

Vom Fachbereich Mathematik der Technischen Universität Kaiserslautern  
zur Verleihung des akademischen Grades  
Doktor der Naturwissenschaften (Doctor rerum naturalium, Dr. rer. nat.)  
genehmigte

**Dissertation**

**Anisotropic Smoothing and  
Image Restoration Facing  
Non-Gaussian Noise**

Tanja Teuber

Gutachter: Prof. Dr. Gabriele Steidl  
Prof. Dr. Martin Burger

Datum der Disputation: 4. Juli 2012

### **Promotionskommission**

Vorsitzender: Prof. Dr. Claus Fieker, TU Kaiserslautern  
Erstgutachterin: Prof. Dr. Gabriele Steidl, TU Kaiserslautern  
Zweitgutachter: Prof. Dr. Martin Burger, Universität Münster  
Weiterer Prüfer: Prof. Dr. Jürgen Franke, TU Kaiserslautern

# Acknowledgment

I would like to thank

**Gabriele Steidl** for her guidance, constant support and encouragement during the past years and for the opportunity to work in this fascinating field,

**Martin Burger** from the Westfälische Wilhelms-Universität Münster for serving as a referee of this thesis and giving me the possibility to present my work at the CIME Summer School on "Level Set and PDE based Reconstruction Methods: Applications to Inverse Problems and Image Processing" in Cetraro/Italy 2008,

my former college **Simon Setzer** from the Saarland University in Saarbrücken for being a great college who I always enjoyed working with and discussing ideas and open problems,

my current colleges **Sören Häuser**, **René Ciak**, **Behrang Shafei**, **Stanislav Harizanov** and our secretary **Kirsten Höffler** at the University of Kaiserslautern for the very good working atmosphere and many interesting discussions,

**Annika Lang** from the ETH Zurich for the fruitful collaboration on nonlocal filtering approaches and similarity measures for data corrupted by multiplicative noise,

**Markus Grasmair** from the University of Vienna for his interest in discussing open problems and for his support in answering one of the theoretical questions in this thesis through his preprint [99],

**Pascal Gwosdek**, **Christian Schmaltz** and **Joachim Weickert** from the Saarland University in Saarbrücken, **Guido Moerkotte** from the University of Mannheim, **Daniel J. Strauß** and **Farah I. Corona-Strauß** from the Neurocenter at the Saarland University Hospital in Homburg as well as **Steffen Remmele** and **Jürgen Hesser** from the University Medical Center Mannheim for fruitful collaborations and discussions on topics beyond this thesis,

many other people including **Jan Lellmann** from the University of Cambridge and **Christoph Brune** from the University of California Los Angeles for many interesting discussions and insights at various conferences and workshops,

---

**Marc Sigelle** and **Florence Tupin** from Telecom ParisTech for providing the radar and optical data used in Figures 3.12 and 4.12, respectively,

the authors of [70], in particular **Mila Nikolova** (CMLA, ENS Cachan), for providing the initial data and the result of their method contained for comparison in Figure 4.11,

last and foremost **my family and friends** for supporting me and being there for me all the time.

I also would like to acknowledge the financial support of

the State Baden-Württemberg through the provided doctoral scholarship and of the German Research Foundation (DFG) through the project Ste 571/9-1.



# Summary

Image restoration and enhancement methods that respect important features such as edges play a fundamental role in digital image processing. In the last decades a large variety of methods have been proposed. Nevertheless, the correct restoration and preservation of, e.g., sharp corners, crossings or texture in images is still a challenge, in particular in the presence of severe distortions. Moreover, in the context of image denoising many methods are designed for the removal of additive Gaussian noise and their adaptation for other types of noise occurring in practice requires usually additional efforts.

The aim of this thesis is to contribute to these topics and to develop and analyze new methods for restoring images corrupted by different types of noise:

First, we present variational models and diffusion methods which are particularly well suited for the restoration of sharp corners and X junctions in images corrupted by strong additive Gaussian noise. For their deduction we present and analyze different tensor based methods for locally estimating orientations in images and show how to successfully incorporate the obtained information in the denoising process. The advantageous properties of the obtained methods are shown theoretically as well as by numerical experiments. Moreover, the potential of the proposed methods is demonstrated for applications beyond image denoising.

Afterwards, we focus on variational methods for the restoration of images corrupted by Poisson and multiplicative Gamma noise. Here, different methods from the literature are compared and the surprising equivalence between a standard model for the removal of Poisson noise and a recently introduced approach for multiplicative Gamma noise is proven. Since this Poisson model has not been considered for multiplicative Gamma noise before, we investigate its properties further for more general regularizers including also nonlocal ones. Moreover, an efficient algorithm for solving the involved minimization problems is proposed, which can also handle an additional linear transformation of the data. The good performance of this algorithm is demonstrated experimentally and different examples with images corrupted by Poisson and multiplicative Gamma noise are presented.

In the final part of this thesis new nonlocal filters for images corrupted by multiplicative noise are presented. These filters are deduced in a weighted maximum likelihood estimation framework and for the definition of the involved weights a new similarity

---

measure for the comparison of data corrupted by multiplicative noise is applied. The advantageous properties of the new measure are demonstrated theoretically and by numerical examples. Besides, denoising results for images corrupted by multiplicative Gamma and Rayleigh noise show the very good performance of the new filters.

# Zusammenfassung

Methoden zur Restaurierung und Verbesserung von digitalen Bildern, die wichtige Strukturen wie zum Beispiel Kanten erhalten, spielen eine grundlegende Rolle in der digitalen Bildverarbeitung. Obwohl im Laufe der letzten Jahrzehnte eine Vielzahl solcher Methoden entwickelt wurde, ist die korrekte Restauration und Erhaltung von beispielsweise scharfen Ecken, Kreuzungen oder Texturen insbesondere in stark degradierten Bildern immer noch eine Herausforderung. Überdies sind viele Methoden zum Entrauschen von Bildern auf additives Gauß'sches Rauschen zugeschnitten und deren Anpassung für andere Arten von in der Praxis auftretendem Rauschen ist typischerweise nicht ohne Weiteres möglich.

Ziel dieser Arbeit ist es einen Beitrag zu diesen Themen zu leisten und neue Methoden zur Restauration von Bildern mit unterschiedlichen Arten von Rauschen aufzuzeigen und zu analysieren:

Im ersten Teil der Arbeit werden Variationsmethoden und Diffusionsverfahren vorgestellt, die insbesondere dafür geeignet sind scharfe Ecken und X-förmige Kreuzungen beim Entrauschen von Bildern zu erhalten, die starkes additives Gauß'sches Rauschen aufweisen. Für deren Herleitung werden verschiedene Tensor-basierte Methoden zur lokalen Bestimmung von Richtungen in Bildern vorgestellt und analysiert. Anschließend wird gezeigt wie die hierdurch gewonnenen Richtungsinformationen erfolgreich zum Entrauschen verwendet werden können. Die sehr guten Eigenschaften der entwickelten Methoden werden theoretisch sowie durch numerische Experimente belegt. Des Weiteren wird das Potenzial der Methoden für Anwendungen über das Entrauschen von Bildern hinaus aufgezeigt.

Anschließend werden Variationsmethoden zur Restauration von Bildern untersucht, die Poisson- und multiplikatives Rauschen aufweisen. Hier werden verschiedene Methoden aus der Literatur verglichen und die überraschende Äquivalenz zwischen einem Standardmodell zum Entfernen von Poisson-Rauschen und einem kürzlich vorgeschlagenen Modell für multiplikatives Gamma-Rauschen bewiesen. Da das für Poisson-Rauschen entwickelte Modell bisher nicht für multiplikatives Gamma-Rauschen in Betracht gezogen wurde, werden die Eigenschaften des Modells für allgemeinere, auch nicht-lokale Regularisierer weiter untersucht. Darüber hinaus wird ein effizienter Algorithmus zum Lösen des im Modell enthaltenen Minimierungsproblems vorgestellt, welcher es auch erlaubt eine zusätzliche lineare Transformation der Daten zu berücksichtigen. Die Effi-

---

zienz des Algorithmus wird experimentell belegt und es werden verschiedene Beispiele für Bilder mit Poisson- und multiplikativem Gamma-Rauschen gezeigt.

Im letzten Teil der Arbeit werden neue nicht-lokale Filter zum Entfernen von multiplikativem Rauschen vorgestellt. Diese Filter basieren auf einem gewichteten Maximum-Likelihood Ansatz und für die Definition der im Ansatz enthaltenen Gewichte wird eine neues Ähnlichkeitsmaß zum Vergleich von mit multiplikativem Rauschen behafteten Daten hergeleitet. Die vorteilhaften Eigenschaften des neuen Maßes werden theoretisch und experimentell belegt. Zudem zeigen Ergebnisse für Bilder mit multiplikativem Gamma- und Rayleigh-Rauschen die hohe Leistungsfähigkeit der neuen Filter.

# Contents

<b>List of Publications</b>	<b>ix</b>
<b>Notation and Symbols</b>	<b>xi</b>
<b>1 Introduction and Overview</b>	<b>1</b>
Contributions . . . . .	2
Outline . . . . .	5
<b>2 Anisotropic Smoothing Facing Additive Gaussian Noise</b>	<b>7</b>
2.1 Methods of Interest . . . . .	8
2.2 Properties of Existing Methods and Motivation . . . . .	13
2.3 Orientation Estimation . . . . .	16
2.3.1 Classical Structure Tensor . . . . .	16
2.3.2 Adaptations of the Structure Tensor for Special Shapes . . . . .	21
2.3.3 Double Orientation Estimation . . . . .	24
2.4 Shape Preserving Anisotropic Regularization . . . . .	29
2.4.1 Shape Preservation . . . . .	30
2.4.2 A New Anisotropic Regularization Method . . . . .	36
2.4.3 Discretization and Minimization . . . . .	38
2.5 A Related Anisotropic Diffusion Model . . . . .	39
2.5.1 The Model . . . . .	39
2.5.2 Discretization . . . . .	42
2.6 Anisotropic Infimal Convolution Regularization . . . . .	43
2.6.1 Infimal Convolution Regularization and Shape Preservation . . . . .	43
2.6.2 Discretization and Minimization . . . . .	50
2.7 Numerical Experiments . . . . .	50
2.7.1 Results for Images with Rotated or Sheared Shapes . . . . .	51
2.7.2 Results for Images with Arbitrary Double Orientations . . . . .	54
<b>3 Variational Approaches Facing Poisson and Multiplicative Noise</b>	<b>61</b>
3.1 Notation and Preliminaries . . . . .	62
3.2 Poisson and Multiplicative Noise in Imaging . . . . .	62
3.2.1 Poisson Noise . . . . .	62

3.2.2	Multiplicative Noise . . . . .	64
3.3	Variational Approaches by Bayesian Modeling . . . . .	66
3.3.1	Variational Approaches for Poisson Noise . . . . .	66
3.3.2	Variational Models for Multiplicative Gamma Noise . . . . .	69
3.4	Equivalence of the SO-Model and IDIV-TV . . . . .	75
3.5	Theoretical Results for the Discrete Setting . . . . .	81
3.6	Minimization Algorithms . . . . .	85
3.6.1	ADMM and Multiple Splittings . . . . .	86
3.6.2	Minimization of the Discrete IDIV Model . . . . .	90
3.7	Numerical Experiments . . . . .	94
3.7.1	Blurred Images Contaminated by Poisson Noise . . . . .	95
3.7.2	Images Corrupted by Multiplicative Noise . . . . .	100
<b>4</b>	<b>Nonlocal Filters for Removing Multiplicative Noise</b>	<b>105</b>
4.1	Nonlocal Means Filter and Motivation . . . . .	106
4.2	The Similarity Measure of Deledalle et al. . . . .	107
4.2.1	Properties in the Presence of Additive Noise . . . . .	109
4.2.2	Properties in the Presence of Multiplicative Noise . . . . .	110
4.3	A New Similarity Measure for Multiplicative Noise . . . . .	115
4.4	Relations to Other Similarity Measures . . . . .	121
4.5	Nonlocal Filtering Facing Multiplicative Noise . . . . .	126
4.5.1	Nonlocal Filters by Maximum Likelihood Estimation . . . . .	126
4.5.2	Definition of the Weights . . . . .	128
4.5.3	Updating the Similarity Neighborhoods . . . . .	131
4.6	Numerical Results . . . . .	132
<b>5</b>	<b>Conclusions and Perspectives</b>	<b>139</b>
<b>A</b>	<b>Densities of Transformed Random Variables</b>	<b>141</b>
	<b>Bibliography</b>	<b>145</b>
	<b>Index</b>	<b>163</b>
	<b>Scientific Career</b>	<b>165</b>
	<b>Wissenschaftlicher Werdegang</b>	<b>167</b>

# List of Publications

- [STSCS] D. J. Strauss, T. Teuber, G. Steidl, and F. I. Corona-Strauss. Exploiting the self-similarity in ERP images by nonlocal means for single trial denoising. *Preprint*. Submitted 2011.
- [GSWT] P. Gwosdek, C. Schmaltz, J. Weickert, and T. Teuber. Fast electrostatic halftoning. *Journal of Real-Time Image Processing*. Accepted for publication 2011.
- [TRHS12] T. Teuber, S. Remmele, J. Hesser, and G. Steidl. Denoising by second order statistics. *Signal Processing*, 92(12):2837–2847, 2012.
- [TL12a] T. Teuber and A. Lang. A new similarity measure for nonlocal filtering in the presence of multiplicative noise. *Computational Statistics & Data Analysis*, 56(12):3821–3842, 2012. [5](#), [122](#)
- [TL12b] T. Teuber and A. Lang. Nonlocal filters for removing multiplicative noise. In A. M. Bruckstein, B. M. ter Haar Romeny, A. M. Bronstein, and M. M. Bronstein, editors, *Scale Space and Variational Methods in Computer Vision, Third International Conference, SSVM 2011*, volume 6667 of *LNCS*, pages 50–61. Springer, 2012. [5](#), [122](#)
- [SST12] S. Setzer, G. Steidl, and T. Teuber. On vector and matrix median computation. *Journal of Computational and Applied Mathematics*, 236(8):2200–2222, 2012. [3](#)
- [SST11] S. Setzer, G. Steidl, and T. Teuber. Infimal convolution regularizations with discrete l1-type functionals. *Communications in Mathematical Sciences*, 9(3):797–827, 2011. [44](#), [50](#), [93](#), [139](#)
- [TSG<sup>+</sup>11] T. Teuber, G. Steidl, P. Gwosdek, C. Schmaltz, and J. Weickert. Dithering by differences of convex functions. *SIAM Journal on Imaging Science*, 4(1):79–108, 2011.
- [SST10] S. Setzer, G. Steidl, and T. Teuber. Deblurring Poissonian images by split Bregman techniques. *Journal of Visual Communication and Image Representation*, 21(3):193–199, 2010. [5](#)

- [ST10] G. Steidl and T. Teuber. Removing multiplicative noise by Douglas-Rachford splitting methods. *Journal of Mathematical Imaging and Vision*, 36(2):168–184, 2010. [5](#), [92](#)
- [SSTM10] S. Setzer, G. Steidl, T. Teuber, and G. Moerkotte. Approximation related to quotient functionals. *Journal of Approximation Theory*, 162(3):545–558, 2010. [5](#)
- [ST09a] G. Steidl and T. Teuber. Anisotropic smoothing using double orientations. In X.-C. Tai, K. Morken, M. Lysaker, and K.-A. Lie, editors, *Scale Space and Variational Methods in Computer Vision, Second International Conference, Proceedings*, volume 5567 of *LNCS*, pages 477–489. Springer, 2009. [3](#)
- [ST09b] G. Steidl and T. Teuber. Diffusion tensors for processing sheared and rotated rectangles. *IEEE Transactions on Image Processing*, 18(12):2640–2648, 2009. [3](#)
- [SST08] S. Setzer, G. Steidl, and T. Teuber. Restoration of images with rotated shapes. *Numerical Algorithms*, 48(1):49–66, 2008. [1](#), [11](#), [15](#), [21](#), [22](#), [34](#), [35](#), [38](#)



# Notation and Symbols

## Sets, Spaces and Norms

$\mathbb{N}$	Set of natural numbers including 0
$\mathbb{R}_{\geq 0}$	Set of all nonnegative real numbers
$\Omega \subset \mathbb{R}^d$	Open, convex subset of $\mathbb{R}^d$ with typically $d = 2$
$\partial\Omega$	Lipschitz boundary of $\Omega$
$\overline{A}$	Closure of the set $A$
$B_\varepsilon(x)$	Open ball with center $x \in \mathbb{R}^n$ and radius $\varepsilon$
$W_\varphi$	Wulff shape of a function $\varphi$
$\mathcal{C}^p(\Omega)$	Space of functions $f : \Omega \rightarrow \mathbb{R}$ which are, together with their partial derivatives of order $\leq p$ , continuous on $\Omega$
$\mathcal{C}^p(\overline{\Omega})$	Space consisting of all those functions $f \in \mathcal{C}^p(\Omega)$ which are, together with their partial derivatives of order $\leq p$ , bounded and uniformly continuous on $\Omega$
$\mathcal{C}^p(\Omega, \mathbb{R}^d)$	Space of functions $f : \Omega \rightarrow \mathbb{R}^d$ , which are, together with their partial derivatives of order $\leq p$ , continuous on $\Omega$
$\mathcal{C}_c^p(\Omega, \mathbb{R}^d)$	Subspace of $\mathcal{C}^p(\Omega, \mathbb{R}^d)$ consisting of all those functions with compact support in $\Omega$
$L_p(\Omega)$	Space of all measurable functions $f : \Omega \rightarrow \mathbb{R}$ such that $\int_\Omega  f(x) ^p dx < \infty$ ( $1 \leq p < \infty$ ), where two functions are identified with each other if they are equal almost everywhere (a.e.) on $\Omega$
$L_\infty(\Omega)$	Space of all measurable functions $f : \Omega \rightarrow \mathbb{R}$ for which a constant $K$ exists such that $ f(x)  \leq K$ a.e. on $\Omega$
$L_{2,w}(\Omega)$	Space of all measurable functions $f : \Omega \rightarrow \mathbb{R}$ such that $\int_\Omega w(x)  f(x) ^2 dx < \infty$ ( $w : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ ), which is equipped with the inner product $\langle f, u \rangle_{L_{2,w}} := \int_\Omega w(x) f(x) u(x) dx$
$BV(\Omega)$	Space of functions on $\Omega$ with bounded variation
$\Pi_n$	Space of polynomials of degree $\leq n$
$ \cdot $	Euclidean norm given in $\mathbb{R}$ by the absolute value and in $\mathbb{R}^2$ by the length of the vector
$2^S$	Power set of the set $S$

## Function(al)s and Operators

$\nabla$	Gradient operator
$\partial_{x_i}, \frac{\partial}{\partial x_i}$	Partial derivative operator in $x_i$ direction
$\text{div}$	Divergence
$\partial f(x)$	Subdifferential of the function $f$ at $x$
$f * g$	Convolution of two functions $f$ and $g$
$f_\sigma$	Function $f$ smoothed by a Gaussian of standard deviation $\sigma$
$K_\rho$	Gaussian with mean 0 and standard deviation $\rho$
$f_1 \square \dots \square f_m$	Infimal convolution of the functions $f_i, i = 1, \dots, m$
$f^*$	Conjugate function of $f$
$1_A$	Characteristic function $1_A : \mathbb{R} \rightarrow \mathbb{R}$ of $A$ defined by
	$1_A(x) = \begin{cases} 1 & x \in A, \\ 0 & \text{otherwise} \end{cases}$
$\iota_A$	Indicator function $\iota_A : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$ of $A$ defined by
	$\iota_A(x) = \begin{cases} 0 & x \in A, \\ \infty & \text{otherwise} \end{cases}$
$\sigma_C$	Support function of the set $C$
$\text{dom } f$	Domain of the function $f$
$\text{ess inf } f$	Essential infimum of the function $f$
$\text{ess sup } f$	Essential supremum of the function $f$
$\text{supp}(f)$	Support of the function $f$

## Vectors and Matrices

$r^\perp$	Vector $r^\perp = (r_2, -r_1)^\text{T}$ orthogonal to $r \in \mathbb{R}^2$
$\text{vec}(A)$	Result of columnwise reshaping the matrix $A$ into a vector
$\text{diag}(v)$	Diagonal matrix with diagonal entries $v \in \mathbb{R}^n$
$\otimes$	Tensor product
$\angle(r_1, r_2)$	Angle between $r_1$ and $r_2$ measured clockwise
$I_N$	Identity matrix of size $N \times N$
$1_N$	Vector of length $N$ where all components are 1
$0_{M,N}$	Zero matrix of size $M \times N$
$0_N$	Zero vector of length $N$
$\det A$	Determinant of the matrix $A$
$\mathcal{N}(A)$	Null space of the matrix $A$
$\text{rank } A$	Rank of the matrix $A$
$\text{span}\{v_1, \dots, v_k\}$	Linear hull of the set $\{v_1, \dots, v_k\}$ with $v_i \in \mathbb{R}^n, i = 1, \dots, k$ , defined as the set of all linear combinations of vectors $v_i$

---

## Terms and Definitions from Probability Theory

$\mathbb{E}(X)$	Expectation value of the random variable $X$
$\mathbb{E}(Y X = x)$	Conditional expectation value of the random variable $Y$ given that $X = x$
$P(X = x)$	Probability that $X$ takes the value $x$
$p_X$	Probability density function of $X$ (if $X$ is a continuous random variable) or the probability mass function of $X$ (for discrete $X$ )
$p_{Y X}(\cdot x)$	Conditional probability density function of $Y$ (if $Y$ is continuous) or the conditional probability mass function of $Y$ (for discrete $Y$ ) given that $X = x$ with $p_X(x) > 0$
$X \sim \text{Poisson}(\lambda)$	$X$ is Poisson distributed with parameter $\lambda$



# CHAPTER 1

## Introduction and Overview

A major goal in digital image processing is the enhancement of images and the removal of noise and other distortions without destroying important image features such as edges, corners or texture. In the last decades a large variety of methods have been proposed in particular for the restoration of images corrupted by additive white Gaussian noise. Such approaches include

- *variational models* like the Rudin, Osher, Fatemi (ROF) model [176], the Bregmanized total variation (TV) method [155], the total generalized variation approach [30] or the nonlocal model [121],
- *PDE-based methods* including the nonlinear diffusion method of Perona and Malik [159], Weickert's anisotropic edge- and coherence-enhancing diffusion [212, 213, 214], the curvature based method of Tschumperlé [206] or approaches within the Beltrami framework [119, 196, 197],
- *nonlinear filtering approaches* such as bilateral filters [158] or the nonlocal means filter [39],
- *sparsity-based methods* incorporating for example wavelets [66] or learned dictionaries [74, 138].

These types of methods are often closely interrelated. There exist for example relations between certain variational models and diffusion methods as outlined in [34, 184, 203, 216, SST08] and PDE-based methods are also closely connected to iterative filtering. Moreover, some methods like for example the BM3D filter [56] belong to more than one group.

In the context of, e.g., variational methods the restored image  $\hat{u}$  is determined as the minimizer of an appropriate energy functional. For a given, possibly noisy or otherwise

degraded measurement  $f$  this energy is typically of the form

$$F(u) = H_f(u) + \lambda J(u), \quad \lambda > 0, \quad (1.1)$$

where the *data fidelity* functional  $H_f$  relates each candidate  $u$  to the given measurement  $f$  and the *regularizer*  $J$  incorporates additional a priori knowledge about the image to be reconstructed. The functional  $H_f$  is in most cases determined by maximum a posteriori (MAP) estimation so that it equals  $-\log p_{F|U}(f|u)$  up to additive terms which do not depend on  $u$ . The involved function  $p_{F|U}$  is directly related to the noise statistics and represents the conditional probability distribution function for observing  $f$  given that  $u$  is the original image. For images corrupted by additive white Gaussian noise the often considered data fidelity term  $H_f(u) = \frac{1}{2} \|f - u\|_{L_2}^2$  is usually deduced in this way.

The choice of the regularizer  $J$  is a challenging part in the design of a variational model, which can significantly influence the quality of the results. Typically,  $J$  is chosen to impose some regularity conditions on the images  $u$ . Since this a priori knowledge favors however certain types of images, it can easily lead to undesired effects in the results. It is for example well known that though the ROF model [176] restores edges very well, it enforces piecewise constant results, leads to a loss of contrast and introduces rounding artifacts at sharp corners in the images, cf. [7, 49]. As we will see in more detail in Section 2.2 rounding artifacts can also be observed for other image restoration methods, especially for images contaminated by strong noise. As illustrated, e.g., in [86] the correct restoration of junctions and fine structures is also often a problem.

In many real-world applications the additional difficulty arises that the occurring noise is neither additive nor Gaussian distributed. Here, the question is typically how to adapt existing methods to other types of noise. Although the data fidelity term of variational methods can for example be determined as described above and some standard regularizer could be used, this must not lead to a good variational model as we will show for images corrupted by multiplicative Gamma noise in Section 3.3. Also the adaptation of nonlinear filters to non-Gaussian noise is not straightforward. In the case of the nonlocal means filter [39] a central question is for example how to determine for two given, noisy pixels with known noise distribution whether their original values have been the same or not.

## Contributions

In the following, we will present solutions for some of these problems. The contributions of this thesis are three-folded:



Figure 1.1: Three different images corrupted by strong additive Gaussian noise.

### Anisotropic image restoration facing additive Gaussian noise

In the first part we concentrate on regularization and diffusion methods for restoring images corrupted by additive Gaussian noise. We show how to define new models which respect not only edges but preserve also corners and crossings like the ones contained in the images in Figure 1.1. In [ST09a, ST09b] parts of these results have been published. To start with, we show that the concepts of anisotropic regularization and diffusion are related to each other under appropriate assumptions. Besides, we analyze in detail the two-dimensional structure tensor of Förstner and Gülch [85] and the double orientation tensors of Aach et al. [2, 148] for locally estimating up to two significant orientations per image pixel. Here, we prove amongst other things that the eigenvectors of the two-dimensional structure tensor can also be computed by appropriate weighted vector means, cf. results in [SST12], and we present modifications of this tensor which allow to determine the shape parameters of rotated and sheared shapes in images. Afterwards, inspired by the work [20] of Berkels et al. and theoretical results of Esedoglu and Osher in [75] we incorporate the obtained orientation information for the definition of new anisotropic regularization methods. Furthermore, we present theoretical results showing that our new methods are well suited to preserve shapes with sharp corners and X junctions in images. Using the presented relation between anisotropic regularization and diffusion models we also deduce related anisotropic diffusion models with very similar properties. As a third alternative we study special discrete infimal convolution regularizers. Here, we show the existence and uniqueness of a solution of the corresponding variational model and its ability to preserve crossings. Moreover, we provide detailed explanations for an efficient implementation of each approach. Due the anisotropy of the proposed methods the discretizations have to be carefully chosen here to guarantee that orientation information is correctly handled. The good performance of the proposed methods are finally demonstrated by numerical experiments. To illustrate the potential of the proposed methods also for applications beyond image denoising, examples for building segmentation in aerial

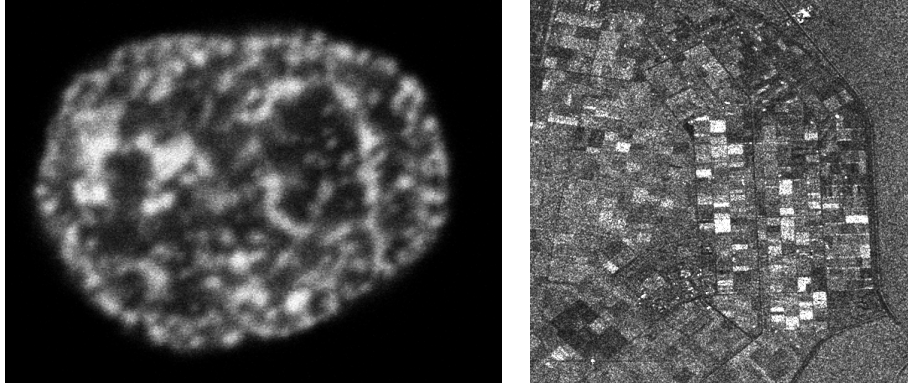


Figure 1.2: Two real-world examples showing a human nucleus which is blurred and corrupted by Poisson noise (left) and a synthetic aperture radar image corrupted by multiplicative noise (right).

images are included. Here, additional use is made of Canny’s edge detector [45] and the windowed Hough transform [53].

### Variational approaches for Poisson and multiplicative noise

In the second part we focus on variational models for restoring images corrupted by Poisson and multiplicative noise. We analyze and compare different approaches for the restoration of images corrupted by multiplicative Gamma noise and prove based on results of Grasmair in [99] the surprising equivalence of a subproblem of the model proposed in [192] and a standard variational method for the removal of Poisson noise considered, e.g., in [15, 114, 129, 180]. This second model involves Csiszár’s I-divergence as data fidelity term and the total variation (TV) semi-norm as a regularizer. We show for the corresponding discrete models and one-dimensional signals that these results can also be generalized to other regularizers like the nonlocal ones studied in [91]. Besides, further good properties of the generalized I-divergence model are proven like the existence of a maximum-minimum principle. For the minimization of the involved variational problems which may also include an additional linear, possibly ill-conditioned transformation of the data we propose an efficient algorithm based on the alternating direction method of multipliers (ADMM). To be able to automatically adapt the parameter involved in ADMM we consider a slightly modified variant of the approach presented in [26, 105]. The good performance of the final algorithm with and without the iterative parameter adaptation strategy is demonstrated in comparison to other existing algorithms for blurred images corrupted by Poisson noise. Moreover, numerical experiments document that the I-divergence model with TV and nonlocal regularizers is well suited for the restoration of images corrupted by multiplicative Gamma noise. To highlight the practical relevance of Poisson and multiplicative noise, restorations results for the microscopy and synthetic aperture radar images displayed



---

in Figure 1.2 are presented. Parts of the results outlined here have been published in [ST10, SST10].

### Nonlocal filtering facing multiplicative noise

In the third part we present a new similarity measure and nonlocal filters for images corrupted by multiplicative noise. Most of these results are contained in [TL12a, TL12b]. The proposed filters are generalizations of the nonlocal means filter [39], which is known to be well suited for the removal of additive Gaussian noise. To adapt this filter to other noise models the involved patch comparison has first of all to be performed by a suitable noise dependent similarity measure. For the deduction of such a measure we study the probabilistic measure recently proposed in [60]. This measure is analyzed in the context of conditional density functions and its properties are examined for data corrupted by additive and multiplicative noise. Since we can show that it has unfavorable properties for multiplicative noise, a new similarity measure is deduced, which consists of a probability density function specially chosen for this type of noise. The properties of this new measure are studied theoretically as well as by numerical experiments. Furthermore, we prove that it is closely related to several of the similarity measures studied very recently in [59, 62]. In contrast to the standard measure for additive Gaussian noise the measure deduced here relies on the quotients of the noisy data values rather than their differences. Best approximations under such optimality conditions have also been studied in [SSTM10], which is however out of the scope of this thesis. To finally obtain our nonlocal filters we apply a weighted maximum likelihood estimation framework as in [60], which incorporates again the noise statistics. The weights occurring in these filters are defined using the new similarity measure and different adaptations are presented to further improve the results. Finally, restoration results for synthetic and real-world examples corrupted by multiplicative Gamma and Rayleigh noise are presented, which demonstrate the very good performance of the new nonlocal filters.

## Outline

The organization of the subsequent chapters is as follows:

In **Chapter 2** we deduce the announced corner and junction preserving anisotropic denoising methods for images corrupted by additive Gaussian noise. We start in Section 2.1 by introducing the variational and diffusion models of interest and investigate the relation between the concepts of anisotropic regularization and diffusion. In Section 2.2 existing image restoration methods are examined with respect to their ability to preserve sharp corners and junctions in images. Afterwards, different tensors for locally estimating orientation information in images are analyzed in Section 2.3, which are then applied in Sections 2.4 – 2.6 for the definition of our new anisotropic models.

The good properties of these methods are proven theoretically and demonstrated by numerical experiments in Section 2.7.

In **Chapter 3** we focus on variational methods for the removal of Poisson and multiplicative noise. Here, we start after some preliminaries in Section 3.1 by highlighting the practical relevance of these types of noise in Section 3.2. In Section 3.3 known variational methods for Poisson and multiplicative Gamma noise are then presented and compared. The equivalence of a subproblem of the model proposed in [192] and the variational model with the I-divergence data fidelity term and the TV regularizer is established in Section 3.4. In addition, further properties of the second model with a more general regularizer are shown for discrete functionals in Section 3.5. To be able to solve the occurring minimization problems in an efficient manner we apply a special variant of the alternating direction method of multipliers. In Section 3.6 this approach, our final algorithm and other algorithms from the literature are presented. The efficiency of our new algorithm is demonstrated for blurred images corrupted by Poisson noise in Section 3.7 and numerical results show the good properties of the I-divergence model with TV and nonlocal regularizers for the removal of multiplicative noise.

Next, we present a new similarity measure and nonlocal filters for general multiplicative noise models in **Chapter 4**. We first recall the definition of the standard nonlocal means filter in Section 4.1 and review on existing approaches for the adaptation of this filter to general noise models. In Section 4.2 we then revisit the similarity measure proposed in [60]. We analyze it in the framework of conditional density functions and study its properties for data corrupted by additive and multiplicative noise. Since it turns out to be well suited for additive noise but to have unfavorable properties for multiplicative noise, we deduce our new measure given by a noise dependent probability density function in Section 4.3. The advantages of this measure are shown theoretically as well as by different examples and experiments. Moreover, relations to other similarity measures recently studied in [59, 62] are investigated in Section 4.4. The proposed nonlocal filters are finally deduced by maximum likelihood estimation in Section 4.5. The involved weights are defined using our new similarity measure and different modifications are presented, which further improve the results. Finally, the very good performance of the new filters is demonstrated for images corrupted by multiplicative Gamma and Rayleigh noise in Section 4.6.

In **Chapter 5** we close with conclusions. Details on the computation of the probability density functions occurring in former chapters are provided in **Appendix A**.

# CHAPTER 2

## Anisotropic Smoothing Facing Additive Gaussian Noise

### Outline

---

2.1	Methods of Interest . . . . .	8
2.2	Properties of Existing Methods and Motivation . . . . .	13
2.3	Orientation Estimation . . . . .	16
2.3.1	Classical Structure Tensor . . . . .	16
2.3.2	Adaptations of the Structure Tensor for Special Shapes . . . . .	21
2.3.3	Double Orientation Estimation . . . . .	24
2.4	Shape Preserving Anisotropic Regularization . . . . .	29
2.4.1	Shape Preservation . . . . .	30
2.4.2	A New Anisotropic Regularization Method . . . . .	36
2.4.3	Discretization and Minimization . . . . .	38
2.5	A Related Anisotropic Diffusion Model . . . . .	39
2.5.1	The Model . . . . .	39
2.5.2	Discretization . . . . .	42
2.6	Anisotropic Infimal Convolution Regularization . . . . .	43
2.6.1	Infimal Convolution Regularization and Shape Preservation . . . . .	43
2.6.2	Discretization and Minimization . . . . .	50
2.7	Numerical Experiments . . . . .	50
2.7.1	Results for Images with Rotated or Sheared Shapes . . . . .	51
2.7.2	Results for Images with Arbitrary Double Orientations . . . . .	54

---

As pointed out in the introduction the correct restoration of advanced image features like sharp corners, junctions or texture is an often occurring problem in image denoising. In this chapter we therefore present new anisotropic regularization and diffusion methods for restoring and enhancing images corrupted by additive white Gaussian noise. The proposed methods have the benefit that in contrast to many state-of-the-art methods they are not only edge-preserving but are also able to restore sharp

corners and X junctions in the presence of strong noise. An outline of the structure of this chapter is provided above.

To fix the notation, let  $\Omega \subset \mathbb{R}^2$  denote an open, convex and possibly bounded set with Lipschitz boundary  $\partial\Omega$ . Throughout this chapter, all continuous gray value images are considered to be functions  $u : \overline{\Omega} \rightarrow \mathbb{R}$ . If  $u$  is partially differentiable, we set  $\nabla u = (\partial_{x_1} u, \partial_{x_2} u)^T$ . Moreover, the MATLAB routine `imagesc` is applied for visualizing the images, which implements an affine gray value scaling setting the minimal image value to black and the maximal gray value to white.

## 2.1 Methods of Interest

To start with, we introduce the anisotropic regularization and diffusion methods we are many interested in and show how these methods can be related to each other.

**Isotropic and anisotropic regularization** A common approach to denoise an image  $f \in L_2(\Omega)$  corrupted by additive white Gaussian noise is to compute a minimizer of the energy minimization problem

$$\operatorname{argmin}_{u \in L_2(\Omega)} \left\{ \frac{1}{2} \|f - u\|_{L_2}^2 + \lambda J(u) \right\}, \quad (2.1)$$

where  $\lambda > 0$  denotes the so-called *regularization parameter* and  $J : L_2(\Omega) \rightarrow \mathbb{R}_{\geq 0} \cup \{+\infty\}$  an appropriate regularization functional. Aside from other properties,  $J$  is usually convex. This guarantees that the whole minimization problem is convex, which makes it in general much easier to solve.

For the first part of this chapter we restrict our attention to functionals  $J$  which can be written in the form

$$J(u) = \int_{\Omega} \phi(\nabla u(x), x) \, dx \quad \forall u \in \mathcal{C}^1(\Omega) \quad (2.2)$$

with  $\phi : \mathbb{R}^2 \times S \rightarrow \mathbb{R}_{\geq 0} \cup \{+\infty\}$ , an open set  $S$  fulfilling  $\overline{\Omega} \subset S$  and an appropriate generalization of  $J$  for  $u \in L_2(\Omega) \setminus \mathcal{C}^1(\Omega)$ . We call  $J$  an *isotropic regularizer* if and only if for every  $x \in \overline{\Omega}$  there exists a function  $\tilde{\phi}_x : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$  so that

$$\phi(p, x) = \tilde{\phi}_x(|p|^2) \quad \forall p = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} \in \mathbb{R}^2 \quad \text{with} \quad |p| := \sqrt{p_1^2 + p_2^2}, \quad (2.3)$$

i.e.,  $\phi$  is radially symmetric with respect to  $p$ . If this is not the case, (2.2) is called an *anisotropic regularizer*.

An interesting class of regularizers fitting into this framework can for example be obtained by setting  $\phi(p, x) = \varphi(p)$  with  $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$  denoting a positively homogeneous,

convex and even function with  $\varphi(0) = 0$  and  $\varphi(x) > 0$  for  $x \neq 0$ . In this case, we have that

$$J(u) = \int_{\Omega} \varphi(\nabla u) dx, \quad (2.4)$$

where the (*anisotropic*) *total variation* (TV) semi-norm

$$\int_{\Omega} \varphi(\nabla u) dx := \sup_{\substack{V \in C_c^1(\Omega, \mathbb{R}^2) \\ V \in W_{\varphi} \text{ a.e.}}} - \int_{\Omega} u(x) \operatorname{div} V(x) dx \quad (2.5)$$

defines  $J$  for general functions  $u \in L_2(\Omega)$ , cf. [75]. The involved set

$$W_{\varphi} := \{x \in \mathbb{R}^2 : \langle x, y \rangle \leq \varphi(y) \ \forall y \in \mathbb{R}^2\}$$

denotes the so-called *Wulff shape* of  $\varphi$ .

If we set, e.g.,  $\varphi(x) := |x| = \sqrt{x_1^2 + x_2^2}$ , the functional  $J$  defined by (2.4) is according to the above definition an isotropic regularizer. Moreover, the Wulff shape of  $\varphi$  is the unit circle  $W_{\varphi} = \{x \in \mathbb{R}^2 : \sqrt{x_1^2 + x_2^2} \leq 1\}$  and thus, the minimization problem (2.1) becomes the well-known Rudin-Osher-Fatemi (ROF) model [50, 176], see also [52].

For  $\varphi(x) := \|x\|_1 = |x_1| + |x_2|$  it holds that  $W_{\varphi} = \{x \in \mathbb{R}^2 : \max\{|x_1|, |x_2|\} \leq 1\}$ , i.e., the Wulff shape of  $\varphi$  is the unit square with edges parallel to the axes, and  $J$  is the anisotropic TV regularizer studied, e.g., in [48, 107].

**Isotropic and anisotropic diffusion** As already mentioned in the introduction there exist also many PDE-based image denoising and enhancement methods. For images  $f$  corrupted by additive Gaussian noise a common approach is, e.g., to use a *diffusion model* of the form

$$\begin{cases} \frac{\partial}{\partial t} u(t, x) = \operatorname{div} (D(\nabla u_{\sigma})(t, x) \nabla u(t, x)) & \forall (t, x) \in (0, T) \times \Omega, \\ (D(\nabla u_{\sigma})(t, x) \nabla u(t, x))^T n(x) = 0 & \forall (t, x) \in (0, T) \times \partial\Omega, \\ u(0, x) = f(x) & \forall x \in \overline{\Omega}, \end{cases} \quad (2.6)$$

cf. [11, 206, 213], where  $n(x)$  denotes the outward normal at  $x$  to  $\partial\Omega$ . The function

$$u_{\sigma}(t, \cdot) := K_{\sigma} * u(t, \cdot) \quad \text{for } t \in (0, T)$$

represents here a smoothed version of  $u(t, \cdot)$  obtained by convolving  $u(t, \cdot)$  with a Gaussian  $K_{\sigma}$  of mean 0 and standard deviation  $\sigma$ . The restored image is finally set to be the diffused image  $u(\hat{t}, \cdot)$  after an appropriate diffusion time  $\hat{t} \in (0, T)$ .

The first equation of this diffusion model is the so-called *diffusion equation* describing the change of the image over time. Its properties depend highly on the choice of the *diffusion tensors*  $D(\nabla u_{\sigma})(t, x)$ ,  $t \in (0, T)$ ,  $x \in \overline{\Omega}$ , which are supposed to be symmetric

and positive definite  $2 \times 2$  matrices. The second equation of (2.6) contains Neumann boundary conditions and guarantees that no flux

$$j := -D(\nabla u_\sigma) \nabla u$$

occurs over the boundary of the image region. The initial state of the diffusion process is given by the corrupted image  $f$  as stated by the last equation of the model.

If the diffusion tensors  $D(\nabla u_\sigma)$  can be substituted by a scalar-valued diffusivity function, i.e.,

$$D(\nabla u_\sigma) \nabla u = g(\nabla u_\sigma) \nabla u$$

for some function  $g(\nabla u_\sigma) : (0, T) \times \bar{\Omega} \rightarrow \mathbb{R}_{>0}$ , we call the diffusion model (2.6) *isotropic*. In this case, the flux directs always in the direction of the gradients  $\nabla u$ . On the contrary, if the direction of the flux varies from the gradient direction, the diffusion model is called *anisotropic*. These definitions equal the ones given in [213] and may vary from others used in the literature. According to these definitions the diffusion model proposed by Perona and Malik in [159] belongs for example to the class of isotropic models, although they called their model originally anisotropic.

An example of an anisotropic diffusion approach according to the above definition is, e.g., the *edge-enhancing diffusion* (EED) model of Weickert, cf. [212, 213]. This method uses the matrices

$$J_0(\nabla u_\sigma) := \nabla u_\sigma \nabla u_\sigma^T, \quad \sigma > 0 \quad (2.7)$$

to define appropriate diffusion tensors. Obviously,  $J_0$  is a rank 1 matrix with spectral decomposition

$$J_0(\nabla u_\sigma) = \underbrace{\begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}}_{=:Q} \begin{pmatrix} |\nabla u_\sigma|^2 & 0 \\ 0 & 0 \end{pmatrix} \underbrace{\begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix}}_{Q^T},$$

where  $\alpha$  is defined such that  $\nabla u_\sigma = |\nabla u_\sigma|(\cos \alpha, \sin \alpha)^T$ . The diffusion tensors for EED are set to

$$D(\nabla u_\sigma) := g(J_0(\nabla u_\sigma)) = Q \begin{pmatrix} g(|\nabla u_\sigma|^2) & 0 \\ 0 & 1 \end{pmatrix} Q^T$$

for a nonnegative monotonically decreasing *diffusivity function*  $g : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{>0}$  with  $g(0) = 1$  as, e.g.,

$$g(x^2) = \begin{cases} 1 & \text{if } x = 0, \\ 1 - \exp\left(\frac{-C_m}{(x^2/\lambda^2)^m}\right) & \text{otherwise.} \end{cases} \quad (2.8)$$

Since  $J_0(\nabla u_\sigma) = J_0(-\nabla u_\sigma)$ , i.e.,  $J_0$  does not distinguish between left and right directions, the same holds true for the diffusion tensors  $D(\nabla u_\sigma)$ . Moreover, the definition of  $D(\nabla u_\sigma)$  leads to the negative flux

$$D(\nabla u_\sigma) \nabla u = g(|\nabla u_\sigma|^2) \frac{\nabla u_\sigma^T \nabla u}{|\nabla u_\sigma|^2} \nabla u_\sigma + 1 \frac{(\nabla u_\sigma^\perp)^T \nabla u}{|\nabla u_\sigma|^2} \nabla u_\sigma^\perp \quad \text{for } |\nabla u_\sigma| \neq 0,$$

which is in general not parallel to  $\nabla u$  and thus confirms that EED is indeed of anisotropic nature.

**Link between regularization and diffusion models** As examined, e.g., in [34, 184, 203, 216, SST08], certain types of regularization and diffusion models are closely related to each other. To establish a relation between the variational model (2.1) with regularizer (2.2) and corresponding diffusion models, we suppose that

- (I)  $\phi$  is a finite function in  $\mathcal{C}^2(\mathbb{R}^2 \times S)$ , where  $S$  is an open set with  $\overline{\Omega} \subset S$ ,
- (II) for all  $p \in \mathbb{R}^2$  and  $x \in \overline{\Omega}$  there exists a symmetric and positive definite matrix  $A(p, x) \in \mathbb{R}^{2,2}$  such that

$$\nabla_p \phi(p, x) = A(p, x) p \quad \text{with } \nabla_p := (\partial p_1, \partial p_2)^T. \quad (2.9)$$

**Remark 2.1.1** The positive definiteness of  $A(p, x)$  guarantees that

$$\nabla_p \phi(p, x) = A(p, x) p \neq 0 \quad \forall p \in \mathbb{R}^2 \setminus \{0\}.$$

Thus,  $\phi(\cdot, x)$  with fixed  $x \in S$  can only have an extremum for  $p = 0$ . Since for  $p = 0$  the Hesse matrix of  $\phi$  with respect to  $p$  is given by

$$H_p \phi(0, x) = A(0, x),$$

we immediately obtain by the positive definiteness of  $A(0, x)$  that any function  $\phi$  fulfilling the above conditions must have a minimum at  $p = 0$  for fixed  $x \in S$ . Beyond that no other local or global extrema can exist.

Now, to see how the considered models are interrelated, we compute for  $u \in \mathcal{C}^2(\overline{\Omega})$  the Euler-Lagrange equation of the energy functional

$$F(u) = \frac{1}{2} \|f - u\|_{L_2}^2 + \lambda \int_{\Omega} \phi(\nabla u(x), x) dx.$$

With  $D(\nabla u)(x) := A(\nabla u(x), x)$  for all  $x \in \overline{\Omega}$  it is given by

$$\begin{aligned} 0 &= u - f - \lambda \operatorname{div}(\nabla_p \phi(\nabla u, \cdot)) \\ &= u - f - \lambda \operatorname{div}(D(\nabla u) \nabla u) \quad \text{on } \Omega \end{aligned} \quad (2.10)$$

with boundary conditions

$$0 = n^T D(\nabla u) \nabla u \quad \text{on } \partial\Omega. \quad (2.11)$$

Here,  $n : \partial\Omega \rightarrow \mathbb{R}^2$  denotes again the function of outward normals to the image boundary  $\partial\Omega$ . Equation (2.10) could here be viewed as a steady-state of a diffusion-reaction equation, cf., e.g., [213, Sec. 1.4]. However, we prefer to rewrite it further in the form

$$\frac{u - f}{\lambda} = \operatorname{div} (D(\nabla u) \nabla u)$$

and to interpret the resulting equation as one time step of an implicit time discretization of the diffusion equation

$$\frac{\partial}{\partial t} u(t, \cdot) = \operatorname{div} (D(\nabla u)(t, \cdot) \nabla u(t, \cdot)), \quad D(\nabla u)(t, \cdot) = A(\nabla u(t, \cdot), \cdot) \quad (2.12)$$

with the initial condition  $u(0, \cdot) = f$ . In this way we finally obtain with (2.11) and  $\sigma = 0$  a diffusion model of the form (2.6). Hence, we showed that whenever  $\phi$  fulfills conditions (I) and (II), we can deduce for the variational model (2.1) with regularizer (2.2) a related diffusion model. By the subsequent considerations we will see that the deduced diffusion model is isotropic, respectively anisotropic if and only if the considered regularizer  $J$  is isotropic, respectively anisotropic for sufficiently smooth functions  $u$ :

To start with, let us assume that  $\phi$  fulfills conditions (I) and (II) and  $J$  defined in (2.2) is isotropic. Then, we know that for every  $x \in \bar{\Omega}$  there exists a function  $\tilde{\phi}_x$  fulfilling (2.3) and consequently,

$$\nabla_p \phi(p, x) = \nabla_p \tilde{\phi}_x(|p|^2) = 2 \tilde{\phi}'_x(|p|^2) p.$$

By (2.9) we have further

$$\begin{aligned} D(\nabla u)(t, x) \nabla u(t, x) &= A(\nabla u(t, x), x) \nabla u(t, x) \\ &= 2 \tilde{\phi}'_x(|\nabla u(t, x)|^2) \nabla u(t, x) \quad \forall (t, x) \in (0, T) \times \bar{\Omega}. \end{aligned}$$

Since  $\tilde{\phi}'_x$  is a scalar-valued function, this implies that the deduced diffusion model with diffusion equation (2.12) is also isotropic.

To show the opposite direction, let us assume that the obtained diffusion model is isotropic. Then, there must exist a scalar-valued function  $g : \mathbb{R}^2 \times \bar{\Omega} \rightarrow \mathbb{R}$  so that for every possible gradient sequence  $\nabla u : (0, T) \times \bar{\Omega} \rightarrow \mathbb{R}^2$  we have

$$D(\nabla u)(t, x) \nabla u(t, x) = g(\nabla u(t, x), x) \nabla u(t, x) \quad \forall (t, x) \in (0, T) \times \bar{\Omega}.$$

By (2.9) and the relation  $D(\nabla u)(t, x) = A(\nabla u(t, x), x)$  this leads to

$$\nabla_p \phi(p, x) = A(p, x) p = g(p, x) p \quad \forall p \in \mathbb{R}^2, x \in \bar{\Omega}.$$



Now, let  $q_1, q_2 \in \mathbb{R}^2$  be two arbitrary points with  $|q_1| = |q_2|$ . Then, there exist constants  $r \in \mathbb{R}_{\geq 0}$  and  $\alpha_1, \alpha_2 \in [0, 2\pi)$  so that  $q_1 = r(\cos \alpha_1, \sin \alpha_1)^T$  and  $q_2 = r(\cos \alpha_2, \sin \alpha_2)^T$  and we define  $\gamma : [\alpha_1, \alpha_2] \rightarrow \mathbb{R}^2$  by

$$\gamma(\alpha) = r \begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix}.$$

Since  $\gamma$  is continuously differentiable, it holds by [127, p. 96/99f] that

$$\begin{aligned} \phi(q_2, x) - \phi(q_1, x) &= \int_{\gamma} \nabla_p \phi(p, x) dp = \int_{\alpha_1}^{\alpha_2} \langle \nabla_p \phi(\gamma(\alpha), x), \gamma'(\alpha) \rangle d\alpha \\ &= \int_{\alpha_1}^{\alpha_2} g(\gamma(\alpha), x) \langle \gamma(\alpha), \gamma'(\alpha) \rangle d\alpha \\ &= \int_{\alpha_1}^{\alpha_2} g(\gamma(\alpha), x) \langle r \begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix}, r \begin{pmatrix} -\sin \alpha \\ \cos \alpha \end{pmatrix} \rangle d\alpha \\ &= 0 \end{aligned}$$

for all  $x \in \overline{\Omega}$  and thus,  $\phi(q_1, \cdot) = \phi(q_2, \cdot)$ . Hence,  $\phi$  is radially symmetric with respect to  $p$ , which finally implies that the regularizer  $J$  in (2.2) is isotropic.

As a last remark we want to point out one further class of closely related methods:

**Remark 2.1.2** As outlined in [199, 200, 217, 218, 219] there exists also a direct relation of the considered regularization and diffusion models to *anisotropic (Haar) wavelet shrinkage*. By establishing this relation coupled shrinkage rules can be deduced which allow also highly anisotropic filtering with simple wavelet filters.

## 2.2 Properties of Existing Methods and Motivation

Apart from the presented methods there exists of course a large variety of other approaches for restoring images corrupted by additive Gaussian noise. In the following, we will comment on some of these approaches including the already mentioned ones and pay special attention to their ability to preserve sharp corners in images corrupted by strong noise.

A first commonly applied variational denoising method is the above mentioned ROF model [50, 176], which is well-known for its edge-preserving properties. A review on this approach can for example be found in [52]. Applied to a noisy image of

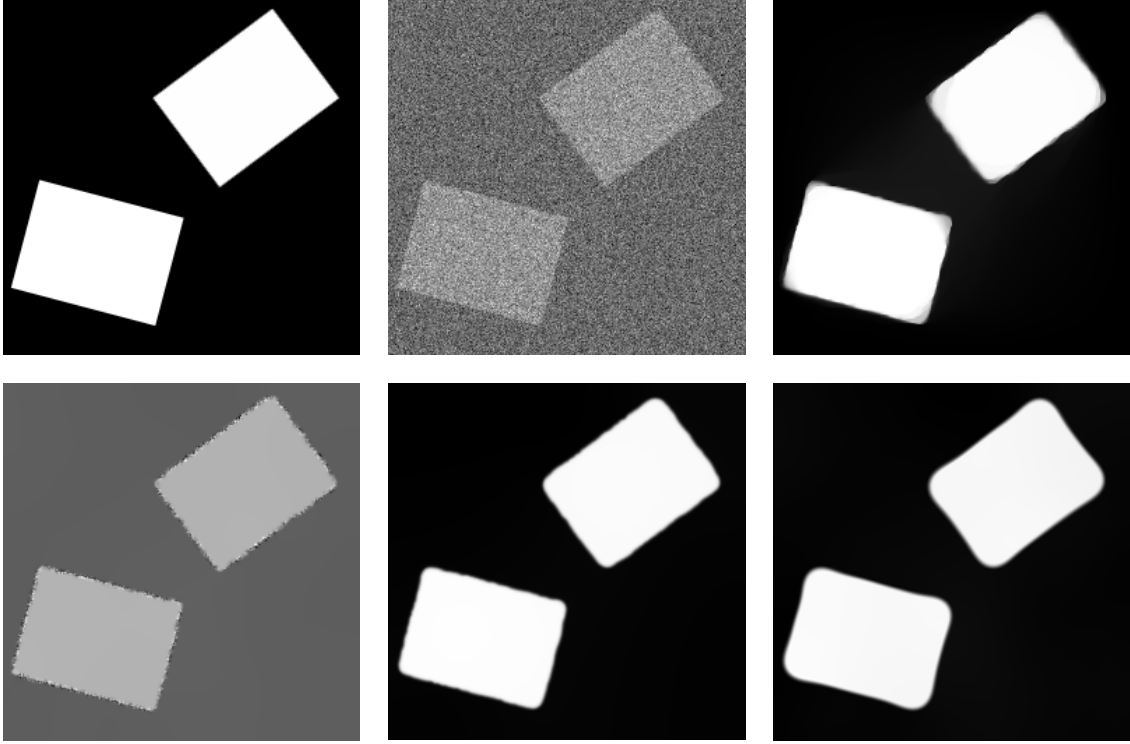


Figure 2.1: Comparison of different denoising methods, see also [203]. *Top left:* Original image. *Top middle:* Corrupted image by strong white Gaussian noise of standard deviation 150. *Top right:* Denoised image by ROF with  $\lambda = 1000$ . This result can be improved at the slightly bumpy edges by increasing  $\lambda$ , but then the vertices become even more rounded. The usually observed loss of contrast has been compensated by the MATLAB routine `imagesc`, here. *Bottom:* Denoised images using the isotropic diffusion equation (2.13) with diffusivity (2.8),  $\sigma = 3$ ,  $m = 4$ ,  $C_4 = 3.31488$ ,  $\lambda = 10$  and stopping times  $t = 400$  (left) and  $t = 3000$  (middle). For the moderate diffusion time  $t = 400$  noisy pixels survive at edges, which is not the case for EED with  $m$ ,  $C_m$ ,  $\lambda$  as set before,  $\sigma = 2.5$  and  $t = 400$  (right). However, both methods suffer from rounding artifacts at vertices.

rotated rectangles as displayed in Figure 2.1 (top middle), it preserves the edges of the rectangles very well, cf. Figure 2.1 (top right). Unfortunately, since the regularization parameter must be chosen rather large to eliminate the noise, it creates rounding artifacts at the corners of the rectangles. A theoretical study of this effect has for example been provided in [7, 49]. In addition, the result suffers typically from a loss of contrast in the gray values. This effect has been compensated by an affine rescaling of the gray values, here. Alternatively, the more sophisticated approach proposed in [42, 43, 155] could be used.

