat{u}_{\mathrm{MAP}}\rangle \\
&= D_E(\hat{u},\hat{u}_{\mathrm{MAP}}) + \alpha D_R(\hat{u},\hat{u}_{\mathrm{MAP}}) - \langle q, K\hat{u} - K\hat{u}_{\mathrm{MAP}}\rangle - \alpha\langle p, \hat{u} - \hat{u}_{\mathrm{MAP}}\rangle,
\end{aligned}$$

where $\hat{q}_{\mathrm{MAP}} \in \partial E(\hat{u}_{\mathrm{MAP}};K,f)$, $\hat{p}_{\mathrm{MAP}} \in \partial R(\hat{u}_{\mathrm{MAP}})$, $q \in \partial E(u;K,f)$ and $p \in \partial R(u)$. Note that the last equality holds due to the optimality condition in (3.1). Now, by using the fact that $\int p(u|f)\mathrm{d}u = 1$, we have

$$\begin{aligned}
\int C(\hat{u},u)p(u|f)\,\mathrm{d}u &= D_E(\hat{u},\hat{u}_{\mathrm{MAP}}) + \alpha D_R(\hat{u},\hat{u}_{\mathrm{MAP}}) \\
&\quad - \int \left(\langle q, K\hat{u} - K\hat{u}_{\mathrm{MAP}}\rangle + \alpha\langle p, \hat{u} - \hat{u}_{\mathrm{MAP}}\rangle\right) p(u|f)\ \mathrm{d}u.
\end{aligned} \tag{3.6}$$

Based on

$$\langle q, K\hat{u}\rangle + \alpha\langle p, \hat{u}\rangle = \langle K^\top q + \alpha p, \hat{u}\rangle = \langle \nabla_u \log p(u|f), \hat{u}\rangle,$$

up to some terms independent on $\hat{u}$, we have

$$\begin{aligned}
&\min_{\hat{u}} \int C(\hat{u}, u) p(u|f)\ \mathrm{d}u \\
\Longleftrightarrow &\min_{\hat{u}} D_E(\hat{u}, \hat{u}_{\mathrm{MAP}}) + \alpha D_R(\hat{u}, \hat{u}_{\mathrm{MAP}}) - \int \left(\langle q, K\hat{u}\rangle + \alpha\langle p, \hat{u}\rangle\right) p(u|f)\ \mathrm{d}u \\
\Longleftrightarrow &\min_{\hat{u}} D_E(\hat{u}, \hat{u}_{\mathrm{MAP}}) + \alpha D_R(\hat{u}, \hat{u}_{\mathrm{MAP}}) - \int \langle \nabla_u \log p(u|f), \hat{u}\rangle p(u|f)\ \mathrm{d}u. \qquad (3.7)
\end{aligned}$$

By the assumption (3.4), we have

$$\begin{aligned}
\left\| \int \nabla_u p(u|f)\ \mathrm{d}u \right\| &= \lim_{r\to\infty} \left\| \int_{\mathcal{B}_r(0)} \nabla_u (p(u|f)) \mathrm{d}u \right\| \\
&= \lim_{r\to\infty} \left\| \int_{\partial\mathcal{B}_r(0)} p(u|f) \cdot \mathbf{n}\ \mathrm{d}s \right\| \\
&\leq \lim_{r\to\infty} \int_{\partial\mathcal{B}_r(0)} p(u|f) \mathrm{d}s \\
&= 0,
\end{aligned}$$

where the second equality follows the Gauss's theorem and $\mathbf{n}$ indicates the outward unit normal to the surface. By using the logarithmic derivative $\nabla_u p(u|f) = (\nabla_u \log p(u|f))\, p(u|f)$, we obtain

$$\begin{aligned}
\int \langle \nabla_u \log p(u|f), \hat{u}\rangle p(u|f)\ \mathrm{d}u &= \left\langle \int \nabla_u \log p(u|f) p(u|f)\ \mathrm{d}u, \hat{u} \right\rangle \\
&= \left\langle \int \nabla_u p(u|f)\ \mathrm{d}u, \hat{u} \right\rangle. \\
&= 0.
\end{aligned}$$