Another well-known image restoration method is the isotropic diffusion approach presented by Perona and Malik in [159]. With a slight modification of Catté et al. in [46] to make the problem well-posed, see also [118, 226], the diffusion equation for this

approach reads

$$\frac{\partial}{\partial t}u(t, x) = \operatorname{div} \left( g(|\nabla u_\sigma(t, x)|^2) \nabla u(t, x) \right), \quad (2.13)$$

where  $g : \mathbb{R}_{\geq 0} \rightarrow [0, 1]$  is again a scalar-valued diffusivity function. Since  $g$  is supposed to be decreasing with  $g(0) = 1$ , the flux becomes smaller for high absolute gradient values. Thus, the diffusion is lowered at important image features such as edges, which results again in an edge preserving method. However, lowering the diffusion at edges also implies that noisy pixels survive there for a long time as demonstrated in Figure 2.1 (bottom left). Of course, if we diffuse sufficiently long, these pixels are smoothed, too, but the vertices of the denoising result become again rounded as visible at Figure 2.1 (bottom middle). At least the first effect can be avoided by an appropriate anisotropic diffusion method like Weickert's EED. Here, the diffusion tensors allow a diffusion perpendicular to the gradient directions and thus, the noisy pixels are also cleaned along the edges. However, we are still confronted with rounded vertices as visible in Figure 2.1 (bottom right).

A different approach for adapting a PDE to the local geometry of an image has been proposed in [205, 206]. Here, Tschumperlé suggests to estimate the local structures by the so-called *structure tensor* originally introduced in [85] and then to perform the diffusion along integral curves deduced from the structure tensors. It has been shown that the solution of the considered curvature-based PDE can be approximated by successively convolving the initial image with one-dimensional Gaussians along these integral curves, which leads to 'curved' filtering. As a result, this method is very flexible with respect to different shapes. However, rounding artifacts occur again at corners in the presence of strong noise.

Apart from the described regularization and PDE-based methods there exist various nonlinear filtering approaches. For example, we observed that bilateral filters [157, 204] cannot cope with this large amount of noise. We obtained much better results by the recently proposed nonlocal means (NL means) filter [39], at least if we apply this method iteratively. Results of this filter will be presented later on in Section 2.7.

To ideally incorporate geometric information into the restoration process further methods have been proposed. For the special case of images with rotated rectangular shapes the topic of vertex preserving image simplification was for example addressed by Berkels et al. in [20], who suggested a regularization approach based on theoretical results of Osher and Esedoglu in [75]. This method leads finally to an alternating two step algorithm, which estimates in the first step the rotation angles of the rectangles by minimizing a functional containing first and second order derivatives of the rotation angles. In the second step the restored image is then compute. A simpler method for finding the rotation angles which includes only first order derivatives has been presented in [203, SST08]. In Subsection 2.3.2 we will also show how this can be done using an adapted structure tensor.

For arbitrary images various papers deal further with the smoothing of normal vectors

through the minimization of certain energy functionals [95, 120, 137, 166, 209, 228] and then incorporating this information into a subsequent denoising step. However, these minimization procedures are in general computationally much more expensive than the orientation estimation approaches we will present in the following. Instead of estimating one normal vector per pixel we will use special tensors to estimate more than one significant orientation per image point. This orientation information will then be used to define new anisotropic regularization and diffusion methods which are capable of preserving sharp corners in images.

Besides, we will also focus on approaches for the estimation of local orientations at crossings and appropriate methods for their preservation. An approach specially designed for such structures has for example been presented by Scharr in [182]. This approach is based on one of the orientation estimation tensors presented in [2, 148] and uses the second order partial derivatives of the image to design an appropriate higher order diffusion process. In [86] another nonlinear diffusion method has been proposed by Franken and Duits. Here, each image is first transformed into a so-called *orientation score* by introducing an additional third dimension, which guarantees that crossed lines are separated from each other. Then, anisotropic diffusion is performed on the orientation score and finally a back transformation is applied to obtain the final denoised image. Apart from 2D image data this approach has also been extended to HARDI (high angular resolution diffusion imaging) data as presented in [68].

## 2.3 Orientation Estimation

For the definition of our new anisotropic models we have to find appropriate methods to capture the local orientations in images first. For this purpose, we will study different tensors for locally estimating orientations in this section. Of course, there exist also other techniques like for example the already mentioned methods for estimating normals proposed in [95, 120, 137, 166, 209, 228], shearlets [102, 126, 225], curvelets [44] or the Riesz-Laplace wavelet transform based method presented in [208] to name only a few others.

For the subsequent considerations, we make a distinction between the terms 'orientations' and 'directions'. If two vectors differ only by their sign, we say that they share the same orientation. However, they are considered to point in opposite directions.

### 2.3.1 Classical Structure Tensor

We start by recalling an orientation estimation approach based on the classical structure tensor of Förstner and Gülch [85]. This approach is especially well suited if there exists only one predominant orientation in each part of an image.

To deduce this tensor theoretically, let  $\Omega \subset \mathbb{R}^2$  be an image part of interest. For simplicity, we set  $\Omega = B_\varepsilon(0)$ , where  $B_\varepsilon(0)$  is the open ball around 0 with radius  $\varepsilon$ .

Moreover, we ideally assume that this part of the image corresponds to a function  $f : \Omega \rightarrow \mathbb{R}$  which has constant values along an orientation  $r \in \mathbb{R}^2$  with  $|r| = 1$ , i.e.,  $f = \varphi(s^\top \cdot)$  for  $s := r^\perp = (r_2, -r_1)^\top$  and a univariate, sufficiently smooth function  $\varphi : [-\varepsilon, \varepsilon] \rightarrow \mathbb{R}$ . This implies that

$$\frac{\partial}{\partial r} f(x) = r^\top \nabla f(x) = r^\top \varphi'(s^\top x) s = 0 \quad \forall x \in \Omega.$$

Hence, we have for every nonnegative weight function  $w : \Omega \rightarrow \mathbb{R}$  that

$$0 = \int_{\Omega} w(y) (r^\top \nabla f(y))^2 dy = r^\top \left( \int_{\Omega} w(y) \nabla f(y) \nabla f(y)^\top dy \right) r. \quad (2.14)$$

If  $\varphi$  is not constant, the symmetric, positive semidefinite  $2 \times 2$  matrix

$$\mathcal{J} := \int_{\Omega} w(y) \nabla f(y) \nabla f(y)^\top dy = \left( \int_{\Omega} w(y) (\varphi'(s^\top y))^2 dy \right) s s^\top$$

has rank one and  $r$  is an eigenvector of  $\mathcal{J}$  with eigenvalue 0.

However, in applications we often deal with noisy images and thus, equation (2.14) may not hold in such situations. To handle this, a presmoothing step with the 2D Gaussian  $K_\sigma$  of standard deviation  $\sigma$  is performed and the gradients in  $\mathcal{J}$  are substituted by  $\nabla f_\sigma := \nabla(K_\sigma * f)$ , where  $*$  represents the ordinary convolution operator. As a consequence, (2.14) holds at least approximately and the orientation vector  $r$  can be considered to be a minimizer of the weighted least squares expression

$$\min_r r^\top \mathcal{J} r \quad \text{subject to } |r| = 1,$$

i.e., a normed eigenvector belonging to the smallest eigenvalue of  $\mathcal{J}$ .

Since in natural images the significant orientations usually vary between image parts, we need to compute  $r$  for every image region. For this purpose, we consider  $\Omega$  to be the whole image domain again and use  $w = K_\rho(\cdot - x)$  (truncated outside  $B_{3\rho}(x)$ ) with  $\rho \geq 0$  to estimate the orientation  $r$  in the neighborhood of each image point  $x$ . In detail, we set

$$\mathcal{J}_\rho(x) := \int_{\Omega} w(x - y) \nabla f_\sigma(y) \nabla f_\sigma(y)^\top dy = (K_\rho * (\nabla f_\sigma \nabla f_\sigma^\top))(x) \quad (2.15)$$

and thus attach to each image point a  $2 \times 2$  matrix, the so-called *structure tensor*. This tensor is a smoothed variant of the tensor  $J_0$  defined in (2.7). Note that the convolution is meant componentwise here and a generalization to nonlinear structure tensors was presented in [33]. Alternatively to the above derivation, a similar definition can be deduced from a least squares problem in the Fourier domain as done in [24, Chapter

10.3]. Like  $\mathcal{J}$  the tensors  $\mathcal{J}_\rho(x)$  are all symmetric and positive semidefinite. Thus, they have an eigendecomposition and the eigenvalues  $\lambda_1, \lambda_2$  are always nonnegative. As the following theorem shows the eigenvectors and eigenvalues can even be explicitly determined:

**Theorem 2.3.1 (Eigenvectors and eigenvalues of  $\mathcal{J}_\rho$ )**

For all  $x \in \Omega$ , let  $\alpha$  be defined by  $\nabla f_\sigma(x) = |\nabla f_\sigma(x)| (\cos \alpha(x), \sin \alpha(x))^T$  for  $|\nabla f_\sigma(x)| \neq 0$  and set  $\alpha(x)$  to an arbitrary value in  $[0, 2\pi)$  otherwise. Moreover, define

$$\nabla_{2\alpha} f_\sigma := |\nabla f_\sigma| \begin{pmatrix} \cos(2\alpha) \\ \sin(2\alpha) \end{pmatrix} = R(\alpha) \nabla f_\sigma \quad \text{with} \quad R(\alpha) = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}, \quad (2.16)$$

i.e.,  $\nabla_{2\alpha} f_\sigma$  are the gradients of  $f$  where the angles of the vectors have been doubled. Then, for all  $x \in \Omega$  the eigenvalues  $\lambda_1(x), \lambda_2(x)$  of  $\mathcal{J}_\rho(x)$  with  $\lambda_1(x) \geq \lambda_2(x)$  are given by

$$\lambda_{1/2}(x) = \frac{1}{2} \left( (K_\rho * |\nabla f_\sigma|^2)(x) \pm |v(x)| \right), \quad \text{where} \quad v(x) := (K_\rho * |\nabla f_\sigma| \nabla_{2\alpha} f_\sigma)(x).$$

If  $\lambda_1(x) > \lambda_2(x)$  and for the components  $v_i(x)$ ,  $i = 1, 2$ , of  $v(x)$  it holds that  $v_1(x) > 0$  or  $v_2(x) \neq 0$ , then any eigenvectors  $r_1(x), r_2(x)$  corresponding respectively to  $\lambda_1(x), \lambda_2(x)$  fulfill

$$r_1(x) \parallel R\left(-\frac{\beta(x)}{2}\right) v(x) \quad \text{and} \quad r_2(x) \perp R\left(-\frac{\beta(x)}{2}\right) v(x) \quad (2.17)$$

with  $\beta(x)$  being the angle of  $v(x)$ , i.e.,  $v = |v| (\cos \beta, \sin \beta)^T$ .

On the other hand, if  $\lambda_1(x) > \lambda_2(x)$ ,  $v_1(x) < 0$  and  $v_2(x) = 0$ , it holds that

$$r_1(x) \parallel (0, 1)^T \quad \text{and} \quad r_2(x) \parallel (1, 0)^T. \quad (2.18)$$

**Proof:** For a symmetric matrix  $X = \begin{pmatrix} x_{1,1} & x_{1,2} \\ x_{1,2} & x_{2,2} \end{pmatrix} \in \mathbb{R}^{2,2}$  straightforward calculations show that the eigenvalues  $\lambda_1 \geq \lambda_2$  of  $X$  are given by

$$\lambda_{1/2} = \frac{1}{2} \left( x_{1,1} + x_{2,2} \pm \sqrt{(x_{1,1} - x_{2,2})^2 + 4x_{1,2}^2} \right). \quad (2.19)$$

Besides, if  $\lambda_1 > \lambda_2$ , a short calculation shows that any eigenvector  $r_1$  to  $\lambda_1$  fulfills

$$r_1 \parallel \begin{cases} (\lambda_1 - x_{2,2}, x_{1,2})^T & \text{if } x_{1,2} \neq 0 \text{ or } x_{1,1} > x_{2,2}, \\ (0, 1)^T & \text{otherwise.} \end{cases} \quad (2.20)$$

Since the convolution in the definition of the structure tensor is meant componentwise, it follows by (2.19) that the eigenvalues  $\lambda_1(x) \geq \lambda_2(x)$  of  $\mathcal{J}_\rho(x)$ ,  $x \in \Omega$ , are given by

$$\begin{aligned} \lambda_{1/2} &= \frac{1}{2} \left( K_\rho * (\partial_{x_1} f_\sigma)^2 + K_\rho * (\partial_{x_2} f_\sigma)^2 \right. \\ &\quad \left. \pm \sqrt{(K_\rho * (\partial_{x_1} f_\sigma)^2 - K_\rho * (\partial_{x_2} f_\sigma)^2)^2 + 4(K_\rho * \partial_{x_1} f_\sigma \partial_{x_2} f_\sigma)^2} \right) \\ &= \frac{1}{2} \left( K_\rho * |\nabla f_\sigma|^2 \pm \sqrt{(K_\rho * ((\partial_{x_1} f_\sigma)^2 - (\partial_{x_2} f_\sigma)^2))^2 + (K_\rho * 2 \partial_{x_1} f_\sigma \partial_{x_2} f_\sigma)^2} \right). \end{aligned}$$

With  $\alpha$  fulfilling  $\nabla f_\sigma = |\nabla f_\sigma| (\cos \alpha, \sin \alpha)^T$  it holds that

$$\begin{aligned} K_\rho * \begin{pmatrix} ((\partial_{x_1} f_\sigma)^2 - (\partial_{x_2} f_\sigma)^2) \\ 2 \partial_{x_1} f_\sigma \partial_{x_2} f_\sigma \end{pmatrix} &= K_\rho * \left( |\nabla f_\sigma| \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \nabla f_\sigma \right) \\ &= K_\rho * |\nabla f_\sigma| \nabla_{2\alpha} f_\sigma \end{aligned}$$

and thus,

$$\lambda_{1/2} = \frac{1}{2} (K_\rho * |\nabla f_\sigma|^2 \pm |v|) \quad \text{with} \quad v := K_\rho * |\nabla f_\sigma| \nabla_{2\alpha} f_\sigma.$$

Now, let us denote by  $v_i(x)$ ,  $i = 1, 2$ , the components of  $v(x)$  and define  $\beta(x)$  by  $v(x) = |v(x)| (\cos \beta(x), \sin \beta(x))^T$ . If  $\lambda_1(x) > \lambda_2(x)$  and the components  $z_{i,j}(x)$ ,  $i, j = 1, 2$ , of  $\mathcal{J}_\rho(x)$  fulfill  $z_{1,2}(x) = \frac{1}{2} v_2(x) \neq 0$  or  $z_{1,1}(x) - z_{2,2}(x) = v_1(x) > 0$ , it follows by (2.20) that for any eigenvector  $r_1(x)$  belonging to  $\lambda_1(x)$  we have

$$r_1(x) \parallel \frac{1}{2} \begin{pmatrix} v_1(x) + |v(x)| \\ v_2(x) \end{pmatrix} = \frac{1}{2} |v(x)| \begin{pmatrix} \cos \beta(x) + 1 \\ \sin \beta(x) \end{pmatrix}. \quad (2.21)$$

Since  $\sin \beta = 2 \sin \frac{\beta}{2} \cos \frac{\beta}{2}$  and  $\cos \beta = (\cos \frac{\beta}{2})^2 - (\sin \frac{\beta}{2})^2$  we obtain further

$$\begin{pmatrix} \cos \beta(x) + 1 \\ \sin \beta(x) \end{pmatrix} = 2 \cos \frac{\beta(x)}{2} \begin{pmatrix} \cos \frac{\beta(x)}{2} \\ \sin \frac{\beta(x)}{2} \end{pmatrix} = 2 \cos \frac{\beta(x)}{2} R\left(-\frac{\beta(x)}{2}\right) \begin{pmatrix} \cos \beta(x) \\ \sin \beta(x) \end{pmatrix}$$

so that inserting this in (2.21) yields

$$r_1(x) \parallel \cos \frac{\beta(x)}{2} |v(x)| R\left(-\frac{\beta(x)}{2}\right) \begin{pmatrix} \cos \beta(x) \\ \sin \beta(x) \end{pmatrix} \parallel R\left(-\frac{\beta(x)}{2}\right) v(x).$$

On the other hand, if  $\lambda_1(x) > \lambda_2(x)$ ,  $z_{1,2}(x) = 0$  and  $z_{1,1}(x) - z_{2,2}(x) = v_1(x) < 0$ , it follows by (2.20) that

$$r_1(x) \parallel (0, 1)^T.$$

Finally, the eigenvector  $r_2(x)$  belonging to  $\lambda_2(x)$  has to fulfill  $r_2(x) \perp r_1(x)$ , which finishes the proof.  $\square$

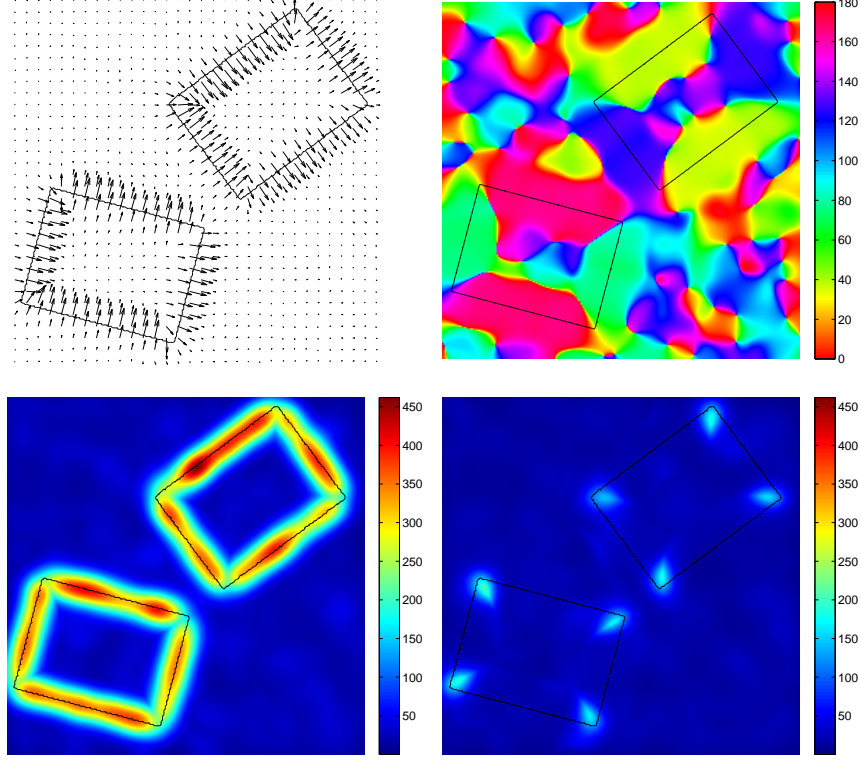


Figure 2.2: Visualization of the information provided by the structure tensors ( $\sigma = 2.5$ ,  $\rho = 8$ ) for the noisy image in Figure 2.1 (top middle). *Top*: Normalized eigenvectors  $r_1$  scaled by the corresponding eigenvalues  $\lambda_1$  (left) and orientation angles of  $r_2$  given by  $\angle(r_2, (1, 0)^T) \bmod \pi$  (right). *Bottom*: Eigenvalues  $\lambda_1$  (left) and  $\lambda_2$  (right).

A particular consequence of this theorem is that if we are only interested in the eigenvectors of  $\mathcal{J}_\rho$  and not its eigenvalues, we do not need the structure tensors. Instead, we can directly compute the weighted vector means  $v$  and apply (2.17)/(2.18) to obtain the orientations  $r_1$  and  $r_2$ .

An example illustrating the information provided by the structure tensor is given in Figure 2.2. Here, we can see that in homogeneous image regions, where  $|\nabla f_\sigma| \approx 0$ , it holds that  $\lambda_1 \approx \lambda_2 \approx 0$  and the eigenvectors may point in any directions. On the other hand, if there is exactly one dominant orientation like in the neighborhood of straight edges, we have  $\lambda_1 \gg \lambda_2 \geq 0$  and the orthogonal eigenvectors  $r_1 = r^\perp$ ,  $r_2 = r$  approximate the orientations of the gradient and the isophote of the edge, respectively. Near one of the vertices it further holds that  $\lambda_1 \geq \lambda_2 \gg 0$ , but since  $r_1 \perp r_2$ , it is not possible to obtain both orientations of the neighboring edges. Instead, the orientations of the eigenvectors are somewhere in between the gradient, respectively the isophote orientations of the neighboring edges. Thus, this can cause rounding artifacts at



vertices if used for image restoration methods.

### 2.3.2 Adaptations of the Structure Tensor for Special Shapes

For images containing only special shapes the information provided at vertices can be improved by estimating directly the parameters of the shapes.

**Rotated rectangular shapes** If we know for example that an image contains only rectangular shapes which may be rotated, one possibility is to modify the classical structure tensor as follows: Defining the angles  $\alpha$  by  $\nabla f_\sigma = |\nabla f_\sigma| (\cos \alpha, \sin \alpha)^T$  for  $|\nabla f_\sigma| \neq 0$ , i.e.,

$$\alpha = \begin{cases} \arccos \frac{\partial_{x_1} f_\sigma}{|\nabla f_\sigma|} & \text{if } \partial_{x_2} f_\sigma \geq 0, \\ 2\pi - \arccos \frac{\partial_{x_1} f_\sigma}{|\nabla f_\sigma|} & \text{if } \partial_{x_2} f_\sigma < 0, \end{cases}$$

and setting  $\nabla_{2\alpha} f_\sigma := |\nabla f_\sigma| (\cos(2\alpha), \sin(2\alpha))^T$  as in (2.16), we can replace the tensors

$$\nabla f_\sigma (\nabla f_\sigma)^T \quad \text{by} \quad \nabla_{2\alpha} f_\sigma (\nabla_{2\alpha} f_\sigma)^T$$

in the definition of the structure tensor (2.15) and obtain the modified tensors

$$J_\rho(\nabla_{2\alpha} f_\sigma) := K_\rho * (\nabla_{2\alpha} f_\sigma (\nabla_{2\alpha} f_\sigma)^T).$$

Since  $\cos(2\cdot)$  and  $\sin(2\cdot)$  are  $\pi$  periodic and  $\nabla_{2\alpha} f_\sigma (\nabla_{2\alpha} f_\sigma)^T = (-\nabla_{2\alpha} f_\sigma) (-\nabla_{2\alpha} f_\sigma)^T$ , the new tensors  $J_\rho(\nabla_{2\alpha} f_\sigma)$  can only distinguish between angles  $\alpha \bmod \frac{\pi}{2}$ . Hence,  $J_\rho(\nabla_{2\alpha} f_\sigma)$  does not distinguish between orthogonal edges and has thus the benefit that the orientations of its eigenvectors do not vary around vertices of rectangular shapes.

The tensors  $J_\rho(\nabla_{2\alpha} f_\sigma)$  can now be easily used to estimate the rotation angles of the shapes. We only need to determine the spectral decomposition of these tensors and reverse the angle doubling process. In detail, we compute for each image point  $x \in \Omega$  the angle  $\alpha_{2,\rho}(x)$  of the eigenvector belonging to the largest eigenvalue of  $J_\rho(\nabla_{2\alpha} f_\sigma)(x)$  and take  $\alpha_{2,\rho}(x)/2$  as smooth approximation of the rotation angle. According to Theorem 2.3.1, these angles can of course also be determined by directly computing

$$v := K_\rho * |\nabla f_\sigma| \nabla_{4\alpha} f_\sigma$$

and setting the rotation angles to one fourth of the angles of these vectors. Here, determining the spectral decomposition of  $J_\rho(\nabla_{2\alpha} f_\sigma)(x)$  has the benefit that the eigenvalues provide additional information about the reliability of the estimates. If we have for example  $\lambda_1(x) \geq \lambda_2(x) \gg 0$  at some image point  $x \in \Omega$ , we know that the image region around  $x$  must also contain differently oriented structures.

Figure 2.3 shows that the results of the proposed local angle smoothing process are at least competitive to ones of the global minimization method proposed in [SST08]. As we will see later on, such angle estimates can be successfully used for vertex-preserving image denoising.

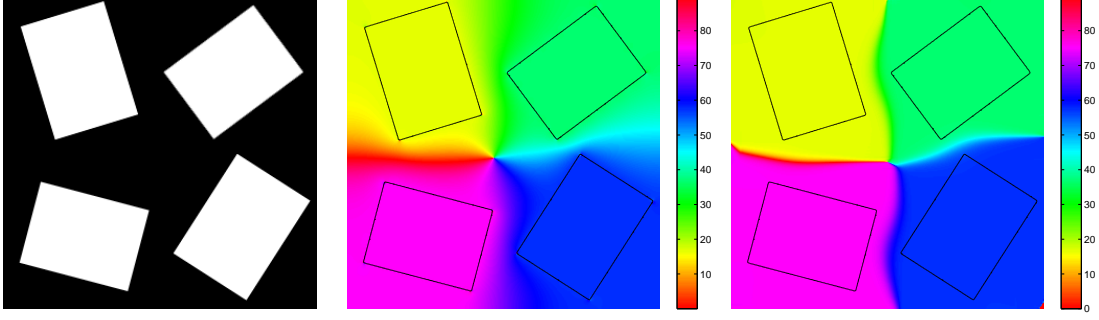


Figure 2.3: *Left*: Original image of different rotated rectangles. *Middle*: Estimated rotation angles by the variational angle adaptation method proposed in [SST08] ( $\sigma = 0.5$ ,  $\mu = 10000$ ). *Right*: Estimated rotation angles by our new structure tensor method ( $\sigma = 0.5$ ,  $\rho = 13$ ).

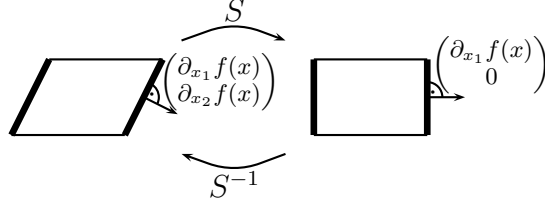


Figure 2.4: Illustration of the considered shear operators.

**Sheared shapes** Next, we focus on images consisting of sheared non-rotated shapes and estimate the corresponding shape operators using again a modified structure tensor. As illustrated in Figure 2.4 we set

$$S = S(\zeta) := \begin{pmatrix} 1 & 0 \\ \zeta & 1 \end{pmatrix}, \quad \zeta \in \mathbb{R} \quad (2.22)$$

to be the shear matrices which transform the gradients of the considered parallelograms back into ordinary rectangles. In other words, if a pixel  $x \in \Omega$  belongs to a non-horizontal edge of a sheared rectangle with shear parameter  $\zeta$ , then it is likely that

$$S(\zeta) \nabla f_\sigma(x) = \begin{pmatrix} 1 & 0 \\ \zeta & 1 \end{pmatrix} \begin{pmatrix} \partial_{x_1} f_\sigma(x) \\ \partial_{x_2} f_\sigma(x) \end{pmatrix} = \begin{pmatrix} \partial_{x_1} f_\sigma(x) \\ \zeta \partial_{x_1} f_\sigma(x) + \partial_{x_2} f_\sigma(x) \end{pmatrix} = \begin{pmatrix} \partial_{x_1} f_\sigma(x) \\ 0 \end{pmatrix},$$

which is equivalent to

$$\zeta = -\frac{\partial_{x_2} f_\sigma(x)}{\partial_{x_1} f_\sigma(x)} \quad \text{if } \partial_{x_1} f_\sigma(x) \neq 0.$$

Hence, we can estimate the shear parameters by the gradients of the two non-horizontal edges of each sheared rectangle. To get only those gradients, we locate the horizontal

edges and the corners of the sheared rectangles and downsize the gradients at those locations to reduce their influence on the structure tensor. In detail, we compute

$$\tilde{\nabla} f_{\tilde{\sigma}} := w_e(\nabla f_{\tilde{\sigma}}) w_v(\nabla f_{\tilde{\sigma}}) \nabla f_{\tilde{\sigma}}, \quad (2.23)$$

using the following weight functions: Let  $\lambda_1, \lambda_2$  be the eigenvalues of the classical structure tensors  $\mathcal{J}_\rho$  with presmoothing of standard deviation  $\tilde{\sigma}$ . In order to *downsize*  $\nabla f_{\tilde{\sigma}}$  at vertices, we multiply these gradients by

$$w_v(\nabla f_{\tilde{\sigma}}) := \begin{cases} \varphi_v \left( \left| \frac{\lambda_2}{\lambda_1} \right| |\nabla f_{\tilde{\sigma}}|^2 \right) & \text{if } \lambda_1 > 0, \\ 1 & \text{otherwise,} \end{cases}$$

where  $\varphi_v$  has the properties of a diffusivity function, i.e., it is decreasing and  $\varphi_v(0) = 1$ . Here,  $\left| \frac{\lambda_2}{\lambda_1} \right| |\nabla f_{\tilde{\sigma}}|^2$  plays the role of a corner detector, since at vertices we typically have  $\lambda_1 \geq \lambda_2 \gg 0$ , while  $\lambda_1 \gg \lambda_2$  at straight edges as outlined in Subsection 2.3.1.

To *enforce*  $\nabla f_{\tilde{\sigma}}$  to become small at horizontal edges we further apply

$$w_e(\nabla f_{\tilde{\sigma}}) := \begin{cases} \varphi_e \left( \left| \frac{\partial_{x_1} f_{\tilde{\sigma}}}{\partial_{x_2} f_{\tilde{\sigma}}} \right| \right) & \text{if } |\partial_{x_2} f_{\tilde{\sigma}}| > 0, \\ 1 & \text{otherwise,} \end{cases}$$

where  $\varphi_e$  is a sigmoidal function, i.e., it is monotone increasing with  $\varphi_e(x) = 0$  for  $x \leq 0$  and  $\varphi_e(x) = 1$  for  $x \geq 1$ .

For the modified gradients (2.23) and a non-radial two-dimensional Gaussian

$$K_{(\rho_x, \rho_y)}(x_1, x_2) := \frac{1}{2\pi\rho_x\rho_y} e^{-\frac{1}{2} \left( \left( \frac{x_1}{\rho_x} \right)^2 + \left( \frac{x_2}{\rho_y} \right)^2 \right)}$$

we now compute the modified structure tensors

$$J_{(\rho_x, \rho_y)}(\tilde{\nabla} f_{\tilde{\sigma}}) := K_{(\rho_x, \rho_y)} * (\tilde{\nabla} f_{\tilde{\sigma}} \tilde{\nabla} f_{\tilde{\sigma}}^T)$$

and determine the eigenvectors  $v = (v_1, v_2)^T$  belonging to the largest eigenvalues. In general, we propose to use an anisotropic Gaussian with  $\rho_x \neq \rho_y$  to better adapt the smoothing to the geometry of the shapes. Finally, we set the shear parameters  $\zeta$  to be

$$\zeta := \begin{cases} -\frac{v_2}{v_1} & \text{if } v_1 > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Examples of estimated shear parameters are given later on in Section 2.7, Figures 2.11 and 2.12.

### 2.3.3 Double Orientation Estimation

If we want to estimate the orientations of two edges at a vertex for arbitrary shapes, we need a more sophisticated approach. For this purpose, we will proceed by investigating the double orientation estimation procedure developed by Aach, Mota and others in [2, 148]. We start by recalling the central ideas of this approach and prove additional properties of the considered tensors. Note that a straightforward generalization for estimating more than two orientations was provided in [149].

In the following, let  $\Omega \subset \mathbb{R}^2$  be an image domain of interest. For simplicity, we set again  $\Omega = B_\varepsilon(0)$  and denote by  $r_i$ ,  $i = 1, 2$ , two orientation vectors with  $|r_i| = 1$  and  $r_1 \nparallel r_2$ . Now, for two sufficiently smooth functions  $\varphi_i : [-\varepsilon, \varepsilon] \rightarrow \mathbb{R}$ ,  $i = 1, 2$ , we assume that an image  $f$  can be decomposed into two functions  $f_i = \varphi_i(s_i^\top \cdot)$  with  $s_i := r_i^\perp$ ,  $i = 1, 2$ , in one of the following ways: As depicted in Figure 2.5, we consider  $f$  to fulfill either the *transparent model*

$$f(x) = f_1(x) + f_2(x) \quad \forall x \in \Omega \quad (2.24)$$

or the *occlusion model* with  $\Omega = \Omega_1 \cup \Omega_2$ ,  $\Omega_1 \cap \Omega_2 = \emptyset$  and

$$f(x) = \begin{cases} f_1(x) & \text{for } x \in \Omega_1, \\ f_2(x) & \text{for } x \in \Omega_2. \end{cases} \quad (2.25)$$

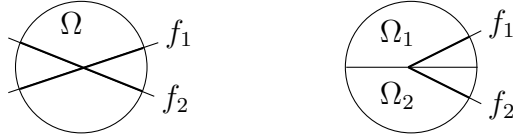


Figure 2.5: Illustration of the transparent model (left) and the occlusion model (right).

**Transparent model** If  $f$  fulfills the transparent model, the definitions of  $f_1$  and  $f_2$  imply for all  $x \in \Omega$  that

$$0 = \frac{\partial^2}{\partial r_1 \partial r_2} (f_1(x) + f_2(x)) = \frac{\partial^2}{\partial r_1 \partial r_2} f(x) = r_2^\top H(x) r_1 = r_1^\top H(x) r_2, \quad (2.26)$$

where  $H(x) := \begin{pmatrix} \partial_{x_1 x_1} f(x) & \partial_{x_1 x_2} f(x) \\ \partial_{x_1 x_2} f(x) & \partial_{x_2 x_2} f(x) \end{pmatrix}$  denotes the *Hessian* of  $f$  at  $x \in \Omega$ . Now, we will rewrite this equation using the tensor (Kronecker) product notation  $\otimes$  and the

following relation: If  $B \in \mathbb{R}^{m,n}$  is a matrix,  $b := \text{vec}(B) \in \mathbb{R}^{mn}$  the corresponding columnwise reshaped vector and  $M_1, M_2$  are matrices of appropriate size, then

$$(M_1 \otimes M_2) b = \text{vec}(M_2 B M_1^T). \quad (2.27)$$

Applying this relation, equation (2.26) becomes

$$0 = (r_1 \otimes r_2)^T h(x) = (r_2 \otimes r_1)^T h(x) \quad (2.28)$$

with  $h := \text{vec}(H) = (\partial_{x_1 x_1} f, \partial_{x_1 x_2} f, \partial_{x_2 x_1} f, \partial_{x_2 x_2} f)^T$ . Since this holds true for all  $x \in \Omega$ , we also have

$$\begin{aligned} 0 &= \int_{\Omega} w(x) (r_2^T H(x) r_1)^2 dx = \int_{\Omega} w(x) (r_1 \otimes r_2)^T h(x) h(x)^T (r_1 \otimes r_2) dx \\ &= (r_1 \otimes r_2)^T T (r_1 \otimes r_2) = (r_2 \otimes r_1)^T T (r_2 \otimes r_1), \end{aligned} \quad (2.29)$$

where  $w : \Omega \rightarrow \mathbb{R}$  denotes again a nonnegative weight function and

$$T := \int_{\Omega} w(x) h(x) h(x)^T dx$$

is a symmetric, positive semidefinite  $4 \times 4$  matrix. By (2.29) and since  $r_1 \nparallel r_2$ , the vectors  $r_1 \otimes r_2$  and  $r_2 \otimes r_1$  are two linearly independent eigenvectors of  $T$  with eigenvalue 0. However, instead of determining the orientations of  $r_1$  and  $r_2$  via the nullspace of  $T$ , Aach et al. proposed to use a *reduced model* by skipping the double entry  $\partial_{x_1 x_2} f$  in  $h$ . With (2.28) this leads to

$$0 = r^T \tilde{h}(x), \quad \text{where} \quad \tilde{h} := (\partial_{x_1 x_1} f, \partial_{x_1 x_2} f, \partial_{x_2 x_2} f)^T \quad (2.30)$$

and

$$r := \begin{pmatrix} r_{11} r_{21} \\ r_{11} r_{22} + r_{12} r_{21} \\ r_{12} r_{22} \end{pmatrix} \quad \text{with} \quad r_i = \begin{pmatrix} r_{i1} \\ r_{i2} \end{pmatrix}, \quad i = 1, 2. \quad (2.31)$$

Hence, our determining equation (2.29) becomes

$$0 = r^T \mathcal{T} r \quad \text{with} \quad \mathcal{T} := \int_{\Omega} w(x) \tilde{h}(x) \tilde{h}(x)^T dx \quad (2.32)$$

and  $r$  is an eigenvector of the symmetric, positive semidefinite  $3 \times 3$  matrix  $\mathcal{T}$  with eigenvalue 0. To describe  $\mathcal{T}$  in more detail we introduce the notation

$$v \tilde{\otimes} v := (v_1^2, v_1 v_2, v_2^2)^T,$$

which is a reduced version of the tensor product of  $v = (v_1, v_2)^T$  with itself. Furthermore, we denote by  $\Pi_n$  the space of polynomials defined on  $[-\varepsilon, \varepsilon]$  with degree  $\leq n$ . Now,  $\mathcal{T}$  can be characterized as follows:

**Proposition 2.3.2** *Let  $f$  fulfill the assumptions of the transparent model. Then, the matrix  $\mathcal{T}$  is of the form  $\mathcal{T} = \mathcal{S} \Phi \mathcal{S}^T$  with  $\mathcal{S} := (s_1 \tilde{\otimes} s_1, s_2 \tilde{\otimes} s_2)$  and*

$$\Phi := \int_{\Omega} w(x) \begin{pmatrix} (\varphi_1''(s_1^T x))^2 & \varphi_1''(s_1^T x) \varphi_2''(s_2^T x) \\ \varphi_1''(s_1^T x) \varphi_2''(s_2^T x) & (\varphi_2''(s_2^T x))^2 \end{pmatrix} dx. \quad (2.33)$$

Moreover, if  $w$  is a positive weight function, it holds that

$$\text{rank } \mathcal{T} = \begin{cases} 0 & \text{if } \varphi_i \in \Pi_1, i = 1, 2, \\ 1 & \text{if } \varphi_i \in \Pi_1 \text{ for exactly one } i \text{ or } \varphi_i \in \Pi_2 \setminus \Pi_1, i = 1, 2, \\ 2 & \text{otherwise.} \end{cases}$$

**Proof:** By

$$\begin{aligned} \partial_{x_1 x_1} f(x) &= s_{11}^2 \varphi_1''(s_1^T x) + s_{21}^2 \varphi_2''(s_2^T x), \\ \partial_{x_1 x_2} f(x) &= s_{11} s_{12} \varphi_1''(s_1^T x) + s_{21} s_{22} \varphi_2''(s_2^T x), \\ \partial_{x_2 x_2} f(x) &= s_{12}^2 \varphi_1''(s_1^T x) + s_{22}^2 \varphi_2''(s_2^T x) \end{aligned}$$

we have that  $\tilde{h}(x) = \mathcal{S} (\varphi_1''(s_1^T x), \varphi_2''(s_2^T x))^T$ , which implies with (2.32) further  $\mathcal{T} = \mathcal{S} \Phi \mathcal{S}^T$ .

Since  $r_1 \nparallel r_2$ , the matrix  $\mathcal{S} \in \mathbb{R}^{3,2}$  has full rank 2 and thus, the rank of  $\mathcal{T}$  is determined by  $\Phi$ . If  $w$  is positive, it holds that  $\Phi$  and consequently  $\mathcal{T}$  are zero matrices if and only if  $\varphi_i \in \Pi_1, i = 1, 2$ . On the other hand, if  $\varphi_i \notin \Pi_1, i = 1, 2$ , we have

$$\det \Phi = \|\varphi_1''(s_1^T \cdot)\|_{L_{2,w}(\Omega)}^2 \|\varphi_2''(s_2^T \cdot)\|_{L_{2,w}(\Omega)}^2 - |\langle \varphi_1''(s_1^T \cdot), \varphi_2''(s_2^T \cdot) \rangle_{L_{2,w}(\Omega)}|^2.$$

Here, Schwarz's inequality guarantees that

$$|\langle \varphi_1''(s_1^T \cdot), \varphi_2''(s_2^T \cdot) \rangle_{L_{2,w}(\Omega)}| \leq \|\varphi_1''(s_1^T \cdot)\|_{L_{2,w}(\Omega)} \|\varphi_2''(s_2^T \cdot)\|_{L_{2,w}(\Omega)},$$

where equality is attained if and only if the involved functions are multiples of each other. Consequently,  $\Phi$  has rank 1 if and only if either exactly one of the functions  $\varphi_i$  fulfills  $\varphi_i''(s_i^T x) = 0$  for all  $x \in \Omega$  or  $\varphi_2''(s_2^T x) = \lambda \varphi_1''(s_1^T x)$  with  $\lambda \in \mathbb{R} \setminus \{0\}$ .

In the first case, we see by setting  $x = t s_i$  that  $\varphi_i''(t) = 0$  holds for all  $t \in [-\varepsilon, \varepsilon]$  if and only if  $\varphi_i$  is linear. In the second case, we can use  $x = t s_2, t \in [-\varepsilon, \varepsilon]$ , to obtain

$$\varphi_2''(t) = \lambda \varphi_1''(t s_1^T s_2) = \lambda \varphi_1''(s_1^T s_1 t s_1^T s_2) = \varphi_2''(t (s_1^T s_2)^2) = \lambda \varphi_1''(t (s_1^T s_2)^3) = \dots \quad (2.34)$$

Since  $r_1 \nparallel r_2$ , we have  $|s_1^T s_2| < 1$  and thus, (2.34) yields  $\varphi_i''(t) = \varphi_i''(0)$  so that the functions  $\varphi_i, i = 1, 2$ , have to be quadratic on  $[-\varepsilon, \varepsilon]$ .  $\square$

Note that matrix  $T$  in (2.29) has a similar structure, namely  $T = S \Phi S^T$  with  $S := (s_1 \otimes s_1, s_2 \otimes s_2)$ . Moreover, this proposition implies that the structure of the nullspace  $\mathcal{N}(\mathcal{T})$  of  $\mathcal{T}$ , i.e., the space of the eigenvectors belonging to the eigenvalue 0, can be described as follows:

**Remark 2.3.3** If  $\text{rank } \mathcal{T} = 2$  (vertex case), then (2.32) yields that  $\mathcal{N}(\mathcal{T}) = \{cr : c \in \mathbb{R}\}$ . Straightforward computations shows further that  $\mathcal{S}^T r = (0, 0)^T$ , i.e.,  $r$  is orthogonal to  $s_1 \tilde{\otimes} s_1$  and  $s_2 \tilde{\otimes} s_2$ , respectively.

If  $\text{rank } \mathcal{T} = 1$  and only  $\varphi_j$  is linear or constant (edge case), we have

$$\mathcal{N}(\mathcal{T}) = \left\{ \begin{pmatrix} r_{i1}c_1 \\ r_{i1}c_2 + r_{i2}c_1 \\ r_{i2}c_2 \end{pmatrix} : c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \in \mathbb{R}^2 \right\} \quad \text{with } i \neq j. \quad (2.35)$$

If  $\text{rank } \mathcal{T} = 1$  and both functions  $\varphi_i$ ,  $i = 1, 2$ , are quadratic, then

$$\mathcal{N}(\mathcal{T}) = \text{span}\{r, \gamma_1(s_1 \tilde{\otimes} s_1) + \gamma_2(s_2 \tilde{\otimes} s_2)\},$$

where  $\gamma_1 := \alpha_1 \langle s_1 \tilde{\otimes} s_1, s_2 \tilde{\otimes} s_2 \rangle + \alpha_2 \|s_2 \tilde{\otimes} s_2\|_2^2$ ,  $\gamma_2 := -\alpha_1 \|s_1 \tilde{\otimes} s_1\|_2^2 - \alpha_2 \langle s_1 \tilde{\otimes} s_1, s_2 \tilde{\otimes} s_2 \rangle$  and  $\alpha_i := \|\varphi_i''(s_i^T \cdot)\|_{L_{2,w}(\Omega)}$ ,  $i = 1, 2$ .

There exist several possibilities to determine the vectors  $r_i$  with  $|r_i| = 1$ ,  $i = 1, 2$ , up to the sign from eigenvectors  $v \in \mathcal{N}(\mathcal{T})$ . It is not hard to check that the following approach from [2] is one possibility:

**Remark 2.3.4** Assume that  $\text{rank } \mathcal{T} = 2$  and we have found an eigenvector  $v$  of  $\mathcal{T}$  with eigenvalue 0, i.e.,  $v = (v_1, v_2, v_3)^T = cr \in \mathcal{N}(\mathcal{T})$ . If  $v_1 \neq 0$ , then

$$r_1 = \frac{1}{\sqrt{v_1^2 + y_1^2}} (v_1, y_1)^T, \quad r_2 = \frac{1}{\sqrt{v_1^2 + y_2^2}} (v_1, y_2)^T \quad (2.36)$$

are the sought orientations, where  $y_1, y_2$  are the solutions of the quadratic equation  $y^2 - v_2 y + v_1 v_3 = 0$ . If  $v_1 = 0$ , then  $y_i = 0$  for one  $i$  and we set  $r_i := (v_2, v_3)^T / \sqrt{v_2^2 + v_3^2}$  and  $r_{3-i} := (0, 1)^T$ .

In the case that  $\text{rank } \mathcal{T} = 1$  with  $\varphi_2$  being linear and  $\varphi_1$  not, let  $v$  be again an eigenvector of  $\mathcal{T}$  with eigenvalue 0, i.e.,  $v$  is an element of the nullspace  $\mathcal{N}(\mathcal{T})$  provided in (2.35). Obviously, this implies that  $v$  has the same structure as (2.31) with  $r_1$  and some second vector  $c \in \mathbb{R}^2 \setminus \{(0, 0)^T\}$ . Thus, we can compute  $r_1, c/|c|$  by the same procedure as just described for  $\text{rank } \mathcal{T} = 2$ , where  $c/|c|$  plays the role of  $r_2$ . To distinguish  $r_1$  from  $c/|c|$ , we may use a second eigenvector  $u \in \mathcal{N}(\mathcal{T})$  which has to be linearly independent from  $v$ . After decomposing this vector into  $r_1$  and some vector  $\tilde{c}/|\tilde{c}|$ , we can identify  $r_1$  as the conform orientation of both decompositions.

**Occlusion model** For the occlusion or covering model (2.25) we can deduce similar results as for the transparent model. Here, the definition of  $f_1$  and  $f_2$  implies that

$$\begin{aligned} 0 &= \frac{\partial}{\partial r_1} f(x) \frac{\partial}{\partial r_2} f(x) = (r_1^T \nabla f(x)) (r_2^T \nabla f(x)) = r_1^T \nabla f(x) (\nabla f(x))^T r_2 \\ &= r_1^T G(x) r_2 = r_2^T G(x) r_1 \quad \forall x \in \Omega, \end{aligned} \quad (2.37)$$

where  $G(x) := \begin{pmatrix} (\partial_{x_1} f(x))^2 & \partial_{x_1} f(x) \partial_{x_2} f(x) \\ \partial_{x_1} f(x) \partial_{x_2} f(x) & (\partial_{x_2} f(x))^2 \end{pmatrix}$ . Rewriting this equation using tensor products we obtain further

$$0 = (r_2 \otimes r_1)^T g(x) = (r_1 \otimes r_2)^T g(x) \quad \text{with } g := ((\partial_{x_1} f)^2, \partial_{x_1} f \partial_{x_2} f, \partial_{x_1} f \partial_{x_2} f, (\partial_{x_2} f)^2)^T.$$

In its reduced form with  $r$  defined as in (2.31) this matches

$$0 = r^T \tilde{g}(x) \quad \text{with } \tilde{g} := ((\partial_{x_1} f)^2, \partial_{x_1} f \partial_{x_2} f, (\partial_{x_2} f)^2)^T.$$

Since this relation holds true for all  $x \in \Omega$ , it finally implies that

$$0 = r^T \mathcal{C} r \quad \text{for } \mathcal{C} := \int_{\Omega} w(x) \tilde{g}(x) \tilde{g}(x)^T dx \quad (2.38)$$

with any nonnegative weight function  $w : \Omega \rightarrow \mathbb{R}$ . Thus,  $r$  is an eigenvector of the symmetric, positive semidefinite  $3 \times 3$  matrix  $\mathcal{C}$  with eigenvalue 0. Moreover, the following proposition holds true:

**Proposition 2.3.5** *If an image  $f$  fulfills the assumptions of the occlusion model, then  $\mathcal{C}$  is of the form*

$$\mathcal{C} = \alpha_1 (s_1 \otimes s_1) (s_1 \otimes s_1)^T + \alpha_2 (s_2 \otimes s_2) (s_2 \otimes s_2)^T \quad \text{with } \alpha_i := \int_{\Omega_i} w(x) (\varphi'_i(s_i^T x))^4 dx,$$

$i = 1, 2$ . For a positive weight function  $w$ , the rank of  $\mathcal{C}$  is  $\nu \in \{0, 1, 2\}$  if and only if exactly  $2 - \nu$  of the functions  $\varphi_i$  are constant on  $\Omega_i$ ,  $i = 1, 2$ .

**Proof:** For  $i = 1, 2$  we obtain by

$$\begin{aligned} (\partial_{x_1} f_i(x))^2 &= (\varphi'_i(s_i^T x))^2 s_{i1}^2, \\ \partial_{x_1} f_i(x) \partial_{x_2} f_i(x) &= (\varphi'_i(s_i^T x))^2 s_{i1} s_{i2}, \\ (\partial_{x_2} f_i(x))^2 &= (\varphi'_i(s_i^T x))^2 s_{i2}^2 \end{aligned}$$

that  $\tilde{g}(x) = (\varphi'_i(s_i^T x))^2 (s_i \otimes s_i)$  for all  $x \in \Omega_i$  and consequently

$$\begin{aligned} \mathcal{C} &= \int_{\Omega_1} w(x) \tilde{g}(x) \tilde{g}(x)^T dx + \int_{\Omega_2} w(x) \tilde{g}(x) \tilde{g}(x)^T dx \\ &= \alpha_1 (s_1 \otimes s_1) (s_1 \otimes s_1)^T + \alpha_2 (s_2 \otimes s_2) (s_2 \otimes s_2)^T. \end{aligned}$$

The rest of the assertion follows immediately by the definition of  $\alpha_i$ ,  $i = 1, 2$ .  $\square$

Due to Proposition 2.3.5 the eigenvectors of  $\mathcal{C}$  with eigenvalue 0 can be described as follows:



**Remark 2.3.6** If  $\text{rank } \mathcal{C} = 2$ , then  $\mathcal{N}(\mathcal{C}) = \{cr : c \in \mathbb{R}\}$ . If  $\text{rank } \mathcal{C} = 1$  and the function  $\varphi_i$  is not constant on  $\Omega_i$ , then

$$\mathcal{N}(\mathcal{C}) = \left\{ \begin{pmatrix} r_{i1}c_1 \\ r_{i1}c_2 + r_{i2}c_1 \\ r_{i2}c_2 \end{pmatrix} : c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \in \mathbb{R}^2 \right\}.$$

The vectors  $r_i$  with  $|r_i| = 1$ ,  $i = 1, 2$ , can again be reconstructed up to the sign from a basis of  $\mathcal{N}(\mathcal{C})$  as described in Remark 2.3.4.

**Double orientation tensors** Since in natural images the significant orientations are not the same in all image parts, we finally want to compute tensors  $\mathcal{T}$  and  $\mathcal{C}$  for every image point. For this purpose, we consider  $\Omega$  to be the whole image domain again and use  $w = K_\rho(\cdot - x)$  (truncated outside  $B_{3\rho}(x)$ ) with  $\rho > 0$  to detect the orientation in the neighborhood of each image point  $x$ . Since in practice we are often confronted with noisy images, we further use  $f_\sigma$  instead of  $f$  in the definition of the tensors. As a result, we obtain the *double orientation tensors*

$$\begin{aligned} \mathcal{T}_\rho(x) &:= \int_{\Omega} K_\rho(x - y) \tilde{h}_\sigma(y) \tilde{h}_\sigma(y)^T dy = \left( K_\rho * (\tilde{h}_\sigma \tilde{h}_\sigma^T) \right)(x) \\ \text{and } \mathcal{C}_\rho(x) &:= \int_{\Omega} K_\rho(x - y) \tilde{g}_\sigma(y) \tilde{g}_\sigma(y)^T dy = \left( K_\rho * (\tilde{g}_\sigma \tilde{g}_\sigma^T) \right)(x) \quad \forall x \in \Omega \end{aligned}$$

with  $\tilde{h}_\sigma := (\partial_{x_1 x_1} f_\sigma, \partial_{x_1 x_2} f_\sigma, \partial_{x_2 x_2} f_\sigma)^T$  and  $\tilde{g}_\sigma := ((\partial_{x_1} f_\sigma)^2, \partial_{x_1} f_\sigma \partial_{x_2} f_\sigma, (\partial_{x_2} f_\sigma)^2)^T$ . These tensors differ from the classical structure tensor (2.15) only by the definition of the vectors  $\tilde{g}_\sigma$  and  $\tilde{h}_\sigma$ . To determine candidates for the significant orientations  $r_1, r_2$  at each image point, we will compute an eigenvector to the smallest eigenvalue of the appropriate tensor and decompose it as described in Remark 2.3.4. An example of estimated orientation vectors is given in Figure 2.6. Since  $\rho$  is chosen rather large here, we obtain the same orientations in the whole image domain.

## 2.4 Shape Preserving Anisotropic Regularization

Now, we want to use these orientation estimation methods in the context of anisotropic regularization. The considered approaches are inspired by the work [20] of Berkels et al. as well as theoretical results of Esedoglu and Osher in [75]. To motivate the applied regularizers, we will first study a particular anisotropic regularizer in detail and prove different results showing which shapes are preserved by this variational approach.

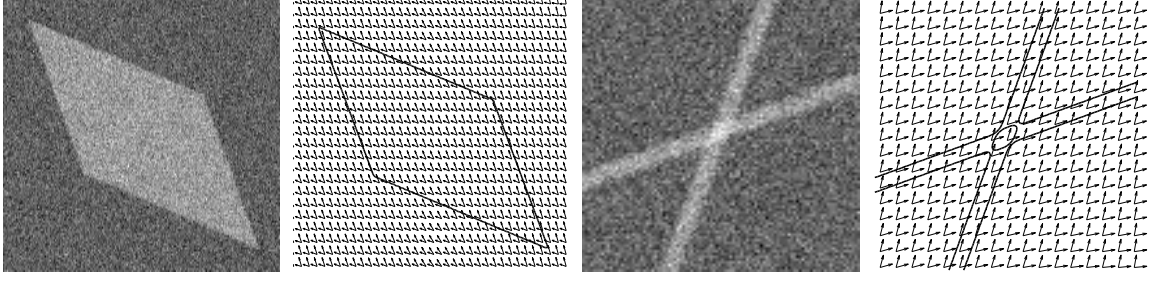


Figure 2.6: Noisy images and the estimated double orientations ( $\sigma = 2$ ,  $\rho = 25$ ) by the occlusion model (left) and the transparent model (right), respectively.

### 2.4.1 Shape Preservation

For the subsequent considerations we set  $\Omega = \mathbb{R}^2$  and consider regularization methods of the form (2.1) with  $J$  defined by (2.4) with  $\varphi(x) = \|x\|_1$ , i.e.,

$$J(u) := \int_{\mathbb{R}^2} |\partial_{x_1} u| + |\partial_{x_2} u| dx. \quad (2.39)$$

The following theorem, cf. [75, Prop. 3.3], provides necessary and sufficient conditions for  $\hat{u}$  to be a minimizer of (2.1) for general regularizers  $J$ :

**Theorem 2.4.1** *For a proper, convex, lower semi-continuous (l.s.c.) and positively homogeneous functional  $J : L_2(\Omega) \rightarrow \mathbb{R}_{\geq 0} \cup \{+\infty\}$  the minimization problem (2.1) has a unique solution and  $\hat{u} \in L_2(\Omega)$  is the solution if and only if*

- i)  $\hat{u} = f - \lambda \hat{v}$ ,
- ii)  $\hat{v} \in C_J := \{v \in L_2(\Omega) : \langle v, w \rangle \leq J(w) \ \forall w \in L_2(\Omega)\}$ ,
- iii)  $\langle \hat{u}, \hat{v} \rangle = J(\hat{u})$ .

The special functional (2.4) with  $\Omega = \mathbb{R}^2$  fulfills the above assumptions and we have that  $\hat{v} \in C_J$  if there exists a vector field  $\hat{V} \in L_\infty(\mathbb{R}^2, \mathbb{R}^2)$  such that

$$\hat{v} := -\operatorname{div} \hat{V} \in L_2(\mathbb{R}^2) \quad \text{and} \quad \hat{V} \in W_\varphi \text{ a.e. on } \mathbb{R}^2.$$

Note that part iii) ensures here that  $J(\hat{u}) < \infty$ . Although the first part of this theorem is somehow standard, we add the proof for completeness.

**Proof:** Let  $F : L_2(\Omega) \rightarrow \mathbb{R} \cup \{+\infty\}$  be defined by

$$F(u) = \frac{1}{2} \|f - u\|_{L_2}^2 + \lambda J(u).$$

Obviously,  $F$  is proper and l.s.c. Since the first term of  $F$  is strictly convex and coercive, and  $J$  is convex and bounded from below, we know that  $F$  is also strictly convex and coercive. Using that  $F$  is defined on the Hilbert space  $L_2(\Omega)$ , it finally follows by [73, p. 35] that the minimization problem (2.1) has a unique solution. By [73, p. 21] we know that  $\hat{u}$  is the minimizer of (2.1) if and only if

$$0 \in \hat{u} - f + \lambda \partial J(\hat{u}), \quad \text{i.e.,} \quad \frac{f - \hat{u}}{\lambda} \in \partial J(\hat{u}).$$

By [73, p. 22] this is equivalent to

$$\hat{u} \in \partial J^* \left( \frac{f - \hat{u}}{\lambda} \right),$$

where  $J^*$  denotes the *conjugate function* of  $J$ . Setting  $\hat{v} := \frac{f - \hat{u}}{\lambda}$  so that  $\hat{u} = f - \lambda \hat{v}$ , we obtain further

$$f - \lambda \hat{v} \in \partial J^*(\hat{v}).$$

This inclusion holds true if and only if

$$\hat{v} = \operatorname{argmin}_{v \in L_2(\Omega)} \left\{ \frac{1}{2} \left\| \frac{f}{\lambda} - v \right\|_{L_2}^2 + \frac{1}{\lambda} J^*(v) \right\}. \quad (2.40)$$

Since  $J$  is proper, convex and positively homogeneous, it follows that  $J^*$  is the indicator function

$$\iota_{C_J}(v) := \begin{cases} 0 & v \in C_J, \\ \infty & \text{otherwise} \end{cases}$$

of the closed, convex set  $C_J := \{v \in L_2(\Omega) : \langle v, w \rangle \leq J(w) \ \forall \ w \in L_2(\Omega)\}$ . Thus, (2.40) can be rewritten as

$$\hat{v} = \operatorname{argmin}_{v \in C_J} \|f - \lambda v\|_{L_2}^2,$$

i.e.,  $\lambda \hat{v}$  is the orthogonal projection of  $f$  onto  $C_{\lambda J}$ . It remains to show that this is equivalent to condition iii). If  $\lambda \hat{v}$  is the orthogonal projection of  $f$  onto  $C_{\lambda J}$ , the projection theorem implies for all  $w \in C_{\lambda J}$  that

$$\langle f - \lambda \hat{v}, w - \lambda \hat{v} \rangle \leq 0 \quad \text{and thus,} \quad \langle f - \lambda \hat{v}, w \rangle \leq \langle f - \lambda \hat{v}, \lambda \hat{v} \rangle.$$

Since  $J(u) = J^{**}(u) = \sup_{w \in C_J} \langle u, w \rangle$ , see, e.g., [73, p. 18], we obtain

$$\lambda J(f - \lambda \hat{v}) = \sup_{w \in C_{\lambda J}} \langle f - \lambda \hat{v}, w \rangle = \langle f - \lambda \hat{v}, \lambda \hat{v} \rangle, \quad \text{i.e.,} \quad J(\hat{u}) = \langle \hat{u}, \hat{v} \rangle.$$

Conversely, let  $\hat{v} \in C_J$  fulfill iii). By the definition of  $C_{\lambda J}$  we have for all  $w \in C_{\lambda J}$  that

$$\langle f - \lambda \hat{v}, w \rangle \leq \lambda J(f - \lambda \hat{v}), \quad \text{and thus,} \quad 0 \leq -\langle f - \lambda \hat{v}, w \rangle + \lambda J(f - \lambda \hat{v}).$$

Since  $J(\hat{u}) = \langle \hat{u}, \hat{v} \rangle$ , it follows that

$$0 \leq \langle f - \lambda \hat{v}, \lambda \hat{v} - w \rangle \quad \forall w \in C_{\lambda J}$$

and the projection theorem finally yields that  $\lambda \hat{v}$  is the orthogonal projection of  $f$  onto  $C_{\lambda J}$ .

For the second part of the assertion, it is easy to see that the regularizer  $J$  defined in (2.4) with (2.5) is nonnegative, proper and positively homogeneous. Following the lines of [3] it is also not hard to prove that (2.4) is convex and weakly lower semi-continuous, which also implies lower semi-continuity. Thus, (2.4) fulfills the assumptions of the first part. The rest of the assertion follows by definition (2.5) and a limit argument in [75].  $\square$

Using Theorem 2.4.1, we can prove the following corollary:

**Corollary 2.4.2** *Let  $1_A : \mathbb{R}^2 \rightarrow \{0, 1\}$  be the characteristic function of the set  $A \subset \mathbb{R}^2$ , i.e.,  $1_A(x) = 1$  for  $x \in A$  and  $1_A(x) = 0$  otherwise. Moreover, denote by  $\hat{u}$  the solution of the minimization problem (2.1) with  $J$  defined by (2.39).*

- i) *If  $f := c 1_R$  is the given initial image with  $c > 0$  and some rectangle  $R := (-a, a) \times (-b, b)$ , where  $a, b > 0$ , then*

$$\hat{u} = \left( c - \lambda \frac{a+b}{ab} \right) 1_R \quad \text{for } \lambda \leq \frac{ab}{a+b} c.$$

- ii) *If  $f := c_1 1_{R_1} + c_2 1_{R_2}$  with  $c_1, c_2 > 0$ ,  $R_1 := (-l, l) \times (-a, a)$ ,  $R_2 := (-b, b) \times (-l, l)$  and  $0 < a, b < l$ , then*

$$\hat{u} = \left( c_1 - \lambda \frac{1}{a} \right) 1_{R_1} + \left( c_2 - \lambda \frac{1}{b} \right) 1_{R_2} \quad \text{for } \lambda \leq \min \{ a c_1, b c_2 \}.$$

**Proof:** The prove of part i) was given for  $c = 1$  in [75, Claim 4.2] and generalizes straightforwardly for  $c \neq 1$ . To show part ii) we define

$$\eta(\xi) = \begin{cases} -1 & \text{if } \xi < -1, \\ \xi & \text{if } -1 \leq \xi \leq 1, \\ 1 & \text{if } \xi > 1. \end{cases}$$

Now, we need to verify that the vector field  $\hat{V} = (\hat{V}^{(1)}, \hat{V}^{(2)}) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  given by

$$\begin{aligned} \hat{V}^{(1)}(x_1, x_2) &= -\eta\left(\frac{x_1}{b}\right) 1_{(-l, l)}(x_2), \\ \hat{V}^{(2)}(x_1, x_2) &= -\eta\left(\frac{x_2}{a}\right) 1_{(-l, l)}(x_1) \end{aligned}$$

fulfills the conditions of Theorem 2.4.1: Obviously,  $\hat{V} \in L_\infty(\mathbb{R}^2, \mathbb{R}^2)$  and for all  $x \in \mathbb{R}^2$  it holds that

$$\hat{V}(x_1, x_2) \in W_\varphi = \{y \in \mathbb{R}^2 : \max\{|y_1|, |y_2|\} \leq 1\}$$

with

$$\hat{v} := -\operatorname{div} \hat{V} = \frac{1}{a} 1_{R_1} + \frac{1}{b} 1_{R_2} \in L_2(\mathbb{R}^2).$$

Hence,  $\hat{v} \in C_J$  and  $\hat{u} = f - \lambda \hat{v}$ . Using [75, Claim 3.2] and the anisotropic coarea formula given in [99, 154] we can show that for  $d_1 := c_1 - \lambda \frac{1}{a}$ ,  $d_2 := c_2 - \lambda \frac{1}{b}$  and  $\lambda \leq \min\{a c_1, b c_2\}$  it holds that

$$\begin{aligned} J(\hat{u}) &= 2 d_1 (2l + 2a) + 2 d_2 (2l + 2b) \\ &= d_1 \frac{1}{a} 2a (2l - 2b) + d_2 \frac{1}{b} 2b (2l - 2a) + (d_1 + d_2) \left( \frac{1}{a} + \frac{1}{b} \right) 2a 2b \\ &= \langle \hat{u}, \hat{v} \rangle, \end{aligned}$$

which finishes the proof.  $\square$

This corollary shows that rectangles with edges parallel to the coordinate axes and + junctions are preserved by the solution of (2.1) with regularizer (2.39) if the regularization parameter  $\lambda$  is small enough. Since the critical values  $\lambda$  depend directly on the shape parameters  $a, b$  and the gray values  $c, c_1, c_2$ , it also implies that for larger values  $\lambda$  thin lines and low contrasted objects are removed first.

**Theoretical results for a discrete setting** Similar results can also be obtained for a discretized version of (2.1) and (2.39). To deduce these results, we consider discrete quadratic images  $\mathbf{f} := (f(x_1, x_2))_{x_1, x_2=0}^{n-1} \in \mathbb{R}^{n,n}$  and their columnwise reshaped vectors  $f := \operatorname{vec}(\mathbf{f}) \in \mathbb{R}^N$  with  $N := n^2$ . Instead of partial derivatives we use *forward differences* now so that the discrete counterpart of the gradient reads

$$D = \begin{pmatrix} D_{x_1} \\ D_{x_2} \end{pmatrix} := \begin{pmatrix} I_N \otimes H_1 \\ H_1 \otimes I_N \end{pmatrix} \quad \text{or} \quad D := \begin{pmatrix} H_0 \otimes H_1 \\ H_1 \otimes H_0 \end{pmatrix}, \quad (2.41)$$

where  $I_N$  denotes the identity matrix of size  $N \times N$  and

$$H_0 := \frac{1}{2} \begin{pmatrix} 1 & 1 & & & \\ & 1 & 1 & & \\ & & \ddots & \ddots & \\ & & & 1 & 1 \\ & & & & 2 \end{pmatrix}, \quad H_1 := \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \\ & & & & 0 \end{pmatrix} \quad (2.42)$$

are Haar matrices with reflecting boundary conditions. Moreover, problem (2.1) becomes

$$\operatorname{argmin}_{u \in \mathbb{R}^N} \left\{ \frac{1}{2} \|f - u\|_2^2 + \lambda J(u) \right\}, \quad (2.43)$$

and the functional (2.39) reads

$$J(u) := \|Du\|_1. \quad (2.44)$$

The solution of (2.43) can be characterized as in the continuous setting, cf. [SST08]:

**Proposition 2.4.3** *Let  $J : \mathbb{R}^N \rightarrow \mathbb{R}_{\geq 0} \cup \{+\infty\}$  be a proper, convex, l.s.c. and positively homogeneous functional. Then, the minimization problem (2.43) has a unique solution. This solution is given by  $\hat{u} \in \mathbb{R}^N$  if and only if conditions i) - iii) of Theorem 2.4.1 hold true, where  $L_2(\Omega)$  has to be replaced by  $\mathbb{R}^N$  with the Euclidean inner product.*

*For the special functional (2.44) we have that  $\hat{v} \in C_J$  if and only if there exists a vector  $\hat{V} \in \mathbb{R}^{2N}$  such that*

$$\hat{v} := D^T \hat{V} \quad \text{and} \quad \|\hat{V}\|_\infty \leq 1.$$

**Proof:** The first part of the proof of Theorem 2.4.1 remains true for general real Hilbert space, i.e., also for the  $N$ -dimensional Euclidean space. The second part of the assertion was proven in [SST08] for general gauge functions  $\varphi$ .  $\square$

As in the continuous case, we can show that for small enough  $\lambda$  rectangles and + junctions are preserved by the solution of (2.43) with  $J$  defined by (2.44). However, due to the image boundaries one has to be careful with the discretization.

**Corollary 2.4.4** *Let  $x_{1,0}, x_{2,0} \geq 0$  and  $x_{1,0} + a, x_{2,0} + b \leq n - 2$ . Moreover, denote by  $\hat{u}$  the solution of minimization problem (2.43) with  $J$  defined by (2.44).*

- i) *For  $f := c 1_R$  with  $c > 0$  and the discrete rectangle  $R := \{x_{1,0} + 1, \dots, x_{1,0} + a\} \times \{x_{2,0} + 1, \dots, x_{2,0} + b\}$ , it holds that*

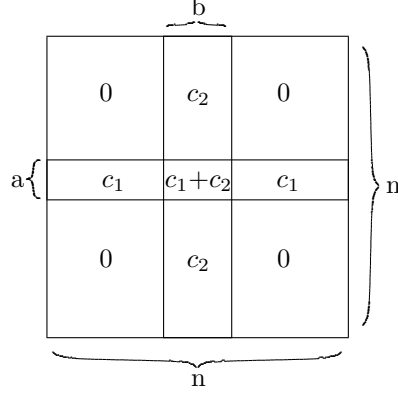
$$\hat{u} = \left( c - \lambda \frac{2(a+b)}{ab} \right) 1_R \quad \text{for } \lambda \leq \frac{ab}{2(a+b)} c$$

*if  $D$  is the second matrix defined in (2.41) and  $H_i, i = 0, 1$ , are modified such that  $H_i(0, 0) = 0, H_i(n-1, n-1) = (-1)^i, i = 0, 1$ .*

- ii) *For  $f := c_1 1_{R_1} + c_2 1_{R_2}$  with  $c_1, c_2 > 0$  and  $R_1 := \{x_{1,0} + 1, \dots, x_{1,0} + a\} \times \{0, \dots, n-1\}, R_2 := \{0, \dots, n-1\} \times \{x_{2,0} + 1, \dots, x_{2,0} + b\}$  we have*

$$\hat{u} = \left( c_1 - \lambda \frac{2}{a} \right) 1_{R_1} + \left( c_2 - \lambda \frac{2}{b} \right) 1_{R_2} \quad \text{for } \lambda \leq \min \left\{ \frac{a}{2} c_1, \frac{b}{2} c_2 \right\}$$

*if  $D$  is one of the matrices defined in (2.41) and  $H_i, i = 0, 1$ , are modified such that  $H_0(n-1, 0) = 1, H_1(0, 0) = 0, H_i(n-1, n-1) = (-1)^i, i = 0, 1$ .*


 Figure 2.7: Sketch of the image  $\mathbf{f}$  examined in Corollary 2.4.4 ii).

**Proof:** Part i) of the assertion was proven in [203, SST08].

The image  $\mathbf{f}$  considered in part ii) is depicted in Figure 2.7. For

$$\hat{V}^{(1)}(x_1, x_2) := \begin{cases} 1 & \text{if } x_1 \in \{0, \dots, x_{1,0}\}, \\ 1 - \frac{2}{a}(x_1 - x_{1,0}) & \text{if } x_1 \in \{x_{1,0} + 1, \dots, x_{1,0} + a\}, \\ -1 & \text{if } x_1 \in \{x_{1,0} + a + 1, \dots, n - 1\} \end{cases}$$

and

$$\hat{V}^{(2)}(x_1, x_2) := \begin{cases} 1 & \text{if } x_2 \in \{0, \dots, x_{2,0}\}, \\ 1 - \frac{2}{b}(x_2 - x_{2,0}) & \text{if } x_2 \in \{x_{2,0} + 1, \dots, x_{2,0} + b\}, \\ -1 & \text{if } x_2 \in \{x_{2,0} + b + 1, \dots, n - 1\} \end{cases}$$

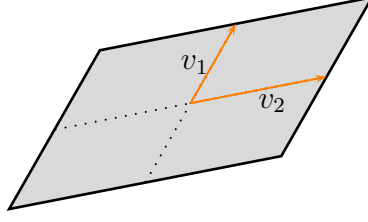
it holds that  $\|\hat{V}\|_\infty \leq 1$ . Moreover,  $\hat{v} = D^T \hat{V} = \frac{2}{a} 1_{R_1} + \frac{2}{b} 1_{R_2}$  for both choices of  $D$  in (2.41), where  $H_0, H_1$  are modified as follows:  $H_0(n - 1, 0) = 1$ ,  $H_1(0, 0) = 0$ ,  $H_i(n - 1, n - 1) = (-1)^i$ ,  $i = 0, 1$ . Thus,

$$\hat{u} = f - \lambda \hat{v} = \left(c_1 - \frac{2}{a}\lambda\right) 1_{R_1} + \left(c_2 - \frac{2}{b}\lambda\right) 1_{R_2} = d_1 1_{R_1} + d_2 1_{R_2}$$

with  $d_1 := c_1 - \frac{2}{a}\lambda$  and  $d_2 := c_2 - \frac{2}{b}\lambda$ . For  $\lambda \leq \min\{\frac{a}{2} c_1, \frac{b}{2} c_2\}$  this implies

$$\begin{aligned} J(\hat{u}) &= 2(n + a) d_1 + 2(n + b) d_2 \\ &= (n - b) a \frac{2}{a} d_1 + (n - a) b \frac{2}{b} d_2 + ab \left(\frac{2}{a} + \frac{2}{b}\right) (d_1 + d_2) \\ &= \langle \hat{u}, \hat{v} \rangle. \end{aligned}$$

Hence, the conditions of Proposition 2.4.3 are fulfilled and  $\hat{u}$  is the sought minimizer.  $\square$


 Figure 2.8: Shape with edge orientations  $v_1$  and  $v_2$ .

### 2.4.2 A New Anisotropic Regularization Method

Having the regularizer (2.39) in mind, we now want to introduce a slightly different functional, which can preserve also other shapes than the ones considered in Corollary 2.4.2. If an image contains for example rotated parallelograms like the one depicted in Figure 2.8 or X junctions with edge orientations  $v_1, v_2$ , the idea is to consider the regularizer (2.4) with

$$\varphi_V(x) = \|Vx\|_1 = |\langle v_1, x \rangle| + |\langle v_2, x \rangle| \quad \text{and} \quad V := (v_1, v_2)^T \in \mathbb{R}^{2,2}$$

instead of  $\varphi(x) = \|x\|_1$ . This modification can be motivated by the following considerations: Assume that  $V$  is invertible, i.e.,  $v_1 \nparallel v_2$ , and  $f$  is an image of a rotated parallelogram or respectively an X junction. If the edges are orientated like  $v_1$  and  $v_2$ , the transformed image  $f_V := f(V^T \cdot)$  shows a rectangle or respectively a + junction with edges parallel to the coordinate axes. By setting  $y := V^T x$ ,  $dy = |\det V| dx$  we obtain

$$\begin{aligned} \int_{\mathbb{R}^2} \frac{1}{2}(f - u)^2 + \lambda \|V \nabla u\|_1 dy &= |\det V| \int_{\mathbb{R}^2} \frac{1}{2}(f(V^T x) - u(V^T x))^2 + \lambda \|V \nabla_y u(V^T x)\|_1 dx \\ &= |\det V| \int_{\mathbb{R}^2} \frac{1}{2}(f_V - u_V)^2 + \lambda \|\nabla u_V\|_1 dx. \end{aligned}$$

Consequently,  $\hat{u}$  minimizes the left-hand side if and only if the transformed image  $\hat{u}_V := \hat{u}(V^T \cdot)$  minimizes

$$\frac{1}{2} \|f_V - u_V\|_{L_2}^2 + \lambda \int_{\mathbb{R}^2} \|\nabla u_V\|_1 dx. \quad (2.45)$$

Thus, if we minimize

$$\frac{1}{2} \|f - u\|_{L_2}^2 + \lambda \int_{\mathbb{R}^2} \|V \nabla u\|_1 dx$$

for sufficiently small  $\lambda$ , Corollary 2.4.2 implies that shapes like the one in Figure 2.8 and appropriate X junctions are preserved. The Wulff shape of  $\varphi_V := \|V \cdot\|_1$  has



exactly the shape depicted in Figure 2.8. Using that  $V^{-1} = \frac{1}{\det(V)}(v_2^\perp, -v_1^\perp)$  we have in detail

$$\begin{aligned} W_{\varphi_V} &= \{x \in \mathbb{R}^2 : \langle x, y \rangle \leq \|Vy\|_1 \quad \forall y \in \mathbb{R}^2\} \\ &= \{x \in \mathbb{R}^2 : \langle x, V^{-1}z \rangle \leq \|z\|_1 \quad \forall z \in \mathbb{R}^2\} \\ &= \{x \in \mathbb{R}^2 : \langle (V^{-1})^T x, z \rangle \leq \|z\|_1 \quad \forall z \in \mathbb{R}^2\} \\ &= \{x \in \mathbb{R}^2 : \max\{|\langle v_1^\perp, x \rangle|, |\langle v_2^\perp, x \rangle|\} \leq |\det(V)|\}. \end{aligned}$$

The orientations of the structures in a natural image  $u : \Omega \rightarrow \mathbb{R}$  may of course vary from image region to image region. Therefore, we will in the following replace the single matrix  $V \in \mathbb{R}^{2,2}$  by a whole tensor field  $V : \Omega \rightarrow \mathbb{R}^{2,2}$ . To determine this tensor field from a possibly noisy initial image  $f : \Omega \rightarrow \mathbb{R}$ , we apply one of the models presented in Section 2.3. Hence, our proposed anisotropic regularization method can be summarized as follows:

**Algorithm 2.4.5** *For a possibly noisy initial image  $f : \Omega \rightarrow \mathbb{R}$  proceed as follows:*

**Step 1: Orientation estimation**

*Estimate the orientation matrices  $V$  from the initial image  $f$  by one of the methods described above:*

- **Case 1: Rotated rectangular shapes** *If the initial image contains up to noise mainly rotated rectangular shapes, estimate the rotation angles  $\alpha$  of the shapes by the adapted structure tensor in Subsection 2.3.2 and set*

$$V(x) = R(-\alpha(x)) = \begin{pmatrix} \cos \alpha(x) & \sin \alpha(x) \\ -\sin \alpha(x) & \cos \alpha(x) \end{pmatrix} \quad \forall x \in \Omega.$$

- **Case 2: Sheared shapes** *If we are confronted with sheared shapes, determine the shear parameters as described in Subsection 2.3.2 and set*

$$V(x) = S(\zeta(x)) = \begin{pmatrix} 1 & 0 \\ \zeta(x) & 1 \end{pmatrix} \quad \forall x \in \Omega.$$

- **Case 3: Arbitrary double orientations** *If the initial image can locally be described by one of the double orientation models in Subsection 2.3.3, estimate the orientations  $r_1, r_2$  by the appropriate double orientation tensor and set*

$$V(x) = (r_1(x), r_2(x))^T \quad \forall x \in \Omega.$$

**Step 2: Anisotropic regularization**

Using these orientation matrices  $V$  and an appropriately chosen regularization parameter  $\lambda$ , determine the regularized image  $\hat{u}$  by solving

$$\operatorname{argmin}_{u \in L_2(\Omega)} \left\{ \frac{1}{2} \|f - u\|_{L_2}^2 + \lambda \int_{\Omega} \|V \nabla u\|_1 dx \right\}. \quad (2.46)$$

As outlined in [100, 132] it is also possible to make the orientation matrices  $V$  in (2.46) directly dependent on  $u$  rather than  $f$ .

### 2.4.3 Discretization and Minimization

To minimize the functional in (2.46) numerically, we compute the minimizer of its discrete counterpart

$$\operatorname{argmin}_{u \in \mathbb{R}^N} \left\{ \frac{1}{2} \|f - u\|_2^2 + \lambda \|\underbrace{\mathcal{V}D}_M u\|_1 \right\}, \quad (2.47)$$

where  $f, u \in \mathbb{R}^N$  are columnwise reshaped image vectors,  $D \in \mathbb{R}^{2N, N}$  is a discrete gradient operator and  $\mathcal{V} \in \mathbb{R}^{2N, 2N}$  a matrix containing the orientation matrices  $V(i, j)$  for all image points  $i \in \{0, \dots, m-1\}$ ,  $j \in \{0, \dots, n-1\}$ . For simplicity we consider only the case that  $m = n$ , since the generalization for  $m \neq n$  is straightforward. We set  $D$  to be second matrix in (2.41), which equals the choice in [SST08]. The matrix  $\mathcal{V}$  is thus defined by

$$\mathcal{V} = \begin{pmatrix} \operatorname{diag}(\operatorname{vec}(v_{1,1})) & \operatorname{diag}(\operatorname{vec}(v_{1,2})) \\ \operatorname{diag}(\operatorname{vec}(v_{2,1})) & \operatorname{diag}(\operatorname{vec}(v_{2,2})) \end{pmatrix} \quad \text{with} \quad V(i, j) = \begin{pmatrix} v_{1,1}(i, j) & v_{1,2}(i, j) \\ v_{2,1}(i, j) & v_{2,2}(i, j) \end{pmatrix}.$$

To avoid possible checkerboard effects, we extend (2.47) by adding the additional term  $\lambda \nu \|(H_1 \otimes H_1)u\|_1$ , where  $\nu \geq 0$  is a very small constant and  $H_1$  is matrix defined in (2.42). This additional term penalizes diagonal differences so that checkerboard patterns are avoided. For more details see [SST08, Section 2.2]. To include this term directly into (2.47) we only need to substitute  $M$  by  $\widetilde{M} = (M^T, \nu(H_1 \otimes H_1)^T)^T$ .

Instead of minimizing (2.47) directly, the minimizer  $\hat{u}$  can also be determined via its dual formulation  $\hat{u} = f - \lambda M^T \hat{W}$ , where  $\hat{W} \in \mathbb{R}^{2N}$  is a minimizer of the quadratic functional with linear constraints

$$\operatorname{argmin}_{W \in \mathbb{R}^{2N}} \|f - \lambda M^T W\|_2^2 \quad \text{subject to} \quad \|W\|_{\infty} \leq 1. \quad (2.48)$$

There exist various numerical methods for solving either (2.47) or (2.48). One possibility is for example to minimize (2.48) by second-order cone programming (SOCP).

For this purpose, (2.48) has to be reformulated appropriately as

$$\min_{t \in \mathbb{R}, W \in \mathbb{R}^{2N}} \begin{pmatrix} t \\ W \end{pmatrix}^T \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{subject to} \quad \begin{pmatrix} f - \lambda M^T W \\ t \\ 1 \end{pmatrix} \in L_r^{N+2},$$

$$-W_i, W_i \leq 1, \quad i \in \{1, \dots, 2N\},$$

where  $L_r^{N+2}$  denotes the *rotated second-order cone* defined by

$$L_r^{N+2} := \{ (\bar{x}, x_{N+1}, x_{N+2})^T \in \mathbb{R}^{N+2} : \|\bar{x}\|_2^2 \leq 2x_{N+1}x_{N+2}, x_{N+2} \geq 0 \}.$$

Now, a software package like for example MOSEK [1] can be applied to solve this problem. For literature on SOCP we refer to [6, 147].

For solving (2.48) directly we can for example use forward-backward splitting (FBS) as, e.g., outlined in [188, Sec. 3, Algorithm 2]. Note that this leads to the same algorithm as the gradient descent reprojection algorithm proposed in [48]. Alternatively, also Nesterov's algorithm [151] or the fast iterative shrinkage thresholding algorithm (FISTA) of Beck and Teboulle [17, 18] can be applied. The primal-dual hybrid gradient method (PDHG) studied in [51, 77, 231] is also a possibility. All these algorithms have the benefit that they are simple to implement and no additional software packages are required.

## 2.5 A Related Anisotropic Diffusion Model

### 2.5.1 The Model

As outlined in Section 2.1 we can establish a relation between a variational method of the form (2.1) with regularizer (2.2) and a corresponding diffusion model if  $\phi$  fulfills the conditions (I) and (II) in Section 2.1, i.e., if  $\phi$  is a finite function in  $\mathcal{C}^2(\mathbb{R}^2 \times S)$ , where  $S$  is an open set with  $\bar{\Omega} \subset S$ , and if for all  $p \in \mathbb{R}^2$  and  $x \in \bar{\Omega}$  there exists a symmetric and positive definite matrix  $A(p, x) \in \mathbb{R}^{2,2}$  such that

$$\nabla_p \phi(p, x) = A(p, x) p \quad \text{with} \quad \nabla_p = (\partial_{p_1}, \partial_{p_2})^T. \quad (2.49)$$

In case of the minimization problem (2.46) we have

$$\phi(p, x) = \|V(x)p\|_1 = |\langle v_1(x), p \rangle| + |\langle v_2(x), p \rangle| \quad \text{with} \quad V = (v_1, v_2)^T = \begin{pmatrix} v_{1,1} & v_{1,2} \\ v_{2,1} & v_{2,2} \end{pmatrix}.$$

Since this function is not in  $\mathcal{C}^2(\mathbb{R}^2 \times \Omega)$ , we approximate it by

$$\phi_\varepsilon(p, x) = \sqrt{|\langle v_1(x), p \rangle|^2 + \varepsilon^2} + \sqrt{|\langle v_2(x), p \rangle|^2 + \varepsilon^2}$$

for a small constant  $\varepsilon > 0$ . If  $V$  can be extended to a tensor field  $\tilde{V} : S \rightarrow \mathbb{R}^{2,2}$  with  $\tilde{V} \in \mathcal{C}^2(S, \mathbb{R}^{2,2})$ , this provides a finite extension of  $\phi_\varepsilon$  so that  $\phi_\varepsilon \in \mathcal{C}^2(\mathbb{R}^2 \times S)$ . Moreover, it holds that

$$\begin{aligned} \nabla_p \phi_\varepsilon(p, x) &= \begin{pmatrix} \frac{v_{1,1}(x) \langle v_1(x), p \rangle}{\sqrt{|\langle v_1(x), p \rangle|^2 + \varepsilon^2}} + \frac{v_{2,1}(x) \langle v_2(x), p \rangle}{\sqrt{|\langle v_2(x), p \rangle|^2 + \varepsilon^2}} \\ \frac{v_{1,2}(x) \langle v_1(x), p \rangle}{\sqrt{|\langle v_1(x), p \rangle|^2 + \varepsilon^2}} + \frac{v_{2,2}(x) \langle v_2(x), p \rangle}{\sqrt{|\langle v_2(x), p \rangle|^2 + \varepsilon^2}} \end{pmatrix} \\ &= V(x)^T \begin{pmatrix} \frac{1}{\sqrt{|\langle v_1(x), p \rangle|^2 + \varepsilon^2}} & 0 \\ 0 & \frac{1}{\sqrt{|\langle v_2(x), p \rangle|^2 + \varepsilon^2}} \end{pmatrix} V(x) p. \end{aligned}$$

Setting

$$A(p, x) := V(x)^T \begin{pmatrix} \frac{1}{\sqrt{|\langle v_1(x), p \rangle|^2 + \varepsilon^2}} & 0 \\ 0 & \frac{1}{\sqrt{|\langle v_2(x), p \rangle|^2 + \varepsilon^2}} \end{pmatrix} V(x)$$

the function  $\phi_\varepsilon$  fulfills obviously condition (2.49). Moreover, for all  $p \in \mathbb{R}^2$  and  $x \in \bar{\Omega}$  there exists a matrix  $M(p, x)$  such that  $A(p, x) = M(p, x)^T M(p, x)$ . Thus,  $A(p, x)$  is symmetric and positive definite if  $v_1 \nparallel v_2$ .

Hence, if  $V$  fulfills the above smoothness conditions and  $v_1 \nparallel v_2$ , then  $\phi_\varepsilon$  has all the required properties and according to Section 2.1 the minimization problem (2.46) is related to the diffusion model

$$\begin{cases} \frac{\partial}{\partial t} u(t, x) = \operatorname{div} (D(\nabla u_\sigma)(t, x) \nabla u(t, x)) & \forall (t, x) \in (0, T) \times \Omega, \\ (D(\nabla u_\sigma)(t, x) \nabla u(t, x))^T n(x) = 0 & \forall (t, x) \in (0, T) \times \partial\Omega, \\ u(0, x) = f(x) & \forall x \in \bar{\Omega} \end{cases} \quad (2.50)$$

with  $\sigma = 0$  and

$$D(\nabla u_\sigma)(t, x) = V(x)^T \begin{pmatrix} \frac{1}{\sqrt{|\langle v_1(x), \nabla u_\sigma(t, x) \rangle|^2 + \varepsilon^2}} & 0 \\ 0 & \frac{1}{\sqrt{|\langle v_2(x), \nabla u_\sigma(t, x) \rangle|^2 + \varepsilon^2}} \end{pmatrix} V(x).$$

Setting  $g : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$  to be the TV-related diffusivity  $g(s) = 1/\sqrt{|s|^2 + \varepsilon^2}$ , see, e.g., [52], and defining  $(V(x)p)_i := \langle v_i, p \rangle$  for  $p \in \mathbb{R}^2$ ,  $i = 1, 2$ , the matrices  $V$  considered in Algorithm 2.4.5 correspond to the following diffusion tensors:

- **Rotated rectangular shapes** For estimated rotation angles  $\alpha$  we have

$$D(\nabla u_\sigma) := R(-\alpha)^T \begin{pmatrix} g((R(-\alpha)\nabla u_\sigma)_1) & 0 \\ 0 & g((R(-\alpha)\nabla u_\sigma)_2) \end{pmatrix} R(-\alpha) \quad (2.51)$$

$$\text{with } R(\alpha) = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}.$$

- **Sheared shapes** For estimated shear parameters  $\zeta$  we obtain

$$D(\nabla u_\sigma) := S(\zeta)^\top \begin{pmatrix} g(\partial_{x_1} u_\sigma) & 0 \\ 0 & g(\zeta \partial_{x_1} u_\sigma + \partial_{x_2} u_\sigma) \end{pmatrix} S(\zeta) \quad (2.52)$$

$$\text{with } S(\zeta) = \begin{pmatrix} 1 & 0 \\ \zeta & 1 \end{pmatrix}.$$

- **Arbitrary double orientations** For estimated orientations  $r_1, r_2$  by one of the double orientation tensors it holds that

$$D(\nabla u_\sigma) := V^\top \begin{pmatrix} g(\langle r_1, \nabla u_\sigma \rangle) & 0 \\ 0 & g(\langle r_2, \nabla u_\sigma \rangle) \end{pmatrix} V \quad (2.53)$$

$$\text{with } V = (r_1, r_2)^\top.$$

Note that the TV-related diffusivity  $g$  is a 'boundary case' between forward and backward diffusion, see [65, p. 57], [213]. The diffusion model (2.50) with the new diffusion tensors can of course also be applied for other diffusivity functions  $g$ . If we use for example the Perona-Malik diffusivity

$$g(s) = \frac{1}{1 + \frac{|s|^2}{\gamma^2}}, \quad \gamma > 0 \quad (2.54)$$

proposed in [159], the diffusion model (2.50) is related to a variational approach with a non-convex regularization term. Since in such cases, (2.50) maybe ill-posed for  $\sigma = 0$ , cf. [226], we propose to choose  $\sigma > 0$  as in [46] so that the resulting diffusion model is again well-posed. As done for EED or coherence-enhancing diffusion [214] it is also possible to estimate the orientation information for the diffusion tensors directly from  $u$  rather than  $f$ . With these additional modifications the deduced anisotropic diffusion model can be summarized as follows:

**Algorithm 2.5.1** For an initial image  $f : \Omega \rightarrow \mathbb{R}$ , a diffusivity function  $g : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ ,  $\sigma \geq 0$  and a stopping time  $\hat{t} \in (0, T)$  proceed as follows:

Compute the diffused image  $u(\hat{t}, \cdot)$  of the anisotropic diffusion model (2.50) using one of the diffusion tensors (2.51)–(2.53) defined above. Here, the involved orientation estimates  $\alpha, \zeta$  and  $r_1, r_2$  are computed with respect to the current image  $u(t, \cdot)$  following the appropriate procedure described in Subsections 2.3.2/2.3.3.

### 2.5.2 Discretization

As before, our numerical implementation is based on a finite difference approach, where  $u \in \mathbb{R}^N$ ,  $N = mn$ , denotes again the columnwise reshaped image vector corresponding to an image  $U \in \mathbb{R}^{m \times n}$ . To discretize the diffusion model (2.50) in time an explicit time discretization is applied. For the partial spatial derivatives we use central differences with an additional smoothing filter of the form  $\frac{1}{16} \begin{pmatrix} 3 & 10 & 3 \end{pmatrix}$  orthogonal to the derivative direction as suggested by Weickert and Schar in [181, 215]. Consequently, we end up with an iterative scheme of the form

$$\begin{aligned} u^{(k+1)} &= u^{(k)} + \tau Q(u^{(k)}) u^{(k)}, \\ u^{(0)} &= f, \end{aligned} \tag{2.55}$$

where  $\tau > 0$  represents a fixed time step size and  $Q(u^{(k)}) \in \mathbb{R}^{N \times N}$  the iteration matrix corresponding to the spatial derivatives. With

$$D(\nabla u_\sigma(i, j)) = \begin{pmatrix} a_{i,j} & b_{i,j} \\ b_{i,j} & c_{i,j} \end{pmatrix}$$

denoting the diffusion tensor at pixel  $(i, j)$ , all non-vanishing entries of the  $(i + nj)^{\text{th}}$  row of  $Q(u^{(k)})$  are represented by the stencil displayed in Figure 2.9. At the boundaries of  $U$  we apply reflecting boundary conditions and set

$$\begin{aligned} a_{i,0} &= a_{i,n-1} = b_{i,0} = b_{i,n-1} = 0 & \forall i \in \{0, \dots, n-1\} \\ b_{0,j} &= b_{n-1,j} = c_{0,j} = c_{n-1,j} = 0 & \forall j \in \{0, \dots, n-1\} \end{aligned}$$

ensuring that  $(D(\nabla u_\sigma(i, j)) \nabla u(i, j))^T n(i, j) = 0$  for all boundary pixels.

An alternative to this scheme is for example the nonnegativity discretization of Weickert [213]. However, this discretization is less accurate and lacks to be rotationally invariant as outlined in [181, 215].

Since the central value of the stencil in Figure 2.9 is the negative sum of all the other values, the sum over all entries of the stencil is zero. Due to the reflecting boundary conditions this guarantees that the row sums of  $Q(u^{(k)})$  are zero. Since Figure 2.9 implies further that  $Q(u^{(k)})$  is symmetric, the column sums of  $Q(u^{(k)})$  are zero, too. Hence, the iterative scheme (2.55) preserves the average gray value, i.e.,

$$\frac{1}{N} \sum_{i=1}^N f_i = \frac{1}{N} \sum_{i=1}^N u_i^{(k)} \quad \forall k \in \mathbb{N}.$$

Since there may appear however negative matrix entries in  $I + \tau Q(u^{(k)})$ , the iterates do not in general fulfill a min-max principle like the nonnegativity discretization with small enough time step size, cf. [181, 213, 215].

$9a_{i-1,j-1}$ $+18b_{i-1,j-1}$ $+9c_{i-1,j-1}$	$30b_{i-1,j-1} + 30b_{i-1,j}$ $+30c_{i-1,j-1} + 30c_{i-1,j}$	$-9a_{i-1,j-1} - 9a_{i-1,j+1}$ $+9c_{i-1,j-1} + 9c_{i-1,j+1}$ $+100c_{i-1,j}$	$-30b_{i-1,j} - 30b_{i-1,j+1}$ $+30c_{i-1,j} + 30c_{i-1,j+1}$	$9a_{i-1,j+1}$ $-18b_{i-1,j+1}$ $+9c_{i-1,j+1}$
$30a_{i-1,j-1} + 30a_{i,j-1}$ $+30b_{i-1,j-1} + 30b_{i,j-1}$	$100b_{i-1,j} + 100b_{i,j-1}$	$-30a_{i-1,j-1} - 30a_{i-1,j+1}$ $-30a_{i,j-1} - 30a_{i,j+1}$ $-30b_{i-1,j-1} + 30b_{i-1,j+1}$ $+30b_{i,j-1} - 30b_{i,j+1}$	$-100b_{i-1,j} - 100b_{i,j+1}$	$30a_{i-1,j+1} + 30a_{i,j+1}$ $-30b_{i-1,j+1} - 30b_{i,j+1}$
$9a_{i-1,j-1} + 9a_{i+1,j-1}$ $+100a_{i,j-1}$ $-9c_{i-1,j-1} - 9c_{i+1,j-1}$	$-30b_{i-1,j-1} + 30b_{i-1,j}$ $+30b_{i+1,j-1} - 30b_{i+1,j}$ $-30c_{i-1,j-1} - 30c_{i-1,j}$ $-30c_{i+1,j-1} - 30c_{i+1,j}$	$-9a_{i-1,j-1} - 9a_{i-1,j+1}$ $-100a_{i,j-1} - 100a_{i,j+1}$ $-9a_{i+1,j-1} - 9a_{i+1,j+1}$ $-18b_{i-1,j-1} + 18b_{i-1,j+1}$ $+18b_{i+1,j-1} - 18b_{i+1,j+1}$ $-9c_{i-1,j-1} - 9c_{i-1,j+1}$ $-100c_{i-1,j} - 100c_{i+1,j}$ $-9c_{i+1,j-1} - 9c_{i+1,j+1}$	$-30b_{i-1,j} + 30b_{i-1,j+1}$ $+30b_{i+1,j} - 30b_{i+1,j+1}$ $-30c_{i-1,j} - 30c_{i-1,j+1}$ $-30c_{i+1,j} - 30c_{i+1,j+1}$	$9a_{i-1,j+1} + 9a_{i+1,j+1}$ $+100a_{i,j+1}$ $-9c_{i-1,j+1} - 9c_{i+1,j+1}$
$30a_{i,j-1} + 30a_{i+1,j-1}$ $-30b_{i,j-1} - 30b_{i+1,j-1}$	$-100b_{i,j-1} - 100b_{i+1,j}$	$-30a_{i,j-1} - 30a_{i,j+1}$ $-30a_{i+1,j-1} - 30a_{i+1,j+1}$ $-30b_{i,j-1} + 30b_{i,j+1}$ $+30b_{i+1,j-1} - 30b_{i+1,j+1}$	$100b_{i,j+1} + 100b_{i+1,j}$	$30a_{i,j+1} + 30a_{i+1,j+1}$ $+30b_{i,j+1} + 30b_{i+1,j+1}$
$9a_{i+1,j-1}$ $-18b_{i+1,j-1}$ $+9c_{i+1,j-1}$	$-30b_{i+1,j-1} - 30b_{i+1,j}$ $+30c_{i+1,j-1} + 30c_{i+1,j}$	$-9a_{i+1,j-1} - 9a_{i+1,j+1}$ $+9c_{i+1,j-1} + 9c_{i+1,j+1}$ $+100c_{i+1,j}$	$30b_{i+1,j} + 30b_{i+1,j+1}$ $+30c_{i+1,j} + 30c_{i+1,j+1}$	$9a_{i+1,j+1}$ $+18b_{i+1,j+1}$ $+9c_{i+1,j+1}$

Figure 2.9: Discretization stencil (multiplied by  $32^2$ ) of  $\text{div}(D(\nabla u_\sigma)\nabla u)$  at location  $(i, j)$  as suggested in [181, 215]. The entries of this stencil are the reshaped, non-vanishing entries of the  $(i + nj)^{\text{th}}$  row of  $Q(u^{(k)})$ . The  $(i + nj)^{\text{th}}$  entry of the vector  $Q(u^{(k)}) u^{(k)}$  can be obtained by summing up the values of the componentwise product of this stencil with the  $5 \times 5$  image patch of  $U$  centered at  $U(i, j)$ .

To avoid possible checkerboard effects, which may appear in the presence of strong noise, the scheme (2.55) can be extended by the additional term  $\nu(\tilde{I} - I)u^{(k)}$  with a very small parameter  $\nu \geq 0$ . The low pass filter matrix  $\tilde{I}$  is set to be a 5-band Toeplitz matrix with band  $\frac{1}{16}(-1, 4, 10, 4, -1)$ , here. Since this is a discretization of the identity filter of consistency order 4, this additional term has no influence on the consistency order of the original scheme, see [181, Section 9.5] for more details. Moreover, it leads again to a symmetric iteration matrix with a row sum of zero. Consequently, the resulting scheme preserves the average gray value, too.

## 2.6 Anisotropic Infimal Convolution Regularization

### 2.6.1 Infimal Convolution Regularization and Shape Preservation

Before we proceed with numerical examples, we want to consider an alternative regularizer to (2.39) given by

$$J(u) = (J_1 \square J_2)(u) := \inf_{\substack{u_1, u_2 \\ u = u_1 + u_2}} \{J_1(u_1) + J_2(u_2)\}. \quad (2.56)$$

The operator  $\square$  denotes here the so-called *infimal convolution* well-known in the field of convex analysis. Possible choices for  $J_1, J_2$  which are closely related to (2.39) are, e.g.,  $\int_{\Omega} |\partial_{x_1} u_1| dx$  and  $\int_{\Omega} |\partial_{x_2} u_2| dx$  as first considered in [50]. With these functionals the already discretized functional  $J$  reads as

$$J(u) = \inf_{\substack{u_1, u_2 \\ u = u_1 + u_2}} \{\|D_{x_1} u_1\|_1 + \|D_{x_2} u_2\|_1\} \quad (2.57)$$

with  $D_{x_i} \in \mathbb{R}^{N,N}$  representing again a discrete derivative operator in direction  $x_i$ ,  $i = 1, 2$ .

In general, assuming that the functions  $J_i : \mathbb{R}^N \rightarrow \mathbb{R} \cup \{+\infty\}$ ,  $i = 1, 2$ , are proper, convex and l.s.c. does not necessarily imply that  $J = J_1 \square J_2$  has these properties. Setting, e.g.,  $N = 1$ ,  $J_1(u) = u$  and  $J_2(u) = c$  with some constant  $c \in \mathbb{R}$ , it follows that  $J(u) = -\infty$ . Hence,  $J$  has not even to be proper if  $J_1$  and  $J_2$  are proper. To show that (2.57) has these properties nevertheless, we define the so-called *recession function* of a proper, convex, l.s.c. function  $\Psi : \mathbb{R}^N \rightarrow \mathbb{R} \cup \{+\infty\}$  by

$$(\Psi 0^+)(v) := \lim_{\lambda \rightarrow \infty} \frac{\Psi(u + \lambda v) - \Psi(u)}{\lambda}, \quad u \in \text{dom } \Psi$$

with  $\text{dom } \Psi := \{u \in \mathbb{R}^N : \Psi(u) < +\infty\}$  denoting the *domain* of  $\Psi$ . Here,  $\Psi 0^+$  is well-defined, i.e.,  $(\Psi 0^+)(v)$  is independent of the choice of  $u$ . With these definitions we can now state the following proposition, cf. [SST11]:

**Proposition 2.6.1** *Let  $J : \mathbb{R}^N \rightarrow \mathbb{R} \cup \{+\infty\}$  be the infimal convolution of functions  $J_i : \mathbb{R}^N \rightarrow \mathbb{R} \cup \{+\infty\}$ ,  $i = 1, \dots, m$ , i.e.,*

$$J(u) = (J_1 \square \dots \square J_m)(u) = \inf_{\substack{u_1, \dots, u_m \\ u = u_1 + \dots + u_m}} \sum_{i=1}^m J_i(u_i) \quad \forall u \in \mathbb{R}^N.$$

*Then,  $J$  has the following properties:*

- i)  $J$  is positively homogeneous, i.e.,  $J(\lambda u) = \lambda J(u)$  for all  $\lambda > 0$ ,  $u \in \mathbb{R}^N$ , if all  $J_i$  for  $i = 1, \dots, m$  are positively homogeneous.
- ii) If all  $J_i$  are proper and convex for  $i = 1, \dots, m$ , then  $J$  is convex.
- iii) If  $J_i$  is proper, convex, l.s.c. for all  $i = 1, \dots, m$  and

$$(J_1 0^+)(u_1) + \dots + (J_m 0^+)(u_m) \leq 0, \quad (2.58)$$

$$(J_1 0^+)(-u_1) + \dots + (J_m 0^+)(-u_m) > 0 \quad (2.59)$$



imply that  $u_1 + \dots + u_m \neq 0$ , then  $J$  is proper, convex and l.s.c. and the infimum in the definition of  $J(u)$  is attained for any  $u \in \mathbb{R}^N$ . In particular, the above implication holds if  $J_i(u) = J_i(-u)$  for all  $u \in \mathbb{R}^N$ .

iv) If  $J_i(u) := \|L_i u\|$  with  $L_i \in \mathbb{R}^{N_i, N}$ ,  $i = 1, \dots, m$ , and some norm  $\|\cdot\|$  in  $\mathbb{R}^{N_i}$ , then  $J$  is continuous.

**Proof:** If the functions  $J_i$ ,  $i = 1, \dots, m$ , are positively homogeneous, it follows straightforwardly that

$$J(\lambda u) = \inf_{\substack{u'_1, \dots, u'_m \\ \lambda u = \lambda u'_1 + \dots + \lambda u'_m}} \sum_{i=1}^m J_i(\lambda u'_i) = \lambda \inf_{\substack{u'_1, \dots, u'_m \\ u = u'_1 + \dots + u'_m}} \sum_{i=1}^m J_i(u'_i) = \lambda J(u)$$

for all  $\lambda > 0$ ,  $u \in \mathbb{R}^N$  and thus,  $J$  is positively homogeneous.

For the proof of ii) we refer to [173, p. 33] and also the first part of iii) can be found in [173, p. 76]. The last part of iii) is clear, since  $J_i(u) = J_i(-u)$  implies  $u \in \text{dom } J_i \Leftrightarrow -u \in \text{dom } J_i$  and we have

$$\begin{aligned} (J_i 0^+)(-v) &= \lim_{\lambda \rightarrow \infty} \frac{J_i(u - \lambda v) - J_i(u)}{\lambda} = \lim_{\lambda \rightarrow \infty} \frac{J_i(-u - \lambda v) - J_i(-u)}{\lambda} \\ &= \lim_{\lambda \rightarrow \infty} \frac{J_i(u + \lambda v) - J_i(u)}{\lambda} = (J_i 0^+)(v). \end{aligned}$$

Thus, in this case there exist no  $u_1, \dots, u_m \in \mathbb{R}^N$  for which (2.58) and (2.59) are fulfilled simultaneously.

Finally, to prove iv) we consider

$$J(u+h) = \inf_{\substack{u_1, \dots, u_m \\ u+h = u_1 + \dots + u_m}} \sum_{i=1}^m \|L_i u_i\| = \inf_{u_1, \dots, u_{m-1}} \left\{ \sum_{i=1}^{m-1} \|L_i u_i\| + \|L_m(u+h - \sum_{i=1}^{m-1} u_i)\| \right\}.$$

Since

$$\|L_m(u - \sum_{i=1}^{m-1} u_i)\| - \|L_m h\| \leq \|L_m(u+h - \sum_{i=1}^{m-1} u_i)\| \leq \|L_m(u - \sum_{i=1}^{m-1} u_i)\| + \|L_m h\|$$

we conclude that

$$J(u) - \|L_m h\| \leq J(u+h) \leq J(u) + \|L_m h\|.$$

This implies that  $|J(u+h) - J(u)| \rightarrow 0$  if  $\|h\| \rightarrow 0$ , which finishes the proof.  $\square$

More properties of the infimal convolution operator can be found in [16, 173, 201]. For example in [16, Prop. 15.7] further conditions for  $J$  to be proper, convex and l.s.c. are provided.

Since the functions  $J_i(u) = \|D_{x_i}u\|_1$ ,  $i = 1, 2$ , are obviously proper, convex, positively homogeneous and l.s.c. with  $J_i(u) = J_i(-u)$ , Proposition 2.6.1 implies immediately that their infimal convolution  $J = J_1 \square J_2$  in (2.57) is also proper, convex, positively homogeneous, l.s.c. and even continuous.

Next, we will use this proposition to show that any minimization problem of the form

$$\operatorname{argmin}_{u \in \mathbb{R}^N} \left\{ \frac{1}{2} \|f - u\|_2^2 + \lambda \underbrace{(J_1 \square \dots \square J_m)}_J(u) \right\} \quad \text{with} \quad J_i(u) = \|L_i u\|_1 \quad (2.60)$$

has a unique solution  $\hat{u}$  and  $\hat{u}$  can be characterized by similar conditions as in Proposition 2.4.3:

**Theorem 2.6.2** *For any matrices  $L_i \in \mathbb{R}^{N_i, N}$ ,  $i = 1, \dots, m$ , the minimization problem (2.60) has a unique solution. This solution is given by  $\hat{u} \in \mathbb{R}^N$  if and only if conditions i) - iii) of Theorem 2.4.1 hold true, where  $L_2(\Omega)$  has to be replaced by  $\mathbb{R}^N$  with the Euclidean inner product. In addition, we have that  $\hat{v} \in C_J$  if and only if there exists a vector  $\hat{V} = ((\hat{V}^{(1)})^T, \dots, (\hat{V}^{(m)})^T)^T$  with  $\hat{V}^{(i)} \in \mathbb{R}^{N_i}$ ,  $i = 1, \dots, m$ , such that*

$$\hat{v} := L_1^T \hat{V}^{(1)} = \dots = L_m^T \hat{V}^{(m)} \quad \text{and} \quad \|\hat{V}\|_\infty \leq 1.$$

**Proof:** Since the functions  $J_i(u) = \|L_i u\|_1$ ,  $L_i \in \mathbb{R}^{N_i, N}$ ,  $i = 1, \dots, m$ , are again proper, convex, positively homogeneous and l.s.c. with  $J_i(u) = J_i(-u)$ , Proposition 2.6.1 implies immediately that their infimal convolution  $J = J_1 \square \dots \square J_m$  is also proper, convex, positively homogeneous and l.s.c. Besides, it is easy to see that  $J$  is nonnegative, since all  $J_i$  are nonnegative. Thus, the first part of the assertion follows immediately by Proposition 2.4.3.

To prove the second part recall that for any proper, convex, l.s.c. and positively homogeneous function  $\psi : \mathbb{R}^N \rightarrow \mathbb{R} \cup \{+\infty\}$  we have that  $\psi = \sigma_{C_\psi} := \sup_{v \in C_\psi} \langle \cdot, v \rangle$ , i.e.,  $\psi$  is the *support function* of the nonempty, closed, convex set

$$C_\psi = \{v \in \mathbb{R}^N : \langle v, w \rangle \leq \psi(w) \ \forall w \in \mathbb{R}^N\},$$

see, e.g., [173, Cor. 13.2.1]. By [173, Thm. 13.2] the conjugate function of  $\psi$  is further the indicator function  $\psi^* = \iota_{C_\psi}$ . Hence, it holds that  $J = \sigma_{C_J}$  and  $J_i = \sigma_{C_{J_i}}$  with  $C_{J_i} := \{v = L_i^T V^{(i)} : \|V^{(i)}\|_\infty \leq 1\}$ ,  $i = 1, \dots, m$ , according to Proposition 2.4.3. Since we know by [173, Thm. 16.4] that  $(J_1 \square \dots \square J_m)^* = J_1^* + \dots + J_m^*$ , we obtain further

$$\iota_{C_J} = J^* = J_1^* + \dots + J_m^* = \iota_{C_{J_1}} + \dots + \iota_{C_{J_m}} = \iota_{C_{J_1} \cap \dots \cap C_{J_m}},$$

which finally implies that

$$C_J = \bigcap_{i=1}^m C_{J_i} = \{v = L_i^T V^{(i)} : \|V^{(i)}\|_\infty \leq 1, \ i = 1, \dots, m\}.$$

□

By this theorem we know in particular that the infimal convolution approach (2.60) with  $L_i = D_{x_i}$ ,  $i = 1, 2$ , has a unique solution. Similarly to Corollary 2.4.4 ii) it can now be shown that this minimization problem preserves + junctions like the one depicted in Figure 2.7 for appropriately chosen discrete derivative operators  $D_{x_1}, D_{x_2}$ . Interestingly, the observed loss of contrast in the result  $\hat{u}$  is significantly reduced compared to the result deduced in Corollary 2.4.4 ii):

**Corollary 2.6.3** *Let  $x_{1,0}, x_{2,0} \geq 0$  and  $x_{1,0}+a, x_{2,0}+b \leq n-2$ . For  $R_1 := \{x_{1,0}+1, \dots, x_{1,0}+a\} \times \{0, \dots, n-1\}$  and  $R_2 := \{0, \dots, n-1\} \times \{x_{2,0}+1, \dots, x_{2,0}+b\}$  set  $f := c_1 1_{R_1} + c_2 1_{R_2}$  with  $c_1, c_2 > 0$  as in Corollary 2.4.4 ii). Then, for the first matrices  $D_{x_1}, D_{x_2}$  in (2.41) with  $H_1(n-1, n-1) = -1$  the solution of (2.60) with  $L_i = D_{x_i}$ ,  $i = 1, 2$ , is given by*

$$\hat{u} = \left( c_1 - \frac{n-b}{n^2-ab} \lambda \right) 1_{R_1} + \left( c_2 - \frac{n-a}{n^2-ab} \lambda \right) 1_{R_2} \quad \text{for } \lambda \leq \min \left\{ \frac{n^2-ab}{n-b} c_1, \frac{n^2-ab}{n-a} c_2 \right\}.$$

If  $D_{x_1}, D_{x_2}$  coincide with one of the choices in (2.41), the minimizer  $\hat{u}$  of (2.60) with  $L_i = D_{x_i}$ ,  $i = 1, 2$ , is given by  $\hat{u} = f$  for all  $\lambda \geq 0$ .