Back to (3.7), we get

$$\min_{\hat{u}} \int C(\hat{u}, u) p(u|f)\ \mathrm{d}u \Longleftrightarrow \min_{\hat{u}} D_E(\hat{u}, \hat{u}_{\mathrm{MAP}}) + \alpha D_R(\hat{u}, \hat{u}_{\mathrm{MAP}}).$$

Since $D_E(\hat{u}, \hat{u}_{\mathrm{MAP}}) + \alpha D_R(\hat{u}, \hat{u}_{\mathrm{MAP}})$ reaches to minimal at $\hat{u} = \hat{u}_{\mathrm{MAP}}$, we have proved that the MAP estimate is a Bayes estimator for $C(\hat{u}, u)$.

□

We finally mention that the cost function defined in [5] for the case of additive Gaussian noise is indeed a special case of (3).

4 Cost Functions for the CM Estimate

The CM estimate has been proved as a Bayes estimator for the mean squared error (2.2) independently on the noise model. In this section, we will show that there exist other cost functions for which the CM estimate is Bayes estimator. The proposed cost functions are the sum of the Bregman distances of any convex functions, and they are independent on the prior probability density and the noise model as well.

4.1 Optimality condition of the CM estimate

In [5, Sect. 3.3], under the additive Gaussian noise model it has been proved that the CM estimate fulfils an optimality condition "on average", i.e. with respect to the average gradient $\hat{p}_{\mathrm{CM}} = \mathbb{E}[\partial R(u)] = \int p\, p(u|f)\, \mathrm{d}u$ with $p \in \partial R(u)$, we have

$$K^\top (K\hat{u}_{\mathrm{CM}} - f) + \alpha \hat{p}_{\mathrm{CM}} = 0.$$

In the following theory, we prove that under more general noise models the CM estimate does not always fulfill the optimality condition "on average", and with some assumptions on the posterior, this optimality condition can be strictly positive.

Theorem 4.1. *Assume that the decay assumption* (3.4) *holds, the operator K is positive, that E is differentiable with respect to u, and the maps $u \mapsto \partial_{u_i} E(u; K, f)$, $i = 1, \dots, n$ are concave for every u. Then we have*

$$\hat{q}_{CM} + \alpha \hat{p}_{CM} \geq 0, \tag{4.1}$$

where $\hat{q}_{CM} \in \partial E(\hat{u}_{CM}; K, f)$. The equality in (4.1) *holds if and only if q is linear or a constant with respect to u.*

Proof. Based on the forward model in (1.1), the data fidelity term E is composed of Ku and f. For convenience, we also use the notation $E(Ku; f)$ instead of $E(u; K, f)$, when the operator K plays an important rule.

According to the linearity of K and the definition of $\hat{u}_{\mathrm{CM}}$, we have

$$K^\top \partial E(K\hat{u}_{\mathrm{CM}}; f) = K^\top \partial E(K\mathbb{E}[u]; f) = K^\top \partial E(\mathbb{E}[Ku]; f)$$

By Jensen's inequality, if ϕ is concave we have

$$\phi(\mathbb{E}(x)) \geq \mathbb{E}(\phi(x)),$$

and the equality holds if and only if ϕ is a constant or a linear function. Thus, by the positivity condition on the operator K, we have

$$\begin{aligned} K^\top \partial E(K\hat{u}_{\mathrm{CM}}; f) + \alpha \hat{p}_{\mathrm{CM}} &= K^\top \partial E(\mathbb{E}[Ku]; f) + \alpha \hat{p}_{\mathrm{CM}} \\ &\geq K^\top \mathbb{E}[\partial E(Ku; f)] + \alpha \int p\, p(u|f)\, \mathrm{d}u \\ &= \mathbb{E}[K^\top \partial E(Ku; f)] + \alpha \int p\, p(u|f)\, \mathrm{d}u \\ &= \int K^\top \partial E(Ku; f) p(u|f)\, \mathrm{d}u + \alpha \int p\, p(u|f)\, \mathrm{d}u \\ &= \int \nabla_u \log p(u|f) p(u|f)\, \mathrm{d}u = \int \nabla_u p(u|f)\, \mathrm{d}u = 0. \end{aligned} \tag{4.2}$$

□

Note that in this paper K indicates the blurring operator, therefore the positivity condition is always satisfied.