**Proof:** To show the first part of the assertion, assume that  $D_{x_1}, D_{x_2}$  are given as on the left of (2.41) with  $H_1(n-1, n-1) = -1$ . Besides, set

$$\hat{V}^{(1)}(x_1, x_2) := \begin{cases} -\frac{n-b}{n^2-ab}(x_1 - x_{1,0}) & \text{if } x_1 \in I_1^{(2)}, x_2 \in I_2^{(1)} \cup I_2^{(3)}, \\ -\frac{n-b}{n^2-ab} a & \text{if } x_1 \in I_1^{(3)}, x_2 \in I_2^{(1)} \cup I_2^{(3)}, \\ -\frac{n-a}{n^2-ab}(x_1 + 1) & \text{if } x_1 \in I_1^{(1)}, x_2 \in I_2^{(2)}, \\ -\frac{n-a}{n^2-ab}(x_1 + 1) - \frac{n-b}{n^2-ab}(x_1 - x_{1,0}) & \text{if } x_1 \in I_1^{(2)}, x_2 \in I_2^{(2)}, \\ -\frac{n-a}{n^2-ab}(x_1 + 1) - \frac{n-b}{n^2-ab} a & \text{if } x_1 \in I_1^{(3)}, x_2 \in I_2^{(2)}, \\ 0 & \text{otherwise,} \end{cases}$$

$$\hat{V}^{(2)}(x_1, x_2) := \begin{cases} -\frac{n-a}{n^2-ab}(x_2 - x_{2,0}) & \text{if } x_1 \in I_1^{(1)} \cup I_1^{(3)}, x_2 \in I_2^{(2)}, \\ -\frac{n-a}{n^2-ab} b & \text{if } x_1 \in I_1^{(1)} \cup I_1^{(3)}, x_2 \in I_2^{(3)}, \\ -\frac{n-b}{n^2-ab}(x_2 + 1) & \text{if } x_1 \in I_1^{(2)}, x_2 \in I_2^{(1)}, \\ -\frac{n-b}{n^2-ab}(x_2 + 1) - \frac{n-a}{n^2-ab}(x_2 - x_{2,0}) & \text{if } x_1 \in I_1^{(2)}, x_2 \in I_2^{(2)}, \\ -\frac{n-b}{n^2-ab}(x_2 + 1) - \frac{n-a}{n^2-ab} b & \text{if } x_1 \in I_1^{(2)}, x_2 \in I_2^{(3)}, \\ 0 & \text{otherwise,} \end{cases}$$

for  $I_1^{(1)} := \{0, \dots, x_{1,0}\}$ ,  $I_1^{(2)} := \{x_{1,0}+1, \dots, x_{1,0}+a\}$ ,  $I_1^{(3)} := \{x_{1,0}+a+1, \dots, n-1\}$ , and  $I_2^{(1)} := \{0, \dots, x_{2,0}\}$ ,  $I_2^{(2)} := \{x_{2,0}+1, \dots, x_{2,0}+b\}$ ,  $I_2^{(3)} := \{x_{2,0}+b+1, \dots, n-1\}$ . Then, straightforward computations yield  $\|\hat{V}^{(i)}\|_\infty \leq 1$ ,  $i = 1, 2$ , and

$$\hat{v} := D_{x_1}^T \hat{V}^{(1)} = D_{x_2}^T \hat{V}^{(2)} = \frac{n-b}{n^2-ab} 1_{R_1} + \frac{n-a}{n^2-ab} 1_{R_2}.$$

Hence,  $\hat{v} \in C_J$  and we have

$$\hat{u} = f - \lambda \hat{v} = (c_1 - \frac{n-b}{n^2-ab} \lambda) 1_{R_1} + (c_2 - \frac{n-a}{n^2-ab} \lambda) 1_{R_2} = d_1 1_{R_1} + d_2 1_{R_2}$$

with  $d_1 := c_1 - \frac{n-b}{n^2-ab} \lambda$ ,  $d_2 := c_2 - \frac{n-a}{n^2-ab} \lambda$  so that

$$\begin{aligned} \langle \hat{u}, \hat{v} \rangle &= (n-b)a \frac{n-b}{n^2-ab} d_1 + (n-a)b \frac{n-a}{n^2-ab} d_2 + ab \left( \frac{n-b}{n^2-ab} + \frac{n-a}{n^2-ab} \right) (d_1 + d_2) \\ &= \frac{a(n-b)^2 + ab(n-b) + ab(n-a)}{n^2-ab} d_1 + \frac{b(n-a)^2 + ab(n-b) + ab(n-a)}{n^2-ab} d_2 \\ &= ad_1 + bd_2. \end{aligned}$$

By the proof of Theorem 2.6.2 we know that

$$J(\hat{u}) = \sup_{v \in C_J} \langle \hat{u}, v \rangle \geq \langle \hat{u}, \hat{v} \rangle = ad_1 + bd_2.$$

On the other hand, for  $\tilde{u}_1 = d_2 1_{R_2}$ ,  $\tilde{u}_2 = d_1 1_{R_1}$  we have that  $\hat{u} = \tilde{u}_1 + \tilde{u}_2$  and

$$J(\hat{u}) = \inf_{\substack{u_1, u_2 \\ \hat{u} = u_1 + u_2}} \{ \|D_{x_1} u_1\|_1 + \|D_{x_2} u_2\|_1 \} \leq \|D_{x_1} \tilde{u}_1\|_1 + \|D_{x_2} \tilde{u}_2\|_1 = bd_2 + ad_1$$

as long as  $\lambda \leq \min\{\frac{n^2-ab}{n-b} c_1, \frac{n^2-ab}{n-a} c_2\}$ . This finally implies that

$$J(\hat{u}) = ad_1 + bd_2 = \langle \hat{u}, \hat{v} \rangle \quad \text{if } \lambda \leq \min\left\{ \frac{n^2-ab}{n-b} c_1, \frac{n^2-ab}{n-a} c_2 \right\}$$

so that all conditions of Theorem 2.6.2 are fulfilled and  $\hat{u}$  is the sought minimizer for  $\lambda \leq \min\left\{ \frac{n^2-ab}{n-b} c_1, \frac{n^2-ab}{n-a} c_2 \right\}$ .

To prove the second part of the assertion assume that  $D_{x_1}, D_{x_2}$  coincide with one of the choices in (2.41). Besides, set  $\hat{u} = f$ ,  $\tilde{u}_1 = c_2 1_{R_2}$  and  $\tilde{u}_2 = c_1 1_{R_1}$  so that  $\hat{u} = \tilde{u}_1 + \tilde{u}_2$ . Then,

$$0 \leq \min_{u \in \mathbb{R}^N} \left\{ \frac{1}{2} \|f - u\|_2^2 + \lambda J(u) \right\} \leq \frac{1}{2} \|f - \hat{u}\|_2^2 + \lambda (\|D_{x_1} \tilde{u}_1\|_1 + \|D_{x_2} \tilde{u}_2\|_1) = 0$$

and thus,  $\hat{u}$  is the sought minimizer for all  $\lambda \geq 0$ .  $\square$

**A new anisotropic infimal convolution approach** Similarly to Subsection 2.4.2 we now want to adapt the regularizer (2.57) so that arbitrary X junctions can be preserved. The idea here is to consider instead of (2.57) the discrete counterpart of the functional

$$J(u) = \inf_{\substack{u_1, u_2 \\ u = u_1 + u_2}} \left\{ \int_{\Omega} |r_1^T \nabla u_1| + |r_2^T \nabla u_2| \, dy \right\}, \quad (2.61)$$

where  $r_1, r_2 \in \mathbb{R}^2$ ,  $r_1 \nparallel r_2$ , are supposed to have the same orientations as the lines of the X junctions which shall be preserved. For  $r_1 = (1, 0)^T$ ,  $r_2 = (0, 1)^T$  we obviously have that (2.57) and the discretized version of (2.61) coincide. For differently oriented vectors  $r_1, r_2$  the regularizer (2.61) can be motivated similarly to the justification of (2.46) in Subsection 2.4.2 as we will see in the following: For this purpose, set  $V = (r_1, r_2)^T$ ,  $f_V = f(V^T \cdot)$  and  $u_V = u(V^T \cdot)$ . If we assume that  $u$  is an image of an X junction with lines parallel to  $r_1, r_2$ , then the coordinate transformation  $y = V^T x$  in  $u_V$  will align the lines of the junction with the coordinate axes so that  $u_V$  is an image of a + junction. Moreover, for any image  $u$  it holds that

$$\frac{1}{|\det V|} \left( \frac{1}{2} \|f - u\|_{L_2}^2 + \lambda \inf_{\substack{u_1, u_2 \\ u = u_1 + u_2}} \left\{ \int_{\mathbb{R}^2} |r_1^T \nabla u_1| + |r_2^T \nabla u_2| dy \right\} \right) \quad (2.62)$$

$$\begin{aligned} &= \frac{1}{2} \|f(V^T \cdot) - u(V^T \cdot)\|_{L_2}^2 + \lambda \inf_{\substack{u_1, u_2 \\ u = u_1 + u_2}} \left\{ \int_{\mathbb{R}^2} |r_1^T \nabla_y u_1(V^T x)| + |r_2^T \nabla_y u_2(V^T x)| dx \right\} \\ &= \frac{1}{2} \|f_V - u_V\|_{L_2}^2 + \lambda \inf_{\substack{u_{1,V}, u_{2,V} \\ u_V = u_{1,V} + u_{2,V}}} \left\{ \int_{\mathbb{R}^2} |r_1^T V^{-1} \nabla u_{1,V}| + |r_2^T V^{-1} \nabla u_{2,V}| dx \right\} \\ &= \frac{1}{2} \|f_V - u_V\|_{L_2}^2 + \lambda \inf_{\substack{u_{1,V}, u_{2,V} \\ u_V = u_{1,V} + u_{2,V}}} \left\{ \int_{\mathbb{R}^2} |(1, 0) \nabla u_{1,V}| + |(0, 1) \nabla u_{2,V}| dx \right\} \\ &= \frac{1}{2} \|f_V - u_V\|_{L_2}^2 + \lambda \inf_{\substack{u_{1,V}, u_{2,V} \\ u_V = u_{1,V} + u_{2,V}}} \left\{ \int_{\mathbb{R}^2} |\partial_{x_1} u_{1,V}| + |\partial_{x_2} u_{2,V}| dx \right\}. \end{aligned} \quad (2.63)$$

Hence,  $\hat{u}$  is the minimizer of (2.62) if and only if  $\hat{u}_V = \hat{u}(V^T \cdot)$  minimizes (2.63). Since we showed, at least in a discrete setting, that minimizing (2.63) preserves in general + junctions, the energy functional (2.62) is a good choice if we want to preserve X junctions with lines parallel to  $r_1, r_2$ .

To estimate appropriate vectors  $r_1, r_2$  for each image point, we can use the orientation tensors deduced from the transparent model in Subsection 2.3.3. However, in contrast to (2.46) interchanging the orientations vectors  $r_1(i, j)$  and  $r_2(i, j)$  for some pixel  $(i, j)$  influences in general the minimizer of (2.62). Therefore, we need to be careful which vectors to assign to  $r_1(i, j)$  or  $r_2(i, j)$ . For images  $f$  which are expected for noise mainly dominated by two superimposed one-dimensional patterns, we will group the orientations in such a way that the vector fields  $r_1, r_2$  become as smooth as possible. For more complicated structures more sophisticated approaches are required.

In any case, note that due to the infimum in (2.62) a general assumption of this model is that the sought minimizer  $\hat{u}$  can be decomposed into two images  $\hat{u}_1$  and  $\hat{u}_2$  so that each  $\hat{u}_i$  is almost constant in direction  $r_i(i, j)$  at pixel  $(i, j)$ . Hence, this assumption should hold for any image to be reconstructed.

### 2.6.2 Discretization and Minimization

The functional (2.62) leads for orientation vectors  $r_1 = (r_{1,1}, r_{1,2})^T$ ,  $r_2 = (r_{2,1}, r_{2,2})^T$  to the discretized minimization problem

$$\operatorname{argmin}_{u \in \mathbb{R}^N} \left\{ \frac{1}{2} \|f - u\|_2^2 + \lambda \inf_{\substack{u_1, u_2 \\ u = u_1 + u_2}} \left\{ \underbrace{\|V_1 D u_1\|_1}_{M_1} + \underbrace{\|V_2 D u_2\|_1}_{M_2} \right\} \right\}, \quad (2.64)$$

where  $D = (D_{x_1}^T, D_{x_2}^T)^T$  is again a discrete gradient operator and the matrices  $V_1, V_2$  are defined by  $V_i = (\operatorname{diag}(\operatorname{vec}(r_{i,1})), \operatorname{diag}(\operatorname{vec}(r_{i,2})))$  for  $i = 1, 2$ .

As in Subsection 2.4.3 we restrict our attention to quadratic images and set  $D$  to be second matrix in (2.41). Moreover, we add the term  $\nu \|(H_1 \otimes H_1)u\|_1$  with  $\nu \geq 0$  to each function  $J_i(u) = \|M_i u\|_1$  to suppress the possible appearance of checkerboard patterns. Again, this is the same as substituting each matrix  $M_i$  by  $\widetilde{M}_i = (M_i^T, \nu(H_1 \otimes H_1)^T)^T$ .

Instead of determining the minimizer  $\hat{u}$  of (2.64) directly, we can again solve the corresponding dual problem

$$\begin{aligned} \hat{v} = \operatorname{argmin}_{v \in \mathbb{R}^N} \frac{1}{2} \|f - \lambda v\|_2^2 \quad \text{subject to} \quad v = M_1^T W_1 = M_2^T W_2, \\ \|W_i\|_\infty \leq 1, \quad i = 1, 2 \end{aligned}$$

and set  $\hat{u} = f - \lambda \hat{v}$ . This problem can for example be solved by SOCP, cf. the explanations in Subsection 2.4.3. Alternatively, we can also rewrite the primal problem (2.64) in the form

$$\begin{aligned} \begin{pmatrix} \hat{u}_1 \\ \hat{u}_2 \end{pmatrix} &= \operatorname{argmin}_{u_1, u_2 \in \mathbb{R}^N} \left\{ \frac{1}{2} \|f - u_1 - u_2\|_2^2 + \lambda (\|M_1 u_1\|_1 + \|M_2 u_2\|_1) \right\} \\ &= \operatorname{argmin}_{u_1, u_2 \in \mathbb{R}^N} \left\{ \frac{1}{2} \|f - (I_N, I_N) \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}\|_2^2 + \lambda \left\| \begin{pmatrix} M_1 & 0_{N,N} \\ 0_{N,N} & M_2 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \right\|_1 \right\}, \quad (2.65) \end{aligned}$$

where  $0_{M,N}$  represents a zero matrix of size  $M \times N$  and  $\hat{u} = \hat{u}_1 + \hat{u}_2$  denotes the sought solution. Here, the vectors  $\hat{u}_1, \hat{u}_2$  are in general not unique, although  $\hat{u}$  has this property. To solve (2.65) we can for example apply the alternating direction method of multipliers (ADMM) [26, 76, 88, 93] or equivalently the alternating split Bregman algorithm [96]. A detailed description of this approach has been provided in [SST11].

## 2.7 Numerical Experiments

Finally, we want to demonstrate the good performance of the proposed models by numerical examples. For this purpose, all methods have been implemented in MATLAB and are applied to test images with gray values in the range  $[0, 255]$ . The parameters of the models are chosen with respect to the best visual results. Moreover, we use

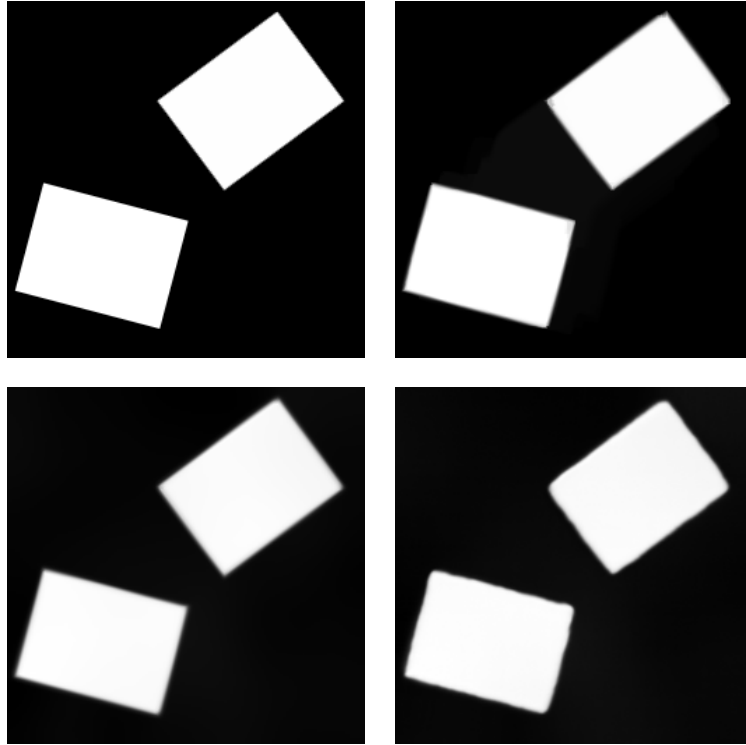


Figure 2.10: Denoising results for the image at Figure 2.1 (top middle). *Top left:* Original image. *Top right:* Denoised image using the variational approach proposed in Subsection 2.4.2 with  $\sigma = 2.5$ ,  $\rho = 20$ ,  $\lambda = 800$  and  $\nu = 0.1$ . *Bottom left:* Denoised image by anisotropic diffusion with diffusion tensors (2.51) adapted for rotated rectangles ( $\sigma = 2.5$ ,  $\rho = 20$ ) and Perona-Malik diffusivity (2.54) ( $\gamma = 1.5$ ). *Bottom right:* Denoised image after iterating two times the NL means filter ( $t = 30$ ,  $f = 10$ ,  $h = 4000$ ) available at [142]. The result is slightly worse than those of our methods.

again the MATLAB routine `imagesc` to visualize the images, which applies an affine gray value scaling setting the minimal image value to black and the maximal gray value to white.

### 2.7.1 Results for Images with Rotated or Sheared Shapes

To start with, we first concentrate on the restoration methods for images containing rotated, respectively sheared shapes in Subsections 2.4.2 and 2.5.1. To discretize the partial derivatives of the involved orientation tensors we apply again central differences with the additional smoothing filter  $\frac{1}{16} \begin{pmatrix} 3 & 10 & 3 \end{pmatrix}$  orthogonal to the derivative direction, cf. Subsection 2.5.2. If not stated otherwise, we use for all diffusion experiments the time step size  $\tau = 0.1$  and 4000 iterations, i.e., a diffusion time of  $\hat{t} = 400$ . Besides, we choose  $\nu = 0.001$  to avoid any checkerboard effects.

For our first example we denoised the initial test image in Figure 2.1 (top middle) with our new anisotropic methods designed for images with rotated shapes. Figure 2.10 (top right) shows the result of the anisotropic regularization approach proposed in Subsection 2.4.2. In contrast to the denoising results in Figure 2.1 the vertices of the rectangles are well preserved here. This is also the case if we use the anisotropic diffusion approach described in Algorithm 2.5.1 with the diffusion tensors (2.51) adapted for rotated rectangles, see Figure 2.10 (bottom left). For comparison we also applied iteratively the NL means filter [39] available at [142]. Since this method uses the information of similar patches in the neighborhood of each pixel, it shows at least after two iterations also a good performance in the presence of strong noise. However, the vertices are slightly more rounded.

For our next examples we applied the results of Section 2.3.2 to denoise images with sheared shapes. The functions  $\varphi_v$ ,  $\varphi_e$  are set to

$$\begin{aligned}\varphi_v(x) &:= e^{-(x/c)^4}, & c &:= 0.06 \max_{\substack{x \in \Omega, \\ \lambda_1(x) > 0}} \left\{ \left| \frac{\lambda_2(x)}{\lambda_1(x)} \right| |\nabla f_{\tilde{\sigma}}|^2(x) \right\}, \\ \varphi_e(x) &:= \sin\left(\frac{\pi}{2}x\right), & x &\in [0, 1]\end{aligned}$$

with  $\lambda_1$ ,  $\lambda_2$  being the eigenvalues of the classical structure tensor  $\mathcal{J}_\rho$  used in Section 2.3.2. As displayed in Figure 2.11 (top right) the structure tensor adapted for sheared shapes yields good estimates for the shear parameters of the noisy image at top left. Moreover, Figure 2.11 (bottom) shows that our new restoration methods preserve the vertices of the parallelograms very well.

Of course, these approaches can also be applied for more general shapes as long as all vertices have one horizontal edge. An example with such shapes is given in Figure 2.12. To test our method also on real-world data we included Figure 2.13. Here, the new variational approach performs much better than ROF, since due to the estimated shear parameters the shapes of the shadows are preserved much more accurately.

Regarding possible applications we applied our new angle estimation method also to an image with differently oriented toy cars. As visible in Figure 2.14 (right), the orientations of the cars are well estimated by our new approach and through the color coding of the angles equally oriented cars are grouped together. In connection with quality control this might for example be helpful to detect wrongly oriented objects very easily.

Another application of our methods for rotated rectangular shapes is presented in Figure 2.15. Here, we have extracted the cartoon of a real-world city area with the anisotropic regularization method of Subsection 2.4.2. The result presented in Figure 2.15 (top right) shows that the estimated rotation angles approximate well the orientations of the houses. For this reason, their shapes are well preserved in the cartoons while the details have been removed. If we now apply the Canny edge detector [45] implemented in the MATLAB `edge` routine to the original image as well



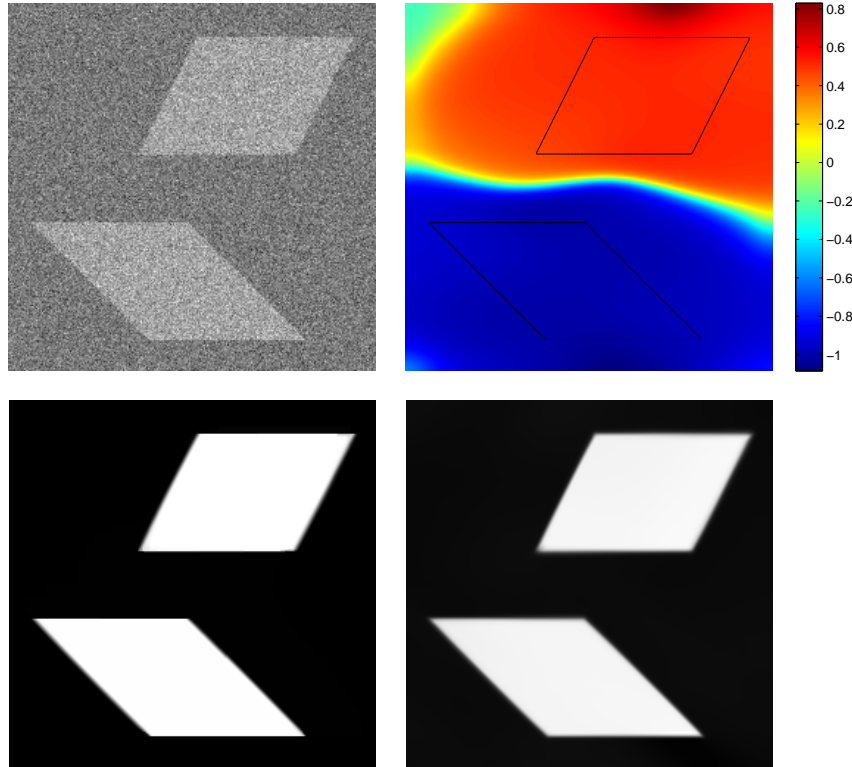


Figure 2.11: *Top left:* Noisy image of two sheared rectangles ( $\zeta_1 = 0.5$  and  $\zeta_2 = -1$ ) corrupted by additive Gaussian noise of standard deviation 150. *Top right:* Estimated shear parameters  $\zeta$  for  $\tilde{\sigma} = 5$ ,  $\rho = 2$ ,  $\rho_x = 35$  and  $\rho_y = 20$ . *Bottom left:* Denoised image using the regularization method proposed in Subsection 2.4.2 with the shear parameters  $\zeta$  depicted at top right,  $\lambda = 1000$  and  $\nu = 0.1$ . *Bottom right:* Denoised image by the anisotropic diffusion approach with diffusion tensors (2.52) adapted for sheared shapes ( $\sigma = 2.5$ ) and Perona-Malik diffusivity (2.54) ( $\gamma = 1.5$ ).

as to our simplified one, we see that the shapes of the houses are well extracted from the simplified image without the details contained in the original one. The obtained edge image can now be used, e.g., for building segmentation. A first overview on the vast literature on this topic can be found in [144]. To be able to provide an example we have implemented the windowed Hough transform for extracting rectangles presented in [53]. The results are depicted in Figures 2.16 and 2.17. By using the edge image of the cartoon in Figure 2.15 (bottom middle) instead of the original image we benefit from a lower risk of detecting false positives as well as a reduction in the time required for the segmentation.

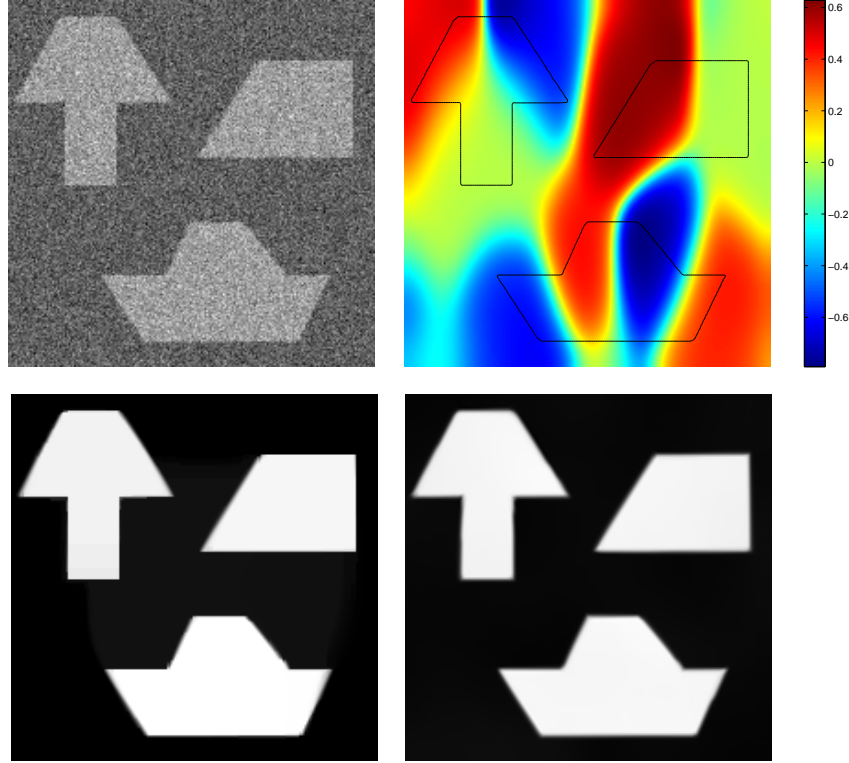


Figure 2.12: *Top left:* Noisy image corrupted by additive Gaussian noise of standard deviation 100. *Top right:* Estimated shear parameters  $\zeta$  for  $\tilde{\sigma} = 2$ ,  $\rho = 1$ ,  $\rho_x = 12$  and  $\rho_y = 25$ . *Bottom left:* Denoised image using the regularization method proposed in Subsection 2.4.2 with the shear parameters at top right,  $\lambda = 600$  and  $\nu = 0.07$ . *Bottom right:* Denoised image by anisotropic diffusion ( $\sigma = 1.5$ , 3000 iterations) with diffusion tensors (2.52) adapted for sheared shapes and Perona-Malik diffusivity (2.54) ( $\gamma = 1.5$ ).

### 2.7.2 Results for Images with Arbitrary Double Orientations

Next, we will focus on images with arbitrary double orientations. To discretize the derivatives occurring in the double orientation tensors examined in Subsection 2.3.3 we apply the optimized  $5 \times 5$  filters suggested by Scharr in [182].

For our first example in Figure 2.18 we used a noisy image with various shapes and restored it by ROF, nonlocal means and by the anisotropic regularization method proposed in Subsection 2.4.2 using orientations estimated by the double orientation tensor following the occlusion model. Similarly as in Figure 2.1 the result by ROF suffers again from rounding artifacts at corners, since to remove all noise the regularization parameter  $\lambda$  has been chosen rather large. A better result is obtained by the nonlocal means filter. Nevertheless, as visible at bottom right the new anisotropic regularization approach performs best due to the incorporated orientation information.

Figure 2.19 shows the performance of these methods also on a real world image of a

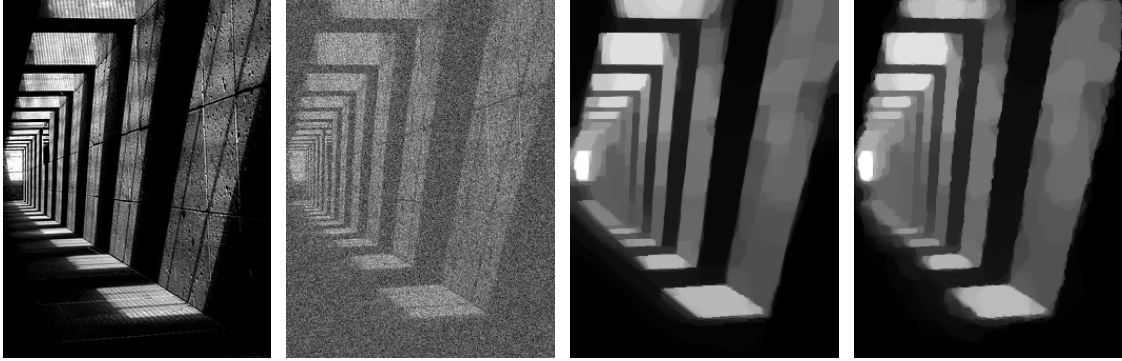


Figure 2.13: *Left to right*: Original image (copyright by [94]), noisy image corrupted by additive Gaussian noise of standard deviation 70, restored image by the anisotropic regularization method proposed for sheared shapes in Subsection 2.4.2 ( $\tilde{\sigma} = 6$ ,  $\rho = 1$ ,  $\rho_x = \rho_y = 20$ ,  $\lambda = 150$ ,  $\nu = 0.1$ ), and a denoised image by ROF ( $\lambda = 200$ ) for comparison.

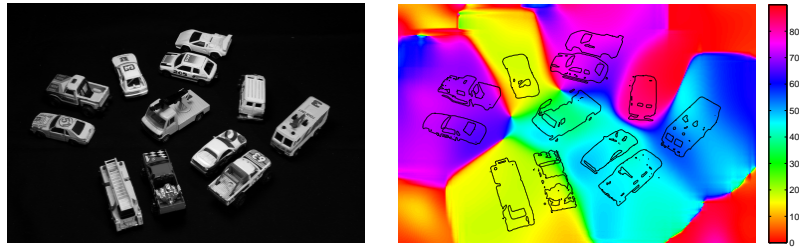


Figure 2.14: *Left*: Real image of differently oriented toy cars. *Right*: Estimated orientations by the new structure tensor method for rotated shapes ( $\sigma = 0.5$ ,  $\rho = 15$ ).

building. As expected, our new approach restores the shape of the building much better than ROF. Unfortunately, both methods suffer from staircasing effects, which are not present in nonlocal means result. On the other hand, due to the large smoothing parameter necessary for this noise level, nonlocal means creates small blur artifacts where our result shows sharp structures.

In our last example we finally want to point out the benefits of the anisotropic infimal convolution approach presented in Subsection 2.6.1. For this purpose, Figure 2.20 shows results for an oriented texture consisting of two superimposed one-dimensional patterns. This image can be well described by the transparent model so that it is nearby to estimate its orientations by the corresponding tensor presented in Subsection 2.3.3. As displayed at bottom right the anisotropic infimal convolution approach outperforms the regularization method proposed in Subsection 2.4.2 significantly for this image. The reason is that the results of the approach of Subsection 2.4.2 are either not smooth enough or for larger values  $\lambda$  too many details have been removed.

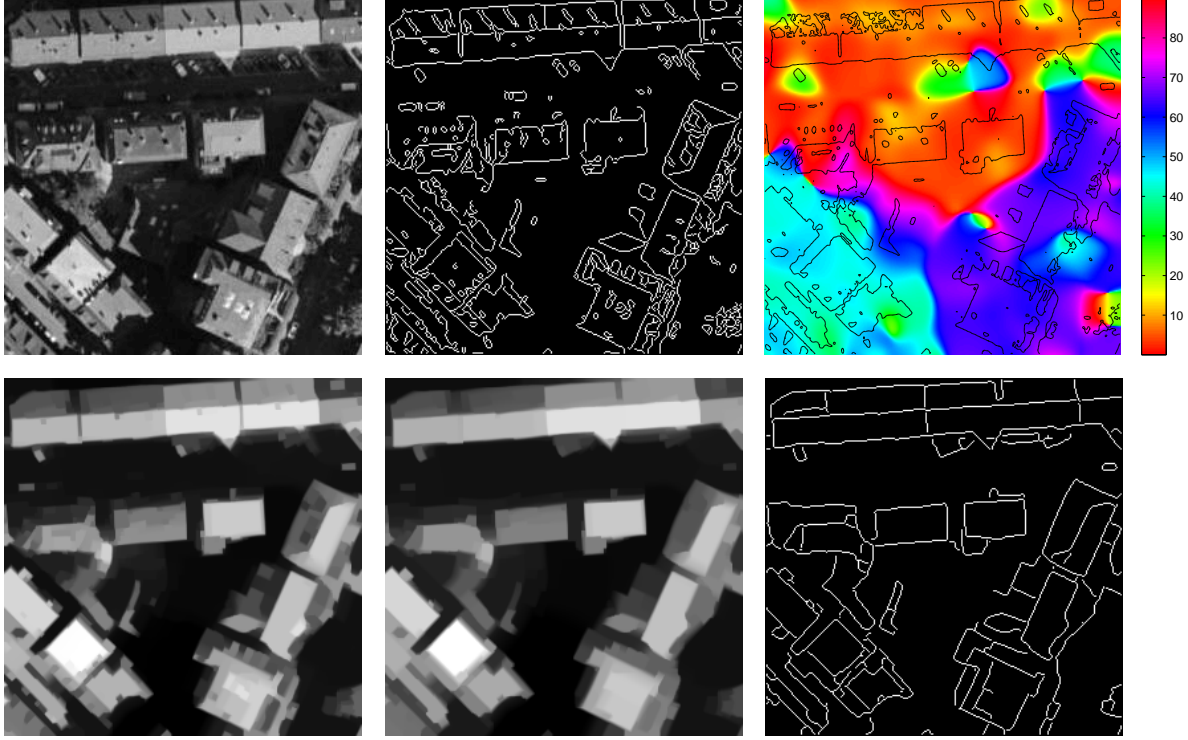


Figure 2.15: *Top left*: Original image of a city area of Stuttgart (Germany). *Top middle*: Edges detected by the Canny edge detector in the original image ( $\sigma = 0.5$ , thresh =  $[0.1, 0.3]$ ). *Top right*: Estimated rotation angles by the structure tensor adapted for rotated rectangles ( $\sigma = 0.5$ ,  $\rho = 10$ ). *Bottom left/middle*: Cartoons generated by anisotropic regularization using the rotation angles at top right,  $\nu = 0.1$  and  $\lambda = 50$  (left), resp.  $\lambda = 100$  (middle). *Bottom right*: Edges detected by the Canny edge detector in the cartoon at bottom middle ( $\sigma = 0.5$ , thresh =  $[0.03, 0.12]$ ).

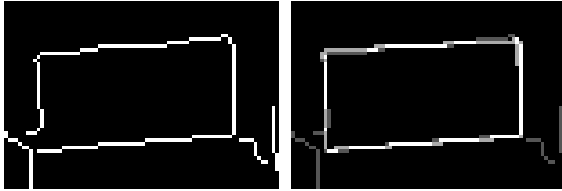


Figure 2.16: *Left*: Part of the detected edges in Figure 2.15 (bottom right). *Right*: Detected edges + segmented building (white line) by the windowed Hough transform presented in [53].

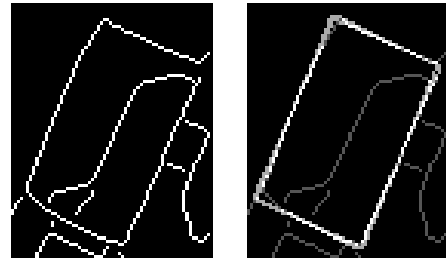


Figure 2.17: *Left*: Another part of the detected edges in Figure 2.15 (bottom right). *Right*: Detected edges + segmented building (white line) by the windowed Hough transform presented in [53].

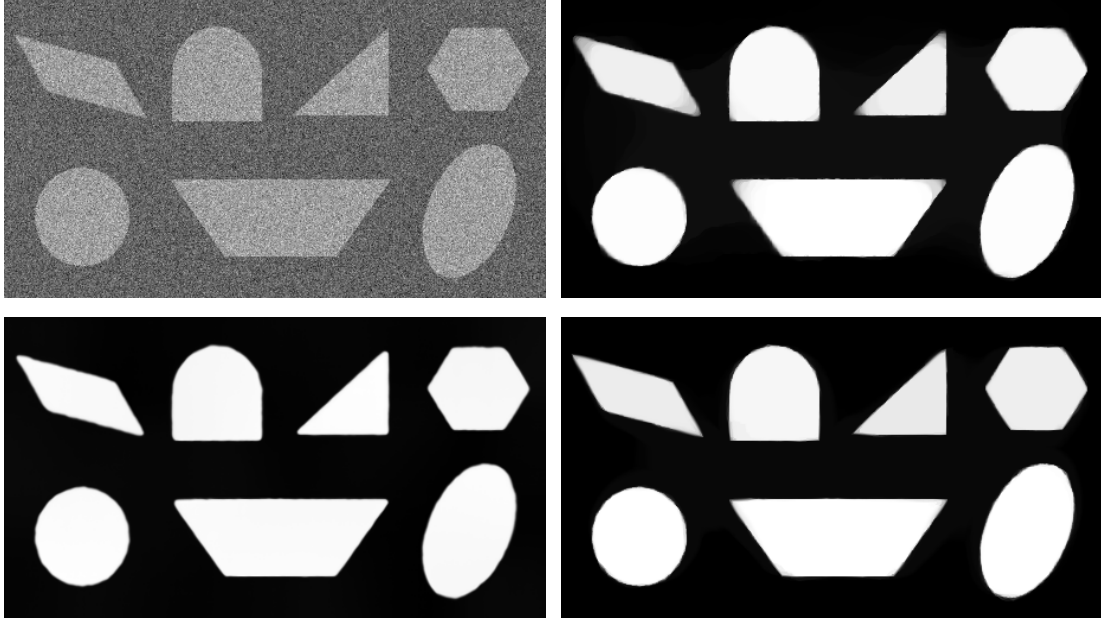


Figure 2.18: *Top*: Noisy image corrupted by additive Gaussian noise of standard deviation 100 and restored image by ROF ( $\lambda = 500$ ). *Bottom left*: Denoised image by iterating two times the nonlocal means filter [142] ( $t = 30$ ,  $f = 5$ ,  $h = 4000$ ). *Bottom right*: Restored image by the anisotropic regularization method proposed in Subsection 2.4.2 with double orientations estimated with respect to the occlusion model ( $\sigma = 2$ ,  $\rho = 6$ ,  $\lambda = 900$ ,  $\nu = 0.02$ ).

For comparison we have also implemented the diffusion method proposed by Scharr in [182]. Nevertheless, the obtained result is also slightly worse than the one by our infimal convolution approach.



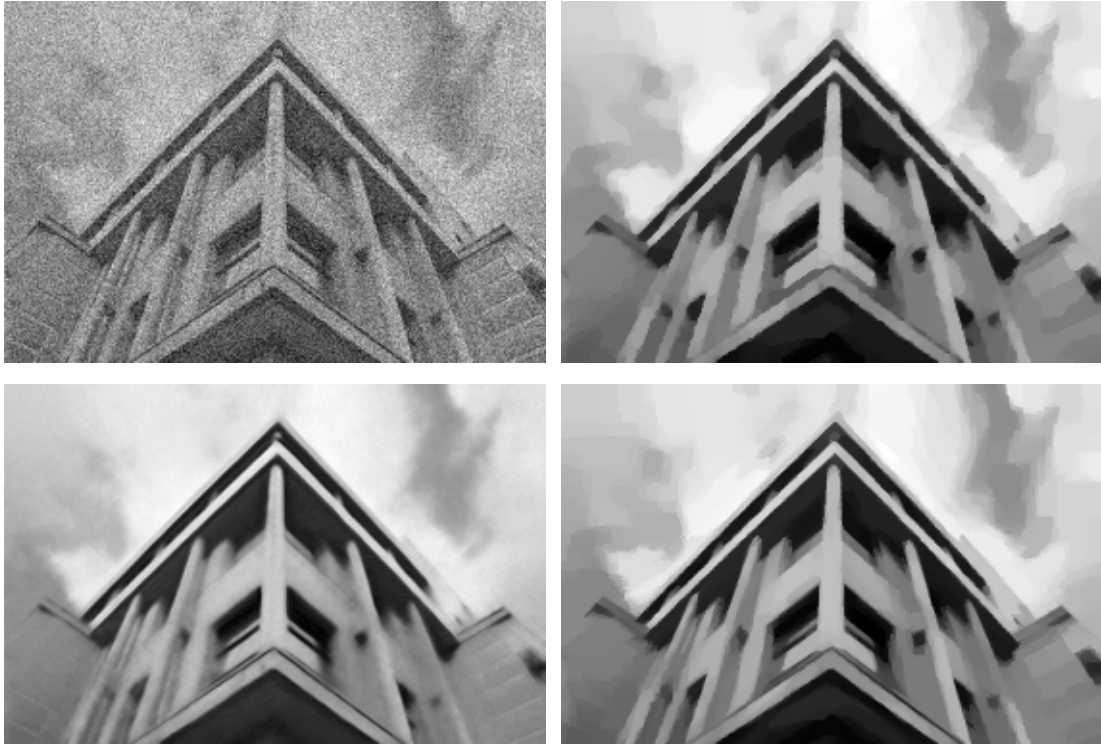


Figure 2.19: *Top*: Noisy image corrupted by additive Gaussian noise of standard deviation 30 and result by ROF ( $\lambda = 50$ ). *Bottom left*: Denoised image by the nonlocal means filter [142] ( $t = 21$ ,  $f = 7$ ,  $h = 1000$ ). *Bottom right*: Result by the anisotropic regularization method proposed in Subsection 2.4.2 with double orientations estimated with respect to the occlusion model ( $\sigma = 0.5$ ,  $\rho = 8$ ,  $\lambda = 50$ ,  $\nu = 0.1$ ).

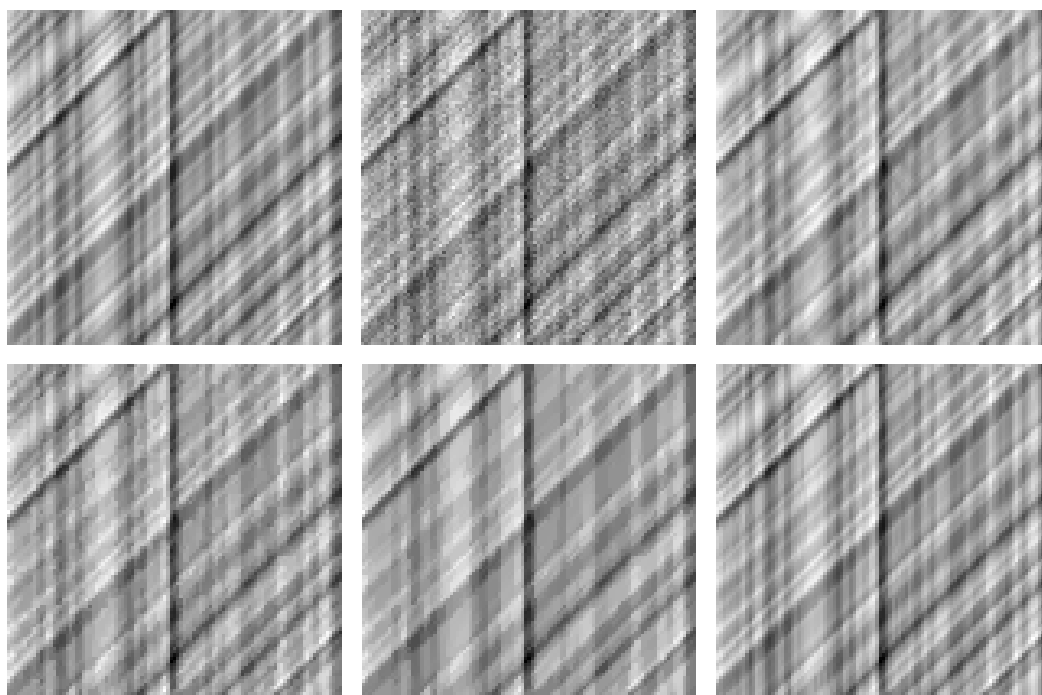


Figure 2.20: *Top, left to right:* Original image (cf. [182]), noisy image corrupted by additive Gaussian noise of standard deviation 10 and the result by the method of Schar proposed in [182]. *Bottom, left to right:* Denoised images by the anisotropic regularization method proposed in Subsection 2.4.2 with double orientations estimated with respect to the transparent model ( $\sigma = 2$ ,  $\rho = 12$ ,  $\nu = 0.15$ ,  $\lambda = 4$  (left), resp.  $\lambda = 10$  (middle)) and finally the result by the infimal convolution approach (2.64) ( $\sigma = 2$ ,  $\rho = 12$ ,  $\lambda = 40$ ,  $\nu = 0.03$ ).





# CHAPTER 3

## Variational Approaches Facing Poisson and Multiplicative Noise

### Outline

3.1	Notation and Preliminaries . . . . .	<b>62</b>
3.2	Poisson and Multiplicative Noise in Imaging . . . . .	<b>62</b>
3.2.1	Poisson Noise . . . . .	62
3.2.2	Multiplicative Noise . . . . .	64
3.3	Variational Approaches by Bayesian Modeling . . . . .	<b>66</b>
3.3.1	Variational Approaches for Poisson Noise . . . . .	66
3.3.2	Variational Models for Multiplicative Gamma Noise . . . . .	69
3.4	Equivalence of the SO-Model and IDIV-TV . . . . .	<b>75</b>
3.5	Theoretical Results for the Discrete Setting . . . . .	<b>81</b>
3.6	Minimization Algorithms . . . . .	<b>85</b>
3.6.1	ADMM and Multiple Splittings . . . . .	86
3.6.2	Minimization of the Discrete IDIV Model . . . . .	90
3.7	Numerical Experiments . . . . .	<b>94</b>
3.7.1	Blurred Images Contaminated by Poisson Noise . . . . .	95
3.7.2	Images Corrupted by Multiplicative Noise . . . . .	100

In real-word applications we are often confronted with noise which is neither additive nor Gaussian distributed. Typical examples are, e.g., Poisson noise or different types of multiplicative noise. In the following, we proceed by outlining the practical relevance of these types of noise in imaging. Afterwards, we study different variational approaches for removing Poisson and multiplicative Gamma noise and show equivalence between certain models. To finally compute the occurring minimizers an efficient algorithm based on the alternating direction method of multipliers is proposed and its very good performance is demonstrated by numerical examples.

### 3.1 Notation and Preliminaries

Before going into detail we fix some notation. First, all random variables occurring in this chapter are named with capital letters and are supposed to be real-valued and defined on a probability space  $(\Omega, \mathcal{F}, P)$ . As usual,  $\Omega$  denotes the sample space,  $\mathcal{F} \subset 2^\Omega$  is a  $\sigma$ -algebra over  $\Omega$  and  $P : \mathcal{F} \rightarrow [0, 1]$  represents the applied probability measure. Moreover, the function  $p_X$  stands either for the *probability density function* of the random variable  $X$  if  $X$  is continuous or the *probability mass function*  $p_X(k) := P(X = k)$  if  $X$  is discrete. Here,  $P(X = k)$  denotes the probability that  $X$  takes the value  $k$ .

For any  $x$  with  $p_X(x) > 0$ , the *conditional density* or respectively the *conditional mass function* of a random variable  $Y$  given  $X = x$  is defined by

$$p_{Y|X}(\cdot | x) := \frac{p_{Y,X}(\cdot, x)}{p_X(x)},$$

see, e.g., [101, p. 67/104]. For a continuous random variable  $Y$  it holds that

$$\int_{-\infty}^y p_{Y|X}(t | x) dt = \lim_{\varepsilon \rightarrow 0^+} P(Y \leq y | X \in (x - \varepsilon, x + \varepsilon]),$$

which shows the connections between a conditional density and the corresponding conditional probability. Various rules for the calculation with conditional densities are summarized in Appendix A, which will be used throughout this and the subsequent chapter.

### 3.2 Poisson and Multiplicative Noise in Imaging

After these preliminaries we now want to outline the practical relevance of Poisson and multiplicative noise in imaging.

#### 3.2.1 Poisson Noise

Typically, Poisson noise appears in all applications where images are formed by counting particles, usually photons, which arrive in the image domain, see, e.g., [22]. Here, the number of particles  $k_i \in \mathbb{N}$  observed during a fix time interval  $[0, T]$  at some position  $i$  can be viewed as a realization of a discrete random variable  $X_i$  with probability mass function

$$p_{X_i}(k_i) = P(X_i = k_i) = \frac{\lambda_i^{k_i} e^{-\lambda_i}}{k_i!}.$$

In short, we will write  $X_i \sim \text{Poisson}(\lambda_i)$ . Since the expectation value of  $X_i$  is  $\mathbb{E}(X_i) = \lambda_i$ , the parameter  $\lambda_i \geq 0$  stands for the expected number of occurrences

at location  $i$ , which is typically proportional to length of the time interval  $T$ . If the random variables  $X_i$  are pairwise independent for all locations  $i = 1, \dots, N$ , the joint probability mass function of  $X = (X_i)_{i=1, \dots, N}$  is given by

$$p_X(k_1, \dots, k_N) = \prod_{i=1}^N \frac{\lambda_i^{k_i} e^{-\lambda_i}}{k_i!}. \quad (3.1)$$

Real-world applications facing data corrupted Poisson noise include for example astronomical imaging [28, 193, 194, 195], microscopy [35, 64, 124, 167], single particle emission computed tomography (SPECT) [156] or positron emission tomography (PET) [35, 179, 180, 191]. Here, the recorded noisy data  $f = (f_i)_{i=1, \dots, M}$  with  $f_i \in \mathbb{N}$  can often be interpreted as a realization of an independent random vector  $\tilde{F} = (\tilde{F}_i)_{i=1, \dots, M}$  such that

$$\tilde{F}_i \sim \text{Poisson}((Ku_0)_i), \quad (Ku_0)_i \geq 0, \quad (3.2)$$

where  $u_0$  denotes the unknown noise-free image data we are interested in. The additional operator  $K$  models a possible transformation of the image data introduced by the image acquisition process. In astronomical imaging or microscopy it is often a convolution kernel corrupting the data even further.

For many applications  $u_0$  is known to be nonnegative, since it represents for example density or intensity information. In addition,  $K$  is usually assumed to be linear and nonnegativity preserving, which guarantees that  $(Ku)_i \geq 0$  holds true for all possible candidates  $u$ . If we interpret each candidate as a realization  $u = (u_i)_{i=1, \dots, N}$  of some random vector  $U = (U_i)_{i=1, \dots, N}$  with  $p_U(u) > 0$ , the probability mass function in (3.1) with  $\lambda_i := (Ku)_i$  can be viewed as the conditional mass function of observing  $\tilde{F} = f$  given that  $U = u$ , i.e.,

$$p_{\tilde{F}|U}(f|u) = \prod_{i=1}^M \frac{(Ku)_i^{f_i} e^{-(Ku)_i}}{f_i!} \quad \forall f \text{ such that } f_i \in \mathbb{N}, i = 1, \dots, N,$$

cf. [21, 22].

For some applications an additional random vector  $B = (B_i)_{i=1, \dots, M}$  needs to be added to  $\tilde{F}$  to account for accidentally detected particles which have for example been emitted from other sources than the object of interest. Examples include sky emission in astronomy [128, 193, 194], reflections of the excitation light and auto-fluorescence of the medium carrying the sample in fluorescence microscopy [22] or accidental coincidences and attenuation in tomography [108, 109, 162, 224]. Here, we typically have again that  $B_i \sim \text{Poisson}(b_i)$ , where  $b_i \geq 0$  stands for the average number of observed background counts at location  $i$ . If all  $\tilde{F}_i, B_i, i = 1, \dots, M$ , are pairwise independent, this implies that

$$F_i := \tilde{F}_i + B_i \sim \text{Poisson}((Ku_0 + b)_i) \quad \text{and} \quad p_{F|U}(f|u) = \prod_{i=1}^M \frac{(Ku + b)_i^{f_i} e^{-(Ku + b)_i}}{f_i!}, \quad (3.3)$$

cf., e.g., [111, p. 119]. Hence, such effects can easily be included into the model.

Sometimes, it occurs that the data is also corrupted by noise which is not Poisson distributed like for example read-out noise in the context of charge-coupled device (CCD) cameras, cf. [193, 194, 195]. However, if the Poisson distributed noise is dominating, the other noise sources can often be neglected, cf., [22]. Nevertheless, there exist also some approaches trying to account for various types of noise like for example Poisson noise and additive Gaussian noise, cf. [19, 128, 194].

#### 3.2.2 Multiplicative Noise

In the case of *multiplicative noise* we assume that our given noisy data values  $f_i$  can be modeled as

$$f_i = u_i v_i, \quad i = 1, \dots, N,$$

where  $u_i$  is again the original noise-free value of  $f_i$  and  $v_i$  denotes the corresponding noise component. To incorporate the noise statistics,  $f_i$ ,  $u_i$  and  $v_i$  are supposed to be realizations of continuous random variables  $F_i$ ,  $U_i$  and  $V_i$ , respectively, with known densities  $p_{V_1} = \dots = p_{V_N}$ . Thus,  $F_i = U_i V_i$ , where all  $U_i$  and  $V_i$  are typically assumed to be pairwise independent.

In practice, multiplicative noise occurs for example often as speckle [25, 97, 98, 207] produced by various imaging systems. Real-world applications include but are not limited to laser, ultrasonic [41, 202, 211] and synthetic aperture radar (SAR) imaging [29, 139]. The exact statistics of the occurring noise represented by the density  $p_{V_i}$  depend on different factors as for example outlined in [98, 123]. In case of fully developed speckle, the magnitudes  $\tilde{A}_i$  of the complex observations of these imaging devices can be modeled as corrupted by multiplicative Rayleigh noise, cf. [98, 139]. To see this in detail, let us assume that  $\tilde{A}_i$  and  $A_i := \sqrt{\bar{I}_i} > 0$  are realizations of random variables  $\tilde{\mathcal{A}}_i$  and  $\mathcal{A}_i$ , respectively, where  $\bar{I}_i > 0$  denotes the expected mean intensity of the radiation measured at location  $i$ . Then, supposing that  $p_{\mathcal{A}_i}(A_i) > 0$  we know from the literature that the conditional density function of  $\tilde{\mathcal{A}}_i$  given that  $\mathcal{A}_i = A_i$  is

$$p_{\tilde{\mathcal{A}}_i|\mathcal{A}_i}(\tilde{A}_i|A_i) = \frac{2\tilde{A}_i}{A_i^2} \exp\left(-\frac{\tilde{A}_i^2}{A_i^2}\right) 1_{\mathbb{R}_{\geq 0}}(\tilde{A}_i).$$

With  $v_i := \tilde{A}_i/A_i$  and assuming that  $v_i$  is a realization of a random variable  $V_i$  which is independent from  $A_i$ , we obviously have that  $\tilde{\mathcal{A}}_i = \mathcal{A}_i V_i$  and Proposition A.7 ii) yields

$$p_{V_i}(v_i) = p_{V_i}\left(\frac{\tilde{A}_i}{A_i}\right) = A_i \cdot p_{\tilde{\mathcal{A}}_i|\mathcal{A}_i}(\tilde{A}_i|A_i) = \frac{v_i}{\theta^2} \exp\left(-\frac{v_i^2}{2\theta^2}\right) 1_{\mathbb{R}_{\geq 0}}(v_i)$$

with  $\theta = 1/\sqrt{2}$ . Consequently,  $V_i$  is a Rayleigh distributed random variable and  $\tilde{\mathcal{A}}_i = \mathcal{A}_i V_i$  can be considered as corrupted by *multiplicative Rayleigh noise*.