Example 4.2. In this example, we consider the Poisson noise model. The corresponding data fidelity term is given in (1.2), which is convex, so Theorem 4.1 can be apply to it. Based on the definition in (1.2), we have

$$q = K^\top v \quad \text{and} \quad v_i = 1 - \frac{f_i}{(Ku)_i} + g_i \in \mathbb{R}^n,$$

where $g \in \partial I_{\{i:(Ku)_i \geq 0\}}(u)$ and $I_{\{i:(Ku)_i \geq 0\}}(u)$ denotes the indicator function. Since q is concave with respect to u, based on Theorem 4.1 we can conclude that

$$\hat{q}_{\mathrm{CM}} + \alpha \hat{p}_{\mathrm{CM}} > 0.$$

In [5], under the Gaussian noise model we have

$$q = K^\top (Ku - f),$$

which is linear with respect to u, Therefore, based on Theorem 4.1 the optimality condition "on average" is satisfied.

4.2 Cost function for the CM estimate

The CM estimate has been proved as a Bayes estimator for the mean squared error (2.2) independently on the noise model. In the following theory, we prove that for any linear combinations of Bregman distances as cost functions the CM estimate is also a Bayes estimator.

Theorem 4.3. *The CM estimate is a Bayes estimator for any cost function, which is a linear combination of Bregman distances of convex functions, i.e.*

$$C_{CM}(\hat{u}, u) = \sum_{i=1}^{N} D_{F_i}^{\hat{q}_i}(u, \hat{u}), \tag{4.3}$$

where $F_1, \ldots, F_N$ are convex functions with finite expectation under $p(\cdot|f)$ and $\hat{q}_i \in \partial F_i(\hat{u})$.

Proof. Based on the definition of Bayes estimator, we need prove that

$$\hat{u}_{\mathrm{CM}} \in \arg\min_{\hat{u}} \int C_{CM}(\hat{u}, u) p(u|f) \mathrm{d}u.$$

By expanding the Bregman distance and collecting constant terms independent of $\hat{u}$, we have

$$\begin{aligned}
\int C_{CM}(\hat{u}, u) p(u|f) \, \mathrm{d}u &= \int \left(\sum_{i=1}^{N} D_{F_i}^{\hat{q}_i}(u, \hat{u}) \right) p(u|f) \, \mathrm{d}u \\
&= \int \left[\sum_{i=1}^{N} (F_i(u) - F_i(\hat{u}) - \langle \hat{q}_i, u - \hat{u} \rangle) \right] p(u|f) \, \mathrm{d}u \\
&= \sum_{i=1}^{N} (-F_i(\hat{u}) - \langle \hat{q}_i, \hat{u}_{\mathrm{CM}} - \hat{u} \rangle) + \mathrm{const} \\
&= \sum_{i=1}^{N} D_{F_i}^{\hat{q}_i}(\hat{u}_{\mathrm{CM}}, \hat{u}) + \mathrm{const}.
\end{aligned}$$

Obviously, it attains the minimal value at $\hat{u} = \hat{u}_{\mathrm{CM}}$. □

In [5], it has been shown that the cost function for the CM estimate can be independent on the prior under the additive Gaussian noise model. In Theorem 4.3 we give a much more general result. We have proved that the cost function for the CM estimate can be completely independent on the posterior density, i.e., independent on the noise model and also prior density. However, in Theorem 3.2 we have shown that the cost function for the MAP estimate depends on the likelihood and the prior density. In the end of this section, we list a few examples of cost functions for $\hat{u}_{\mathrm{CM}}$.