Setting in a next step  $\tilde{V}_i := V_i^2$ ,  $\bar{\mathcal{I}}_i := \mathcal{A}_i^2$  and  $\mathcal{I}_i := \tilde{\mathcal{A}}_i^2 = \bar{\mathcal{I}}_i \tilde{V}_i$  to obtain the corresponding *intensity* data, it follows by Theorem A.1 that

$$p_{\tilde{V}_i}(v_i) = \frac{1}{2\sqrt{|v_i|}} p_{V_i}\left(\sqrt{|v_i|}\right) 1_{\mathbb{R}_{\geq 0}}(v_i) = \exp(-v_i) 1_{\mathbb{R}_{\geq 0}}(v_i), \quad (3.4)$$

i.e., the transformed random variables  $\tilde{V}_i$  are exponentially distributed with parameter  $\lambda = 1$ . The intensities  $\mathcal{I}_i = \tilde{\mathcal{A}}_i^2$  are thus corrupted by *multiplicative exponentially distributed noise* and by Proposition A.7 ii) we have

$$p_{\mathcal{I}_i|\bar{\mathcal{I}}_i}(I_i|\bar{I}_i) = \frac{1}{\bar{I}_i} \exp\left(-\frac{I_i}{\bar{I}_i}\right) 1_{\mathbb{R}_{\geq 0}}(I_i) \quad \text{whenever } p_{\bar{\mathcal{I}}_i}(\bar{I}_i) > 0.$$

**Multilook data** To improve the quality of such amplitude and intensity data, a common approach in, e.g., SAR imaging is to average independent intensity observations  $\mathcal{I}_i^{(j)} = \bar{\mathcal{I}}_i \tilde{V}_i^{(j)}$ ,  $j = 1, \dots, L$ , of the same scene to obtain so-called *multilook data*, where  $L$  denotes the *number of looks*, see, e.g., [25, 98, 139]. Setting

$$\mathcal{I}_i := \frac{1}{L} \sum_{j=1}^L \mathcal{I}_i^{(j)} = \frac{1}{L} \sum_{j=1}^L \bar{\mathcal{I}}_i \tilde{V}_i^{(j)} = \bar{\mathcal{I}}_i \frac{1}{L} \sum_{j=1}^L \tilde{V}_i^{(j)} = \bar{\mathcal{I}}_i \tilde{V}_i$$

with  $\tilde{V}_i := \frac{1}{L} \sum_{j=1}^L \tilde{V}_i^{(j)}$  we can easily see that such multilook data is still corrupted by multiplicative noise. If all  $\tilde{V}_i^{(j)}$  are independent and exponentially distributed with  $\lambda = 1$  as in (3.4), we have by Lemma A.6 that  $\tilde{V}_i$  is further Gamma distributed with density

$$p_{\tilde{V}_i}(v_i) = \frac{L^L}{\Gamma(L)} v_i^{L-1} \exp(-Lv_i) 1_{\mathbb{R}_{\geq 0}}(v_i), \quad (3.5)$$

where  $\Gamma$  denotes the Gamma function. Thus,  $\mathcal{I}_i = \bar{\mathcal{I}}_i \tilde{V}_i$  is contaminated by *multiplicative Gamma noise* and Proposition A.7 ii) implies that

$$p_{\mathcal{I}_i|\bar{\mathcal{I}}_i}(I_i|\bar{I}_i) = \frac{L^L}{\Gamma(L)} \frac{I_i^{L-1}}{\bar{I}_i^L} \exp\left(-L\frac{I_i}{\bar{I}_i}\right) 1_{\mathbb{R}_{\geq 0}}(I_i) \quad \text{whenever } p_{\bar{\mathcal{I}}_i}(\bar{I}_i) > 0. \quad (3.6)$$

For the corresponding amplitude data  $\tilde{\mathcal{A}}_i := \sqrt{\mathcal{I}_i} = \mathcal{A}_i V_i$  with  $\mathcal{A}_i := \sqrt{\bar{\mathcal{I}}_i}$  and  $V_i := \sqrt{\tilde{V}_i}$  we obtain by Corollary A.2 that

$$p_{V_i}(v_i) = 2 v_i p_{\tilde{V}_i}(v_i^2) 1_{\mathbb{R}_{\geq 0}}(v_i) = \frac{2 L^L}{\Gamma(L)} v_i^{2L-1} \exp(-Lv_i^2) 1_{\mathbb{R}_{\geq 0}}(v_i). \quad (3.7)$$

Hence,  $V_i$  is Nakagami distributed and the amplitude data  $\tilde{\mathcal{A}}_i = \mathcal{A}_i V_i$  is corrupted by *multiplicative Nakagami distributed noise*. Moreover, by Proposition A.7 ii) it holds

that the conditional density of  $\tilde{\mathcal{A}}_i$  given that  $\mathcal{A}_i = A_i$  is

$$p_{\tilde{\mathcal{A}}_i|\mathcal{A}_i}(\tilde{A}_i|A_i) = \frac{2L^L}{\Gamma(L)} \frac{\tilde{A}_i^{2L-1}}{A_i^{2L}} \exp\left(-L\frac{\tilde{A}_i^2}{A_i^2}\right) 1_{\mathbb{R}_{\geq 0}}(\tilde{A}_i) \quad \text{whenever } p_{\mathcal{A}_i}(A_i) > 0.$$

### 3.3 Variational Approaches by Bayesian Modeling

Now, let us assume that some noisy and possibly blurred or differently transformed data  $f$  is given, where the noise statistics and the linear transformation are known. In this case, a standard way for determining an estimate  $\hat{u}$  of the original image data is by maximum a posteriori (MAP) estimation, see, e.g., [103, 115, 153]. To be able to apply this framework at least an estimate of the likelihood function  $p_{F|U}(\cdot|u)$  encoding the noise statistics and the data formation process needs to be known. Moreover, some a priori knowledge about the data to be reconstructed is required so that a prior distribution  $p_U$  can be defined. This prior encodes for each possible candidate  $u$  how likely it is to actually observe it. Here,  $p_{F|U}(\cdot|u)$  and  $p_U$  stand either for conditional probability density functions if  $F$ , respectively  $U$  is a continuous random vector or conditional mass functions if  $F$ , respectively  $U$  is discrete. If all the required information is available, the a posteriori distribution  $p_{U|F}(\cdot|f)$  is also defined and an estimate  $\hat{u}$  of the noise-free image data can be determined from the noisy observation  $f$  by solving

$$\operatorname{argmax}_u \{p_{U|F}(u|f)\} = \operatorname{argmin}_u \{-\log p_{U|F}(u|f)\}. \quad (3.8)$$

Hence,  $\hat{u}$  is set to be one of the most likely candidates given the observation  $F = f$ . Following Bayesian rules and neglecting constant terms which do not depend on  $u$  problem (3.8) is equivalent to

$$\operatorname{argmin}_u \{-\log p_{F|U}(f|u) - \log p_U(u)\} \quad (3.9)$$

with  $\log 0 := -\infty$ .

#### 3.3.1 Variational Approaches for Poisson Noise

For the conditional mass function (3.3) corresponding to data corrupted by Poisson noise it holds that

$$-\log p_{F|U}(f|u) = \sum_{i=1}^N \left( (Ku + b)_i - f_i \log((Ku + b)_i) + \log(f_i!) \right). \quad (3.10)$$

With respect to (3.9) the additive term  $\log(f_i!)$  can here be neglected. In the following, we will assume that  $f : \Sigma \rightarrow \mathbb{R}_{\geq 0}$ ,  $u : \Omega \rightarrow \mathbb{R}$  and  $b : \Sigma \rightarrow \mathbb{R}_{\geq 0}$  are functions with

open, bounded domains  $\Omega, \Sigma \subset \mathbb{R}^2$  and  $u \in \mathcal{U}(\Omega)$ ,  $f, b \in \mathcal{V}(\Sigma)$  for some Banach spaces  $\mathcal{U}(\Omega)$  and  $\mathcal{V}(\Sigma)$ . Since for most applications the noise-free image data is known to be nonnegative, we set

$$p_U(u) = \begin{cases} \exp(-\lambda J(u)) & \text{if } u \geq 0, \\ 0 & \text{otherwise} \end{cases} \quad (3.11)$$

for  $\lambda > 0$  and some functional  $J : \mathcal{U}(\Omega) \rightarrow \mathbb{R}_{\geq 0}$ . Moreover,  $K : \mathcal{U}(\Omega) \rightarrow \mathcal{V}(\Sigma)$  is supposed to be a linear, bounded and nonnegativity preserving operator so that  $Ku \geq 0$  for all  $u \geq 0$ . The resulting minimization problem corresponding to (3.9) with the term (3.10) is given by

$$\hat{u} = \operatorname{argmin}_{\substack{u \in \mathcal{U}(\Omega) \\ u \geq 0}} \left\{ \int_{\Sigma} Ku + b - f \log(Ku + b) \, dx + \lambda J(u) \right\}, \quad (3.12)$$

cf. [22, 23], where  $K$  and  $b$  are supposed to be known. Thus,  $\hat{u}$  denotes an estimate of the original data  $u_0$  assuming that the relation between  $u_0$  and  $f$  is correctly described by (3.3). To insured that (3.12) is well-defined,  $K$ ,  $f$  and in particular  $J$  have to fulfill some additional conditions. For details on this topic we refer to [35, 178] and references therein.

In (3.12) the functional  $J$  plays usually the role of a regularizer with regularization parameter  $\lambda$  and the integral term

$$H_f(u) := \int_{\Sigma} Ku + b - f \log(Ku + b) \, dx$$

is viewed as a data fidelity term. It is not hard to prove that  $H_f$  is convex. Moreover, with  $v := Ku + b$  it is up to additive terms that do not depend on  $u$  the same as

$$I(f, v) := \int_{\Sigma} f \log \frac{f}{v} - f + v \, dx = \int_{\Sigma} v - f \log v + f \log f - f \, dx,$$

which is known as Csiszár's *I-divergence* [55] or *generalized Kullback-Leibler divergence*. If  $I$  is finite, it coincides with the *Bregman distance* [32] of the (negative) *Boltzmann-Shannon entropy*  $E(v) := \int_{\Sigma} \log v \, dx$ , cf. [169]. Therefore, it shares the useful properties of a Bregman distance, in particular  $I(f, v) \geq 0$ , which shows that the functional  $H_f$  is bounded from below. In [169, 170] further properties of  $I$  have been proven.

For choosing the regularizer  $J$  there exist many possibilities. A classical restoration method related to  $J = 0$  is for example the *expectation-maximization (EM) algorithm* described in [63, 150, 170, 191] and references therein. This method can sometimes also

be found under the name *Richardson-Lucy algorithm* due to [136, 171], and different variants including for example [156] have been proposed.

A nontrivial regularizer which is frequently applied is the *total variation* (TV) semi-norm

$$|u|_{BV} := \sup_{\substack{V \in C_c^1(\Omega, \mathbb{R}^2) \\ \| |V| \|_\infty \leq 1}} \int_{\Omega} u(x) \operatorname{div} V(x) dx \quad \forall u \in L_1(\Omega). \quad (3.13)$$

This regularizer is a special case of the anisotropic total variation (2.5) and reads for  $L_1(\Omega)$  functions with weak first derivatives in  $L_1(\Omega)$  as

$$|u|_{BV} = \int_{\Omega} |\nabla u(x)| dx.$$

The corresponding *space of functions of bounded variation* is defined by

$$BV(\Omega) = \{u \in L_1(\Omega) : |u|_{BV} < \infty\}$$

and forms a real Banach space with respect to the norm  $\|u\|_{BV} := \|u\|_{L_1} + |u|_{BV}$ , cf. [3, 8, 92, 145, 183, 210, 232]. Setting  $J := |\cdot|_{BV}$  and  $U(\Omega) := BV(\Omega)$  problem (3.12) leads to the variational model

$$\hat{u} = \operatorname{argmin}_{\substack{u \in BV(\Omega) \\ u \geq 0}} \left\{ \int_{\Sigma} Ku + b - f \log(Ku + b) dx + \lambda |u|_{BV} \right\}, \quad (3.14)$$

which can for example be found in [15, 114, 129]. To keep terms simple we will refer to this approach as the IDIV-TV method.

Alternatively to the total variation, *nonlocal regularization* terms of the form

$$J(u) = \int_{\Omega} |\nabla_w u| dx, \quad |\nabla_w u(x)| := \left( \int_{\Omega} (u(y) - u(x))^2 w(x, y) dy \right)^{1/2} \quad (3.15)$$

can be applied for some weight function  $w : \Omega \times \Omega \rightarrow \mathbb{R}_{\geq 0}$ . These regularizers have been introduced and studied for variational models with  $L_1$  and  $L_2$  data fidelity terms in [89, 90, 91, 121, 135, 160, 161, 229, 230]. Here, the involved weight functions are typically chosen similarly to the weights of the nonlocal (NL) means filter [39] with some modifications to fit the application at hand. In [178] a nice overview and analysis of nonlocal regularizers with non-symmetric weight functions has been provided. Moreover, the variational model (3.12) has been applied with a nonlocal regularizer to solve the inverse problem (3.2) facing Poisson noise.

As we will see in Section 3.4 model (3.12) with the total variation or nonlocal regularizers is not only well suited for restoring data corrupted by Poisson noise but also a good choice for removing multiplicative Gamma noise. Before we explain this point in more detail, we will have a closer look at some of the variational approaches suggested in the literature for multiplicative Gamma noise.



### 3.3.2 Variational Models for Multiplicative Gamma Noise

For the restoration of images  $f$  corrupted by multiplicative noise various methods have been proposed in the literature like for example local linear minimum mean square error approaches [125, 130] or anisotropic diffusion methods [4, 123, 227]. Recently, also a couple of variational approaches for removing multiplicative Gamma noise have been presented. In this subsection, we will concentrate on variational methods including the total variation as a regularizer. To examine the differences between the investigated models we will have a look at their properties and apply them to the simplest possible discrete signals  $f := (f_1, f_2)^T$  consisting of only two pixels so that the corresponding discrete models are of the form

$$(\hat{u}_1, \hat{u}_2)^T = \underset{u=(u_1, u_2)^T}{\operatorname{argmin}} \{H_{f_1}(u_1) + H_{f_2}(u_2) + \lambda|u_2 - u_1|\}. \quad (3.16)$$

**Log-model** Facing multiplicative noise we know that  $f = uv$ . If the assumptions  $u, v > 0$  are reasonable, a first nearby idea is often to use a logarithmic transformation to obtain transformed data  $\log f = \log u + \log v$  which is corrupted by additive noise. With  $w := \log u$  we may now apply for example the standard ROF model [52, 176] to the noisy data  $\tilde{f} := \log f \in L_2(\Omega)$ ,  $f > 0$ , leading to the convex minimization problem

$$\hat{w} := \underset{w \in BV(\Omega)}{\operatorname{argmin}} \left\{ \frac{1}{2} \int_{\Omega} (w - \log f)^2 dx + \lambda |w|_{BV} \right\} \quad \text{with} \quad \hat{u} := e^{\hat{w}}. \quad (3.17)$$

Here,  $\Omega$  is again considered to be an open, bounded and connected set with Lipschitz boundary  $\partial\Omega$ . In the following, we set further

$$\begin{aligned} f_{\min} &:= \operatorname{ess\,inf} f = \sup \{a \in \mathbb{R} : \mu(\{x \in \Omega : f(x) < a\}) = 0\}, \\ f_{\max} &:= \operatorname{ess\,sup} f = \inf \{b \in \mathbb{R} : \mu(\{x \in \Omega : f(x) > b\}) = 0\}, \end{aligned}$$

i.e.,  $f_{\min}$  and  $f_{\max}$  are respectively the *essential infimum* and *essential supremum* of the given noisy image function  $f : \Omega \rightarrow \mathbb{R}_{>0}$  with  $\mu$  denoting the Lebesgue measure on  $\mathbb{R}^2$ . Then, it is not hard to show that the minimizer of (3.17) has the following properties:

**Proposition 3.3.1** *For the minimization problem (3.17) with  $\tilde{f} = \log f$  we have*

i) *Maximum-minimum principle:*

$$\log f_{\min} \leq \hat{w} \leq \log f_{\max} \quad a.e.$$

and therefore,

$$f_{\min} \leq \hat{u} \leq f_{\max} \quad a.e. \quad (3.18)$$

ii) (*Logarithmic*) *mean value preservation*:

$$\int_{\Omega} \hat{w} \, dx = \int_{\Omega} \tilde{f} \, dx \quad \text{and thus,} \quad \int_{\Omega} \log \hat{u} \, dx = \int_{\Omega} \log f \, dx.$$

iii)  $L_1$ -convergence to a constant solution:

$$\lim_{\lambda \rightarrow \infty} \|\hat{w} - \text{mean}(\tilde{f})\|_{L_1} = 0 \quad \text{with} \quad \text{mean}(\tilde{f}) := \frac{1}{\mu(\Omega)} \int_{\Omega} \tilde{f} \, dx.$$

Part i) of this proposition follows by standard results for the total variation, see, e.g., [122, Lem. 1] and the proof of [10, Thm. 4.1]. The proof of ii) can be found in [50] and to prove iii) we can follow the lines of [11, p. 89] using the Poincaré inequality [9, p. 302] and part ii).

In a discrete setting with an appropriate discretization the mean value preservation  $\sum_{i=1}^N \hat{w}_i = \sum_{i=1}^N \tilde{f}_i$  implies that

$$\prod_{i=1}^N \hat{u}_i = \prod_{i=1}^N f_i.$$

Thus, the *log-model* (3.17) *preserves the geometric mean rather than the arithmetic mean*. For  $\lambda \rightarrow \infty$  we even have that  $\hat{w}_i \rightarrow \frac{1}{N} \sum_{j=1}^N \tilde{f}_j$  for all  $i = 1, \dots, N$ , and therefore,  $\hat{u}$  converges for increasing  $\lambda$  to a constant signal of value  $(\prod_{j=1}^N f_j)^{1/N}$ . Due to the assumption  $f = uv$  we know that

$$\left( \prod_{j=1}^N f_j \right)^{1/N} = \left( \prod_{j=1}^N u_j \right)^{1/N} \left( \prod_{j=1}^N v_j \right)^{1/N}.$$

Consequently, the preservation of the geometric mean of  $f$  is only reasonable if we can assume that  $(\prod_{j=1}^N v_j)^{1/N} = 1$ . However, if we consider all pixels  $f_i, u_i, v_i$  to be respectively realizations of corresponding random variables  $F_i, U_i, V_i$  fulfilling the assumptions stated at the beginning of Subsection 3.2.2, it holds that

$$\mathbb{E}\left(\frac{1}{N} \sum_{i=1}^N F_i\right) = \mathbb{E}\left(\frac{1}{N} \sum_{i=1}^N U_i V_i\right) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}(U_i) \mathbb{E}(V_i) = \mathbb{E}\left(\frac{1}{N} \sum_{i=1}^N U_i\right) \mathbb{E}(V_k)$$

for  $k = 1, \dots, N$ . Since we know for, e.g., multiplicative Gamma noise that  $\mathbb{E}(V_k) = 1$ ,  $k = 1, \dots, N$ , this implies that  $\mathbb{E}\left(\frac{1}{N} \sum_{i=1}^N F_i\right) = \mathbb{E}\left(\frac{1}{N} \sum_{i=1}^N U_i\right)$  and thus, the mean value of  $f$  is likely to equal the mean value of  $u$ . However, since the geometric mean of  $f$  is always smaller than its arithmetic mean provided that  $f$  is not constant, see,

e.g., [112, p. 30], the solution  $\hat{u}$  of (3.17) does in general not preserve the mean value of  $f$ . Especially if  $f$  is corrupted by strong multiplicative noise, we must expect that the mean value of  $\hat{u}$  is a lot smaller than the mean of  $f$ . To examine this point in more detail let us have a look at the following two pixel example:

**Example** (Two pixel signals)

The two pixel model (3.16) corresponding to (3.17) is given by

$$(\hat{w}_1, \hat{w}_2)^T = \underset{w=(w_1, w_2)^T}{\operatorname{argmin}} \left\{ \frac{1}{2}(w_1 - \log f_1)^2 + \frac{1}{2}(w_2 - \log f_2)^2 + \lambda |w_2 - w_1| \right\} \quad (3.19)$$

with  $(\hat{u}_1, \hat{u}_2)^T = (e^{\hat{w}_1}, e^{\hat{w}_2})^T$ . For simplicity, let us assume that  $f_1 \geq f_2 > 0$ . Then, setting the subdifferential of the involved energy function to zero shows that with increasing  $\lambda$  the solution  $(\hat{u}_1, \hat{u}_2)^T$  approaches  $(\sqrt{f_1 f_2}, \sqrt{f_1 f_2})^T$  as follows:

$$\begin{aligned} \hat{u}_1 &= f_1 e^{-\lambda}, \quad \hat{u}_2 = f_2 e^{\lambda} & \text{for } 0 \leq \lambda < \frac{1}{2} \log \frac{f_1}{f_2}, \\ \hat{u}_1 &= \hat{u}_2 = \sqrt{f_1 f_2} & \text{for } \frac{1}{2} \log \frac{f_1}{f_2} \leq \lambda. \end{aligned}$$

Supposing that  $f_1 = c_1 u$  and  $f_2 = c_2 u$  with  $c_1 > c_2 > 0$  are two pixels corrupted by multiplicative noise we obtain for  $\lambda \geq \frac{1}{2} \log \frac{f_1}{f_2} = \frac{1}{2} \log \frac{c_1}{c_2}$  the constant solution  $\hat{u}_1 = \hat{u}_2 = \sqrt{f_1 f_2} = \sqrt{c_1 c_2} u$ . Here, the critical value  $\lambda = \log \frac{f_1}{f_2}$  depends fortunately only on the distortions  $c_1, c_2$  and not on  $u$ . However, if  $c_1, c_2, c_1 > c_2$  are chosen such that  $\frac{f_1 + f_2}{2} = u$ , then  $\sqrt{f_1 f_2} < u$ .

Since the function  $g : [0, \frac{1}{2} \log \frac{f_1}{f_2}] \rightarrow \mathbb{R}$  defined by  $g(\lambda) = \frac{1}{2}(f_1 e^{-\lambda} + f_2 e^{\lambda})$  is monotonically decreasing with  $g(0) = \frac{1}{2}(f_1 + f_2)$  and  $g(\frac{1}{2} \log \frac{f_1}{f_2}) = \sqrt{f_1 f_2}$ , the mean value of  $(\hat{u}_1, \hat{u}_2)^T$  decreases monotonically to  $\sqrt{f_1 f_2}$  with increasing  $\lambda \in [0, \frac{1}{2} \log \frac{f_1}{f_2}]$  as depicted in Figure 3.1. Here, the difference between the arithmetic and geometric mean of  $(f_1, f_2)^T$  can be arbitrarily large as we can easily verify by setting, e.g.,  $f_1 := cf$ ,  $f_2 := \frac{1}{c}f$  with fixed  $f > 0$  and increasing  $c \geq 1$ .

Due to these observations model (3.17) does not seem to be an optimal choice for removing multiplicative (Gamma) noise in images. A more sophisticated approach for log-transformed data has for example been presented in [69, 70]. This method combines curvelet shrinkage and TV regularization for the removal of multiplicative Gamma noise.

**AA-model** Based on the MAP estimate (3.9) a different approach for multiplicative Gamma noise has been proposed by Aubert and Aujol in [10]. For the deduction of this model let us again assume that the discretized functions  $f, u$  and  $v$  with  $f = uv$  can be considered as realizations of random vectors  $F = (F_1, \dots, F_N)^T$ ,  $U = (U_1, \dots, U_N)^T$

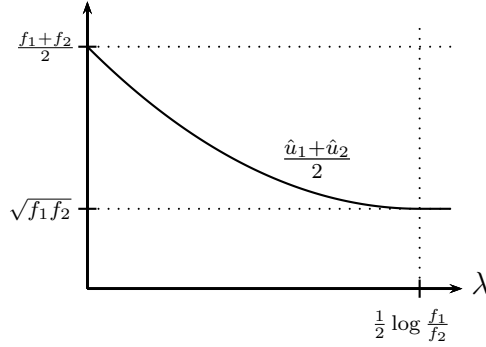


Figure 3.1: Graph illustrating the change in the mean value of the two-pixel signal  $(\hat{u}_1, \hat{u}_2)^T = (e^{\hat{w}_1}, e^{\hat{w}_2})^T$  solving (3.19) for increasing  $\lambda$ .

and  $V = (V_1, \dots, V_N)^T$ , respectively, where all  $U_i, V_i, i = 1, \dots, N$ , are pairwise independent and each  $V_i$  is a Gamma distributed random variable with density (3.5). Then, the conditional density (3.6) implies that

$$p_{F|U}(f|u) = \prod_{i=1}^N p_{F_i|U_i}(f_i|u_i) = \left( \frac{L^L}{\Gamma(L)} \right)^N \prod_{i=1}^N \frac{f_i^{L-1}}{u_i^L} \exp\left(-L \frac{f_i}{u_i}\right) 1_{\mathbb{R}_{\geq 0}}(f_i)$$

whenever  $p_U(u) > 0$  and thus,

$$-\log p_{F|U}(f|u) = L \sum_{i=1}^N \left( \log u_i + \frac{f_i}{u_i} \right) - (L-1) \sum_{i=1}^N \log f_i + C \quad \forall f \geq 0$$

with  $C := N(\log(\Gamma(L)) - L \log L)$ . Returning to the continuous setting and supposing that  $p_U$  is defined as in (3.11) with  $J(u) = |u|_{BV}$  and  $u \in BV(\Omega)$ , the MAP approach (3.9) finally leads for  $f > 0$  to the non-convex minimization problem

$$\operatorname{argmin}_{\substack{u \in BV(\Omega) \\ u > 0}} \left\{ \int_{\Omega} \log u + \frac{f}{u} dx + \lambda |u|_{BV} \right\} \quad (3.20)$$

proposed in [10], where all terms which do not depend on  $u$  have been neglected.

As shown by Aubert and Aujol this minimization problem has for all  $f \in L_{\infty}(\Omega)$  with  $f_{\min} > 0$  at least one minimizer  $\hat{u}$  which fulfills the maximum-minimum principle (3.18). Moreover, restricting our attention to two pixel signals, further properties of this minimizer can be deduced:

**Example** (Two pixel signals)

Let us assume that  $f = (f_1, f_2)^T$  with  $0 < f_2 \leq f_1 \leq 3f_2$ , which is for example the

case if  $f_1 = (1 + \nu)u$  and  $f_2 = (1 - \nu)u$  with  $u > 0$ ,  $0 \leq \nu \leq 1/2$ . Then, the minimizer of the two pixel model (3.16) corresponding to (3.20) reads

$$\begin{aligned} \hat{u}_1 &= \frac{-1 + \sqrt{1 + 4\lambda f_1}}{2\lambda}, \quad \hat{u}_2 = \frac{1 - \sqrt{1 - 4\lambda f_2}}{2\lambda} & \text{for } 0 < \lambda < \frac{2(f_1 - f_2)}{(f_1 + f_2)^2}, \\ \hat{u}_1 &= \hat{u}_2 = \frac{f_1 + f_2}{2} & \text{for } \frac{2(f_1 - f_2)}{(f_1 + f_2)^2} \leq \lambda. \end{aligned}$$

For  $f_1 = (1 + \nu)u$  and  $f_2 = (1 - \nu)u$  representing to noisy pixels with the same initial value  $u$  this means that we have to choose  $\lambda \geq \frac{\nu}{u}$  to get back the original constant signal  $(u, u)^T$ . Since here the critical value  $\lambda = \frac{\nu}{u}$  is inversely proportional to  $u$ , we see that with increasing  $\lambda$  noise survives unfortunately much longer for smaller values  $u$ . Regarding the mean value of  $(\hat{u}_1, \hat{u}_2)^T$  it is not hard to show that

$$g : \left(0, \frac{2(f_1 - f_2)}{(f_1 + f_2)^2}\right] \rightarrow \mathbb{R} \quad \text{defined by} \quad g(\lambda) = \frac{1}{4\lambda} (\sqrt{1 + 4\lambda f_1} - \sqrt{1 - 4\lambda f_2})$$

is maximal for  $\lambda \rightarrow 0$  and  $\lambda = \frac{2(f_1 - f_2)}{(f_1 + f_2)^2}$ , where  $\lim_{\lambda \rightarrow 0} g(\lambda) = g\left(\frac{2(f_1 - f_2)}{(f_1 + f_2)^2}\right) = \frac{f_1 + f_2}{2}$ . Furthermore,  $g$  reaches its minimum  $g(\lambda) = \sqrt{f_1 f_2}$  for  $\lambda = \frac{f_1 - f_2}{4f_1 f_2}$ . Therefore, we have again that

$$\sqrt{f_1 f_2} \leq \frac{1}{2}(\hat{u}_1 + \hat{u}_2) \leq \frac{1}{2}(f_1 + f_2) \quad \forall \lambda \geq 0$$

and  $\frac{1}{2}(\hat{u}_1 + \hat{u}_2)$  varies with  $\lambda$  as depicted in Figure 3.2 (left).

Regarding these findings the AA-model (3.20) has thus also its shortcomings. Especially the non-convexity of (3.20) makes its minimization a lot more challenging compared to other methods.

**SO-model** While the data fidelity term of the AA-model follows canonically from the MAP approach related to multiplicative Gamma noise, the choice of the regularization term is quite flexible. Replacing the regularizer  $|u|_{BV}$  in (3.20) by  $|\log u|_{BV}$  and setting  $w := \log u$  as for the log-model we obtain the alternative minimization problem

$$\hat{w} = \operatorname{argmin}_{w \in BV(\Omega)} \left\{ \int_{\Omega} w + f e^{-w} dx + \lambda |w|_{BV} \right\} \quad \text{with} \quad \hat{u} = e^{\hat{w}}. \quad (3.21)$$

This model is a special case of the approach considered by Shi and Osher in [192]. Indeed, their method involves the more general data fidelity term

$$H_f(u) = \int_{\Omega} (a + b)w + a f e^{-w} + \frac{b}{2} f^2 e^{-2w} dx, \quad a, b \geq 0,$$

which also includes the model proposed in [175]. A slight modification of (3.21) has also been considered in [110], where the authors added a quadratic term to obtain a

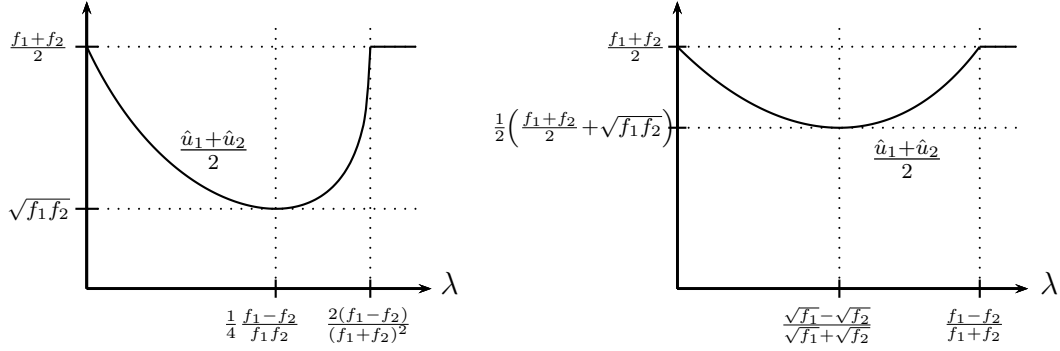


Figure 3.2: Graphs illustrating the changes in the mean values of the two pixel solutions  $(\hat{u}_1, \hat{u}_2)^T$  of the AA-model (left) and SO-model (right) with increasing  $\lambda$ .

simple alternating minimization algorithm.

Compared to (3.20) the minimization problem (3.21) has the big advantage that it is strictly convex and attains a unique minimum for  $f \in L_\infty(\Omega)$  with  $f_{\min} > 0$  as we will see in the next section. Moreover,  $\hat{u}$  fulfills again the maximum-minimum principle (3.18), cf. [113]. To examine further properties of the model let us have again a look at the two pixel case:

**Example** (Two pixel signals)

For  $f = (f_1, f_2)^T$  with  $f_1 \geq f_2 > 0$  the solution of the two pixel model (3.16) corresponding to (3.21) is given by

$$\begin{aligned} \hat{u}_1 &= \frac{f_1}{1+\lambda}, \quad \hat{u}_2 = \frac{f_2}{1-\lambda} & \text{for } 0 \leq \lambda < \frac{f_1 - f_2}{f_1 + f_2}, \\ \hat{u}_1 &= \hat{u}_2 = \frac{f_1 + f_2}{2} & \text{for } \frac{f_1 - f_2}{f_1 + f_2} \leq \lambda. \end{aligned}$$

Supposing that  $f_1 = (1 + \nu)u$ ,  $f_2 = (1 - \nu)u$ , the original constant signal of value  $u$  is reconstructed if  $\lambda \geq \nu$ , where in contrast to the AA-model the critical value  $\lambda = \nu$  is now independent of  $u$ .

Having a look at the function  $g : [0, \frac{f_1 - f_2}{f_1 + f_2}]$  defined by  $g(\lambda) = \frac{1}{2}(\frac{f_1}{1+\lambda} + \frac{f_2}{1-\lambda})$  it is not hard to show that  $g$  is convex and reaches its minimum  $g(\lambda) = \frac{1}{2}(\frac{f_1 + f_2}{2} + \sqrt{f_1 f_2})$  for  $\lambda = \frac{\sqrt{f_1} - \sqrt{f_2}}{\sqrt{f_1} + \sqrt{f_2}}$ . Besides,  $g$  has a maximum of  $g(\lambda) = \frac{f_1 + f_2}{2}$  for  $\lambda = 0$  and  $\lambda = \frac{f_1 - f_2}{f_1 + f_2}$ . Consequently, we know that

$$\frac{1}{2} \left( \frac{f_1 + f_2}{2} + \sqrt{f_1 f_2} \right) \leq \frac{1}{2}(\hat{u}_1 + \hat{u}_2) \leq \frac{1}{2}(f_1 + f_2) \quad \forall \lambda \geq 0,$$

which shows that although the mean value of  $f$  is not preserved, the maximal deviation of  $\frac{1}{2}(\hat{u}_1 + \hat{u}_2)$  from  $\frac{1}{2}(f_1 + f_2)$  is significantly less compared to the AA-model. The changes in  $\frac{1}{2}(\hat{u}_1 + \hat{u}_2)$  with respect to  $\lambda$  are illustrated in Figure 3.2 (right).

Regarding these observations the SO-model (3.21) seems to be a promising candidate for restoring images corrupted by multiplicative (Gamma) noise.

### 3.4 Equivalence of the SO-Model and IDIV-TV

Interestingly, we can show that the minimizer

$$\hat{u} = \operatorname{argmin}_{\substack{u \in BV(\Omega) \\ u \geq 0}} \left\{ \int_{\Omega} u - f \log u \, dx + \lambda |u|_{BV} \right\} \quad (3.22)$$

of the IDIV-TV model (3.14) with  $K(u) = u$ ,  $b = 0$  and the minimizer  $\hat{w}$  of the SO-model (3.21) coincide in the sense that  $\hat{u} = e^{\hat{w}}$  for

- (I)  $\Omega$  being an open, bounded and connected set with Lipschitz boundary  $\partial\Omega$ ,
- (II)  $f \in L_{\infty}(\Omega)$  with  $f_{\min} > 0$ , which includes that  $f_{\max} < \infty$ .

This is quite surprising, since these models have been designed to restore images corrupted by noise of totally different distributions. An intuitive justification for this claim can be given by the following heuristic arguments:

**Remark 3.4.1** (A heuristic explanation)

Let us assume for a moment that the minimization in (3.21) and (3.22) is restricted to sufficiently smooth functions. Then, the convexity of the functionals in (3.21) and (3.22) implies that the minimizers  $\hat{w}$  of (3.21) and  $\hat{u} > 0$  of (3.22) can be characterized by the Euler-Lagrange equations

$$0 = 1 - f e^{-\hat{w}} - \lambda \operatorname{div} \frac{\nabla \hat{w}}{|\nabla \hat{w}|} \quad \text{and} \quad 0 = 1 - \frac{f}{\hat{u}} - \lambda \operatorname{div} \frac{\nabla \hat{u}}{|\nabla \hat{u}|}, \quad \text{respectively}$$

whenever  $|\nabla \hat{w}| \neq 0$  and  $|\nabla \hat{u}| \neq 0$ . Since  $\nabla e^w = e^w \nabla w$ , we have for  $u := e^w$  that  $\nabla u(x) = 0$  if and only if  $\nabla w(x) = 0$ . Hence, setting  $\tilde{u} := e^{\hat{w}}$  implies that

$$\frac{\nabla \hat{w}}{|\nabla \hat{w}|} = \frac{e^{\hat{w}} \nabla \hat{w}}{e^{\hat{w}} |\nabla \hat{w}|} = \frac{\nabla \tilde{u}}{|\nabla \tilde{u}|},$$

which shows with the above Euler-Lagrange equations that  $\hat{u} = e^{\hat{w}}$ . These arguments work also for the anisotropic TV regularizers introduced in Section 2.1.

Of course, this justification has several gaps. To be able to give a precise proof we need to take into account that the minimization in (3.21) and (3.22) is performed over the whole of  $BV(\Omega) \subset L_1(\Omega)$ . Unfortunately, the involved data fidelity terms are

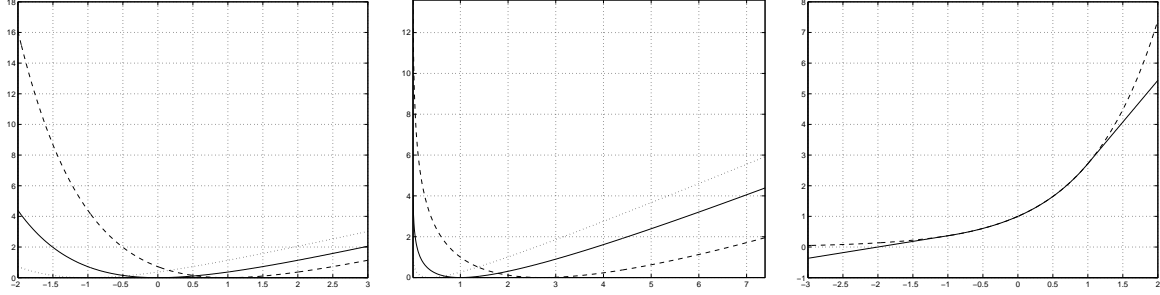


Figure 3.3: *Left:* The integrand  $\varphi(x, \cdot)$  for  $f(x) = e^{-1}$  (dotted),  $f(x) = 1$  (solid) and  $f(x) = e$  (dashed). *Middle:* The integrand  $\psi(x, \cdot)$  for the same values of  $f(x)$ . *Right:* The exponential function (dashed) and the function  $g_k$  defined in (3.30) for  $a_k = -1$  and  $b_k = 1$  (solid).

not continuous over  $L_p(\Omega)$ ,  $1 \leq p < \infty$ , since their domains have an empty interior, compare, e.g., [168, Lem. 4.1(ii)]. In the rest of this section we will show how to deal with these facts.

For this purpose, we consider the integrands  $\varphi, \psi : \Omega \times \mathbb{R} \rightarrow [0, +\infty]$  defined by

$$\varphi(x, s) := s + f(x) e^{-s} - \log f(x) - 1, \quad (3.23)$$

$$\psi(x, s) := \begin{cases} s - f(x) \log s + f(x) \log f(x) - f(x) & \text{for } s > 0, \\ +\infty & \text{otherwise,} \end{cases} \quad (3.24)$$

for which properties like continuity, lower semi-continuity, convexity and subdifferentiability are always understood with respect to the second variable. It is not hard to see that  $\varphi$  and  $\psi$  are proper, strictly convex and l.s.c. with respect to  $s$  as well as measurable with respect to  $x$ . Since the sets  $\text{dom}\varphi(x, \cdot)$  and  $\text{dom}\psi(x, \cdot)$  have in addition a non-empty interior for every  $x$ , we know by [174, p. 176] that  $\varphi$  and  $\psi$  are also *normal* in the sense defined in [174]. Besides,  $\varphi$  and  $\psi$  are nonnegative and for fixed  $x \in \Omega$  they attain their minimum 0 for  $s = \log f(x)$  and  $s = f(x)$ , respectively. Figure 3.3 (left/middle) shows  $\varphi(x, \cdot), \psi(x, \cdot)$  for  $f(x) \in \{e^{-1}, 1, e\}$ .

The integral functionals  $S_\varphi, S_\psi : L_1(\Omega) \rightarrow [0, +\infty]$  corresponding to our data fidelity terms in (3.21) and (3.22) are now defined by

$$S_\varphi(w) := \int_{\Omega} \varphi(x, w(x)) dx \quad \text{and} \quad S_\psi(u) := \int_{\Omega} \psi(x, u(x)) dx, \quad \text{respectively.}$$

Obviously,  $S_\varphi$  and  $S_\psi$  are proper. Since  $\varphi$  and  $\psi$  are nonnegative and normal, we know by [73, p. 239] that  $S_\varphi$  and  $S_\psi$  are also nonnegative and l.s.c. Moreover, the fact that  $\varphi$  and  $\psi$  are normal and strictly convex implies that  $S_\varphi$  and  $S_\psi$  are strictly convex,



too. By [174, p. 189] we thus have that

$$\begin{aligned}\partial S_\varphi(w) &= \{w^* \in L_\infty(\Omega) : w^*(x) \in \partial\varphi(x, w(x)) \text{ a.e.}\} \\ &= \begin{cases} \{1 - fe^{-w}\} & \text{for } 1 - fe^{-w} \in L_\infty(\Omega), \\ \emptyset & \text{otherwise,} \end{cases} \\ \partial S_\psi(u) &= \{u^* \in L_\infty(\Omega) : u^*(x) \in \partial\psi(x, u(x)) \text{ a.e.}\} \\ &= \begin{cases} \{1 - \frac{f}{u}\} & \text{for } u > 0 \text{ a.e. and } 1 - \frac{f}{u} \in L_\infty(\Omega), \\ \emptyset & \text{otherwise.} \end{cases}\end{aligned}$$

Hence, the equality

$$\partial S_\varphi(w) = \partial S_\psi(e^w)$$

holds true.

Next, we define the functionals  $T_\varphi, T_\psi : L_1(\Omega) \rightarrow [0, +\infty]$  corresponding to (3.21) and (3.22) by

$$T_\varphi(w) := S_\varphi(w) + \lambda |w|_{BV}, \quad T_\psi(u) := S_\psi(u) + \lambda |u|_{BV}. \quad (3.25)$$

The properties of  $S_\varphi, S_\psi$  and  $|\cdot|_{BV}$ , see, e.g., [3, 79, 183], immediately imply that  $T_\varphi$  and  $T_\psi$  are proper, l.s.c. and strictly convex. Since we will show in the next lemma that  $T_\varphi$  and  $T_\psi$  are also  $BV$ -coercive, the same arguments as, e.g., in [35, Thm. 5.4.7] imply that there exist minimizers of both functionals and these minimizers are unique due to the strict convexity of  $T_\varphi$  and  $T_\psi$ , respectively.

For the proof of the subsequent lemma we need the existence of constants  $C_\phi, C_\psi > 0$  and  $D_\phi, D_\psi > 0$  such that

$$\varphi(x, s) \geq C_\varphi |s| - D_\varphi \quad \text{and} \quad \psi(x, s) \geq C_\psi |s| - D_\psi. \quad (3.26)$$

To obtain such constants set for example

$$\begin{aligned}a &< \min\{0, \log(f_{\min})\}, & b &> \max\{0, \log(f_{\max})\}, \\ C_a &:= 1 - f_{\min} e^{-a} < 0, & C_b &:= 1 - f_{\max} e^{-b} > 0, \\ w_a &:= a + f_{\min} e^{-a} - \log(f_{\min}) - 1, & w_b &:= b + f_{\max} e^{-b} - \log(f_{\max}) - 1, \\ u_b &:= e^b - f_{\max} b + f_{\max} \log(f_{\max}) - f_{\max}\end{aligned}$$

and take

$$\begin{aligned}C_\varphi &= \min\{|C_a|, C_b\}, & C_\psi &= C_b, \\ D_\varphi &= \max\{C_a a - w_a, C_b b - w_b\}, & D_\psi &= C_b e^b - u_b.\end{aligned}$$

**Lemma 3.4.2** *Let  $\Omega$  and  $f$  fulfill conditions (I), (II). Then,  $T_\varphi$  and  $T_\psi$  are  $BV$ -coercive.*

**Proof:** In the following, we will concentrate on  $T_\psi$ , since the arguments are the same for  $T_\varphi$  and  $T_\psi$ . By the Poincaré inequality [9, p. 302] we know that there exists a constant  $C_P$  such that

$$\|u - \bar{u}\|_{L_1} \leq C_P |u|_{BV}, \quad \bar{u} := \frac{1}{\mu(\Omega)} \int_{\Omega} u(x) dx$$

with  $\mu$  denoting again the Lebesgue measure on  $\mathbb{R}^2$ . Hence, we have

$$\begin{aligned} \|u\|_{L_1} &\leq \|u - \bar{u}\|_{L_1} + |\bar{u}| \mu(\Omega) \leq C_P |u|_{BV} + |\bar{u}| \mu(\Omega), \\ \|u\|_{BV} &\leq (C_P + 1) |u|_{BV} + |\bar{u}| \mu(\Omega) \leq C T_\psi(u) + |\bar{u}| \mu(\Omega) \end{aligned} \quad (3.27)$$

for  $C := (C_P + 1)/\lambda$ . By (3.26) it holds that

$$S_\psi(u) \geq C_\psi \int_{\Omega} |u(x)| dx - D_\psi \mu(\Omega) \geq \mu(\Omega)(C_\psi |\bar{u}| - D_\psi)$$

and consequently

$$|\bar{u}| \leq \frac{1}{\mu(\Omega)C_\psi} S_\psi(u) + \frac{D_\psi}{C_\psi} \leq \frac{1}{\mu(\Omega)C_\psi} T_\psi(u) + \frac{D_\psi}{C_\psi}.$$

Thus, we finally obtain with (3.27) that

$$\|u\|_{BV} \leq (C + 1/C_\psi) T_\psi(u) + \mu(\Omega) D_\psi / C_\psi,$$

which implies that  $T_\psi(u) \rightarrow \infty$  whenever  $\|u\|_{BV} \rightarrow \infty$ .  $\square$

For another proof of the  $BV$ -coercivity of  $T_\psi$  see [35, 178]. To finally show that the solutions of (3.21) and (3.22) coincide we need the following theorem:

**Theorem 3.4.3** i) *Let  $u \in BV(\Omega)$  and assume that  $h : \mathbb{R} \rightarrow \mathbb{R}$  is a non-decreasing and Lipschitz continuous function. Then,  $h(u) \in BV(\Omega)$  and  $\partial|u|_{BV} \subset \partial|h(u)|_{BV}$ .*

ii) *Let  $\phi : \Omega \times \mathbb{R} \rightarrow (-\infty, +\infty]$  be a measurable function and assume that there exists a nonnegative function  $\gamma \in L_1(\Omega)$  and constants  $C > 0$  and  $1 \leq p < \infty$  such that*

$$|\phi(x, s)| \leq C|s|^p + \gamma(x) \quad \text{a.e. } x \in \Omega, \forall s \in \mathbb{R}. \quad (3.28)$$

*Then, the functional  $S_\phi$  is  $L_p$ -continuous if and only if (up to equivalent integrands)  $\phi(x, \cdot)$  is continuous for almost every  $x \in \Omega$ .*

The proof of i) can be found in [183, p. 148] and for ii) see [84, p. 442]. Besides, part i) with non-decreasing and locally absolutely continuous functions  $h$  has been considered in [99].

The final theorem of this section is a special instance of a more general theorem proved by Grasmair in [99]. For the sake of completeness we give the prove for our special setting. For a generalization including also anisotropic TV regularizers we refer to [99].

**Theorem 3.4.4 (Equivalence of the SO-model and IDIV-TV)**

Let  $T_\varphi$  and  $T_\psi$  be given by (3.25) and assume that conditions (I) and (II) are fulfilled. Then,  $\hat{w}$  is the minimizer of  $T_\varphi$  if and only if  $\hat{u} = e^{\hat{w}}$  is the minimizer of  $T_\psi$ .

**Proof:** In the following we show that  $\hat{w} = \operatorname{argmin}_w T_\varphi(w)$  implies  $e^{\hat{w}} = \operatorname{argmin}_u T_\psi(u)$ . The reverse direction follows immediately since the minimizers of  $T_\varphi$  and  $T_\psi$  are unique.

The idea for the prove is to approximate the integrands  $\varphi$ ,  $\psi$  and the exponential function by 'nicer' functions for which the 'adapted' assertion follows immediately by Theorem 3.4.3 and then to apply  $\Gamma$ -convergence arguments to get the final result.

1. To this end, choose a sequence of increasing intervals  $[a_k, b_k] \subset [a_{k+1}, b_{k+1}]$  such that  $\bigcup_k [a_k, b_k] = \mathbb{R}$  and  $a_k < \min\{0, \log(f_{\min})\}$ ,  $b_k > \max\{0, \log(f_{\max})\}$  for all  $k \in \mathbb{N}$ . Moreover, let

$$\xi_k(x) := 1 - f(x)e^{-a_k} < 0, \quad \zeta_k(x) := 1 - f(x)e^{-b_k} > 0.$$

and define the truncated continuous integrands corresponding to  $\varphi$ ,  $\psi$  for  $x \in \Omega$  by

$$\begin{aligned} \varphi_k(x, s) &:= \begin{cases} \varphi(x, a_k) + \xi_k(x)(s - a_k) & \text{if } s < a_k, \\ \varphi(x, s) & \text{if } s \in [a_k, b_k], \\ \varphi(x, b_k) + \zeta_k(x)(s - b_k) & \text{if } s > b_k, \end{cases} \\ \psi_k(x, s) &:= \begin{cases} \psi(x, e^{a_k}) + \xi_k(x)(s - e^{a_k}) & \text{if } s < e^{a_k}, \\ \psi(x, s) & \text{if } s \in [e^{a_k}, e^{b_k}], \\ \psi(x, e^{b_k}) + \zeta_k(x)(s - e^{b_k}) & \text{if } s > e^{b_k}. \end{cases} \end{aligned}$$

The resulting functionals

$$T_{\varphi,k}(w) := S_{\varphi_k}(w) + \lambda|w|_{BV}, \quad T_{\psi,k}(u) := S_{\psi_k}(u) + \lambda|u|_{BV}$$

are proper, convex and l.s.c. for any  $k \in \mathbb{N}$ . Besides, by the same arguments as in the proof of Lemma 3.4.2 we see that

$$T_{\varphi,k}(w) \geq C_1\|w\|_{BV} - D_1, \quad T_{\psi,k}(u) \geq C_2\|u\|_{BV} - D_2 \quad \forall k \in \mathbb{N} \quad (3.29)$$

for some constants  $C_1, C_2, D_1, D_2 > 0$  so that the sequences  $(T_{\varphi,k})_{k \in \mathbb{N}}$  and  $(T_{\psi,k})_{k \in \mathbb{N}}$  are equi-coercive on  $BV(\Omega)$ . Here, recall that a sequence  $(T_k)_{k \in \mathbb{N}}$  is *equi-coercive* on a metric space  $X$  if and only if there exists a l.s.c., coercive function  $F : X \rightarrow (-\infty, +\infty]$  such that  $T_k \geq F$  for all  $k$ , see [57, Prop. 7.7]. Due to these properties of  $T_{\varphi,k}$  and  $T_{\psi,k}$  there exist thus minimizers  $\hat{w}_k$  and  $\hat{u}_k$  of  $T_{\varphi,k}$  and  $T_{\psi,k}$ , respectively, cf. the proof of [35, Thm. 5.4.7].

Since  $f \in L_\infty(\Omega)$  with  $f_{\min} > 0$ , the functions  $\varphi_k$  fulfill condition (3.28) with  $C_k := \max\{|\text{ess inf } \xi_k|, \text{ess sup } \zeta_k\}$ ,  $p := 1$  and  $\gamma_k := \varphi(\cdot, 0) \in L_1(\Omega)$ . Therefore, we obtain by Theorem 3.4.3 ii) that  $S_{\varphi_k}$  is continuous on  $L_1(\Omega)$  (and on  $BV(\Omega)$ ). Now, set

$$g_k(s) := \begin{cases} e^{a_k} + e^{a_k}(s - a_k) & \text{if } s < a_k, \\ e^s & \text{if } s \in [a_k, b_k], \\ e^{b_k} + e^{b_k}(s - b_k) & \text{if } s > b_k \end{cases} \quad (3.30)$$

as exemplified in Figure 3.3 (right). This truncated exponential function is a non-decreasing, Lipschitz continuous function so that we know by Theorem 3.4.3 i) that  $\partial|w|_{BV} \subset \partial|g_k(w)|_{BV}$  for all  $w \in BV(\Omega)$ . Furthermore, one can check by straightforward computation that  $\partial\varphi_k(x, s) = \partial\psi_k(x, g_k(s))$  and thus,  $\partial S_{\varphi_k}(w) = \partial S_{\psi_k}(g_k(w))$ , cf. [174, p. 189]. Consequently, it follows by [73, p. 26] that

$$\partial T_{\varphi,k}(w) = \partial S_{\varphi_k}(w) + \lambda \partial|w|_{BV} \subset \partial S_{\psi_k}(g_k(w)) + \lambda \partial|g_k(w)|_{BV} \subset \partial T_{\psi,k}(g_k(w)).$$

Now  $\hat{w}_k$  is a minimizer of  $T_{\varphi,k}$  if and only if  $0 \in \partial T_{\varphi,k}(\hat{w}_k)$ . Therefore, if  $\hat{w}_k$  is a minimizer of  $T_{\varphi,k}$ , then  $\hat{u}_k = g_k(\hat{w}_k)$  is a minimizer of  $T_{\psi,k}$ .

2. The sequences  $(\varphi_k(x, w(x)))_{k \in \mathbb{N}}$  and  $(\psi_k(x, u(x)))_{k \in \mathbb{N}}$  are non-decreasing, nonnegative and

$$\lim_{k \rightarrow \infty} \varphi_k(x, w(x)) = \varphi(x, w(x)), \quad \lim_{k \rightarrow \infty} \psi_k(x, u(x)) = \psi(x, u(x)) \quad \forall x \in \Omega$$

so that by the Monotone Convergence Theorem

$$\lim_{k \rightarrow \infty} S_{\varphi_k} = S_\varphi, \quad \lim_{k \rightarrow \infty} S_{\psi_k} = S_\psi.$$

The fact that  $(\varphi_k(x, w(x)))_{k \in \mathbb{N}}$  and  $(\psi_k(x, u(x)))_{k \in \mathbb{N}}$  are non-decreasing also implies that  $(T_{\varphi,k})_{k \in \mathbb{N}}$  and  $(T_{\psi,k})_{k \in \mathbb{N}}$  are non-decreasing. Therefore, we have by [27, p. 35] that

$$\Gamma\text{-}\lim_{k \rightarrow \infty} T_{\varphi,k} = \lim_{k \rightarrow \infty} T_{\varphi,k} = T_\varphi, \quad \Gamma\text{-}\lim_{k \rightarrow \infty} T_{\psi,k} = \lim_{k \rightarrow \infty} T_{\psi,k} = T_\psi. \quad (3.31)$$

3. Let  $\hat{w} = \text{argmin}_{w \in BV(\Omega)} T_\varphi(w)$  and  $\hat{u} = \text{argmin}_{u \in BV(\Omega)} T_\psi(u)$ . Using that the sequences  $(T_{\varphi,k})_{k \in \mathbb{N}}$  and  $(T_{\psi,k})_{k \in \mathbb{N}}$  are non-decreasing and bounded from above by  $T_\varphi$  and  $T_\psi$ , respectively, it holds that

$$K_\varphi := \overline{\{w \in BV(\Omega) : T_{\varphi,1}(w) \leq T_\varphi(\hat{w})\}} \neq \emptyset, \\ K_\psi := \overline{\{u \in BV(\Omega) : T_{\psi,1}(u) \leq T_\psi(\hat{u})\}} \neq \emptyset$$

and for any minimizers  $\hat{w}_k$  of  $T_{\varphi,k}$  and  $\hat{u}_k$  of  $T_{\psi,k}$  we have

$$\hat{w}_k \in \{w \in BV(\Omega) : T_{\varphi,k}(w) \leq T_\varphi(\hat{w})\} \subset K_\varphi, \\ \hat{u}_k \in \{u \in BV(\Omega) : T_{\psi,k}(u) \leq T_\psi(\hat{u})\} \subset K_\psi.$$

Hence,

$$\min_{w \in BV(\Omega)} T_{\varphi,k}(w) = \min_{w \in K_{\varphi}} T_{\varphi,k}(w), \quad \min_{u \in BV(\Omega)} T_{\psi,k}(u) = \min_{u \in K_{\psi}} T_{\psi,k}(u) \quad \forall k \in \mathbb{N}$$

and since the sets  $K_{\varphi}$  and  $K_{\psi}$  are compact due to the coercivity of  $T_{\varphi,1}$  and  $T_{\psi,1}$ , this implies that the sequences  $(T_{\varphi,k})_{k \in \mathbb{N}}$  and  $(T_{\psi,k})_{k \in \mathbb{N}}$  are *equi-mildly coercive* according to the definition provided in [27, p. 29]. Moreover, any sets of minimizers  $\{\hat{w}_k\}_{k \in \mathbb{N}} \subset K_{\varphi}$  and  $\{\hat{u}_k\}_{k \in \mathbb{N}} \subset K_{\psi}$  are precompact and thus, the sequences  $(\hat{w}_k)_{k \in \mathbb{N}}$  and  $(\hat{u}_k)_{k \in \mathbb{N}}$  admit convergent subsequences. With the  $\Gamma$ -convergence (3.31) and [27, Thm. 1.21] we obtain further that

$$\lim_{k \rightarrow \infty} T_{\varphi,k}(\hat{w}_k) = T_{\varphi}(\hat{w}), \quad \lim_{k \rightarrow \infty} T_{\psi,k}(\hat{u}_k) = T_{\psi}(\hat{u})$$

and all convergent subsequences of  $(\hat{w}_k)_{k \in \mathbb{N}}$  and  $(\hat{u}_k)_{k \in \mathbb{N}}$  converge in  $L_1(\Omega)$  to  $\hat{w}$  and  $\hat{u}$ , respectively.