Example 4.4. From Theorem 4.3, we have the following functions that are suitable cost functions for the CM estimate,

$$\begin{aligned} C^1_{CM}(\hat{u}, u) &= E(u; K, f) - E(\hat{u}; K, f) - \partial E(\hat{u}; K, f)(Ku - K\hat{u}) = D_E(u, \hat{u}) \\ C^2_{CM}(\hat{u}, u) &= D_R(u, \hat{u}) \\ C^3_{CM}(\hat{u}, u) &= D_E(u, \hat{u}) + \alpha D_R(u, \hat{u}). \end{aligned}$$

Remark 4.5. *Note that $C^3_{CM}(\hat{u}, u)$ resembles the form of the cost function for the MAP estimate in* (3.5)*, but is different since the Bregman distance is not symmetric, except for the case of Gaussian posterior, when MAP and CM estimate coincide.*

5 Comparison of the CM and MAP Estimates

In [5, Thm. 2], it has been shown that under the additive Gaussian noise model the CM estimate performs better than the MAP estimate when the error is measured in a quadratic distance, $\|L(u-\hat{u})\|_2^2$ with any linear operator L. But if the error is measured by the Bregman distance $D_R(\hat{u}, u)$, the MAP estimate outperforms the CM estimate. In this section, based on the results in Theorem 4.3, we give another comparison result under more general noise models, for instance the Laplacian noise or the Poisson noise.

Theorem 5.1. *The CM estimate outperforms the MAP estimate when the error is measured in the Bregman distance $D_R(u, \hat{u})$, i.e.*

$$\mathbb{E}[D_R(u, \hat{u}_{CM})] \leq \mathbb{E}[D_R(u, \hat{u}_{MAP})].$$

This inequality directly follows from the fact that $\hat{u}_{\mathrm{CM}}$ is a Bayes estimator for $C^2_{CM}(\hat{u}, u)$. Note that the above result is exactly the opposite of Theorem 2 in [5] under the additive Gaussian noise model, but with flipped $\hat{u}$ and u in the definition of cost function.

6 Conclusions

In this paper, based on image restoration problem with the more general noise models instead of additive Gaussian noise, we study the two typical point estimators for the posterior probability density: the conditional mean (CM) estimate and the maximum a posteriori (MAP) estimate. The only assumption that we need is that the considered noise model has to lead to a convex data fidelity term. Based on the Bregman distance, we propose new cost functions for which the MAP and the CM estimate are Bayes estimators. Further, we give a new comparison result on these two estimates. In addition, we give the posterior in a MAP-centred form and study the optimality condition on average of the CM estimate.

References

[1] Abergel, R., Louchet, C., Moisan, L., and Zeng, T. (2015). *Total Variation Restoration of Images Corrupted by Poisson Noise with Iterated Conditional Expectations*, In Scale Space and Variational Methods in Computer Vision (pp. 178-190). Springer International Publishing, 2015.

[2] G. Aubert and J. Aujol, *A variational approach to removing multiplicative noise*, SIAM J. Appl. Math., 68 (2008), pp. 925–946.

[3] A. Bovik, *Handbook of Image and Video Processing*, New York: Academic, 2010.

[4] L.M. Bregman, *The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming*, USSR computational mathematics and mathematical physics 7.3 (1967): 200-217.

[5] M. Burger and F. Lucka, *Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper Bayes estimators*, Inverse Problems 30 (2014).

[6] M. Burger and S. Osher, *Convergence rates of convex variational regularization* Inverse problems 20.5 (2004): 1411.

[7] M. Burger, E. Resmerita, and L. He, *Error estimation for Bregman iterations and inverse scale space methods in image restoration*, Computing 81.2-3 (2007): 109-135.

[8] Calatroni, L., Chung, C., Reyes, J. C. D. L., Schönlieb, C. B., and Valkonen, T. , *Bilevel approaches for learning of variational imaging models*, arXiv preprint arXiv:1505.02120, (2015).