4. Let  $(\hat{w}_n)_{n \in I}$ ,  $(\hat{u}_n)_{n \in I} := (g_n(\hat{w}_n))_{n \in I}$ ,  $I \subset \mathbb{N}$ , be subsequences converging in  $L_1(\Omega)$  to  $\hat{w}$  and  $\hat{u}$ , respectively. By [5, Theorem 13.6] there exists a subsequence  $(\hat{w}_m)_{m \in I_1}$ ,  $I_1 \subset I$ , which converges a.e. to  $\hat{w}$  and the construction of  $g_m$  yields that  $(g_m(\hat{w}_m))_{m \in I_1}$  converges a.e. to  $e^{\hat{w}}$ . On the contrary, we know that  $(g_m(\hat{w}_m))_{m \in I_1}$  converges in  $L_1(\Omega)$  to  $\hat{u}$ . Hence, there exists a subsequence  $(g_k(\hat{w}_k))_{k \in I_2}$ ,  $I_2 \subset I_1$  which converges a.e. to  $\hat{u}$  and consequently  $\hat{u} = e^{\hat{w}}$  a.e.  $\square$

By this theorem we finally showed that the SO-model (3.21) and the IDIV-TV model (3.22) provide exactly the same minimizers. Since the IDIV-TV model is well-examined and has not been considered for multiplicative noise before, we will concentrate on approaches related to this model in the subsequent sections.

## 3.5 Theoretical Results for the Discrete Setting

Next, we have a closer look at the discrete counterpart of the IDIV model (3.12) with  $K(u) = u$  and  $b = 0$ . For this purpose, we work with columnwise reshaped image vectors  $f, u \in \mathbb{R}^N$ . Operations like the multiplication or division of vectors or taking the logarithm are meant componentwise, here. Moreover, we denote by  $D \in \mathbb{R}^{mN, N}$  a matrix whose rows are zero or contain exactly one 1 and one  $-1$ . This matrix represents either

- i) some discretization of the gradient operator with  $m = 2$  as for example the first matrix introduced in (2.41), see also [47, 198], or
- ii) a nonlocal operator with binary weights as introduced in [91], where  $m$  is associated to the number of permitted neighbors, see Section 3.7 for more details.

For vectors  $p := (p_1, \dots, p_m)^T \in \mathbb{R}^{mN}$  with  $p_k := (p_{j+(k-1)N})_{j=1}^N$  we set  $|p| := (p_1^2 + \dots + p_m^2)^{1/2} \in \mathbb{R}^N$  and focus on the minimizers  $\hat{u}$  of the problem

$$\min_{u \in \mathbb{R}^N} \left\{ H_f(u) + \lambda \underbrace{\phi(Du)}_{J(u)} \right\} \quad \text{with } f > 0, \quad (3.32)$$

where

$$H_f(u) := \begin{cases} \langle 1, u - f \log u \rangle & \text{if } u > 0, \\ +\infty & \text{otherwise} \end{cases} \quad \text{and} \quad \phi(p) := \| |p| \|_1. \quad (3.33)$$

If  $D$  is defined according to i), this model is a discretized version of the IDIV-TV model (3.22). If  $D$  is determined as in ii), the resulting functional  $J$  is a discretization of the nonlocal regularizer (3.15) and we refer to (3.32) as the IDIV-NL model.

Independent of the chosen matrix  $D$  the functional in (3.32) is proper, l.s.c., coercive and strictly convex and thus, it has a unique minimizer. The corresponding dual problem of (3.32) is given by

$$- \min_{p \in \mathbb{R}^{mN}} \left\{ H_f^*(-D^T p) + \lambda \phi^*(\lambda^{-1} p) \right\}, \quad (3.34)$$

cf. [12, p. 154f], with the conjugate functionals

$$H_f^*(v) = \begin{cases} -\langle f, \log(1-v) \rangle + \langle f, \log f - 1 \rangle & \text{if } v < 1, \\ +\infty & \text{otherwise} \end{cases} \quad \text{and} \quad \phi^*(p) = \iota_C(p)$$

for  $C := \{p \in \mathbb{R}^{mN} : \| |p| \|_\infty \leq 1\}$ . Fortunately, there exists no duality gap, i.e., (3.32) and (3.34) take the same value, and the minimizers  $\hat{u}$  and  $\hat{p}$  are related via

$$-D^T \hat{p} \in \partial H_f(\hat{u}), \quad \hat{p} \in \lambda \partial \phi(D\hat{u}) \quad \text{so that} \quad \hat{u} = \frac{f}{1 + D^T \hat{p}}. \quad (3.35)$$

Moreover,  $\hat{u}$  has a couple of desirable properties as the following proposition shows:

**Proposition 3.5.1** *The minimizer  $\hat{u}$  of (3.32) has the following properties:*

i) *Maximum-minimum principle:*

$$f_{\min} \leq \hat{u}_i \leq f_{\max} \quad \forall i = 1, \dots, N,$$

where  $f_{\min}$  and  $f_{\max}$  are the values of the smallest and largest components of  $f$ .

ii) *Averaging property:*

$$\frac{1}{N} \sum_{i=1}^N \frac{f_i}{\hat{u}_i} = 1.$$

iii) *Convergence to the mean gray value:*

If the nullspace of  $D$  is  $\mathcal{N}(D) = \{v \in \mathbb{R}^N : v_1 = \dots = v_N\}$ , then

$$\lim_{\lambda \rightarrow \infty} \hat{u} = c \mathbf{1}_N \quad \text{with} \quad c = \frac{1}{N} \sum_{j=1}^N f_j.$$

**Proof:** i) The first property follows in the same way as in [10, Theorem 4.1]. We only have to verify that

$$J(\min(u, f_{\max})) \leq J(u), \quad J(\max(u, f_{\min})) \leq J(u).$$

By the structure of  $\phi$  and  $D$  we see that this is the case whenever

$$\begin{aligned} |\min(u_i, f_{\max}) - \min(u_j, f_{\max})| &\leq |u_i - u_j|, \\ |\max(u_i, f_{\min}) - \max(u_j, f_{\min})| &\leq |u_i - u_j|. \end{aligned}$$

These inequalities are trivially fulfilled for  $u_i = u_j$  and thus, we can assume that  $u_i > u_j$ . If  $f_{\max} \geq u_i$  or  $u_j \geq f_{\max}$  and  $f_{\min} \geq u_i$  or  $u_j \geq f_{\min}$ , both inequalities are fulfilled. For  $u_i > f_{\max} > u_j$  the first inequality becomes  $|f_{\max} - u_j| \leq |u_i - u_j|$ , which is also true. Similarly, we get for  $u_i > f_{\min} > u_j$  the correct inequality  $|u_i - f_{\min}| \leq |u_i - u_j|$ .

ii) The second property follows by (3.35) and since  $(\mathbf{1}_N^T D^T)^T = D \mathbf{1}_N = \mathbf{0}_N$ . Namely, we have

$$\frac{1}{N} \sum_{i=1}^N \frac{f_i}{\hat{u}_i} = \frac{1}{N} \sum_{i=1}^N \frac{f_i(1 + (D^T \hat{p})_i)}{f_i} = 1 + \frac{1}{N} \sum_{i=1}^N (D^T \hat{p})_i = 1 + \frac{1}{N} \mathbf{1}_N^T D^T \hat{p} = 1.$$

iii) To prove the last part of the assertion we can proceed as in [11, p. 89]: Let  $\hat{u}(\lambda) := \operatorname{argmin}_{u \in \mathbb{R}^N} \{H_f(u) + \lambda \|Du\|_1\}$ . Obviously, it holds that

$$0 \leq \lambda \|D\hat{u}(\lambda)\|_1 \leq H_f(\hat{u}(\lambda)) + \lambda \|D\hat{u}(\lambda)\|_1 \leq \min_{u \in \mathbb{R}^N} \{H_f(u) + \lambda \|Du\|_1\}.$$

For  $\tilde{u} := \tilde{c} \mathbf{1}_N$  with some constant  $\tilde{c} > 0$  we have  $\|D\tilde{u}\|_1 = 0$ , which yields

$$0 \leq \lambda \|D\hat{u}(\lambda)\|_1 \leq H_f(\tilde{u}).$$

Consequently, we obtain

$$0 \leq \lim_{\lambda \rightarrow \infty} \|D\hat{u}(\lambda)\|_1 \leq \lim_{\lambda \rightarrow \infty} \frac{1}{\lambda} H_f(\tilde{u}) = 0$$

so that  $\lim_{\lambda \rightarrow \infty} \|D\hat{u}(\lambda)\|_1 = 0$ . Since  $\phi = \|\cdot\|_1$  is a norm, it follows that  $\lim_{\lambda \rightarrow \infty} \hat{u}(\lambda) \in \mathcal{N}(D)$ , i.e.,  $\lim_{\lambda \rightarrow \infty} \hat{u}(\lambda) = c \mathbf{1}_N$  for some constant  $c \in \mathbb{R}$ . Using now part ii) we finally see that

$$1 = \lim_{\lambda \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \frac{f_i}{\hat{u}_i(\lambda)} = \frac{1}{N} \sum_{i=1}^N \frac{f_i}{c} = \frac{1}{c} \frac{1}{N} \sum_{i=1}^N f_i$$

and thus,  $c = \frac{1}{N} \sum_{i=1}^N f_i$ .  $\square$

Facing multiplicative Gamma noise we know that  $f_i = \hat{u}_i v_i$ . In this case the second property implies that  $\frac{1}{N} \sum_{i=1}^N v_i = 1$ , which is desirable, since the expectation value of any Gamma distributed random variable with density (3.5) is 1.

Property iii) shows that whenever the nullspace of  $D$  consists of all constant vectors,  $f$  and  $\hat{u}$  have at least for  $\lambda \rightarrow \infty$  the same mean gray value. Note that a matrix  $D$  with this property is for example the first gradient discretization in (2.41).

As a last result of this section we will prove that the SO-model and IDIV-TV are also equivalent in the discrete setting if  $m = 1$ :

**Proposition 3.5.2** *Let  $D \in \mathbb{R}^{mN, N}$  be a matrix, where  $m = 1$ . Then, the minimizer  $\hat{u}$  of (3.32) equals  $e^{\hat{w}}$  with  $\hat{w}$  being the minimizer of the discrete SO-model*

$$\hat{w} := \operatorname{argmin}_{w \in \mathbb{R}^N} \{ \langle 1, w + f e^{-w} \rangle + \lambda \phi(Dw) \}. \quad (3.36)$$

**Proof:** The vectors  $\hat{u}$  and  $\hat{w}$  are the minimizers of (3.32) and (3.36), respectively if and only if

$$0 \in 1 - \frac{f}{\hat{u}} + \lambda D^T \partial \phi(D\hat{u}) \quad \text{and} \quad 0 \in 1 - f e^{-\hat{w}} + \lambda D^T \partial \phi(D\hat{w}).$$

Since  $1 - \frac{f}{\hat{u}} = 1 - f e^{-\hat{w}}$  for  $\hat{u} = e^{\hat{w}}$ , we have to show that  $D^T \partial \phi(D e^{\hat{w}}) = D^T \partial \phi(D \hat{w})$ . It is well-known, see, e.g., [9, Prop. 1.10] for the continuous case, that  $v \in D^T \partial \phi(Dw)$  if and only if  $v = D^T p$  for some  $p \in \mathbb{R}^{mN}$  with  $\|p\|_\infty \leq 1$  and

$$\langle p, Dw \rangle = \phi(Dw).$$

The last equation can also be written as

$$\sum_{i=1}^N \sum_{j=0}^{m-1} p_{i+jN} (Dw)_{i+jN} = \sum_{i=1}^N \left( \sum_{j=0}^{m-1} (Dw)_{i+jN}^2 \right)^{1/2}. \quad (3.37)$$

Setting  $d_i := ((Dw)_{i+jN})_{j=0}^{m-1}$ ,  $i = 1, \dots, N$ , and applying the Cauchy-Schwarz inequality to the inner sums on the left hand side, we see with  $\|p\|_\infty \leq 1$  that

$$\sum_{i=1}^N \left| \sum_{j=0}^{m-1} p_{i+jN} (Dw)_{i+jN} \right| \leq \sum_{i=1}^N \left( \sum_{j=0}^{m-1} p_{i+jN}^2 \right)^{1/2} \|d_i\|_2 \leq \sum_{i=1}^N \|d_i\|_2.$$



Here and thus in (3.37), equality holds true if and only if for each  $i \in \{1, \dots, N\}$  we either have

- i)  $\|d_i\|_2 = 0$  and  $(p_{i+jN})_{j=0}^{m-1}$  is an arbitrary vector with  $\sum_{j=0}^{m-1} p_{i+jN}^2 \leq 1$  or
- ii)  $\|d_i\|_2 \neq 0$  and  $(p_{i+jN})_{j=0}^{m-1} = \alpha d_i$  with  $\sum_{j=0}^{m-1} p_{i+jN}^2 = 1$ . The last two equalities imply that  $\alpha = 1/\|d_i\|_2$  so that  $p_{i+jN} = (Dw)_{i+jN}/\|d_i\|_2$ ,  $j = 0, \dots, m-1$ .

Since the exponential function is strictly monotone and each row of  $D$  is either zero or contains exactly one 1 and one  $-1$ , case i) appears for  $D\hat{w}$  if and only if it appears for  $De^{\hat{w}}$ . In the second case, if for  $j = 0, \dots, m-1$  the  $(i+jN)$ -th row of  $D$  is 1 in column  $k_{i+jN}$  and  $-1$  in column  $h_{i+jN}$ , we get

$$p_{i+jN} = \frac{\hat{w}_{k_{i+jN}} - \hat{w}_{h_{i+jN}}}{\left( \sum_{j=0}^{m-1} (\hat{w}_{k_{i+jN}} - \hat{w}_{h_{i+jN}})^2 \right)^{1/2}} \quad \text{for } w := \hat{w}$$

$$\text{and } p_{i+jN} = \frac{e^{\hat{w}_{k_{i+jN}}} - e^{\hat{w}_{h_{i+jN}}}}{\left( \sum_{j=0}^{m-1} (e^{\hat{w}_{k_{i+jN}}} - e^{\hat{w}_{h_{i+jN}}})^2 \right)^{1/2}} \quad \text{for } w := e^{\hat{w}}.$$

For  $m = 1$  the right-hand sides are just the signs of  $\hat{w}_{k_{i+jN}} - \hat{w}_{h_{i+jN}}$  and  $e^{\hat{w}_{k_{i+jN}}} - e^{\hat{w}_{h_{i+jN}}}$ , respectively and coincide, since the exponential function is strictly increasing. Thus, we finally showed for  $m = 1$  and any  $p \in \mathbb{R}^{mN}$  with  $\|p\|_\infty \leq 1$  that

$$\langle p, D\hat{w} \rangle = \phi(D\hat{w}) \quad \text{if and only if} \quad \langle p, De^{\hat{w}} \rangle = \phi(De^{\hat{w}}),$$

which finishes the proof.  $\square$

Note that this proof implies for the anisotropic setting where  $\phi(p) := \|p\|_1$  that the discrete models provide also the same minimizers for  $m \geq 2$ .

## 3.6 Minimization Algorithms

To actually compute minimizers of the discrete IDIV model

$$\min_{\substack{u \in \mathbb{R}^N \\ u \geq 0}} \{H_f(Ku) + \lambda \phi(Du)\} \quad (3.38)$$

involving possibly also a linear, nonnegativity preserving transformation  $K$ , an efficient algorithm is required. In the following, we will deduce such an algorithm based on the alternating direction method of multipliers (ADMM). In this context, we will also show how it is possible to incorporate multiple splittings into the standard ADMM algorithm, compare also [187].

### 3.6.1 ADMM and Multiple Splittings

In recent years splitting methods became increasingly popular for solving convex minimization problems of the form

$$\operatorname{argmin}_{u \in \mathbb{R}^N} \{\Psi(u) + \Phi(Cu)\} \quad (3.39)$$

for some matrix  $C \in \mathbb{R}^{M,N}$  and proper, convex and l.s.c. functionals  $\Psi : \mathbb{R}^N \rightarrow \mathbb{R} \cup \{+\infty\}$  and  $\Phi : \mathbb{R}^M \rightarrow \mathbb{R} \cup \{+\infty\}$ . One prominent member of this class is the alternating direction method of multipliers (ADMM), which can be traced back to [87, 88, 93]. A nice review article on this algorithm has been provided recently by Boyd et al. in [26]. For problems of the form (3.39) this algorithm is equivalent to the recently proposed alternating split Bregman algorithm [96] as shown by Esser in [76] and to the Douglas-Rachford splitting algorithm [67, 134] applied to the dual problem of (3.39), cf. [71, 87, 186, 188].

To derive the ADMM algorithm for solving (3.39) we need to consider the constrained version of (3.39) given by

$$\operatorname{argmin}_{u \in \mathbb{R}^N, w \in \mathbb{R}^M} \{\Psi(u) + \Phi(w)\} \quad \text{subject to} \quad Cu = w. \quad (3.40)$$

The *method of multipliers* applied to this constrained problem leads to the iterative scheme

$$\begin{aligned} (u^{(k+1)}, w^{(k+1)}) &= \operatorname{argmin}_{u \in \mathbb{R}^N, w \in \mathbb{R}^M} \left\{ \Psi(u) + \Phi(w) + \frac{1}{2\gamma} \|\gamma p^{(k)} + Cu - w\|_2^2 \right\}, \\ p^{(k+1)} &= p^{(k)} + \frac{1}{\gamma} (Cu^{(k+1)} - w^{(k+1)}). \end{aligned}$$

Minimizing alternately with respect to  $u$  and  $w$  and setting  $b^{(k)} := \gamma p^{(k)}$  yields the following ADMM algorithm:

**Algorithm 3.6.1 (Scaled ADMM algorithm)**

*Initialization:* Choose  $w^{(0)}$ ,  $b^{(0)}$  and  $\gamma > 0$ .

*For*  $k = 1, \dots$  *repeat until a stopping criterion is reached:*

$$u^{(k+1)} = \operatorname{argmin}_{u \in \mathbb{R}^N} \left\{ \Psi(u) + \frac{1}{2\gamma} \|b^{(k)} + Cu - w^{(k)}\|_2^2 \right\}, \quad (3.41)$$

$$w^{(k+1)} = \operatorname{argmin}_{w \in \mathbb{R}^M} \left\{ \Phi(w) + \frac{1}{2\gamma} \|b^{(k)} + Cu^{(k+1)} - w\|_2^2 \right\}, \quad (3.42)$$

$$b^{(k+1)} = b^{(k)} + Cu^{(k+1)} - w^{(k+1)}. \quad (3.43)$$

By applying this algorithm we can exploit that it is often much easier to solve the subproblems (3.41) and (3.42) instead of minimizing directly (3.39). Note here that if the minimizer of (3.41) or (3.42) is not unique, an arbitrary one is chosen.

To ensure the convergence of this algorithm we will assume that

(I) the primal problem (3.39) has a solution,

(II)  $0 \in \text{int}(C\text{dom}\Psi - \text{dom}\Phi)$  and  $0 \in \text{int}(\text{dom}\Psi^* + C^T\text{dom}\Phi^*)$ .

The set  $\text{int } A$  denotes the so-called *interior* of the set  $A$  defined by

$$\text{int } A = \{x \in \mathbb{R}^N : \exists \varepsilon > 0 \text{ so that } B_\varepsilon(x) \subset A\}$$

for  $B_\varepsilon(x) = \{y \in \mathbb{R}^N : \|y - x\|_2 < \varepsilon\}$ . The assumptions (I) and (II) guarantee that the primal problem (3.39) is well-posed, there also exists a solution of the corresponding dual problem

$$\underset{p \in \mathbb{R}^M}{\text{argmin}} \{\Psi^*(-C^T p) + \Phi^*(p)\} \quad (3.44)$$

and the duality gap is 0, cf., e.g., [12, p. 155]. With these preliminaries we can state the following theorem. For a proof based on convergence results for Douglas-Rachford splitting we refer to [72, 87, 187, 188].

### Theorem 3.6.2 (Convergence of ADMM)

Let conditions (I) and (II) be fulfilled. Then, the sequences  $(b^{(k)})_{k \in \mathbb{N}}$  and  $(w^{(k)})_{k \in \mathbb{N}}$  generated by the ADMM Algorithm 3.6.1 converge for all starting values and any step sizes  $\gamma > 0$ . Denoting by  $\hat{b}$  and  $\hat{w}$  the limits of these sequences, the vector  $\hat{p} := \frac{1}{\gamma}\hat{b}$  is a solution of the dual problem (3.44). Furthermore, the sequence  $(u^{(k)})_{k \in \mathbb{N}}$  calculated by (3.41) converges to a solution of the primal problem (3.39) if one of the following conditions is fulfilled:

- i) The primal problem has a unique solution.
- ii) The problem  $\underset{u \in \mathbb{R}^N}{\text{argmin}} \left\{ \Psi(u) + \frac{1}{2\gamma} \|\hat{b} + Cu - \hat{w}\|_2^2 \right\}$  has a unique solution.

Whenever  $\hat{u}$  converges, we have of course  $\hat{w} = C\hat{u}$ . Moreover, since  $\hat{p} = \frac{1}{\gamma}\hat{b}$  is a solution of the dual problem, we know, e.g., by [12, Thm. 5.2.1] that  $\hat{u}$ ,  $\hat{w}$  and  $\hat{p}$  are related via

$$-C^T \hat{p} \in \partial \Psi(\hat{u}) \Leftrightarrow \hat{u} \in \partial \Psi^*(-C^T \hat{p}) \quad \text{and} \quad \hat{p} \in \partial \Phi(\hat{w}) \Leftrightarrow \hat{w} \in \partial \Phi^*(\hat{p}).$$

**Multiple splittings** In many applications more general problems of the form

$$\operatorname{argmin}_{u \in \mathbb{R}^N} \left\{ \Psi(u) + \sum_{i=1}^r f_i(C_i u) \right\} \quad (3.45)$$

need to be solved, where  $C_i \in \mathbb{R}^{M_i, N}$ ,  $\Psi : \mathbb{R}^N \rightarrow \mathbb{R} \cup \{+\infty\}$  and  $f_i : \mathbb{R}^{M_i} \rightarrow \mathbb{R} \cup \{+\infty\}$ . In these cases we can apply the splitting idea even further. Rewriting problem (3.45) as

$$\operatorname{argmin}_{u \in \mathbb{R}^N, w_i \in \mathbb{R}^{M_i}} \left\{ \Psi(u) + \sum_{i=1}^r f_i(w_i) \right\} \quad \text{subject to} \quad w_i = C_i u$$

and setting  $\Phi(w) := \sum_{i=1}^r f_i(w_i)$  and  $C := (C_1^T, \dots, C_r^T)^T$ , we can apply again Algorithm 3.6.1. Here, the subproblem (3.42) becomes

$$w^{(k+1)} = \operatorname{argmin}_{w_i \in \mathbb{R}^{M_i}} \left\{ \sum_{i=1}^r f_i(w_i) + \frac{1}{2\gamma} \left\| \begin{pmatrix} b_1^{(k)} \\ \vdots \\ b_r^{(k)} \end{pmatrix} + \begin{pmatrix} C_1 \\ \vdots \\ C_r \end{pmatrix} u^{(k+1)} - \begin{pmatrix} w_1 \\ \vdots \\ w_r \end{pmatrix} \right\|^2 \right\}.$$

Since the variables  $w_i$  are not coupled, we can minimize separately with respect to each  $w_i$ ,  $i = 1, \dots, r$ . Hence, the resulting iterative scheme reads as follows:

**Algorithm 3.6.3 (Scaled ADMM with multiple splittings)**

*Initialization:* Choose  $w_1^{(0)}, \dots, w_r^{(0)}$ ,  $b_1^{(0)}, \dots, b_r^{(0)}$  and  $\gamma > 0$ .

*For*  $k = 1, \dots$  *repeat until a stopping criterion is reached:*

$$\begin{aligned} u^{(k+1)} &= \operatorname{argmin}_{u \in \mathbb{R}^N} \left\{ \Psi(u) + \frac{1}{2\gamma} \left\| \begin{pmatrix} b_1^{(k)} \\ \vdots \\ b_r^{(k)} \end{pmatrix} + \begin{pmatrix} C_1 \\ \vdots \\ C_r \end{pmatrix} u - \begin{pmatrix} w_1^{(k)} \\ \vdots \\ w_r^{(k)} \end{pmatrix} \right\|^2 \right\}, \\ w_1^{(k+1)} &= \operatorname{argmin}_{w_1 \in \mathbb{R}^{M_1}} \left\{ f_1(w_1) + \frac{1}{2\gamma} \|b_1^{(k)} + C_1 u^{(k+1)} - w_1\|^2 \right\}, \\ &\vdots \\ w_r^{(k+1)} &= \operatorname{argmin}_{w_r \in \mathbb{R}^{M_r}} \left\{ f_r(w_r) + \frac{1}{2\gamma} \|b_r^{(k)} + C_r u^{(k+1)} - w_r\|^2 \right\}, \\ b^{(k+1)} &= b^{(k)} + \begin{pmatrix} C_1 \\ \vdots \\ C_r \end{pmatrix} u^{(k+1)} - \begin{pmatrix} w_1^{(k+1)} \\ \vdots \\ w_r^{(k+1)} \end{pmatrix}. \end{aligned}$$

As already described in [76] the here applied splitting can also be useful for functionals consisting of only two additive terms where each is a concatenation of a proper, convex and l.s.c. functional with a linear operator.

It is not hard to show that the conjugate of  $\Phi$  is given by

$$\Phi^*(p_1, \dots, p_r) = \sum_{i=1}^r f_i^*(p_i).$$

Hence, we can see by (3.44) that the dual problem of (3.45) with  $p^T = (p_1^T, \dots, p_r^T)$  is

$$\operatorname{argmin}_{p \in \mathbb{R}^M} \{ \Psi^*(-C^T p) + \sum_{i=1}^r f_i^*(p_i) \}. \quad (3.46)$$

Due to Theorem 3.6.2 we know that if conditions (I) and (II) are fulfilled, the sequences  $(b^{(k)})_{k \in \mathbb{N}}$  and  $(w^{(k)})_{k \in \mathbb{N}}$  generated by the ADMM Algorithm 3.6.3 converge, in particular,  $p^{(k)} := \frac{1}{\gamma} b^{(k)}$  converges to a solution of the dual problem (3.46). Moreover, the sequence  $(u^{(k)})_{k \in \mathbb{N}}$  converges to the solution of the primal problem (3.45) if either the primal problem or

$$\operatorname{argmin}_{u \in \mathbb{R}^N} \left\{ \Psi(u) + \frac{1}{2\gamma} \|\hat{b} + Cu - \hat{w}\|_2^2 \right\}$$

has a unique solution. In these cases we know that the limits  $\hat{u}$ ,  $\hat{w}$  and  $\hat{p}$  fulfill  $C_i \hat{u} = \hat{w}_i$  for  $i = 1, \dots, r$  and are related via

$$-C^T \hat{p} \in \partial \Psi(\hat{u}) \Leftrightarrow \hat{u} \in \partial \Psi^*(-C^T \hat{p}) \quad \text{and} \quad \hat{p}_i \in \partial f_i(\hat{w}_i) \Leftrightarrow \hat{w}_i \in \partial f_i^*(\hat{p}_i).$$

**Iterative adaptation of the parameter  $\gamma$**  In Algorithms 3.6.1 and 3.6.3 the quantity  $\gamma$  is a fixed parameter for all iterations, which has to be chosen in advance. Although Theorem 3.6.2 guarantees that ADMM converges for any fixed  $\gamma$ , the speed of convergence depends in practice crucially on its value.

For a small value we see in (3.41) and (3.42) that the functionals  $\Psi$  and  $\Phi$  have nearly no effect. Rewriting (3.41) and (3.42) with  $b^{(k)} = \gamma p^{(k)}$  in the form

$$u^{(k+1)} = \operatorname{argmin}_{u \in \mathbb{R}^N} \left\{ \Psi(u) + \frac{1}{2\gamma} \|Cu - w^{(k)}\|_2^2 + \langle p^{(k)}, Cu - w^{(k)} \rangle \right\}, \quad (3.47)$$

$$w^{(k+1)} = \operatorname{argmin}_{w \in \mathbb{R}^M} \left\{ \Phi(w) + \frac{1}{2\gamma} \|Cu^{(k+1)} - w\|_2^2 + \langle p^{(k)}, Cu^{(k+1)} - w \rangle \right\} \quad (3.48)$$

shows that instead the optimality constraint  $Cu^{(k+1)} = w^{(k+1)}$  is enforced. On the contrary, for a large value of  $\gamma$  the constraint  $Cu^{(k+1)} = w^{(k+1)}$  is nearly ignored. Here, the conditions  $-C^T p^{(k)} \in \partial \Psi(u^{(k+1)})$  and  $p^{(k)} \in \partial \Phi(w^{(k+1)})$  are enforced and thus the optimality condition  $0 \in \partial \Psi(u^{(k+1)}) + C^T \partial \Phi(w^{(k+1)})$ . To balance these two extremes,  $\gamma$  has to be carefully chosen as we will also see later on in Section 3.7.

To improve the performance of ADMM in general and to make the convergence speed less dependent on the initially chosen  $\gamma$ , an interesting strategy for an iterative adaptation of  $\gamma$  has been reported in [26, 105]. Here, the idea is to evaluate the quantities

$$\begin{aligned} r^{(k+1)} &:= Cu^{(k+1)} - w^{(k+1)} = b^{(k+1)} - b^{(k)} = \gamma(p^{(k+1)} - p^{(k)}), \\ s^{(k+1)} &:= -\frac{1}{\gamma}C^T(w^{(k+1)} - w^{(k)}) \end{aligned}$$

at the end of each iteration and to adapt  $\gamma$  such that the  $\ell_2$  norms of these two vectors remain within a reasonable range of each other. By (3.47) and (3.48) it is not hard to show that

$$s^{(k+1)} \in \partial\Psi(u^{(k+1)}) + C^T\partial\Phi(w^{(k+1)}).$$

Since the terms  $\frac{1}{2}\|Cu - w\|_2^2$  in (3.47) and (3.48) are however scaled by  $\frac{1}{\gamma}$  compared to  $\Psi(u)$  and  $\Phi(w)$ , we propose here to use

$$\tilde{r}^{(k+1)} := \frac{1}{\gamma}(Cu^{(k+1)} - w^{(k+1)}) = \frac{1}{\gamma}(b^{(k+1)} - b^{(k)}) = p^{(k+1)} - p^{(k)}$$

instead of  $r^{(k+1)}$  and to choose

$$\gamma^{(k+1)} = \begin{cases} 2\gamma^{(k)} & \text{if } \|s^{(k+1)}\|_2 > 10\|\tilde{r}^{(k+1)}\|_2, \\ \frac{1}{2}\gamma^{(k)} & \text{if } \|\tilde{r}^{(k+1)}\|_2 > 10\|s^{(k+1)}\|_2, \\ \gamma^{(k)} & \text{otherwise} \end{cases}$$

starting from an initial guess  $\gamma^{(0)} > 0$  similarly as in [26, 105]. Since  $b^{(k+1)}$  is related to the dual iterates  $p^{(k+1)}$  via  $b^{(k+1)} = \gamma^{(k)}p^{(k+1)}$  so far, this requires to rescale  $b^{(k+1)}$  by the factor  $\gamma^{(k+1)}/\gamma^{(k)}$  at the end of each iteration.

Although the convergence of ADMM with this updating scheme is no longer guaranteed by Theorem 3.6.2, we will show later on that it performs well in practice for our problem at hand. Sets of conditions for which the convergence is still guaranteed theoretically for adapted  $\gamma$  can be found in [105].

### 3.6.2 Minimization of the Discrete IDIV Model

Now, we want to apply the ADMM Algorithm 3.6.3 to the IDIV model (3.38), where  $H_f$  and  $\phi$  have been defined in (3.33),  $K \in \mathbb{R}^{M,N}$  is a positive matrix and  $f > 0$ . For this purpose, we consider the equivalent unconstrained problem

$$\operatorname{argmin}_{u \in \mathbb{R}^N} \left\{ \underbrace{H_f(Ku)}_{=:f_1(Ku)} + \underbrace{\lambda\phi(Du)}_{=:f_2(Du)} + \underbrace{\iota_{\mathbb{R}_{\geq 0}^N}(u)}_{=:f_3(u)} \right\} \quad (3.49)$$

and set

$$\Psi(u) = \langle 0, u \rangle, \quad \Phi(w) = \sum_{i=1}^3 f_i(w_i) \quad \text{and} \quad C = \begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} = \begin{pmatrix} K \\ D \\ I_N \end{pmatrix}.$$

The ADMM scheme resulting from Algorithm 3.6.3 is given as follows. To guarantee here that a solution exists and conditions (I) and (II) of Theorem 3.6.2 are fulfilled, we impose the additional assumptions that  $K \neq 0_{M,N}$  and  $\mathcal{N}(K) \cap \mathcal{N}(D) = \{0\}$  similarly as in [83]. If  $K$  is further injective, it is even ensured that (3.49) has a unique solution.

**Algorithm 3.6.4 (PIDSplit+)**

*Initialization:*  $w_1^{(0)} = Kf$ ,  $w_2^{(0)} = Df$ ,  $w_3^{(0)} = f$ ,  $b_1^{(0)} = 0$ ,  $b_2^{(0)} = 0$ ,  $b_3^{(0)} = 0$  and  $\gamma > 0$ .  
*For*  $k = 1, \dots$  *repeat until a stopping criterion is reached:*

$$\begin{aligned} u^{(k+1)} &= \underset{u \in \mathbb{R}^N}{\operatorname{argmin}} \left\{ \|b_1^{(k)} + Ku - w_1^{(k)}\|_2^2 + \|b_2^{(k)} + Du - w_2^{(k)}\|_2^2 + \|b_3^{(k)} + u - w_3^{(k)}\|_2^2 \right\}, \\ w_1^{(k+1)} &= \underset{w_1 \in \mathbb{R}^M}{\operatorname{argmin}} \left\{ H_f(w_1) + \frac{1}{2\gamma} \|b_1^{(k)} + Ku^{(k+1)} - w_1\|_2^2 \right\}, \\ w_2^{(k+1)} &= \underset{w_2 \in \mathbb{R}^{mN}}{\operatorname{argmin}} \left\{ \lambda \phi(w_2) + \frac{1}{2\gamma} \|b_2^{(k)} + Du^{(k+1)} - w_2\|_2^2 \right\}, \\ w_3^{(k+1)} &= \underset{w_3 \in \mathbb{R}^N}{\operatorname{argmin}} \left\{ \iota_{\mathbb{R}_{\geq 0}^N}(w_3) + \frac{1}{2\gamma} \|b_3^{(k)} + u^{(k+1)} - w_3\|_2^2 \right\}, \\ b_1^{(k+1)} &= b_1^{(k)} + Ku^{(k+1)} - w_1^{(k+1)}, \\ b_2^{(k+1)} &= b_2^{(k)} + Du^{(k+1)} - w_2^{(k+1)}, \\ b_3^{(k+1)} &= b_3^{(k)} + u^{(k+1)} - w_3^{(k+1)}. \end{aligned}$$

*Output:*  $\hat{u}_{\text{approx}} := w_3^{(k+1)}$ .

Due to the constraint  $w = Cu$  we know that the limits have to fulfill  $w_3 = u$ . Thus, we can choose  $w_3^{(k+1)}$  for  $\hat{u}_{\text{approx}}$  instead of  $u^{(k+1)}$ . Since  $w_3^{(k)}$  is nonnegative for all  $k$ , this guarantees that  $\hat{u}_{\text{approx}} \geq 0$ . In case that  $K$  is the identity, note that the nonnegativity of  $u$  is already ensured by  $H_f$  so that the terms and variables corresponding to the nonnegativity constraint can be neglected.

The great benefit of PIDSplit+ is that the involved subproblems are easy to solve. The iterates  $w_1^{(k+1)}, w_2^{(k+1)}, w_3^{(k+1)}$  can be computed by the subsequent direct formulas and since the matrix  $C^T C = I + K^T K + D^T D$  is always positive definite,  $u^{(k+1)}$  is the solution of the following linear system of equations:

$$\begin{aligned} u^{(k+1)} &= (I + K^T K + D^T D)^{-1} \left( K^T (w_1^{(k)} - b_1^{(k)}) + D^T (w_2^{(k)} - b_2^{(k)}) + (w_3^{(k)} - b_3^{(k)}) \right), \\ w_1^{(k+1)} &= \frac{1}{2} \left( b_1^{(k)} + Ku^{(k+1)} - \gamma + \sqrt{\left( b_1^{(k)} + Ku^{(k+1)} - \gamma \right)^2 + 4\gamma f} \right), \\ w_2^{(k+1)} &= \operatorname{shrink}_{\gamma\lambda}(b_2^{(k)} + Du^{(k+1)}), \end{aligned} \tag{3.50}$$

$$w_3^{(k+1)} = \begin{cases} b_3^{(k)} + u^{(k+1)} & \text{if } b_3^{(k)} + u^{(k+1)} \geq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (3.51)$$

The operator  $\text{shrink}_\lambda : \mathbb{R}^{mN} \rightarrow \mathbb{R}^{mN}$  is usually referred to as *coupled shrinkage* and is given componentwise with  $p := (p_1, \dots, p_m)^T$ ,  $p_k := (p_{\ell+(k-1)N})_{\ell=1}^N$ , as

$$\text{shrink}_\lambda(p_k(i)) := \begin{cases} p_k(i) - \lambda \frac{p_k(i)}{\sqrt{\sum_{j=1}^m p_j^2(i)}} & \text{if } \sqrt{\sum_{j=1}^m p_j^2(i)} \geq \lambda, \\ 0 & \text{otherwise} \end{cases}$$

for  $k \in \{1, \dots, m\}$ ,  $i \in \{1, \dots, N\}$ , compare, e.g., [186]. Supposing reflecting boundary conditions for our image, a separable blur kernel  $K$  and  $D$  to be the first gradient discretization in (2.41), the matrix  $I + K^T K + D^T D$  can be diagonalized using the discrete cosine transform DCT-II, cf., e.g., [104, 164]. Thus, solving the corresponding linear system of equations requires in this case only  $\mathcal{O}(N \log N)$  arithmetic operations. In all other cases a (preconditioned) conjugate gradient (PCG) method can be applied due to the symmetry and positive definiteness of  $I + K^T K + D^T D$ .

**Alternative algorithms** For determining a minimizer of (3.49) also other algorithms have been proposed in the literature. For example in [ST10] an alternative splitting algorithm has been provided for the case  $K = I_N$ . Moreover, in [82] Figueiredo and Bioucas-Dias first proposed an algorithm for the IDIV model (3.49) without the constraint  $u \geq 0$ , which has been added in [83]. The final algorithm of the authors can be deduced from the ADMM Algorithm 3.6.3 for  $\Psi(u) = \langle 0, u \rangle$  and  $\Phi(w) = \sum_{i=1}^3 f_i(w_i)$  with

$$\operatorname{argmin}_{u \in \mathbb{R}^N} \left\{ \underbrace{\langle 1, Ku - f \log(Ku) \rangle}_{=: f_1(Ku)} + \underbrace{\lambda \phi(Du)}_{=: f_2(u)} + \underbrace{\iota_{\mathbb{R}_{\geq 0}^N}(u)}_{=: f_3(u)} \right\},$$

i.e.,  $(C_1, C_2, C_3) := (K, I_N, I_N)$ . This splitting differs from the one used for the deduction of PIDSplit+ by the fact that the operator  $D$  is contained in  $f_2$  rather than the matrix  $C$ . The resulting algorithm is given as follows:

#### Algorithm 3.6.5 (PIDAL)

*Initialization:*  $w_1^{(0)} = Kf$ ,  $w_2^{(0)} = f$ ,  $w_3^{(0)} = f$ ,  $b_1^{(0)} = 0$ ,  $b_2^{(0)} = 0$ ,  $b_3^{(0)} = 0$  and  $\gamma > 0$ .

*For*  $k = 1, \dots$  *repeat until a stopping criterion is reached:*

$$\begin{aligned} u^{(k+1)} &= \operatorname{argmin}_{u \in \mathbb{R}^N} \left\{ \|b_1^{(k)} + Ku - w_1^{(k)}\|_2^2 + \|b_2^{(k)} + u - w_2^{(k)}\|_2^2 + \|b_3^{(k)} + u - w_3^{(k)}\|_2^2 \right\}, \\ w_1^{(k+1)} &= \operatorname{argmin}_{w_1 \in \mathbb{R}^M} \left\{ \langle 1, w_1 - f \log w_1 \rangle + \frac{1}{2\gamma} \|b_1^{(k)} + Ku^{(k+1)} - w_1\|_2^2 \right\}, \end{aligned} \quad (3.52)$$



$$w_2^{(k+1)} = \operatorname{argmin}_{w_2 \in \mathbb{R}^N} \left\{ \lambda \phi(Dw_2) + \frac{1}{2\gamma} \|b_2^{(k)} + u^{(k+1)} - w_2\|_2^2 \right\}, \quad (3.53)$$

$$w_3^{(k+1)} = \operatorname{argmin}_{w_3 \in \mathbb{R}^N} \left\{ \iota_{\mathbb{R}_{\geq 0}^N}(w_3) + \frac{1}{2\gamma} \|b_3^{(k)} + u^{(k+1)} - w_3\|_2^2 \right\}, \quad (3.54)$$

$$\begin{aligned} b_1^{(k+1)} &= b_1^{(k)} + Ku^{(k+1)} - w_1^{(k+1)}, \\ b_2^{(k+1)} &= b_2^{(k)} + u^{(k+1)} - w_2^{(k+1)}, \\ b_3^{(k+1)} &= b_3^{(k)} + u^{(k+1)} - w_3^{(k+1)}. \end{aligned}$$

Output:  $\hat{u}_{\text{approx}} := w_3^{(k+1)}$ .

Here, setting  $\hat{u}_{\text{approx}}$  to  $w_3^{(k+1)}$  instead of  $u^{(k+1)}$  guarantees again that  $\hat{u}_{\text{approx}} \geq 0$ . Besides, the steps (3.52) and (3.54) of this algorithm are the same as for PIDSplit+ and can be solved by the explicit formulas (3.50) and (3.51). For the first step a linear system of equations has again to be solved:

$$u^{(k+1)} = (2I + K^T K)^{-1} \left( K^T(w_1^{(k)} - b_1^{(k)}) + (w_2^{(k)} - b_2^{(k)}) + (w_3^{(k)} - b_3^{(k)}) \right).$$

The matrix  $2I + K^T K$  is positive definite and therefore invertible. Supposing again reflecting boundary conditions for our images and a spatially invariant, separable blur kernel  $K$  this linear system of equations can be solved efficiently by the discrete cosine transform DCT-II equally as the corresponding step of PIDSplit+. However, for the computation of  $w_2^{(k+1)}$  we have to solve a ROF-like denoising problem. Since here no closed form solution exists, the authors of [82, 83] suggest to use Chambolle's iterative algorithm [47] with a fixed number of iterations and the internal variables initialized with the results of the former iteration cycle. Alternatively, also Nesterov's algorithm [151], FISTA [17, 18] of Beck and Teboulle or the alternating split Bregman (ADMM) algorithm [96] could be used.

Aside from the ADMM algorithms PIDSplit+ and PIDAL the primal-dual hybrid gradient method (PDHG) [51, 77, 231] provides another possibility for the minimization of the IDIV model (3.49), see also [SST11]. This algorithm can be particularly interesting in cases where  $u^{(k+1)}$  has to be computed iteratively and the convergence is too slow.

Another option is to apply the EM-TV algorithm presented by Sawatzky et al. in [179, 180]. This algorithm is an extension of the standard EM algorithm [63, 150, 170, 191] and consists of alternating an EM step with a TV step as follows. The starting value  $u^{(0)} = f$  is chosen analogously to the other algorithms, here.

**Algorithm 3.6.6 (EM-TV)**

Initialization:  $u^{(0)} = f$ .

For  $k = 1, \dots$  repeat until a stopping criterion is reached:

$$u^{(k+\frac{1}{2})} = u^{(k)} \frac{1}{K^T \mathbf{1}_M} K^T \left( \frac{f}{K u^{(k)}} \right) \quad (\text{EM step}), \quad (3.55)$$

$$u^{(k+1)} = \operatorname{argmin}_{v \in \mathbb{R}^N} \left\{ \frac{1}{2} \sum_{i=1}^N \frac{(K^T \mathbf{1}_M)_i (v_i - u_i^{(k+\frac{1}{2})})^2}{u_i^{(k)}} + \lambda \phi(Dv) \right\} \quad (\text{TV step}). \quad (3.56)$$

As indicated by its name the TV step requires the minimization of the ROF-like functional

$$\operatorname{argmin}_{w \in \mathbb{R}^N} \left\{ \frac{1}{2} \|w - h\|_2^2 + \lambda \phi(D\Lambda w) \right\}$$

$$\text{with } w := \sqrt{\frac{K^T \mathbf{1}_M}{u^{(k)}}} v, \quad h := \sqrt{\frac{K^T \mathbf{1}_M}{u^{(k)}}} u^{(k+\frac{1}{2})}, \quad \Lambda := \operatorname{diag} \left( \sqrt{\frac{u^{(k)}}{K^T \mathbf{1}_M}} \right).$$

This must again be done iteratively, e.g., by one of the above mentioned algorithms. As for PIDAL the authors of [179, 180] suggest using Chambolle's algorithm [47].

To study the convergence of EM-TV theoretically, a damped EM-TV variant was introduced in [35, 36]. Here, the TV-step was extended by additional damping parameters  $\omega_k \in (0, 1]$  as follows:

$$u^{(k+1)} = \operatorname{argmin}_{v \in \mathbb{R}^N} \left\{ \frac{1}{2} \sum_{i=1}^N \frac{(K^T \mathbf{1}_M)_i (v_i - (\omega_k u^{(k+\frac{1}{2})} + (1 - \omega_k) u_i^{(k)}))^2}{u_i^{(k)}} + \omega_k \lambda \phi(Dv) \right\}.$$

It could be shown that damped EM-TV is a forward-backward splitting algorithm, where the convergence is ensured whenever the  $\omega_k$  are small enough. Nevertheless, EM-TV, which corresponds to the case  $\omega_k = 1$ , was reported to perform well in practice as long as the regularization parameter  $\lambda$  was not too large, cf. [35, 36].

## 3.7 Numerical Experiments

In the following, we compare the presented algorithms in terms of computational speed and show restoration results of the IDIV model (3.38) for images corrupted by different types of noise. To this end, all algorithms have been implemented in MATLAB and the computations are performed on a computer with an Intel Core i5-2520M Processor (3M Cache, 2.50 GHz) and 8 GB physical memory. If not stated otherwise, the values of the original images lie in the interval  $[0, 255]$  and the images are visualized accordingly. Moreover, to guarantee a consistent coloring for each example the corrupted images and results are depicted with the same color map.

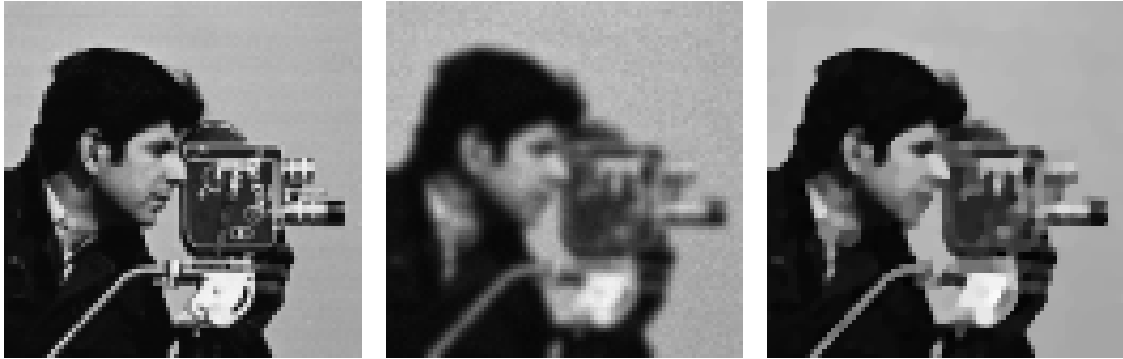


Figure 3.4: *Left:* Part of the 'cameraman' image of size  $84 \times 84$ . *Middle:* Image corrupted by a Gaussian blur (standard deviation  $\sigma = 1$ ) and Poisson noise ( $I_{\max} = 3000$ ). *Right:* Deblurred image by IDIV-TV with  $\lambda = 0.008$ .

### 3.7.1 Blurred Images Contaminated by Poisson Noise

To start with, we focus on blurred images corrupted by Poisson noise, which we restore using the discretized IDIV-TV model. Therefore, the discrete gradient  $D$  is set to be the first matrix in (2.41), which involves reflecting boundary conditions for the images. In the presented examples the blur kernels are two-dimensional symmetric Gaussians and thus, separable. Details on the definition of the corresponding matrices  $K$  can, e.g., be found in [104]. In those cases where Poisson noise is added manually to a blurred image  $U$ , the image is first scaled by  $I_{\max}/(10^{12} U_{\max})$ , where  $U_{\max}$  is the maximal value of  $U$  and  $I_{\max} \in [0, 10^{12}]$  denotes a specified maximal intensity. Then, the MATLAB routine `imnoise(I, 'poisson')` is applied and afterwards we scale back again.

Due to the structure of  $K$  and  $D$  both PIDSplit+ and PIDAL can be implemented efficiently using the discrete cosine transform (DCT-II) as outlined in the former section. For solving the ROF-like subproblems (3.53) and (3.56) of PIDAL and EM-TV we use the gradient descent reprojection algorithm presented by Chambolle in [48] with a fixed step size of 0.249. In [48] this algorithm was reported to perform better than the standard Chambolle algorithm from [47]. To initialize the starting values of the resulting inner iteration loops the final values of the former iteration cycle are used as in [37, 83]. Moreover, we perform a fixed number of inner iterations per loop.

Our first example in Figure 3.4 shows a part of the 'cameraman' image, which has been blurred by a Gaussian kernel of standard deviation 1 and contaminated by Poisson noise. The restored image by IDIV-TV is displayed in Figure 3.4 (right). Furthermore, the computation times of the different algorithms for computing this result up to a specified maximal pixel difference are reported in Table 3.1. Except for the starting value  $\gamma^{(0)}$  of PIDSplit+ with iteratively adapted  $\gamma$  all parameters have been optimized

### 3. Variational Approaches Facing Poisson and Multiplicative Noise

Algorithms	Parameter $\gamma$	Number of inner iterations	Number of outer iterations	Computation times
$\ \hat{u}^{(k)} - \hat{u}_{\text{ref}}\ _{\infty} < 3$				
EM-TV	–	2	567	0.65 sec
PIDAL	745	6	98	0.35 sec
PIDSplit+	1460	–	176	<b>0.31 sec</b>
PIDSplit+ ( $\gamma$ adapt.)	( $\gamma^{(0)} = 1$ )	–	(370)	(0.86 sec)
$\ \hat{u}^{(k)} - \hat{u}_{\text{ref}}\ _{\infty} < 1$				
EM-TV	–	3	1278	1.92 sec
PIDAL	680	10	161	0.85 sec
PIDSplit+	1460	–	257	<b>0.45 sec</b>
PIDSplit+ ( $\gamma$ adapt.)	( $\gamma^{(0)} = 1$ )	–	(490)	(1.17 sec)
$\ \hat{u}^{(k)} - \hat{u}_{\text{ref}}\ _{\infty} < 0.2$				
EM-TV	–	9	2892	10.8 sec
PIDAL	670	33	240	3.55 sec
PIDSplit+	1330	–	407	<b>0.72 sec</b>
PIDSplit+ ( $\gamma$ adapt.)	( $\gamma^{(0)} = 1$ )	–	(657)	(1.55 sec)

Table 3.1: Computation times required by the algorithms to compute the sufficiently converged reference result  $\hat{u}_{\text{ref}}$  in Figure 3.4 up to the specified maximal pixel differences.

to provide the shortest possible times and are listed in the table. Regarding the stopping criterion we stopped when the maximal pixel difference to the reference result became smaller than the specified bound. Although PIDSplit+ required more (outer) iterations than PIDAL, it always performed best for fixed but optimized  $\gamma$ . Moreover, PIDAL was always faster than EM-TV. Especially for higher accuracies the factors between the three algorithms increase.

Since it is in general not clear how to choose an optimal value  $\gamma$  for PIDSplit+ and PIDAL, we also included the times required by PIDSplit+ with iteratively adapted  $\gamma$  for a non-optimized starting value of 1. Note that this value is on purpose far away from the fixed, optimized values. As the experiment shows the computation time is here always larger than the time required by PIDSplit+ with fixed but optimized  $\gamma$ . Nevertheless, for a maximal pixel error of 1 it is already faster than EM-TV and for an accuracy of 0.2 it also outperformed PIDAL.

To compare the two variants of PIDSplit+ in more detail we included Figure 3.5. Here,

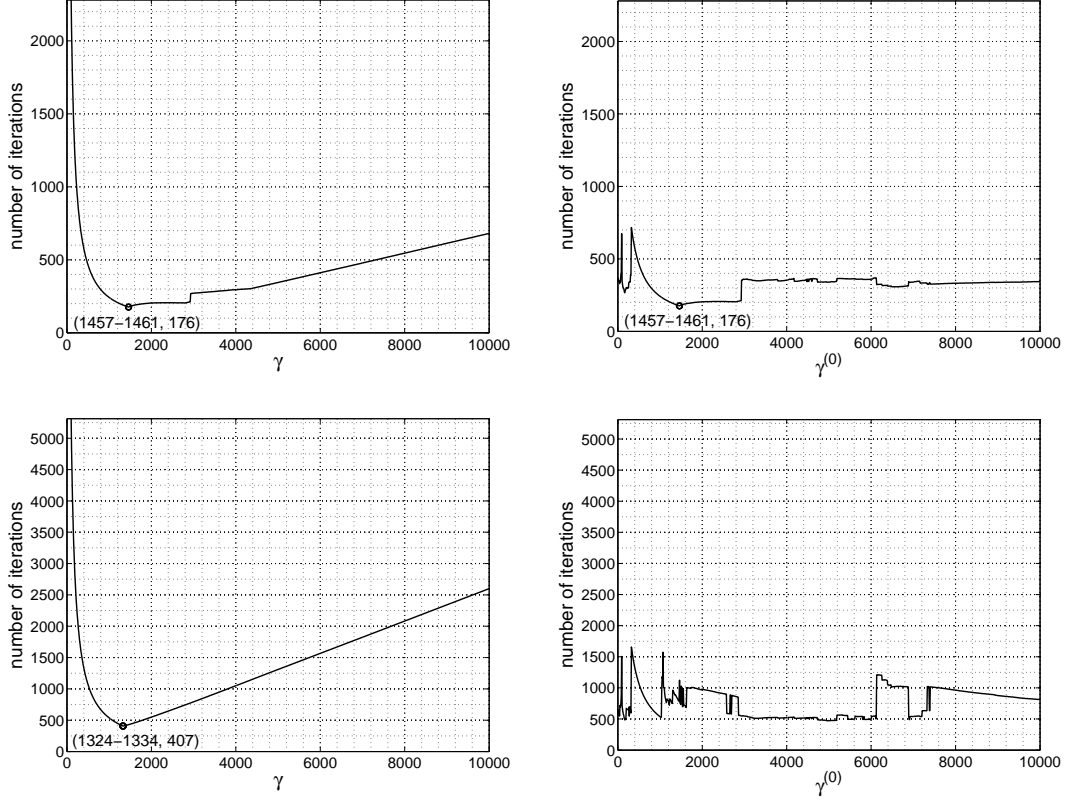


Figure 3.5: Influence of  $\gamma$  on the number of iterations required by PIDSplit+ to compute the restored image of Figure 3.4 up to a maximal difference per pixel of 3 (top) and 0.2 (bottom). *Left:* Fixed  $\gamma$  for all iterations. *Right:* Iteratively adapted  $\gamma$  with  $\gamma^{(0)}$  as initial starting value.

the required number of iterations by PIDSplit+ is reported for different accuracies in dependence on a fixed value  $\gamma$  and iteratively adapted  $\gamma$  with varying starting values  $\gamma^{(0)}$ , respectively. The results show that with fixed  $\gamma$  a too small value or a much too large value  $\gamma$  can significantly affect the performance of PIDSplit+, especially if more accurate results are required. Although the presented adaptation for  $\gamma$  does not help to decrease the number of iterations in general, it efficiently bounds the maximally required number of iterations so that the choice of an initial value  $\gamma^{(0)}$  is far less critical than choosing a fixed value  $\gamma$ . If we chose for example the value  $\gamma^{(0)} = 1$  used in Table 3.1 for the non-adapted variant, we would not get a result in a reasonable amount of time.

A restoration result of IDIV-TV and a comparison of the speed of the algorithms for the whole 'cameraman' image is given in Figure 3.6 and Table 3.2. Also here, PIDSplit+ outperformed the other algorithms for fixed but optimized  $\gamma$ . With iteratively adapted  $\gamma$  and the initial value  $\gamma^{(0)} = 1$  PIDSplit+ is roughly only half as fast but still



Figure 3.6: *Left:* Original 'cameraman' image of size  $256 \times 256$ . *Middle:* Image corrupted by a Gaussian blur (standard deviation  $\sigma = 1.3$ ) and Poisson noise ( $I_{\max} = 1000$ ). *Right:* Deblurred image by IDIV-TV with  $\lambda = 0.01$ .

Algorithms	Parameter $\gamma$	Number of inner iterations	Number of outer iterations	Computation times
$\ \hat{u}^{(k)} - \hat{u}_{\text{ref}}\ _{\infty} < 1$				
EM-TV	—	5	1640	25.2 sec
PIDAL	705	16	191	11.2 sec
PIDSplit+	1245	—	297	<b>6.0 sec</b>
PIDSplit+ ( $\gamma$ adapt.)	( $\gamma^{(0)} = 1$ )	—	(550)	(13.3 sec)

Table 3.2: Computation times required by the algorithms to compute the sufficiently converged reference result in Figure 3.6 up to a maximal pixel difference of 1.

significantly faster than EM-TV with an optimized number of inner iterations.

To apply PIDSplit+ also to real confocal data Figure 3.7 shows at top left an image of a blurred human cell nucleus of size  $401 \times 311$  kindly provided by Prof. T. Cremer (LMU Munich). The corresponding blur kernel has been estimated by S. Remmele (Medical Centre Mannheim) and fits well a symmetric Gaussian of standard deviation  $\sigma = 3.8$ . In contrast to the former examples much more iterations are necessary here to obtain equally small errors as for the 'cameraman' image. However, we get already after a few iterations a visually pleasant result as illustrated in Figure 3.7. An interesting observation is also that if we solved the IDIV-TV model for, e.g.,  $\lambda = 0.002$  without the additional nonnegativity constraint  $u \geq 0$ , the result would contain negative value down to  $-249$ , although the blurred image was nonnegative and all entries of  $K$  are nonnegative, too.

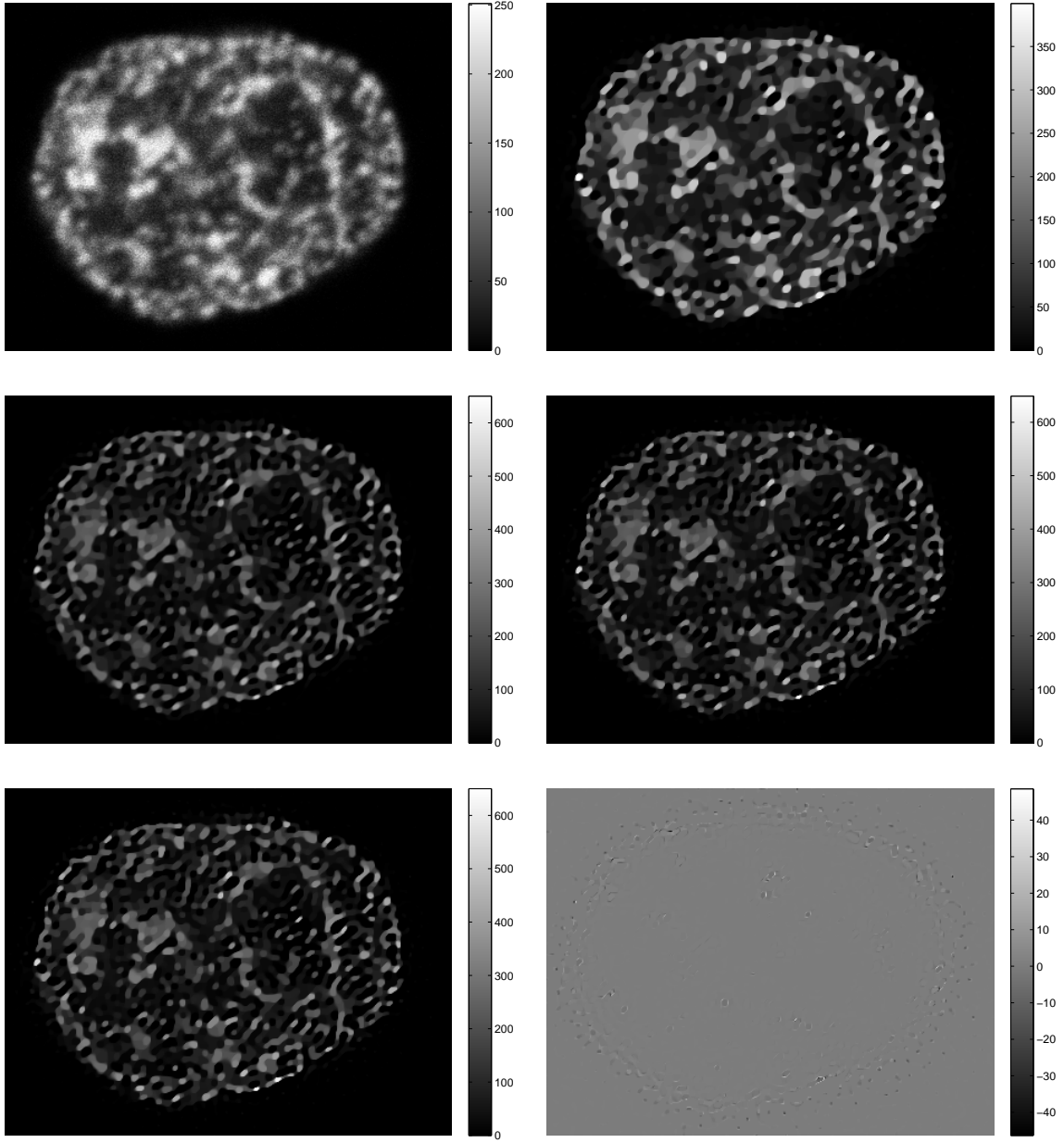


Figure 3.7: *Top:* Blurry cell image of a human nucleus and result by PIDSplit+ with  $\lambda = 0.01$  and  $\gamma = 1000$  after 10.000 iterations. *Middle/bottom:* Results by PIDSplit+ with  $\lambda = 0.002$  and  $\gamma = 2000$  after 200 (middle left), 3000 (middle right) and 10.000 iterations (bottom left). Already the result after 200 iterations is visually very close to the other images, although there are still some differences between the results after 3000 and 10.000 iterations as the difference image at bottom right shows.

### 3.7.2 Images Corrupted by Multiplicative Noise

Next, we will focus on denoising results for data corrupted by multiplicative noise, where the noise is mostly Gamma distributed with density (3.5). For the restoration we will apply aside from the discrete IDIV-TV model with  $K = I_N$  also the discrete IDIV-NL approach introduced in Section 3.5. For this purpose, the involved matrices  $D$  are defined as follows:

- i) For the IDIV-TV model and 2D images we set again  $D$  to be the first gradient discretization in (2.41). In case of 1D signals  $D$  is just the forward-difference matrix  $H_1$  defined in (2.42).
- ii) For the discrete IDIV-NL model the following construction is applied, cf. [91]: Initially, we start with a zero weight matrix  $w \in \mathbb{R}^{N,N}$ . For every image pixel  $i$  we compute for all  $j$  within a search window of size  $\omega \times \omega$  around  $i$  the distances

$$d_a(i, j) := \sum_{h_1 = -\lceil \frac{l-1}{2} \rceil}^{\lceil \frac{l-1}{2} \rceil} \sum_{h_2 = -\lceil \frac{l-1}{2} \rceil}^{\lceil \frac{l-1}{2} \rceil} g_a(h_1, h_2) \left( \tilde{f}(i + (h_1, h_2)) - \tilde{f}(j + (h_1, h_2)) \right)^2,$$

where  $\tilde{f} := \log(f)$  and  $g_a$  represents a discrete normalized Gaussian of mean 0 and standard deviation  $a$ . The parameter  $l$  controls here the size of the image parts being compared. For a predefined bound  $\tilde{m}$  we select the  $k \leq \tilde{m}$  'neighbors'  $j \neq i$  of  $i$  for which  $d_a(i, j)$  takes the smallest values and the number of nonzero elements in the row  $w(j, \cdot)$  is smaller than  $2\tilde{m}$ . Here, we set  $w(i, j) = w(j, i) = 1$ , which causes several weights  $w(j, \cdot)$  to be already non-zero before we actually reach pixel  $j$ . To avoid that the number of non-zero weights becomes too large, we set the number of chosen neighbors to  $k := \min\{\tilde{m}, 2\tilde{m} - r\}$  with  $r$  being the number of non-zero weights  $w(i, \cdot)$  before the selection. With regard to (3.15) and (3.32) we finally construct the matrix  $D \in \mathbb{R}^{mN, N}$  with  $m = 2\tilde{m}$  so that  $D$  consists of  $m$  blocks of size  $N \times N$ , each having maybe some zero rows and rows with  $-1$  as diagonal element plus one additional nonzero value 1 whose position is determined by the nonzero weights  $w(i, j)$ .

*Remark:* Compared to [91] the weights are computed here with respect to the logarithmically transformed image  $f$ . This is again motivated by the multiplicative nature of the noise similarly as the logarithmic transformation for the deduction of the log-model in Subsection 3.3.2.

Our first example in Figure 3.8 shows a restored 1D signal by IDIV-TV, which was corrupted by multiplicative Gamma noise with  $L = 25$ . Here, the signal and noise level were chosen in accordance with the experiments in [192]. Although the IDIV-TV model was originally designed for Poisson noise, it restores the signal quite nicely



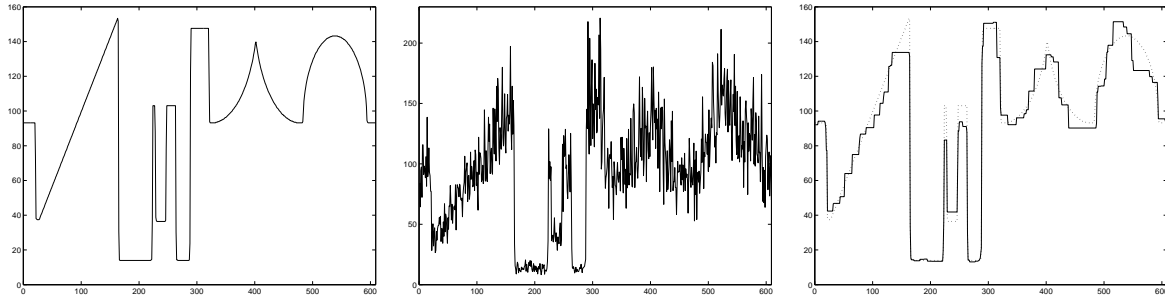


Figure 3.8: *Left*: Original signal. *Middle*: Noisy signal corrupted by multiplicative Gamma noise ( $L = 25$ ). *Right*: Denoised signal by IDIV-TV with  $\lambda = 0.52$ .

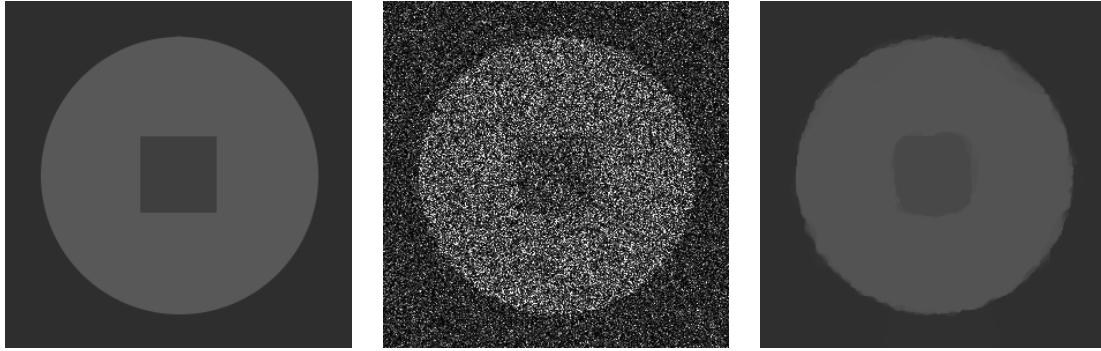


Figure 3.9: *Left*: Reproduction of a test image in [10]. *Middle*: Noisy image corrupted by multiplicative Gamma noise ( $L = 1$ ). *Right*: Denoised image by IDIV-TV with  $\lambda = 2.5$ .

except for the usual staircasing artifacts typical for TV regularization.

Similar observations can also be made for the 2D example in Figure 3.9. Here, we restored a reproduction of a synthetic test image from [10]. Compared to the denoising results presented in [10] our result looks very promising. All noise has been removed and the original image has been mostly recovered. Only the edges are not perfect due to the severe degradation of the noisy image.

Next, we want to compare results of the IDIV-TV model and the IDIV-NL model for natural images. For this purpose, we applied both methods to a part of the 'Barbara' image corrupted by multiplicative Gamma noise with  $L = 25$  as shown in Figure 3.10. The figure and the noise level were chosen so that the experiments are comparable with the ones in [192]. For this image IDIV-TV removed again all noise. However, a lot of staircasing artifacts are clearly visible. The denoised image by the IDIV-NL model is significantly better, here. Due to the semi-local adaptivity of the matrix  $D$  especially the fine structures of the textures are much better restored.

These observations are similarly true for the example provided in Figure 3.11. Here, a



Figure 3.10: *Top:* Original image (left) and noisy version corrupted by multiplicative Gamma noise with  $L = 25$  (right). *Bottom:* Denoised image by IDIV-TV with  $\lambda = 0.2$  (left) and the result by the IDIV-NL model with  $\lambda = 0.12$ ,  $l = 9$ ,  $\omega = 11$ ,  $a = 3$  and  $\tilde{m} = 5$  (right).

part of the 'cameraman' image has been restored, which was corrupted by multiplicative noise with  $L = 4$ , i.e., a significantly higher noise level.

To conclude this chapter we apply our methods also to a real-world SAR image, which is contaminated by multiplicative Nakagami distributed noise with density (3.7). Although our methods have not been designed for this type of noise, we get quite good results by both models as visible in Figure 3.12. As before, the IDIV-NL method provides the best result, here.



Figure 3.11: *Top*: Original image (left) and noisy version corrupted by multiplicative Gamma noise with  $L = 4$  (right). *Bottom*: Restored image by IDIV-TV with  $\lambda = 0.9$  (left) and the result by the IDIV-NL model with  $\lambda = 0.6$ ,  $l = 5$ ,  $\omega = 21$ ,  $a = 1.5$  and  $\tilde{m} = 5$  (right).

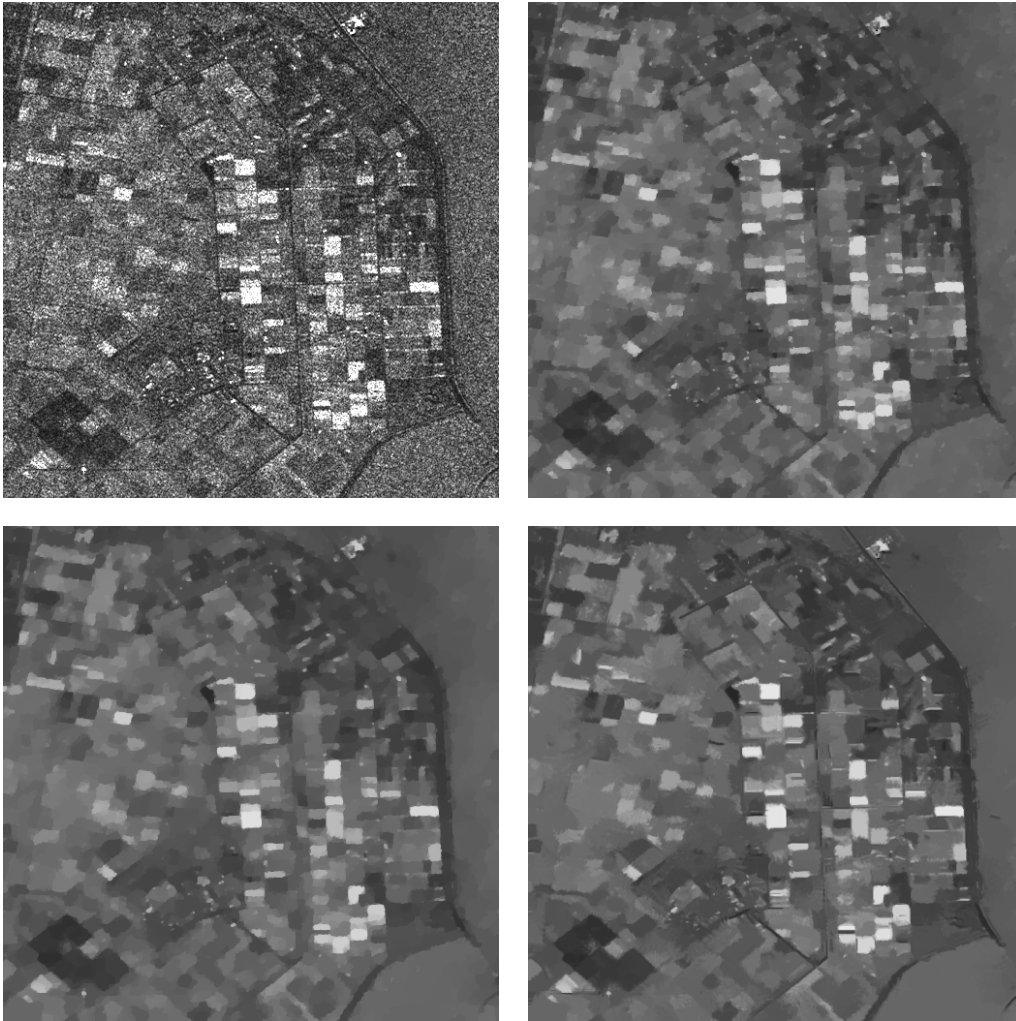


Figure 3.12: *Top left:* Original multi-look SAR image (copyright by [78]). *Top right/bottom left:* Restored images by IDIV-TV with  $\lambda = 0.4$  (top right) and  $\lambda = 0.5$  (bottom left). *Bottom right:* Result by the IDIV-NL approach with  $\lambda = 0.2$ ,  $l = 9$ ,  $\omega = 13$ ,  $a = 3$  and  $\tilde{m} = 5$ .

# CHAPTER 4

## Nonlocal Filters for Removing Multiplicative Noise

### Outline

---

4.1	Nonlocal Means Filter and Motivation . . . . .	<b>106</b>
4.2	The Similarity Measure of Deledalle et al. . . . .	<b>107</b>
4.2.1	Properties in the Presence of Additive Noise . . . . .	<b>109</b>
4.2.2	Properties in the Presence of Multiplicative Noise . . . . .	<b>110</b>
4.3	A New Similarity Measure for Multiplicative Noise . . . . .	<b>115</b>
4.4	Relations to Other Similarity Measures . . . . .	<b>121</b>
4.5	Nonlocal Filtering Facing Multiplicative Noise . . . . .	<b>126</b>
4.5.1	Nonlocal Filters by Maximum Likelihood Estimation . . . . .	<b>126</b>
4.5.2	Definition of the Weights . . . . .	<b>128</b>
4.5.3	Updating the Similarity Neighborhoods . . . . .	<b>131</b>
4.6	Numerical Results . . . . .	<b>132</b>

---

Up to now, we have examined different variational models for the removal of Poisson and multiplicative Gamma noise. In this chapter we now turn our attention to nonlocal filtering approaches and show how it is possible to define efficient nonlocal filters for images corrupted by multiplicative noise. Here, a sticking point is the definition of a suitable similarity measure for determining whether the original noise-free values of two given noisy pixels have been the same or not.

To start with, we shortly introduce the nonlocal means filter [39] of Buades et al. and report on existing approaches for adapting this filter for images corrupted by non-Gaussian noise.

## 4.1 Nonlocal Means Filter and Motivation

If small enough regions, so-called *patches*, of an image are compared with each other, we will typically find a lot of similar ones in a natural image. The idea of the *nonlocal (NL) means filter* [39] proposed by Buades et al. in 2005 is to exploit these self-similarities for image denoising. For this purpose, each pixel  $f_i$  of a given noisy image  $f$  is compared together with its neighboring pixels to other image patches and for each comparison a weight is assigned depending on the similarity of the patches. The restored pixel  $\hat{u}_i$  is finally the weighted average of the central pixels of these patches using the obtained weights. For a discrete image  $f \in \mathbb{R}^{m,n}$ ,  $N = mn$ , this means in detail that

$$\hat{u}_i = \frac{1}{C_i} \sum_{j=1}^N w_{NL}(i, j) f_j \quad (4.1)$$

with  $C_i := \sum_{j=1}^N w_{NL}(i, j)$ . If the image patches with centers  $f_i, f_j$  are given by  $f_{i+I}$ , resp.  $f_{j+I}$  for  $I$  denoting an appropriate index set, the weights are set to

$$w_{NL}(i, j) = \exp \left( -\frac{1}{h} \sum_{k \in I} g_{a,k} |f_{i+k} - f_{j+k}|^2 \right).$$

The parameter  $h > 0$  is used here to control the amount of filtering. The vector  $g_a = (g_{a,k})_{k \in I}$  usually represents a sampled two dimensional Gaussian kernel with mean zero and standard deviation  $a$ , which steers the influence of neighboring pixels on the weights.

To exploit patch similarities for image denoising is also the central idea of the UINTA filter [13, 14], which has been proposed by Awate and Whitaker simultaneously with the NL means filter. However, due to its simplicity, the NL means filter is much better known and has been extensively studied and further improved in various directions in the past years. An analysis and comparison with other state-of-the-art image denoising methods as well as an overview of recent developments in this area can for example be found in [40].

Among other improvements, several authors proposed different approaches to adapt the nonlocal means filter to noise statistics. In [140, 141] Mäkitalo et al. studied for example variance-stabilizing transformations and their inverses which can be applied to data corrupted by Poisson and multiplicative Rayleigh noise before and after the standard NL means approach is applied. Kervrann et al. proposed the so-called *Bayesian NL means filter* [117], which provided a first possibility to incorporate the statistics of the noise directly into the design of the filter. In [54] this filter was applied for the removal of speckle noise in ultrasound images. An approach for Rician noise has been presented in [223]. Another generalization of the original NL means filter in a probabilistic framework was given by Deledalle et al. in [60]. Here, a central step was to incorporate the noise statistics in a suitable way into the weight definition of



their filters. To illustrate the basic idea, the weights of the NL means filter can be written in the form

$$w_{NL}(i, j) = \prod_{k \in I} s_{NL}(f_{i+k}, f_{j+k})^{\frac{g_{a,k}}{h}} \quad \text{with} \quad s_{NL}(x, y) := \exp(-|x - y|^2). \quad (4.2)$$

Consequently, they can be constructed by taking the product over  $s_{NL}(f_{i+k}, f_{j+k})^{\frac{g_{a,k}}{h}}$  for all pairs of pixels  $f_{i+k}$  and  $f_{j+k}$  of the two image patches. The function  $s_{NL} : \mathbb{R} \times \mathbb{R} \rightarrow (0, 1]$  can be viewed as a similarity measure, where  $s_{NL}(f_{i+k}, f_{j+k})$  is supposed to be close to one if the original noise free pixels belonging to  $f_{i+k}$  and  $f_{j+k}$  have been the same and it should be close to zero if not. For images corrupted by additive Gaussian noise,  $s_{NL}$  is known to perform well. Unfortunately, it can be far from optimal for other types of noise. Hence, the challenge is here to find a suitable noise adapted similarity measure, which can cope with different types of noise. The similarity measure proposed for general noise models in [60] was demonstrated to perform well for images corrupted by additive Gaussian noise, noise following a Nakagami-Rayleigh distribution and Poisson noise studied in [61]. However, since we show in the next section that it has unfavorable properties for multiplicative noise, the first major challenge here will be the definition of an appropriate similarity measure for this type of noise. If such a measure has been found, we can proceed with the definition of appropriate nonlocal filters, which can again be based on maximum likelihood estimation as the filters in [60].

## 4.2 The Similarity Measure of Deledalle et al.

In this section we revisit the similarity measure applied by Deledalle, Denis and Tupin in [60]. To fix the notation, all random variables are again named with capital letters and are supposed to be real-valued, continuous and defined on a fixed probability space  $(\Omega, \mathcal{F}, P)$ . Hence,  $p_X : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  stands for the probability density function of the random variable  $X$  and  $p_{Y|X}(\cdot | x)$  represents the conditional density of the random variable  $Y$  given  $X = x$ , which has been introduced in Section 3.1. To compute the densities of differently transformed random variables we will frequently apply results from probability theory reported in Appendix A.

For the subsequent considerations we suppose again that the noisy pixels  $f_i$  are realizations of independent continuous random variables  $F_i$  and the corresponding original noise free pixels  $u_i$  are realizations of independent identically distributed random variables  $U_i$ ,  $i = 1, \dots, N$ . Since we need only two pixels for the following considerations, we set  $N = 2$  for simplicity. Moreover, we assume that all  $f_i$  are contaminated by the same noise model with equal parameters.

Now, to measure by two noisy observations  $f_1, f_2$  whether  $u_1 = u_2$ , Deledalle et al. suggest in [60] to use a so-called ‘*similarity probability*’ denoted by  $p(\theta_1 = \theta_2 | f_1, f_2)$ .

In their paper,  $\theta_i$  is a parameter depending deterministically on  $u_i$  and we set  $\theta_i = u_i$ ,  $i = 1, 2$ . Since it is in general not clear what the probability or even conditional density function of  $U_1 = U_2$  given  $F_1 = f_1$ ,  $F_2 = f_2$  is, see, e.g., [101, p. 111], we start here by interpreting the 'similarity probability' as a conditional density: In [60] it is set to be

$$p(u_1 = u_2 | f_1, f_2) := \int_S p_{U_1|F_1}(u | f_1) p_{U_2|F_2}(u | f_2) du, \quad (4.3)$$

where we need to have  $p_{F_i}(f_i) > 0$ ,  $i = 1, 2$  and define  $S := \text{supp}(p_{U_i})$ , i.e.,  $S$  is the support of  $p_{U_i}$ . By the definition of the conditional density it holds that

$$\int_S p_{U_1|F_1}(u | f_1) p_{U_2|F_2}(u | f_2) du = \frac{\int_S p_{U_1,F_1}(u, f_1) p_{U_2,F_2}(u, f_2) du}{p_{F_1}(f_1) p_{F_2}(f_2)}. \quad (4.4)$$

Furthermore, we obtain by Theorem A.1 and the independence of  $(U_1, F_1)$  and  $(U_2, F_2)$  that

$$p_{U_1-U_2, U_2, F_1, F_2}(x, u, f_1, f_2) = p_{U_1, F_1}(x + u, f_1) p_{U_2, F_2}(u, f_2)$$

and thus,

$$\begin{aligned} p_{U_1-U_2, F_1, F_2}(0, f_1, f_2) &= \int_{-\infty}^{\infty} p_{U_1-U_2, U_2, F_1, F_2}(0, u, f_1, f_2) du \\ &= \int_S p_{U_1, F_1}(u, f_1) p_{U_2, F_2}(u, f_2) du. \end{aligned}$$

Inserting this in (4.4) shows with (4.3) and the independence of  $F_1, F_2$  that

$$p(u_1 = u_2 | f_1, f_2) = p_{U_1-U_2|(F_1, F_2)}(0 | f_1, f_2). \quad (4.5)$$

For this reason, we will in the following refer to the 'similarity probability' by the conditional density on the right hand side. By (4.4) it can also be expressed in the form

$$p_{U_1-U_2|(F_1, F_2)}(0 | f_1, f_2) = \frac{\int_S p_{U_1}(u) p_{U_2}(u) p_{F_1|U_1}(f_1 | u) p_{F_2|U_2}(f_2 | u) du}{p_{F_1}(f_1) p_{F_2}(f_2)}. \quad (4.6)$$

Since in general no knowledge about the distribution of the random variables  $U_i$  is given, Deledalle et al. propose to neglect the densities  $p_{U_i}$  and  $p_{F_i}$ ,  $i = 1, 2$  and to consider only

$$s_{DDT}(f_1, f_2) := \int_S p_{F_1|U_1}(f_1 | u) p_{F_2|U_2}(f_2 | u) du. \quad (4.7)$$

This measure is very close to the one investigated for block matching in [143]. Here, one may ask whether  $s_{DDT}$  can also be interpreted in terms of a conditional density function similarly to (4.5). For the case of additive noise with  $S = \mathbb{R}$  the answer is yes as we will see in the next subsection.



### 4.2.1 Properties in the Presence of Additive Noise

In this subsection suppose additionally that  $V_i$ ,  $i = 1, 2$ , are independent identically distributed random variables, which follow some noise distribution. Moreover, let  $u_i$  be corrupted by *additive noise*, i.e.,  $f_i := u_i + v_i$  and

$$F_i := U_i + V_i, \quad i = 1, 2,$$

where each  $v_i$  is a realization of the random variable  $V_i$ . Consider further all  $U_i$ ,  $V_i$ ,  $i = 1, 2$ , to be pairwise independent. Under these conditions, we can show that  $s_{DDT}$  has the following properties:

**Proposition 4.2.1** *For the described additive noise model with  $S = \text{supp}(p_{U_i}) = \mathbb{R}$  we have*

$$s_{DDT}(f_1, f_2) = p_{V_1-V_2}(f_1 - f_2) = p_{F_1-F_2|U_1-U_2}(f_1 - f_2|0) \quad \forall f_1, f_2 \in \mathbb{R}. \quad (4.8)$$

Moreover,  $s_{DDT}$  is symmetric and has the following properties:

- i)  $s_{DDT}(f, f) = \text{const.}$  for all  $f \in \mathbb{R}$ ,
- ii)  $0 \leq s_{DDT}(f_1, f_2) \leq s_{DDT}(f, f) = p_{V_1-V_2}(0)$  for all  $f_1, f_2, f \in \mathbb{R}$ .

**Proof:** By Proposition A.7 i) and Corollary A.4 i) it holds that

$$s_{DDT}(f_1, f_2) = \int_{-\infty}^{\infty} p_{V_1}(f_1 - u) p_{V_2}(f_2 - u) du = p_{V_1-V_2}(f_1 - f_2).$$

Now, applying again Proposition A.7 i) yields

$$p_{V_1-V_2}(f_1 - f_2) = p_{F_1-F_2|U_1-U_2}(f_1 - f_2|0).$$

Therefore, the listed properties follow directly from Lemma A.5. □

The last property guarantees that  $s_{DDT}(f_1, f_2)$  is maximal whenever  $f_1 = f_2$  and  $s_{DDT}$  is bounded so that it can be scaled to the interval  $[0, 1]$ , i.e., the range of  $s_{NL}$ . For the special case that  $V_i$ ,  $i = 1, 2$ , are Gaussian distributed with standard deviation  $\sigma$ , it follows that

$$s_{DDT}(f_1, f_2) = \frac{1}{2\sqrt{\pi}\sigma} \exp\left(-\frac{|f_1 - f_2|^2}{4\sigma^2}\right) = \frac{1}{2\sqrt{\pi}\sigma} (s_{NL}(f_1, f_2))^{\frac{1}{4\sigma^2}}.$$

Hence, normalizing  $s_{DDT}(f_1, f_2)$  by its maximal value  $c := \max_{x,y \in \mathbb{R}} s_{DDT}(x, y) = \frac{1}{2\sqrt{\pi}\sigma}$  leads to the weight definition

$$w(i, j) = \prod_{k \in I} \left( \frac{s_{DDT}(f_{i+k}, f_{j+k})}{c} \right)^{\frac{g_{a,k}}{h}} = \prod_{k \in I} s_{NL}(f_{i+k}, f_{j+k})^{\frac{g_{a,k}}{h'}} \quad (4.9)$$

with  $h' = 4\sigma^2 h$ . This is just the definition of the original NL means weights with a scaled filtering parameter  $h$  as similarly deduced in [60].

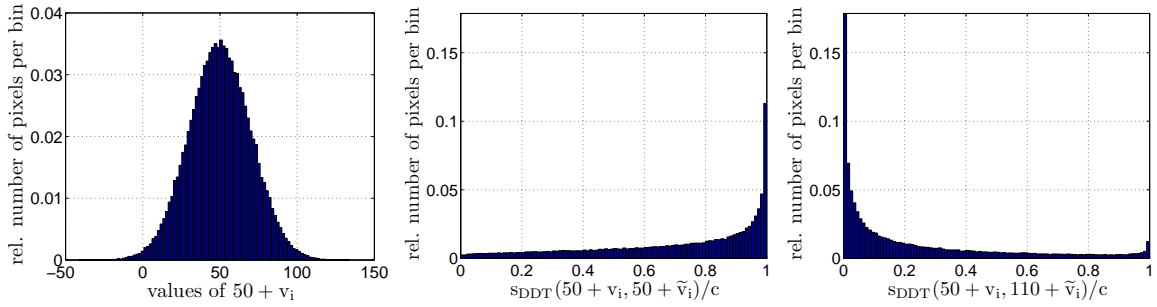


Figure 4.1: *Left*: Histogram of a constant image with gray value 50, which is corrupted by additive Gaussian noise of standard deviation 20. *Middle*: Histogram of  $(s_{DDT}(f_i, \tilde{f}_i)/c)_{i=1}^N$ , where  $f, \tilde{f}$  are images with gray value distributions as on the left. *Right*: Same as in the middle, but now  $\tilde{f}$  represents a constant image of gray value 110 corrupted by equally distributed noise.

The behavior of the similarity measure  $s_{DDT}$  for additive Gaussian noise is illustrated in Figure 4.1. The histogram on the left shows the distribution of the gray values of a constant image of gray value 50 corrupted by additive Gaussian noise of standard deviation 20. Next, the distribution of the values  $s_{DDT}(f_i, \tilde{f}_i)/c$ ,  $i = 1, \dots, N$ , is depicted for the case that both images are corrupted versions of the same constant gray value image. As expected, most values are close to one, i.e.,  $s_{DDT}/c$  detected that the corresponding noisy pixels belong to the same noise free pixel. Only a few values are close to zero, which means that the measure did not recognize that also these noisy pixels had the same initial gray value. For the histogram on the right, different gray values have been used to generate the noisy images. Here, most values  $s_{DDT}(f_i, \tilde{f}_i)/c$  are close to zero and only few pixels are falsely detected to correspond to the same noise free pixel.

#### 4.2.2 Properties in the Presence of Multiplicative Noise

Next, we want to investigate the case of multiplicative noise. For this reason, suppose again that  $V_i$ ,  $i = 1, 2$ , are independent identically distributed random variables, which follow some noise distribution. Moreover, let  $f_i$  be corrupted by *multiplicative noise*,

i.e.,  $f_i := u_i \cdot v_i$  and

$$F_i := U_i V_i, \quad i = 1, 2, \quad (4.10)$$

where each  $v_i$  is again a realization of the random variable  $V_i$ . All  $U_i, V_i, i = 1, 2$ , are further considered to be pairwise independent and we suppose that  $p_{U_i}(x) = 0, p_{V_i}(x) = 0$  for  $x < 0$ , i.e.,  $F_i > 0$  almost surely, as it is usually the case in imaging applications facing multiplicative noise.

Under these preliminaries, we obtain by Proposition A.7 ii) that for  $f_1, f_2$  with  $p_{F_i}(f_i) > 0, i = 1, 2$ , and  $S = \text{supp}(p_{U_i}) \subseteq \mathbb{R}_{\geq 0}$  the 'similarity probability' of Deledalle et al. is given by

$$\begin{aligned} p_{U_1 - U_2 | (F_1, F_2)}(0 | f_1, f_2) &= \int_S \frac{p_{U_1}(u) p_{U_2}(u)}{p_{F_1}(f_1) p_{F_2}(f_2)} p_{F_1 | U_1}(f_1 | u) p_{F_2 | U_2}(f_2 | u) du \\ &= \int_S \frac{p_{U_1}(u) p_{U_2}(u)}{p_{F_1}(f_1) p_{F_2}(f_2)} \frac{1}{u^2} p_{V_1}\left(\frac{f_1}{u}\right) p_{V_2}\left(\frac{f_2}{u}\right) du, \end{aligned} \quad (4.11)$$

which will be investigated further in the examples later on. First, we deduce the following properties of  $s_{DDT}$ :

**Proposition 4.2.2** *For the described multiplicative noise model with  $S = \mathbb{R}_{\geq 0}$  it holds that*

$$s_{DDT}(f_1, f_2) = \int_0^\infty \frac{1}{u^2} p_{V_1}\left(\frac{f_1}{u}\right) p_{V_2}\left(\frac{f_2}{u}\right) du = p_{f_2 V_1 - f_1 V_2}(0) \quad \forall f_1, f_2 > 0. \quad (4.12)$$

In this case,  $s_{DDT}$  is symmetric and has the following properties:

- i)  $s_{DDT}(f, f) = \frac{1}{f} p_{V_1 - V_2}(0)$  for all  $f > 0$ ,
- ii)  $s_{DDT}$  is not bounded from above.

**Proof:** Equation (4.12) follows directly by the definition of  $s_{DDT}$ , Proposition A.7 ii), and Corollary A.4 i). By Corollary A.4 iv) we have for  $f = f_1 = f_2 > 0$  that

$$s_{DDT}(f, f) = p_{f(V_1 - V_2)}(0) = \frac{1}{f} p_{V_1 - V_2}(0)$$

and thus,  $s_{DDT}(f, f)$  tends to infinity for  $f \rightarrow 0$ . □

These properties stand in sharp contrast to the additive case. The first property implies that by  $s_{DDT}$ , small values  $f = f_1 = f_2$  are always considered more likely to be

generated by the same noise free pixel than bigger ones. Moreover, the unboundedness is not desirable with regard to the weight definition of a nonlocal filter, since a single pixel could get an arbitrarily large weight and dominate all others.

To see what we get for  $s_{DDT}$  for concrete noise distributions and to compare its behavior to  $p_{U_1-U_2|(F_1,F_2)}(0 | f_1, f_2)$ , we will consider different examples. In analogy to additive Gaussian noise, it may seem nearby to start with multiplicative Gaussian noise with mean one. However, in this case the assumption  $p_{V_i}(x) = 0$  for  $x < 0$  is violated. Only for a very small standard deviation it can be considered at least very unlikely that a realization  $v_i < 0$  occurs as it has also been pointed out in [10]. For this reason, we will not consider this example further. Instead, we start by studying multiplicative uniform noise as the simplest possible example.

**Example 4.2.3 (Multiplicative uniform noise)**

For  $i = 1, 2$  assume that  $U_i$  is uniformly distributed on the interval  $[0, n]$  and  $V_i$  is uniformly distributed on  $[1 - m, 1 + m]$ ,  $m \in (0, 1)$ , i.e.,

$$p_{U_i}(u) = \begin{cases} \frac{1}{n} & u \in [0, n], \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad p_{V_i}(v) = \begin{cases} \frac{1}{2m} & v \in [1 - m, 1 + m], \\ 0 & \text{otherwise.} \end{cases}$$

By applying (4.11) as well as Corollary A.4 ii) for computing  $p_{F_i}(f_i)$  we obtain by some technical computations

$$\begin{aligned} p_{U_1-U_2|(F_1,F_2)}(0 | f_1, f_2) &= \begin{cases} \frac{(1+m) \min\{\frac{1}{f_1}, \frac{1}{f_2}\} - (1-m) \max\{\frac{1}{f_1}, \frac{1}{f_2}, \frac{1}{(1-m)n}\}}{\log(\min\{\frac{(1+m)n}{f_1}, \frac{1+m}{1-m}\}) \log(\min\{\frac{(1+m)n}{f_2}, \frac{1+m}{1-m}\})} & \text{if } f_1, f_2 \in [0, (1+m)n] \\ & \text{and } \frac{f_1}{f_2} \in [\frac{1-m}{1+m}, \frac{1+m}{1-m}], \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

If we assume in contrast that the distribution of  $U_i$ ,  $i = 1, 2$ , is not known and we set  $S = \mathbb{R}_{\geq 0}$ , it follows that

$$\begin{aligned} s_{DDT}(f_1, f_2) &= \begin{cases} \frac{1}{4m^2} \left( (1+m) \min\left\{\frac{1}{f_1}, \frac{1}{f_2}\right\} - (1-m) \max\left\{\frac{1}{f_1}, \frac{1}{f_2}\right\} \right) & \text{if } \frac{f_1}{f_2} \in \left[\frac{1-m}{1+m}, \frac{1+m}{1-m}\right], \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

These functions have both the property that for fixed  $f_1$  they are maximal if  $f_2 = f_1$ . Moreover, they tend to infinity for  $f_1 = f_2 \rightarrow 0$ , i.e., they are both unbounded.

To analyze the performance of these measures with regard to our specific application we included Figure 4.2. The diagrams on the left show the histogram for each measure

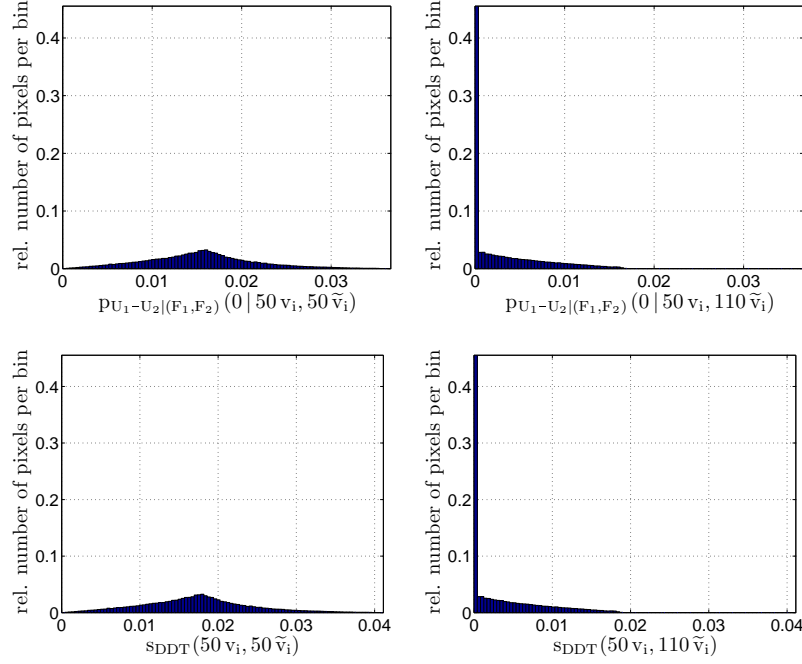


Figure 4.2: *Left*: Histograms of  $(p_{U_1-U_2|(F_1, F_2)}(0 | f_i, \tilde{f}_i))_{i=1}^N$  (top) and  $(s_{DDT}(f_i, \tilde{f}_i))_{i=1}^N$  (bottom), where  $f, \tilde{f}$  are both constant images of gray value 50 corrupted by multiplicative uniform noise with  $m = 0.4$ . *Right*: Same as on the left hand side, but now  $\tilde{f}$  represents a constant image of gray value 110 corrupted by equally distributed noise.

applied to two constant images of the same gray value corrupted by multiplicative uniform noise with  $m = 0.4$ . On the right, the same has been repeated with two constant images of significantly different gray value. As we can see here, except for a scaling factor the results of the two measures are quite similar. Moreover, the histograms for the images with different initial gray values have again a significant peak at zero meaning that most pixels have been correctly detected to belong to different noise free pixels. In contrast to Figure 4.1 (middle), the peaks of the histograms on the left are not at the largest obtained values of the measures, but at some intermediate values. This is not desirable with respect to a weight definition of a nonlocal filter, since it indicates that the measures cannot definitely determine for a large number of pixels whether the true pixels have been the same or not. However, these observations can also be confirmed for data corrupted by differently distributed noise:

#### Example 4.2.4 (Multiplicative Gamma noise)

For this example let us assume that the distribution of  $U_i$  is unknown and the noise components  $V_i$  are Gamma distributed with density

$$p_{V_i}(v) = \frac{L^L}{\Gamma(L)} v^{L-1} \exp(-Lv) \mathbf{1}_{\mathbb{R}_{\geq 0}}(v), \quad L \geq 1, \quad i = 1, 2. \quad (4.13)$$

For this noise distribution we obtain for  $f_1, f_2 > 0$  and  $S = \mathbb{R}_{\geq 0}$  that

$$s_{DDT}(f_1, f_2) = \frac{L^{2L}}{\Gamma(L)^2} (f_1 f_2)^{L-1} \int_0^\infty \frac{1}{u^{2L}} \exp\left(-L \frac{f_1 + f_2}{u}\right) du.$$

By the definition of the Gamma function, see also [60], it holds that

$$\int_0^\infty \frac{c_1}{t^b} \exp\left(-\frac{c_2}{t}\right) dt = \frac{c_1}{c_2^{b-1}} \Gamma(b-1) \quad \forall c_1 \in \mathbb{R}, c_2 > 0, b > 1. \quad (4.14)$$

Hence, we finally have

$$s_{DDT}(f_1, f_2) = L \frac{\Gamma(2L-1)}{\Gamma(L)^2} \frac{(f_1 f_2)^{L-1}}{(f_1 + f_2)^{2L-1}} = L \frac{\Gamma(2L-1)}{\Gamma(L)^2} \frac{1}{f_1 + f_2} \frac{1}{\left(2 + \frac{f_1}{f_2} + \frac{f_2}{f_1}\right)^{L-1}}.$$

One may again expect that for fixed  $f_1$  this similarity measure is maximal if  $f_2 = f_1$ . However, for  $L > 1$  and a given value  $f_1$  it is maximal for  $f_2 = \frac{L-1}{L} f_1$ . This is again in sharp contrast to the properties of  $s_{DDT}$  in the additive case. For the special case  $L = 1$  we have  $s_{DDT}(f_1, f_2) = \frac{1}{f_1 + f_2}$ . This implies that for fixed  $f_1$  the measure  $s_{DDT}(f_1, f_2)$  is large whenever  $f_2$  is small.

Figure 4.3 investigates the suitability of this measure for the weight definition of a non-local filter. As already indicated the performance is here very similar to the histograms in Figure 4.2.

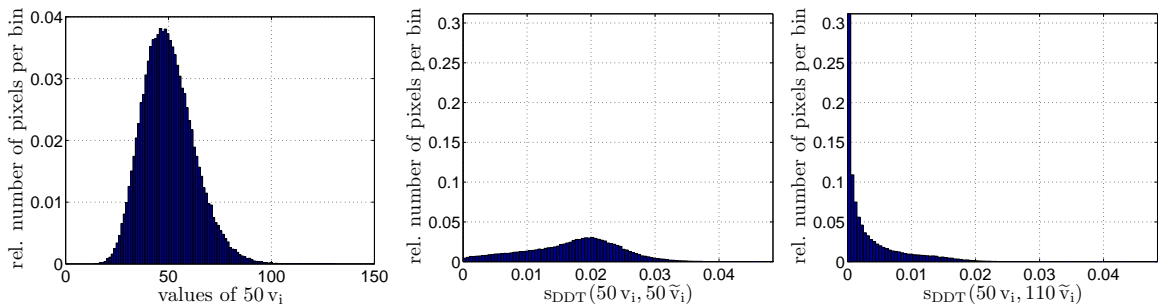


Figure 4.3: *Left*: Histogram of a constant image with gray value 50 corrupted by multiplicative Gamma noise with  $L = 16$ . *Middle*: Histogram of  $(s_{DDT}(f_i, \tilde{f}_i))_{i=1}^N$ , where  $f, \tilde{f}$  have gray value distributions as on the left. *Right*: Same as in the middle, but now  $\tilde{f}$  represents a constant image of gray value 110 corrupted by equally distributed noise.

**Example 4.2.5 (Multiplicative Rayleigh noise)**

Finally, let the distribution of  $U_i$ ,  $i = 1, 2$ , be again unknown and suppose now that the noise components  $v_i$  are realizations of Rayleigh distributed random variables  $V_i$  with

$$p_{V_i}(v) = \frac{v}{\theta^2} \exp\left(-\frac{v^2}{2\theta^2}\right) 1_{\mathbb{R}_{\geq 0}}(v), \quad \theta > 0. \quad (4.15)$$

Here, using again (4.14) yields for  $S = \mathbb{R}_{\geq 0}$  and  $f_1, f_2 > 0$  that

$$s_{DDT}(f_1, f_2) = \frac{\sqrt{2}}{\theta} \Gamma\left(\frac{3}{2}\right) \frac{f_1 f_2}{(f_1^2 + f_2^2)^{\frac{3}{2}}} = \sqrt{\frac{\pi}{2}} \frac{1}{\theta} \frac{f_1 f_2}{(f_1^2 + f_2^2)^{\frac{3}{2}}}.$$

For fixed  $f_1$  we have in this case that  $s_{DDT}(f_1, f_2)$  is maximal for  $f_2 = \frac{1}{\sqrt{2}} f_1$ , which is again in contrast to the additive case.

Altogether,  $s_{DDT}(f_1, f_2) = p_{f_2 V_1 - f_1 V_2}(0)$  does not seem to be an optimal similarity measure for the comparison of data corrupted by multiplicative noise.

## 4.3 A New Similarity Measure for Multiplicative Noise

To deduce a different measure for the multiplicative noise model introduced in Subsection 4.2.2, let us consider the logarithmically transformed random variables  $\tilde{F}_i = \log(F_i)$ , where

$$\underbrace{\log(F_i)}_{\tilde{F}_i} = \log(U_i V_i) = \underbrace{\log(U_i)}_{\tilde{U}_i} + \underbrace{\log(V_i)}_{\tilde{V}_i}, \quad i = 1, 2.$$

The new random variables  $\tilde{F}_i$  follow an additive noise model now and the supports of  $p_{\tilde{U}_i}$ ,  $p_{\tilde{V}_i}$  may be the whole of  $\mathbb{R}$ . Interestingly, computing (4.5) for these random variables leads to  $p_{\frac{U_1}{U_2}|(F_1, F_2)}(1 | f_1, f_2)$  rather than  $p_{U_1 - U_2|(F_1, F_2)}(0 | f_1, f_2)$  as the following lemma shows:

**Lemma 4.3.1** *For  $f_1, f_2 > 0$  with  $p_{F_i}(f_i) > 0$  and  $\tilde{S} = \text{supp}(p_{\tilde{U}_i})$  it holds that*

$$\begin{aligned} p_{\tilde{U}_1 - \tilde{U}_2|(\tilde{F}_1, \tilde{F}_2)}(0 | \log(f_1), \log(f_2)) \\ = \int_{\tilde{S}} \frac{p_{\tilde{U}_1}(t) p_{\tilde{U}_2}(t) p_{\tilde{F}_1|\tilde{U}_1}(\log(f_1) | t) p_{\tilde{F}_2|\tilde{U}_2}(\log(f_2) | t)}{p_{\tilde{F}_1}(\log(f_1)) p_{\tilde{F}_2}(\log(f_2))} dt \end{aligned} \quad (4.16)$$

$$= p_{\frac{U_1}{U_2}|(F_1, F_2)}(1 | f_1, f_2). \quad (4.17)$$

**Proof:** The first equality follows directly by (4.6). Besides, we have by Proposition A.7 i) and Corollary A.3 that

$$\begin{aligned}
 & \int_{\tilde{S}} \frac{p_{\tilde{U}_1}(t) p_{\tilde{U}_2}(t) p_{\tilde{F}_1|\tilde{U}_1}(\log(f_1) | t) p_{\tilde{F}_2|\tilde{U}_2}(\log(f_2) | t)}{p_{\tilde{F}_1}(\log(f_1)) p_{\tilde{F}_2}(\log(f_2))} dt \\
 &= \int_{\tilde{S}} \frac{p_{\tilde{U}_1}(t) p_{\tilde{U}_2}(t) p_{\tilde{V}_1}(\log(f_1) - t) p_{\tilde{V}_2}(\log(f_2) - t)}{p_{\tilde{F}_1}(\log(f_1)) p_{\tilde{F}_2}(\log(f_2))} dt \\
 &= \int_{\tilde{S}} \frac{e^t p_{U_1}(e^t) e^t p_{U_2}(e^t) f_1 e^{-t} p_{V_1}(f_1 e^{-t}) f_2 e^{-t} p_{V_2}(f_2 e^{-t})}{f_1 p_{F_1}(f_1) f_2 p_{F_2}(f_2)} dt \\
 &= \frac{\int_S \frac{1}{u} p_{U_1}(u) p_{U_2}(u) p_{V_1}\left(\frac{f_1}{u}\right) p_{V_2}\left(\frac{f_2}{u}\right) du}{p_{F_1}(f_1) p_{F_2}(f_2)} \quad (4.18)
 \end{aligned}$$

with  $S = \text{supp}(p_{U_i})$ . Next, we set  $X = (U_1, U_2, V_1, V_2)$ ,  $Y = (\frac{U_1}{U_2}, U_2, F_1, F_2)$ ,  $T_1 = T_2 = \mathbb{R}_{>0}^4$  and define  $g : T_1 \rightarrow T_2$  by  $g(u_1, u_2, v_1, v_2) = (\frac{u_1}{u_2}, u_2, u_1 v_1, u_2 v_2)$ . Then, Theorem A.1 yields

$$p_{\frac{U_1}{U_2}, U_2, F_1, F_2}(x, u, f_1, f_2) = \begin{cases} \frac{1}{xu} p_{U_1, U_2, V_1, V_2}(xu, u, \frac{f_1}{xu}, \frac{f_2}{u}) & \text{if } (x, u, f_1, f_2)^T \in T_2, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, by the pairwise independence of  $U_1, U_2, V_1, V_2$  it follows that

$$\begin{aligned}
 p_{\frac{U_1}{U_2}, F_1, F_2}(1, f_1, f_2) &= \int_{-\infty}^{\infty} p_{\frac{U_1}{U_2}, U_2, F_1, F_2}(1, u, f_1, f_2) du \\
 &= \int_S \frac{1}{u} p_{U_1}(u) p_{U_2}(u) p_{V_1}\left(\frac{f_1}{u}\right) p_{V_2}\left(\frac{f_2}{u}\right) du.
 \end{aligned}$$

Inserting this in (4.18) leads by the definition of the conditional density and the independence of  $F_1, F_2$  to the assumption.  $\square$

Now, similarly to Section 4.2 we omit the terms  $p_{\tilde{U}_i}, p_{\tilde{F}_i}, i = 1, 2$ , in (4.16) and suppose that  $\tilde{S} = \mathbb{R}$ , which is equivalent to  $S = \mathbb{R}_{\geq 0}$ . With (4.8) this leads to

$$\begin{aligned}
 \int_{\tilde{S}} p_{\tilde{F}_1|\tilde{U}_1}(\log(f_1) | t) p_{\tilde{F}_2|\tilde{U}_2}(\log(f_2) | t) dt &= p_{\tilde{V}_1 - \tilde{V}_2}(\log(f_1) - \log(f_2)) \\
 &= p_{\tilde{F}_1 - \tilde{F}_2|\tilde{U}_1 - \tilde{U}_2}(\log(f_1) - \log(f_2) | 0).
 \end{aligned}$$

Setting

$$s(f_1, f_2) := p_{\tilde{V}_1 - \tilde{V}_2}(\log(f_1) - \log(f_2)), \quad (4.19)$$

we obtain a new similarity measure.



**Remark 4.3.2** For  $p_{\tilde{U}_i} \equiv 1$  it follows by Corollary A.4 i) that  $p_{\tilde{F}_i} \equiv 1$ . Hence, omitting  $p_{\tilde{U}_i}$  and  $p_{\tilde{F}_i}$  is equivalent to setting  $p_{\tilde{U}_i} \equiv 1$ , disregarding that this is not a proper density. By the relation  $\tilde{U}_i = \log(U_i)$  and Corollary A.3 the prior  $p_{\tilde{U}_i} \equiv 1$  implies that

$$p_{U_i}(u) = \begin{cases} \frac{1}{u} p_{\tilde{U}_i}(\log(u)) & \text{if } u > 0, \\ 0 & \text{otherwise,} \end{cases} = \begin{cases} \frac{1}{u} & \text{if } u > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (4.20)$$

At the first glance, this choice of  $p_{U_i}$  may seem a bit odd. However, it can be justified by *Jeffreys' prior*, which is a so-called *non-informative prior* trying to minimize the influence of the prior on the reasoning if no prior information is available, see, e.g., [172]. For multiplicative noise with an arbitrary density  $p_{V_i}$  and  $u > 0$  this prior is given by

$$\begin{aligned} \tilde{p}_{U_i}(u) &:= \sqrt{\int_{-\infty}^{\infty} \left( \frac{\partial}{\partial u} \log p_{F_i|U_i}(x|u) \right)^2 p_{F_i|U_i}(x|u) dx} \\ &= \sqrt{\int_0^{\infty} \left( \frac{\partial}{\partial u} \log \left( \frac{1}{u} p_{V_i}\left(\frac{x}{u}\right) \right) \right)^2 \frac{1}{u} p_{V_i}\left(\frac{x}{u}\right) dx} \\ &= \sqrt{\int_0^{\infty} \left( \frac{\partial}{\partial u} (-\log u) + \frac{\partial}{\partial u} \log p_{V_i}\left(\frac{x}{u}\right) \right)^2 \frac{1}{u} p_{V_i}\left(\frac{x}{u}\right) dx} \\ &= \sqrt{\int_0^{\infty} \left( -\frac{1}{u} - \frac{v}{u} \frac{\partial}{\partial v} \log p_{V_i}(v) \right)^2 \frac{1}{u} p_{V_i}(v) u dv} \\ &= \frac{1}{u} \sqrt{\underbrace{\int_0^{\infty} \left( 1 + v \frac{\partial}{\partial v} \log p_{V_i}(v) \right)^2 p_{V_i}(v) dv}_{=const.}} \end{aligned} \quad (4.21)$$

Setting  $\tilde{p}_{U_i}(u) = 0$  for  $u \leq 0$  it turns out to be proportional to (4.20).