[9] M. Dashti, K.J.H. Law, A.M. Stuart and J. Voss, *MAP estimators and posterior consistency in Bayesian nonparametric inverse problems*, Inverse Problems, 29(2013) 095017.

[10] Y. Dong and T. Zeng, *A convex variational model for restoring blurred images with multiplicative noise*, SIAM J. Imaging Sci., 6 (2013), pp. 1598–1625.

[11] B. Figueiredo and J. Bioucas-Dias, *Restoration of Poissonian images using alternating direction optimization*, IEEE Trans. Image Process., 19 (2010), pp. 3133-3145.

[12] A. Foi, *Noise estimation and removal in MR imaging: the variance-stabilization approach*, IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 18091814 (2011)

[13] T. Goldstein and S. Osher, *The split Bregman method for L1-regularized problems* SIAM journal on imaging sciences 2.2 (2009): 323-343.

[14] G. Grimmett and D. Welsh, *Probability: an Introduction*, Oxford Science Pubblications, London, 1986.

[15] H. Gudbjartsson and S. Patz, *The Rician distribution of noisy MRI data*, J.Magn. Reson. Med. 34(6), 910914 (1995)

[16] T. Helin and M. Burger, *Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems*, Inverse Problems 31.8 (2015): 085009.

[17] A. Lanza, S. Morigi, F. Sgallari, and Y. W. Wen, *Image restoration with PoissonGaussian mixed noise*. Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization, 2(1), 12-24, 2014.

[18] T. Le, T. Chartrand, and T. Asaki, *A variational approach to reconstructing images corrupted by Poisson noise*, J.Math. Imaging Vis., 27 (2007), pp. 257–263.

[19] C. Louchet, and L. Moisan, *Posterior expectation of the total variation model: properties and experiments*, SIAM Journal on Imaging Sciences 6.4 (2013): 2640-2684.

[20] F. Luisier, T. Blu, and M. Unser, *Image denoising in mixed Poisson-Gaussian noise* Image Processing, IEEE Transactions on 20.3 (2011): 696-708.

[21] M. Nikolova, *A variational approach to remove outliers and impulse noise*, J. Math. Imaging Vis., 20 (2004), pp. 90–120.

[22] J. Kaipio and E. Somersalo, *Statistical and computational inverse problems*, Vol. 160. Springer Science and Business Media, 2006.

[23] L. Rudin, P. Lions, and S. Osher, *Multiplicative denoising and deblurring: theory and algorithms*, S. Osher and N. Paragios, eds, Geometric Level Sets in Imaging, Vision and Graphics, pp. 103–119, Springer, 2003.

[24] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.

[25] H. Rue and M. A. Hurn, *Loss functions for Bayesian image analysis*, Scand. J. Statist., 24 (1997), pp. 103114.

[26] F. Sciacchitano, Y. Dong, and T. Zeng, *Variational approach for restoring blurred images with cauchy noise*, SIAM Journal on Imaging Sciences 8.3 (2015): 1894-1922.

[27] O. Scherzer, *Handbook of mathematical methods in imaging* Springer Science and Business Media, 2010.

[28] C.Schwab, A.M.Stuart, *Sparse deterministic approximation of Bayesian inverse problems*, Inverse Problems, Vol.28 (2012), 045003.

[29] A.M. Stuart, *Inverse problems: a Bayesian perspective*, Acta Numerica 19 (2010), 451–559.

[30] T. Chan and J. Shen, *Image Processing And Analysis: Variational, PDE, Wavelet, And Stochastic Methods*, SIAM Publisher, 2005.

[31] J. Yang, Y. Zhang, and W. Yin, *An efficient TVL1 algorithm for deblurring multichannel images corrupted by impulsive noise*, SIAM J. Sci. Comput., 31 (2009), pp. 2842–2865.

[32] P. Weiss, L. Blanc-Féraud, and G. Aubert, *Efficient schemes for total variation minimization under constraints in image processing*, SIAM J. Sci. Comput., 31 (2009), pp. 2047–2080.