Now, we want to examine the properties of our new similarity measure (4.19):

**Proposition 4.3.3** *For the considered multiplicative noise model it holds that*

$$\begin{aligned}
 s(f_1, f_2) &= p_{\frac{f_2}{f_1} \frac{v_1}{v_2}}(1) = \frac{f_1}{f_2} p_{\frac{F_1}{F_2} | \frac{U_1}{U_2}} \left( \frac{f_1}{f_2} | 1 \right) \\
 &= \int_0^\infty \frac{f_1 f_2}{u^3} p_{V_1} \left( \frac{f_1}{u} \right) p_{V_2} \left( \frac{f_2}{u} \right) du \\
 &= \int_0^\infty \frac{f_1}{f_2} t p_{V_1} \left( \frac{f_1}{f_2} t \right) p_{V_2}(t) dt \quad \forall f_1, f_2 > 0.
 \end{aligned} \tag{4.22}$$

Moreover,  $s(\cdot, \cdot)$  is symmetric and has the following properties:

- i)  $s(f, f) = \text{const.}$  for all  $f > 0$ ,
- ii)  $0 \leq s(f_1, f_2) \leq s(f, f) = p_{\frac{v_1}{v_2}}(1)$  for all  $f_1, f_2, f > 0$ .

**Proof:** Corollaries A.4 i), A.3 and different variable transformations yield for  $f_1, f_2 > 0$  that

$$\begin{aligned}
 s(f_1, f_2) &= \int_{-\infty}^\infty p_{\tilde{V}_1}(\log(f_1) - t) p_{\tilde{V}_2}(\log(f_2) - t) dt = \int_0^\infty \frac{f_1 f_2}{u^3} p_{V_1} \left( \frac{f_1}{u} \right) p_{V_2} \left( \frac{f_2}{u} \right) du \\
 &= \int_0^\infty \frac{f_1}{f_2} t p_{V_1} \left( \frac{f_1}{f_2} t \right) p_{V_2}(t) dt.
 \end{aligned}$$

Moreover, it follows by Corollary A.4 iii), iv) and Proposition A.7 ii) that

$$s(f_1, f_2) = p_{\frac{f_2}{f_1} \frac{v_1}{v_2}}(1) = \frac{f_1}{f_2} p_{\frac{v_1}{v_2}} \left( \frac{f_1}{f_2} \right) = \frac{f_1}{f_2} p_{\frac{F_1}{F_2} | \frac{U_1}{U_2}} \left( \frac{f_1}{f_2} | 1 \right).$$

The properties of  $s(\cdot, \cdot)$  follow by (4.19), Lemma A.5 and  $s(f_1, f_2) = p_{\frac{f_2}{f_1} \frac{v_1}{v_2}}(1)$ .  $\square$

Obviously, our new measure  $s(\cdot, \cdot)$  has similarly good properties as  $s_{DDT}$  in the additive case with  $S = \mathbb{R}$ , although (4.22) differs from (4.12) only by the factor  $\frac{f_1 f_2}{u}$  within the integral. Regarding (4.17) and  $s_{DDT}$  for additive noise given in (4.8), our new measure is not exactly  $p_{\frac{F_1}{F_2} | \frac{U_1}{U_2}} \left( \frac{f_1}{f_2} | 1 \right)$  but a scaled version of it. The following examples show further advantageous properties of  $s(\cdot, \cdot)$ :

**Example 4.3.4** (Example 4.2.3 continued)

For  $f_1, f_2 > 0$  and uniformly distributed random variables  $U_i, V_i, i = 1, 2$ , as in Example 4.2.3, we obtain by technical computations using (4.18) and Corollary A.4 ii) that

$$p_{\frac{U_1}{U_2} | (F_1, F_2)}(1 | f_1, f_2) = \frac{\int \frac{1}{u} p_{U_1}(u) p_{U_2}(u) p_{V_1}\left(\frac{f_1}{u}\right) p_{V_2}\left(\frac{f_2}{u}\right) du}{p_{F_1}(f_1) p_{F_2}(f_2)}$$

$$= \begin{cases} \frac{\log\left(\frac{1+m}{1-m}\right) - \log\left(\max\left\{\frac{f_1}{f_2}, \frac{f_2}{f_1}, \frac{f_1}{(1-m)n}, \frac{f_2}{(1-m)n}\right\}\right)}{\log(\min\{\frac{(1+m)n}{f_1}, \frac{1+m}{1-m}\}) \log(\min\{\frac{(1+m)n}{f_2}, \frac{1+m}{1-m}\})} & \text{if } f_1, f_2 \in [0, (1+m)n] \\ & \text{and } \frac{f_1}{f_2} \in \left[\frac{1-m}{1+m}, \frac{1+m}{1-m}\right], \\ 0 & \text{otherwise.} \end{cases}$$

This measure is still not bounded as it tends to infinity for  $f_1 = f_2 \rightarrow (1+m)n$ . Nevertheless,

$$s(f_1, f_2) = \begin{cases} \frac{1}{8m^2} \left( (1+m)^2 \min\left\{\frac{f_1}{f_2}, \frac{f_2}{f_1}\right\} - (1-m)^2 \max\left\{\frac{f_1}{f_2}, \frac{f_2}{f_1}\right\} \right) & \text{if } \frac{f_1}{f_2} \in \left[\frac{1-m}{1+m}, \frac{1+m}{1-m}\right], \\ 0 & \text{otherwise,} \end{cases}$$

is bounded and has a maximum of  $c = \frac{1}{2m}$ .

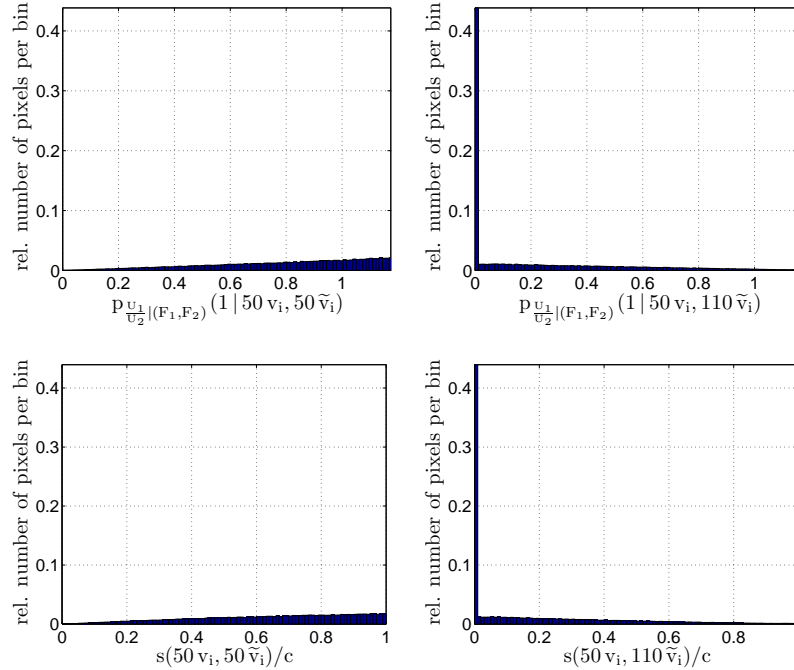


Figure 4.4: *Left*: Histograms of  $(p_{U_1/U_2 | (F_1, F_2)}(1 | f_i, \tilde{f}_i))_{i=1}^N$  (top) and  $(s(f_i, \tilde{f}_i)/c)_{i=1}^N$  (bottom), where  $f, \tilde{f}$  are both constant images of gray value 50 corrupted by multiplicative uniform noise with  $m = 0.4$ . *Right*: Same as on the left hand side, but now  $f$  represents a constant image of gray value 110 corrupted by equally distributed noise.

Figure 4.4 shows that except for a scaling factor the histograms for  $p_{U_1/U_2|(F_1, F_2)}(1|\cdot, \cdot)$  and  $s(\cdot, \cdot)/c$  are quite similar for the considered images. In particular, for the two images with the same initial gray value the histograms have now their maximum at the largest obtained values. As our next example shows, this still remains true if we consider  $s(\cdot, \cdot)/c$  for multiplicative Gamma noise:

**Example 4.3.5** (Example 4.2.4 continued)

Let  $V_i$ ,  $i = 1, 2$ , be Gamma distributed random variables. For  $f_1, f_2 > 0$  Equation (4.14) yields

$$\begin{aligned} s(f_1, f_2) &= \frac{L^{2L}}{\Gamma(L)^2} (f_1 f_2)^L \int_0^\infty \frac{1}{u^{2L+1}} \exp\left(-L \frac{f_1 + f_2}{u}\right) du \\ &= \frac{\Gamma(2L)}{\Gamma(L)^2} \frac{(f_1 f_2)^L}{(f_1 + f_2)^{2L}} = \frac{\Gamma(2L)}{\Gamma(L)^2} \frac{1}{\left(2 + \frac{f_1}{f_2} + \frac{f_2}{f_1}\right)^L}, \end{aligned}$$

which has a maximum value of  $c = \frac{\Gamma(2L)}{\Gamma(L)^2} \frac{1}{4^L}$ .

As illustrated by Figure 4.5, for  $L = 16$  this new measure gives similar histograms as initially obtained for additive Gaussian noise in Figure 4.1. Hence, a similar good performance can be expected if applied for nonlocal filtering.

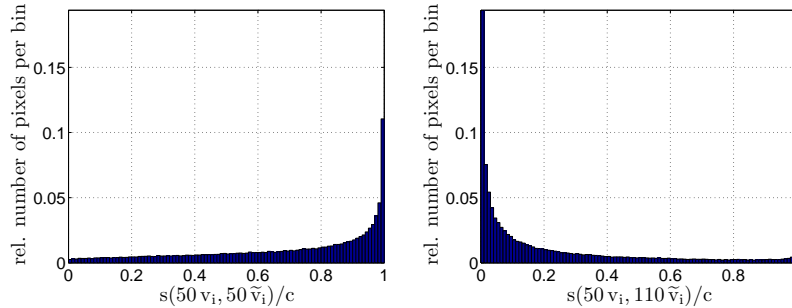


Figure 4.5: *Left:* Histogram of  $(s(f_i, \tilde{f}_i)/c)_{i=1}^N$ , where  $f, \tilde{f}$  are both constant images of gray value 50 corrupted by multiplicative Gamma noise with  $L = 16$ . *Right:* Same as on the left, but now  $\tilde{f}$  represents a constant image of gray value 110 corrupted by equally distributed noise.

**Example 4.3.6** (Example 4.2.5 continued)

Finally, let us assume that  $V_i$ ,  $i = 1, 2$ , follow a Rayleigh distribution. In this case, we obtain by (4.14) that

$$s(f_1, f_2) = \frac{f_1^2 f_2^2}{\theta^4} \int_0^\infty \frac{1}{u^5} \exp\left(-\frac{1}{2\theta^2} \frac{f_1^2 + f_2^2}{u^2}\right) du$$

$$\begin{aligned}
 &= \frac{f_1^2 f_2^2}{\theta^4} \int_0^\infty \frac{1}{2t^3} \exp\left(-\frac{1}{2\theta^2} \frac{f_1^2 + f_2^2}{t}\right) dt \\
 &= 2 \left( \frac{f_1 f_2}{f_1^2 + f_2^2} \right)^2 = \frac{2}{\left( \frac{f_1}{f_2} + \frac{f_2}{f_1} \right)^2}
 \end{aligned}$$

with a maximum of  $c = \frac{1}{2}$ . We will use this measure for our numerical experiments later on.

By all these findings we showed that our new similarity measure has many favorable properties, which are similar to the properties of  $s_{DDT}$  facing additive noise with  $S = \mathbb{R}$ .

## 4.4 Relations to Other Similarity Measures

Interestingly, there exist different relations of our new measure to other similarity measures known in the literature. To start with, we state the following observation:

**A relation to  $s_{NL}$**  If we had initially no idea how to define similarity measures for data corrupted by multiplicative noise, a first nearby approach would have been to transform  $f_1, f_2 > 0$  logarithmically to obtain data corrupted by additive noise and to use

$$s_{NL}(\log(f_1), \log(f_2)) = \exp\left(-(\log(f_1) - \log(f_2))^2\right) = \exp\left(-\left(\log\left(\frac{f_1}{f_2}\right)\right)^2\right) \quad (4.23)$$

as a similarity measure for any  $f_1, f_2$  corrupted by multiplicative noise. Here, the logarithmic transformation can also be considered as a variance-stabilizing transformation, which makes the variance of the transformed noisy data signal independent. Interestingly, the similarity measure (4.23) can be related to the our new measure for data corrupted of multiplicative Gamma noise with  $L \approx 4$ . In detail, using the Taylor approximation

$$\log(x) = 2 \sum_{k=0}^{\infty} \frac{1}{2k+1} \left( \frac{x-1}{x+1} \right)^{2k+1} \approx 2 \frac{x-1}{x+1} \quad \text{for } x > 0,$$

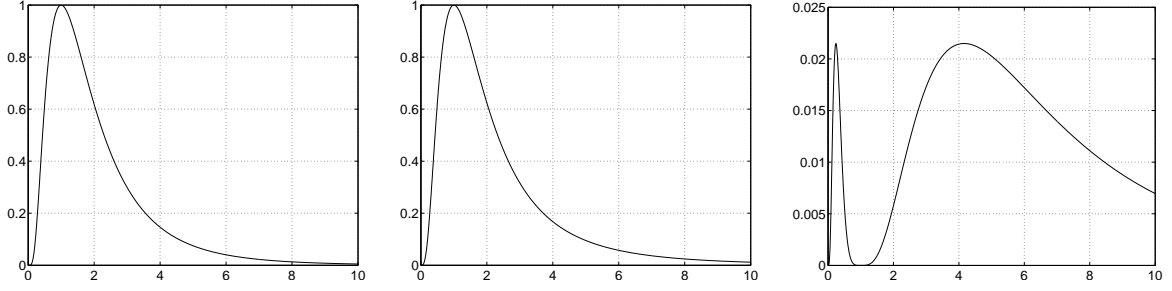


Figure 4.6: Graphs of the functions  $s_1, s_2 : \mathbb{R}_{>0} \rightarrow \mathbb{R}$  defined by  $s_1(x) = \exp(-(\log(x))^2)$  (left) and  $s_2(x) = \left(\frac{4}{2+x+\frac{1}{x}}\right)^4$  (middle) as well as the graph of the difference  $s_2 - s_1$  (right).

see, e.g., [112, p. 137], we can show that

$$\begin{aligned}
 s_{NL}(\log(f_1), \log(f_2)) &= \exp\left(-\left(\log\left(\frac{f_1}{f_2}\right)\right)^2\right) = \exp\left(\int_1^{f_1/f_2} -\frac{2\log(t)}{t} dt\right) \\
 &\approx \exp\left(\int_1^{f_1/f_2} -\frac{4(t-1)}{t(t+1)} dt\right) = \exp\left(\left[4\log\left(\frac{4t}{(t+1)^2}\right)\right]_1^{f_1/f_2}\right) \\
 &= \left(\frac{4\frac{f_1}{f_2}}{\left(\frac{f_1}{f_2}+1\right)^2}\right)^4 = \left(\frac{4}{2+\frac{f_1}{f_2}+\frac{f_2}{f_1}}\right)^4.
 \end{aligned}$$

This last term equals exactly  $\frac{s(f_1, f_2)}{c}$  as given for  $L = 4$  in Example 4.3.5. Thus,  $s_{NL}(\log(f_1), \log(f_2))$  approximates  $\frac{s(f_1, f_2)}{c}$  for data corrupted by multiplicative Gamma noise with  $L \approx 4$ . The quality of the approximation is shown in Figure 4.6. As we can see here, the graphs are very close, although there are small differences visible.

In the very recent works [59, 62], which were published after the completion of [TL12a, TL12b], Deledalle et al. studied a couple of other similarity measures taken from the fields of image processing, detection theory and machine learning. For the here considered case of multiplicative noise we want to show in the following that some of these measures are closely related to each other or even coincide under appropriate assumptions. In addition, close relations to our proposed measure will be proven.

Since we assumed throughout this chapter that all  $U_i$  as well as all  $V_i$  are independent and identically distributed, we will drop in the following the indices  $i$  of the random variables characterizing the densities and write, e.g., only  $p_U$  instead of  $p_{U_i}$ .

**Bayesian likelihood ratio and mutual information kernel** With

$$Q_B(f_1, f_2) := \int_S p_{F|U}(f_1|u) p_{F|U}(f_2|u) p_U(u) du, \quad S := \text{supp}(p_U)$$

two of the measures examined in [59, 62] are the *Bayesian likelihood ratio (BLR)*

$$L_B(f_1, f_2) = \frac{Q_B(f_1, f_2)}{D_B(f_1) D_B(f_2)} \quad \text{with} \quad D_B(f) = \int_S p_{F|U}(f|u) p_U(u) du = p_F(f),$$

cf. [146, 147, 172], and the *mutual information kernel*

$$K_B(f_1, f_2) = \frac{Q_B(f_1, f_2)}{\sqrt{Q_B(f_1, f_1) Q_B(f_2, f_2)}}$$

proposed in [185]. These measures have in common that they both require the definition of a prior  $p_U$ . In cases where the density  $p_U$  is unknown, Jeffreys' prior can again be used. As already determined in Remark 4.3.2 this prior is given by

$$p_U(u) = \begin{cases} d \frac{1}{u} & \text{for } u > 0, \\ 0 & \text{otherwise} \end{cases} \quad \text{with} \quad d := \sqrt{\int_0^\infty \left(1 + v \frac{\partial}{\partial v} \log p_V(v)\right)^2 p_V(v) dv} \quad (4.24)$$

for any type of multiplicative noise which fits into the multiplicative noise model introduced in Subsection 4.2.2. Interestingly, using (4.24) we can prove the following relations between  $L_B(f_1, f_2)$ ,  $K_B(f_1, f_2)$  and  $s(f_1, f_2)$ :

**Proposition 4.4.1** *For the multiplicative noise model introduced in Subsection 4.2.2 and  $p_U$  defined by Jeffreys' prior (4.24) it holds that*

$$L_B(f_1, f_2) = \frac{s(f_1, f_2)}{d} \quad \text{and} \quad K_B(f_1, f_2) = \frac{s(f_1, f_2)}{c}$$

with  $c = \max_{f_1, f_2} s(f_1, f_2)$ .

**Proof:** Using Jeffreys' prior (4.24) and Proposition A.7 ii) we obtain that

$$\begin{aligned} D_B(f) &= \int_S p_{F|U}(f|u) p_U(u) du = d \int_0^\infty \frac{1}{u^2} p_V\left(\frac{f}{u}\right) du = d \int_0^\infty \frac{v^2}{f^2} p_V(v) \frac{f}{v^2} dv \\ &= d \frac{1}{f} \int_0^\infty p_V(v) dv = d \frac{1}{f}. \end{aligned}$$

Since

$$c = \max_{f_1, f_2} s(f_1, f_2) = s(f, f) = \int_0^\infty \frac{f}{f} t p_V\left(\frac{f}{f} t\right) p_V(t) dt = \int_0^\infty t (p_V(t))^2 dt,$$

this implies that

$$\begin{aligned} Q_B(f, f) &= d \int_0^\infty \frac{1}{u} p_{F|U}(f|u) p_{F|U}(f|u) du = d \int_0^\infty \frac{1}{u^3} p_V\left(\frac{f}{u}\right) p_V\left(\frac{f}{u}\right) du \\ &= d \int_0^\infty \frac{v^3}{f^3} p_V(v) p_V(v) \frac{f}{v^2} dv = d \frac{1}{f^2} \underbrace{\int_0^\infty v (p_V(v))^2 dv}_{=c} \\ &= c d \frac{1}{f^2}. \end{aligned}$$

Hence, with (4.22) it follows that

$$\begin{aligned} L_B(f_1, f_2) &= \frac{Q_B(f_1, f_2)}{D_B(f_1) D_B(f_2)} = \frac{f_1 f_2}{d^2} d \int_0^\infty \frac{1}{u} p_{F|U}(f_1|u) p_{F|U}(f_2|u) du \\ &= \frac{1}{d} \int_0^\infty \frac{f_1 f_2}{u^3} p_V\left(\frac{f_1}{u}\right) p_V\left(\frac{f_2}{u}\right) du = \frac{s(f_1, f_2)}{d} \end{aligned}$$

and

$$\begin{aligned} K_B(f_1, f_2) &= \frac{Q_B(f_1, f_2)}{\sqrt{Q_B(f_1, f_1) Q_B(f_2, f_2)}} = \frac{f_1 f_2}{c d} d \int_0^\infty \frac{1}{u} p_{F|U}(f_1|u) p_{F|U}(f_2|u) du \\ &= \frac{1}{c} \int_0^\infty \frac{f_1 f_2}{u^3} p_V\left(\frac{f_1}{u}\right) p_V\left(\frac{f_2}{u}\right) du = \frac{s(f_1, f_2)}{c}. \end{aligned}$$

□

By this proposition we showed that our normalized measure  $s(f_1, f_2)/c$  coincides with the mutual information kernel  $K_B$  and it is proportional to the Bayesian likelihood ratio  $L_B$  if Jeffreys' prior is applied in both cases.

**Generalized likelihood ratio** A related similarity measure studied in [59, 62] is the *generalized likelihood ratio (GLR)*

$$L_G(f_1, f_2) = \frac{\sup_{u \in S} p_{F|U}(f_1|u) p_{F|U}(f_2|u)}{\sup_{u \in S} p_{F|U}(f_1|u) \sup_{u \in S} p_{F|U}(f_2|u)},$$



cf., e.g., [116, 133, 172]. In contrast to  $L_B$  and  $K_B$ , no prior distribution  $p_U$  is incorporated here. For the considered multiplicative noise model and  $S := \mathbb{R}_{>0}$  Proposition A.7 ii) yields

$$\sup_{u \in S} p_{F|U}(f|u) = \sup_{u>0} \frac{1}{u} p_V\left(\frac{f}{u}\right) = \sup_{v>0} \frac{v}{f} p_V(v) = \frac{1}{f} \underbrace{\sup_{v>0} v p_V(v)}_{=: b} = b \frac{1}{f}$$

so that

$$\begin{aligned} L_G(f_1, f_2) &= \frac{\sup_{u>0} p_{F|U}(f_1|u) p_{F|U}(f_2|u)}{\sup_{u>0} p_{F|U}(f_1|u) \sup_{u>0} p_{F|U}(f_2|u)} \\ &= \frac{f_1 f_2}{b^2} \sup_{u>0} \frac{1}{u^2} p_V\left(\frac{f_1}{u}\right) p_V\left(\frac{f_2}{u}\right) \\ &= \frac{1}{b^2} \frac{f_1}{f_2} \sup_{t>0} t^2 p_V\left(\frac{f_1}{f_2} t\right) p_V(t). \end{aligned}$$

By the last term we can see that for the examined multiplicative noise model the value of this measure depends like  $s(f_1, f_2)$  only on the quotient  $f_1/f_2$  rather than the values  $f_1$  and  $f_2$  as such. Although  $L_G$  differs in general from  $s(f_1, f_2)/c$ , we can show that for our standard examples of multiplicative Gamma and Rayleigh noise the measures do again coincide:

**Proposition 4.4.2** *Let  $S := \mathbb{R}_{>0}$ . For multiplicative Gamma and Rayleigh noise with densities  $p_V$  given by (4.13) and (4.15), respectively it holds that*

$$L_G(f_1, f_2) = \frac{s(f_1, f_2)}{c} \quad \text{with} \quad c = \max_{f_1, f_2} s(f_1, f_2).$$

**Proof:** For multiplicative Gamma noise with noise density (4.13) it follows that

$$b = \sup_{v>0} v p_V(v) = \sup_{v>0} \frac{L^L}{\Gamma(L)} v^L \exp(-Lv) = \frac{L^L}{\Gamma(L)} \exp(-L)$$

and thus, with Example 4.3.5 we obtain

$$\begin{aligned} L_G(f_1, f_2) &= \frac{1}{b^2} \frac{f_1}{f_2} \sup_{t>0} t^2 p_V\left(\frac{f_1}{f_2} t\right) p_V(t) \\ &= \frac{1}{\exp(-2L)} \frac{f_1}{f_2} \sup_{t>0} t^{2L} \left(\frac{f_1}{f_2}\right)^{L-1} \exp\left(-L\left(\frac{f_1}{f_2} + 1\right)t\right) \\ &= \frac{1}{\exp(-2L)} \left(\frac{f_1}{f_2}\right)^L \left(\frac{2}{1 + \frac{f_1}{f_2}}\right)^{2L} \exp(-2L) \\ &= \frac{4^L}{\left(2 + \frac{f_1}{f_2} + \frac{f_2}{f_1}\right)^L} = \frac{s(f_1, f_2)}{c}. \end{aligned}$$

On the other hand, for multiplicative Rayleigh noise the density (4.15) yields

$$b = \sup_{v>0} v p_V(v) = \sup_{v>0} \frac{v^2}{\theta^2} \exp\left(-\frac{v^2}{2\theta^2}\right) = 2 \exp(-1).$$

With Example 4.3.6 this finally implies that

$$\begin{aligned} L_G(f_1, f_2) &= \frac{1}{b^2} \frac{f_1}{f_2} \sup_{t>0} t^2 p_V\left(\frac{f_1}{f_2}t\right) p_V(t) \\ &= \frac{1}{4 \exp(-2)} \frac{f_1}{f_2} \sup_{t>0} \frac{1}{\theta^4} t^4 \frac{f_1}{f_2} \exp\left(-\frac{1}{2\theta^2} \left(\left(\frac{f_1}{f_2}\right)^2 + 1\right) t^2\right) \\ &= \frac{1}{4 \theta^4 \exp(-2)} \left(\frac{f_1}{f_2}\right)^2 \frac{(2\theta)^4}{\left(\left(\frac{f_1}{f_2}\right)^2 + 1\right)^2} \exp(-2) \\ &= \frac{4}{\left(\frac{f_1}{f_2} + \frac{f_2}{f_1}\right)^2} = \frac{s(f_1, f_2)}{c}. \end{aligned}$$

□

These findings show that although our proposed similarity measures has been deduced in a totally different way, it is closely connected to other existing approaches. Since  $s_{NL}$  with an appropriate variance stabilization and  $K_B$  and  $L_G$  with Jeffreys' prior performed best among all similarity measures compared in [59, 62], these observations highlight once more that our deduced similarity measure is a very good choice for the task at hand.

## 4.5 Nonlocal Filtering Facing Multiplicative Noise

In the following, we will now use this similarity measure to define appropriate nonlocal filters for the removal of multiplicative noise in images.

### 4.5.1 Nonlocal Filters by Maximum Likelihood Estimation

As in [60] our nonlocal filters are based on weighted maximum likelihood estimation. For further literature on this topic we refer to [81, 163, 172, 190]. Ideally, we would like to determine an estimate  $\tilde{u}_i$  of the true noise free pixels  $u_i$  such that

$$\tilde{u}_i = \operatorname{argmax}_{t \in \mathbb{R}} \sum_{j \in S_i} \log p_{F_j|U_j}(f_j | t) \quad \text{subject to } p_{U_j}(t) > 0$$

with  $S_i$  being the index set of those pixels, which were generated from the same noise free pixel as  $f_i$ , i.e.,  $f_j = u_i v_j$  for all  $j \in S_i$ . Since the set  $S_i$  is not known, we assume

that estimates  $w(i, j) \in [0, 1]$  of the values of the characteristic functions  $1_{S_i}(j)$  are given for all  $j = 1, \dots, N$  and we compute

$$\hat{u}_i := \operatorname{argmax}_{t \in \mathbb{R}} \sum_{j=1}^N w(i, j) \log p_{F_j|U_j}(f_j | t) \quad \text{subject to } p_{U_j}(t) > 0. \quad (4.25)$$

If we set  $w = w_{NL}$ , the estimates  $\hat{u}_i$  for additive Gaussian noise and  $p_{U_i} > 0$  are given by the weighted means (4.1). For multiplicative Gamma and Rayleigh noise the resulting filters are deduced in the subsequent examples:

#### Example 4.5.1 (Multiplicative Gamma noise)

For our multiplicative noise model described in Subsection 4.2.2 and noise following a Gamma distribution we have according to (4.13) and Proposition A.7 ii) that

$$p_{F_j|U_j}(f_j|u_i) = \frac{L^L}{\Gamma(L)} \frac{f_j^{L-1}}{u_i^L} \exp\left(-L \frac{f_j}{u_i}\right) \quad \text{for } f_j, u_i > 0 \text{ with } p_{U_j}(u_i) > 0, \quad (4.26)$$

compare also (3.6). Hence, it follows for  $f_j > 0$ ,  $j = 1, \dots, N$ , that

$$\hat{u}_i = \operatorname{argmax}_{t>0, p_{U_i}(t)>0} \sum_{j=1}^N w(i, j) \log p_{F_j|U_j}(f_j | t) = \operatorname{argmin}_{t>0, p_{U_i}(t)>0} \sum_{j=1}^N w(i, j) \left( \log(t) + \frac{f_j}{t} \right).$$

In [10] Aubert and Aujol deduced similarly  $H(f, u) := \sum_{i=1}^N \log(u_i) + \frac{f_i}{u_i}$  as a data fidelity term for a variational approach to remove multiplicative Gamma noise.

If  $p_{U_i}(t) > 0$  for all  $t > 0$  or the distributions of the  $U_i$  are not known, we omit the restriction  $p_{U_i}(t) > 0$  and obtain for  $f_j > 0$ ,  $j = 1, \dots, N$ , by the first order optimality condition that

$$\hat{u}_i = \frac{1}{C_i} \sum_{j=1}^N w(i, j) f_j \quad \text{with} \quad C_i := \sum_{j=1}^N w(i, j). \quad (4.27)$$

Note that this is again an ordinary weighted average of the  $f_j$ ,  $j = 1, \dots, N$ , like the original NL means filter (4.1).

#### Example 4.5.2 (Multiplicative Rayleigh noise)

For our multiplicative noise model and Rayleigh distributed noise, Equation (4.15) and Proposition A.7 ii) yield

$$p_{F_j|U_j}(f_j|u_i) = \frac{1}{\theta^2} \frac{f_j}{u_i^2} \exp\left(-\frac{1}{2\theta^2} \left(\frac{f_j}{u_i}\right)^2\right) \quad \text{for } f_j, u_i > 0 \text{ with } p_{U_j}(u_i) > 0.$$

Hence, we obtain for  $f_j > 0$ ,  $j = 1, \dots, N$ , that

$$\hat{u}_i = \underset{t > 0, p_{U_i}(t) > 0}{\operatorname{argmin}} \sum_{j=1}^N w(i, j) \left( 2 \log(t) + \frac{1}{2\theta^2} \frac{f_j^2}{t^2} \right).$$

For a variational approach,  $H(f, u) := \sum_{i=1}^N 2 \log(t) + \frac{1}{2\theta^2} \frac{f_j^2}{t^2}$  could also be used as a data fidelity term, where an appropriate regularization term has to be added.

If  $p_{U_i}(t) > 0$  for all  $t > 0$  or the distributions of the  $U_i$  are not known, we omit again the restriction  $p_{U_i}(t) > 0$ . For  $f_j > 0$ ,  $j = 1, \dots, N$ , we finally obtain by the first order optimality condition

$$\hat{u}_i = \sqrt{\frac{1}{2\theta^2 C_i} \sum_{j=1}^N w(i, j) f_j^2} \quad \text{with} \quad C_i := \sum_{j=1}^N w(i, j). \quad (4.28)$$

#### 4.5.2 Definition of the Weights

For multiplicative noise and random variables  $U_i$  with unknown distribution, the weights can now be defined similarly to (4.9) as

$$w(i, j) = \prod_{k \in I} \left( \frac{s(f_{i+k}, f_{j+k})}{c} \right)^{\frac{g_{a,k}}{h}}, \quad (4.29)$$

where  $s(f_1, f_2) = p_{\frac{f_2}{f_1} \frac{V_1}{V_2}}(1)$  and  $c = p_{\frac{V_1}{V_2}}(1)$  as defined in Section 4.3. As before,  $g_a = (g_{a,k})_{k \in I}$  represents a sampled two dimensional Gaussian kernel with mean zero and standard deviation  $a$ , where we normalize  $g_a$  such that  $\sum_{k \in I} g_{a,k} = 1$ . The parameter  $h > 0$  controls again the amount of filtering and the index set  $I$  is set to be a squared grid of size  $l \times l$  centered around 0 using reflecting boundary conditions for  $f$ . In the same way as the ordinary NL means filter the definition in (4.29) relies on the fact that a natural image contains usually many very similar image patches, which have to be detected by the applied similarity measure.

Figure 4.7 (top) shows the histograms of the weights defined in (4.29) for different constant image patches of size  $5 \times 5$  corrupted by multiplicative Gamma noise with  $L = 16$ . As we can see here, multiplying the values of the similarity measure over a whole patch significantly changes the histograms compared to Figure 4.5. Now, the supports of the two histograms do no longer overlap, i.e., the weights for the noisy images generated from the same constant image are always larger than those computed for the constant images of significantly different gray values. Unfortunately, the histogram on the left is no longer maximal at one. Even worse, weights close to one have never been assigned.

To overcome this drawback we propose an additional adaptation of the weights inspired

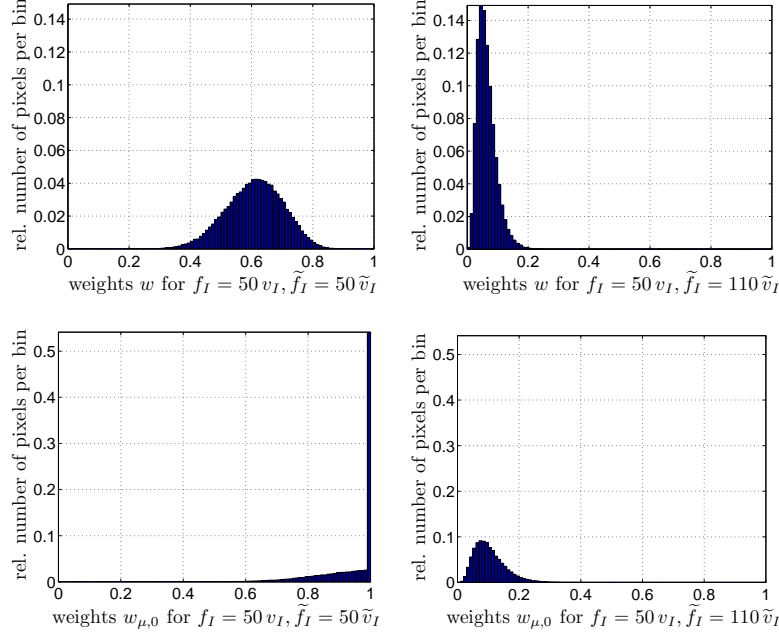


Figure 4.7: Histograms of the weights (4.29) (top) and (4.30) (bottom) computed for  $N$  different image patches  $f_I, \tilde{f}_I$  ( $l = 5, a = 1.5, h = 1, q = 0$ ). *Left*: Both  $f_I$  and  $\tilde{f}_I$  are patches of gray value 50 corrupted by multiplicative Gamma noise with  $L = 16$ . *Right*: Same as on the left, but now  $\tilde{f}_I$  represents an image patch of gray value 110 corrupted by noise.

by the implementation of the NL means filter described at [38]. Here, we use that for random variables  $X, Y$  and a continuous function  $b$ , where  $\mathbb{E}(b(Y))$  exists, the *conditional expectation* of  $b(Y)$  given  $X = x$  is

$$\mathbb{E}(b(Y)|X = x) := \int_{-\infty}^{\infty} b(y) p_{Y|X}(y|x) dy \quad \forall x \in \mathbb{R} \text{ with } p_X(x) > 0,$$

see, e.g., [174, p. 168]. In detail, for two sets of random variables  $F_{i+k} = U_{i+k}V_{i+k}$ ,  $F_{j+k} = U_{j+k}V_{j+k}$ ,  $k \in I$ , fulfilling the assumptions in Subsection 4.2.2, we set

$$b_k\left(\frac{f_{i+k}}{f_{j+k}}\right) := \left(\frac{p_{\frac{f_{j+k}}{f_{i+k}} V_{i+k}}(1)}{c}\right)^{\frac{g_{a,k}}{h}} = \left(\frac{s(f_{i+k}, f_{j+k})}{c}\right)^{\frac{g_{a,k}}{h}}.$$

Assuming that the index sets  $i+I, j+I$  have an empty intersection, i.e., the considered image patches are non-overlapping, we have

$$\mu := \mathbb{E}\left(\prod_{k \in I} b_k\left(\frac{F_{i+k}}{F_{j+k}}\right) \mid \left(\frac{U_{i+k}}{U_{j+k}} = 1\right)_{k \in I}\right) = \prod_{k \in I} \mathbb{E}\left(b_k\left(\frac{F_{i+k}}{F_{j+k}}\right) \mid \frac{U_{i+k}}{U_{j+k}} = 1\right).$$

The definition of the conditional expectation and Proposition A.7 ii) yield

$$\mathbb{E} \left( b_k \left( \frac{F_{i+k}}{F_{j+k}} \right) \mid \frac{U_{i+k}}{U_{j+k}} = 1 \right) = \mathbb{E} \left( b_k \left( \frac{V_{i+k}}{V_{j+k}} \right) \right) = \int_0^\infty b_k(t) p_{\frac{V_{i+k}}{V_{j+k}}}(t) dt$$

and thus, we finally obtain

$$\mu = \prod_{k \in I} \int_0^\infty b_k(t) p_{\frac{V_{i+k}}{V_{j+k}}}(t) dt.$$

Since  $w(i, j)$  is a realization of  $\prod_{k \in I} b_k \left( \frac{F_{i+k}}{F_{j+k}} \right)$ , the variable  $\mu$  describes the value that we can expect for  $w(i, j)$  considering (non-overlapping) image patches which have been generated from the same noise free patch. By some technical computations we obtain for *multiplicative Gamma noise*

$$\mu = \prod_{k \in I} 4^{Lg_{a,k}/h} \frac{\Gamma(2L) \Gamma(L(1 + \frac{g_{a,k}}{h}))^2}{\Gamma(L)^2 \Gamma(2L(1 + \frac{g_{a,k}}{h}))}$$

and for *multiplicative Rayleigh noise*

$$\mu = \prod_{k \in I} 4^{g_{a,k}/h} \frac{\Gamma(1 + \frac{g_{a,k}}{h})^2}{\Gamma(2(1 + \frac{g_{a,k}}{h}))}.$$

Next, we set

$$w_{\mu,q}(i, j) := \begin{cases} 1 & \text{if } w(i, j) \geq \mu, \\ \frac{w(i,j)}{\mu} & \text{if } q\mu \leq w(i, j) < \mu, \\ 0 & \text{otherwise} \end{cases} \quad (4.30)$$

with  $q \in [0, 1)$  and incorporate these weights in our nonlocal filters deduced from (4.25). Note that for overlapping image patches,  $\mu$  is used as an approximation of the true expectation value here. The effect of this modification in contrast to the weights (4.29) can be seen in Figure 4.7 (bottom). The histogram for the image patches generated from the same noise free patch has now a significant peak at one. By setting, e.g.,  $q = 0.5$  we could additionally achieve that all weights of the right histogram are set to zero and thus, the corresponding patches would have no effect if used in a nonlocal filter. On the contrary, the weights of the left histogram would not be effected. For our numerical examples in the subsequent section,  $q$  has been set by hand. Alternatively, a statistical estimate for  $q$  could be obtained by considering  $1 - F^{-1}(1 - \beta)$  for a value  $\beta \in (0, 1)$  very close to one. Here,  $F$  is the cumulative distribution function of  $r_\mu \left( \prod_{k \in I} b_k \left( \frac{V_{i+k}}{V_{j+k}} \right) \right)$  with

$$r_\mu(x) = \begin{cases} 1 & \text{if } x \geq \mu, \\ \frac{x}{\mu} & \text{otherwise} \end{cases}$$

and  $F^{-1}(\alpha) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}$  denotes the corresponding  $\alpha$ -quantile. Thus,  $q$  would be set to the maximal value such that the weights  $w_{\mu,q}(i, j)$  do not change compared to  $w_{\mu,0}(i, j)$  for  $100 \cdot \beta$  percent of the image patches  $f_{i+I}, f_{j+I}$  obtained from the same noise free patch.

As usually done, we finally restrict the number of patches being compared to a so-called *search window*. Thus, we set all weights  $w(i, j), w_{\mu,q}(i, j)$  automatically to zero if pixel  $j$  is outside of a squared image region of size  $\omega \times \omega$  centered at pixel  $i$ . This reduces the computational costs as well as the risk of falsely assigning nonzero weights to a large number of patches.

### 4.5.3 Updating the Similarity Neighborhoods

In [60] Deledalle et al. suggest to refine the weights of their nonlocal filters iteratively using the previous result  $u^{(r-1)}$ . To get the next iterate  $u^{(r)}$ , the filter is again applied to the initial noisy image using the new weights. In the following, we apply a variant of this updating strategy. The first major difference is that we perform only one updating step. For this second step we use within the search windows for  $i \neq j$  the weights

$$\tilde{w}_{i,j}(u^{(1)}) = \exp \left( -\frac{1}{d} \sum_{k \in \tilde{I}} g_{\tilde{a},k} K_{\text{sym}} \left( p_{F_{i+k}|U_{i+k}}(\cdot | u_{i+k}^{(1)}), p_{F_{j+k}|U_{j+k}}(\cdot | u_{j+k}^{(1)}) \right) \right)$$

instead of the ones defined in the former subsection and set  $\tilde{w}_{i,i}(u^{(1)}) = \max_j \tilde{w}_{i,j}(u^{(1)})$ . Here,  $d$  is a positive parameter and  $g_{\tilde{a}} = (g_{\tilde{a},k})_{k \in \tilde{I}}$  represents again a sampled two dimensional Gaussian kernel with mean zero but with standard deviation  $\tilde{a}$  now. As before,  $g_{\tilde{a}}$  is normalized such that  $\sum_{k \in \tilde{I}} g_{\tilde{a},k} = 1$ . Moreover, the index set  $\tilde{I}$  of size  $\tilde{l} \times \tilde{l}$  may vary from  $I$ . Usually, we choose  $\tilde{a} < a$  and  $\tilde{l} < l$ . Furthermore,

$$K_{\text{sym}}(p_X, p_Y) := \int_{-\infty}^{\infty} (p_X(t) - p_Y(t)) \log \left( \frac{p_X(t)}{p_Y(t)} \right) dt$$

denotes the *symmetric Kullback-Leibler divergence* of the densities  $p_X$  and  $p_Y$ . The idea for this updating scheme was originally taken from [163]. Here, Polzehl and Spokoiny used the ordinary Kullback-Leibler divergence of  $p_{F_i|U_i}(\cdot | u_i^{(r-1)})$  and  $p_{F_j|U_j}(\cdot | u_j^{(r-1)})$  to test for the hypotheses  $u_i = u_j$  using estimates  $u_i^{(r-1)}, u_j^{(r-1)}$  of  $u_i, u_j$ , respectively.

Since our numerical experiments in the next section deal with multiplicative Gamma and Rayleigh noise, we compute the symmetric Kullback-Leibler divergence for these two examples:

#### Example 4.5.3 (Multiplicative Gamma noise)

Let us assume that  $p_{U_i}(x) > 0$  for all  $x > 0$ . Then, we obtain by straightforward calculation using (4.26) and (4.14) that the sought symmetric Kullback-Leibler divergence

is given by

$$K_{\text{sym}}\left(p_{F_i|U_i}(\cdot|u_i^{(1)}), p_{F_j|U_j}(\cdot|u_j^{(1)})\right) = L \frac{(u_i^{(1)} - u_j^{(1)})^2}{u_i^{(1)} u_j^{(1)}} \quad \text{for } u_i^{(1)}, u_j^{(1)} > 0.$$

**Example 4.5.4 (Multiplicative Rayleigh noise)**

Assume again that  $p_{U_i}(x) > 0$  for all  $x > 0$ . In case of multiplicative Rayleigh noise similar calculations yield that the symmetric Kullback-Leibler divergence is

$$K_{\text{sym}}\left(p_{F_i|U_i}(\cdot|u_i^{(1)}), p_{F_j|U_j}(\cdot|u_j^{(1)})\right) = \frac{\left((u_i^{(1)})^2 - (u_j^{(1)})^2\right)^2}{(u_i^{(1)})^2 (u_j^{(1)})^2} \quad \text{for } u_i^{(1)}, u_j^{(1)} > 0.$$

## 4.6 Numerical Results

In the following, we present different examples demonstrating the very good performance of our nonlocal filters. These filters have been implemented in MATLAB and the parameters are chosen with respect to the best visual results. Good choices for the patch and search window sizes are usually  $l \in \{5, 7, 9\}$ ,  $\tilde{l} \in \{3, 5\}$  and  $\omega \in \{21, 29\}$  with  $a \in [0.5, 4]$  and  $\tilde{a} \in [0.5, 1.5]$ . Furthermore, all images, especially the noisy ones, are displayed with the colormap of the original image, if available, to have a consistent coloring for each example.

**Results for multiplicative Gamma noise** Our first three examples show different reconstructions of images contaminated by multiplicative Gamma noise. The original and noisy images presented in Figure 4.8 are the same as those considered in Figures 3.10 and 3.11 so that the results are directly comparable.

To obtain the restored image in Figure 4.9 (right) we used the weighted average filter derived in (4.27) with weights (4.30). As we see here, already without an additional updating of the weights we obtain a very good reconstruction, which is superior to the results by the IDIV-TV (left) and IDIV-NL models (middle) studied in Chapter 3.

Also in our next example the reconstructions in Figure 4.10 (bottom middle and right) outperform the result by the IDIV-TV method at top left. At top middle we included a reconstruction by the original NL means filter using  $s_{NL}(\log(\cdot), \log(\cdot))$  instead of  $s_{NL}$  for the patch comparison. Here, the nonzero weights have again been restricted to a search window and the patches are chosen in the same way as for our filters. As predicted in Section 4.4 this result is nearly the same as the one by our nonlocal filter (4.27) using the weights (4.29), which can be found at top right. Note that we have chosen slightly different values  $h$  to get even more similar results. By the definition of the weights we can see that for appropriate values  $h$  this can also be achieved for images corrupted by multiplicative Gamma noise with  $L \neq 4$ .





Figure 4.8: *Left*: Original images with values in  $[0, 255]$ . *Right*: Noisy versions corrupted by multiplicative Gamma noise with  $L = 25$  (top) and  $L = 4$  (bottom).

The results at Figure 4.10 (bottom left and middle) show the effect of applying (4.27) with weights  $w_{\mu,0}(i, j)$  or  $w_{\mu,q}(i, j)$  instead of  $w(i, j)$ . By replacing  $w(i, j)$  by  $w_{\mu,0}(i, j)$  we achieved an additional suppression of the noise especially in the background. Choosing further  $w_{\mu,q}(i, j)$  with an appropriate value  $q$  helped to improve the contrast, e.g., visible at the camera. Besides, it led to sharper edges and contours. By the final updating step applied at Figure 4.10 (bottom right) we further improved the contrast and small amounts of possibly remained noise are finally removed.

Next, we restored the noisy image of [70, Fig. 8] shown in Figure 4.11 (top right). Note that the corrupted image is displayed in a different way here. For an additional quantitative comparison with the results in [70] we use the *peak signal to noise ratio* (PSNR) and *mean absolute error* (MAE) defined by

$$\text{PSNR} = 10 \log_{10} \frac{|\max u - \min u|^2}{\frac{1}{N} \|\hat{u} - u\|_2^2} \quad \text{and} \quad \text{MAE} = \frac{1}{N} \|\hat{u} - u\|_1.$$

Here,  $u$  denotes the original noise free image,  $\hat{u}$  the reconstruction and  $N$  stands for the number of pixels of the images. To detect fluctuations in the quality of the results we generated 500 noisy realizations of the original image and averaged the PSNR and



Figure 4.9: Results for the noisy image at Figure 4.8 (top right) by the IDIV-TV (left) and IDIV-NL model (middle) as shown in Figure 3.10 and a result by our nonlocal filter (4.27) using the weights (4.30) with  $l = 7$ ,  $\omega = 21$ ,  $a = 2.5$ ,  $h = 1$ ,  $q = 0.6$  (right).



Figure 4.10: Results for the noisy image at Figure 4.8 (bottom right). *Top*: Restored images by IDIV-TV as presented in Figure 3.11 (left), by the original NL means filter using  $s_{NL}(\log(\cdot), \log(\cdot))$  instead of  $s_{NL}$  with  $l = 7$ ,  $\omega = 29$ ,  $a = 1.5$ ,  $h = 1.3$  (middle) and by our new nonlocal filter (4.27) using the weights (4.29) with  $l = 7$ ,  $\omega = 29$ ,  $a = 1.5$ ,  $h = 1$  (right). *Bottom*: Results by our nonlocal filter using (4.30) with additional parameter  $q = 0$  (left),  $q = 0.35$  (middle) and finally after an additional updating step with  $\tilde{l} = 3$ ,  $\tilde{a} = 0.5$ ,  $d = 0.25$  (right).

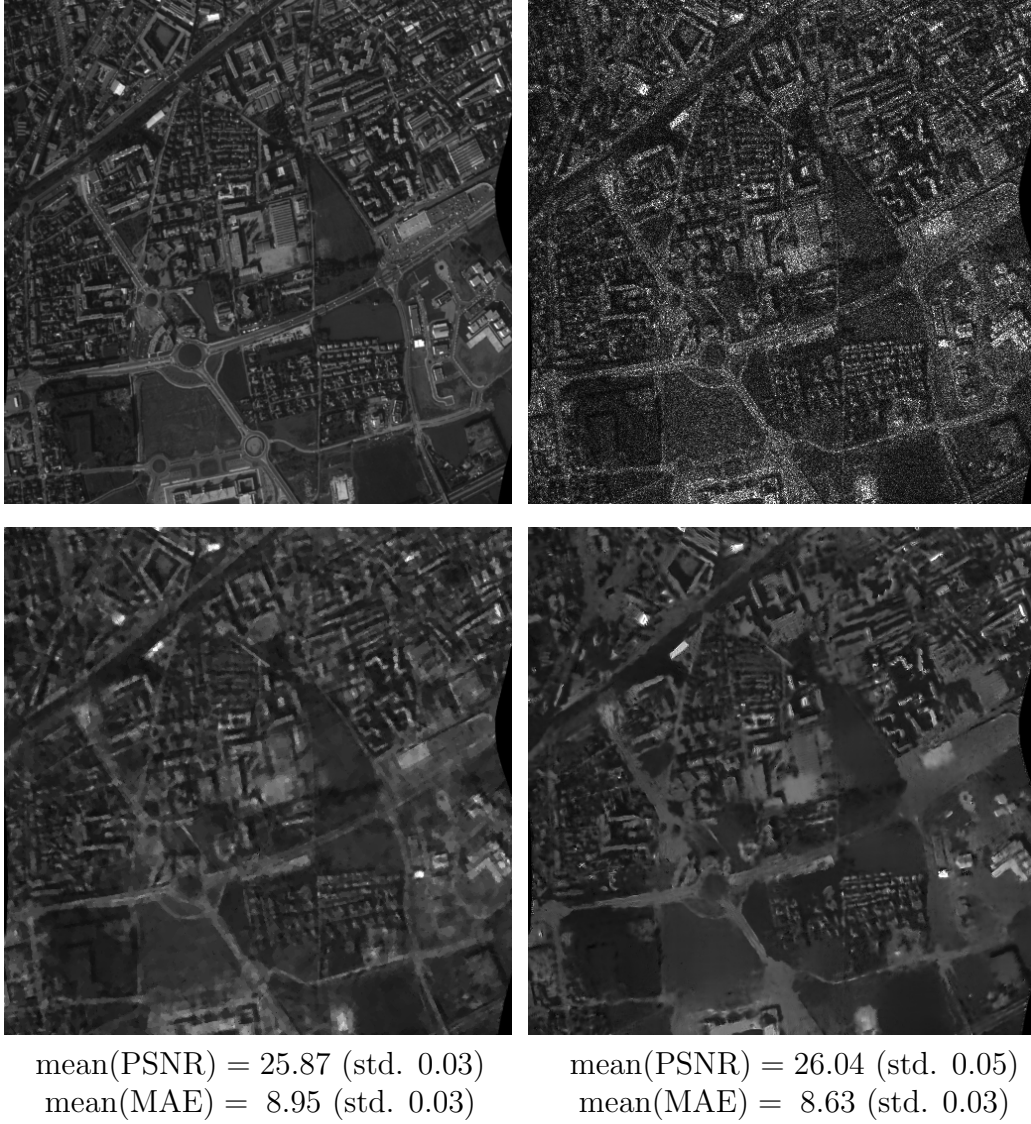


Figure 4.11: *Top*: Original image of the French city of Nîmes ( $512 \times 512$ ) with values in  $[1, 256]$  (left), cf. [70], and noisy image corrupted by multiplicative Gamma noise with  $L = 4$  (right). *Bottom*: Result by the hybrid method of Durand et al. shown in [70, Fig. 8] (left) as well as the result by our nonlocal filter (4.27) using (4.30) and an additional updating step with  $l = 7$ ,  $\omega = 29$ ,  $a = 2$ ,  $h = 0.5$ ,  $q = 0.7$ ,  $\tilde{l} = 5$ ,  $\tilde{a} = 1$ ,  $d = 0.1$  (right). The PSNR and MAE values have been averaged over the results for 500 noisy realizations of the original image.

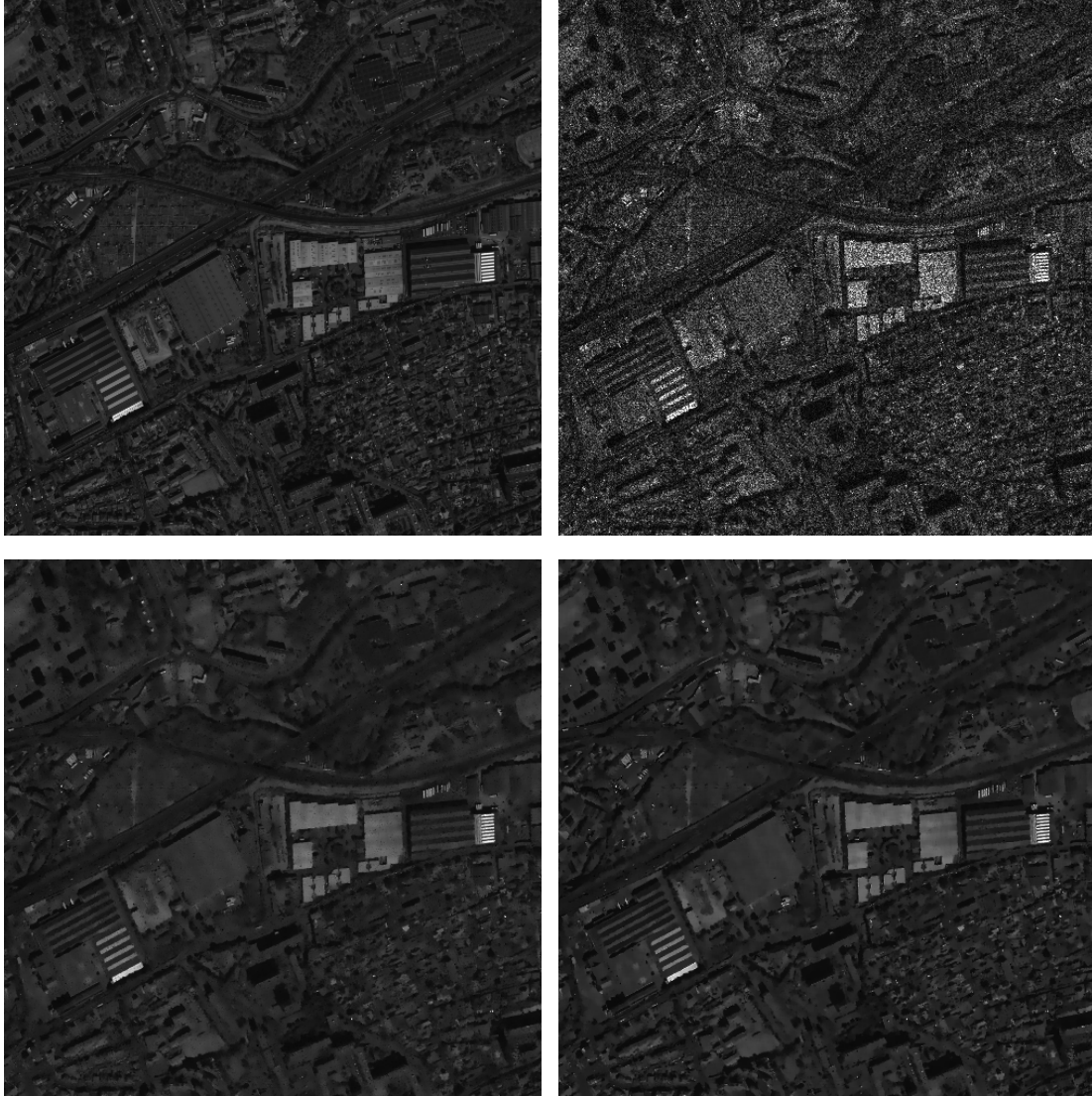
MAE values of all these results. Additionally, we indicate the estimated standard deviations (std.) to quantify their variability.

As a direct comparison shows, our obtained reconstruction is superior or at least competitive to the results obtained by different methods in [70, Fig. 8]. In this

paper the best result was obtained by the proposed hybrid multiplicative noise removal method, which combines variational and sparsity-based shrinkage methods involving curvelets as well as TV regularization. The result of this method is presented at Figure 4.11 (bottom left). For computing the corresponding average PSNR and MAE values we used the implementation of [70] available at [80].

**Results for multiplicative Rayleigh noise** To conclude this section, we present reconstructions of images corrupted by multiplicative Rayleigh noise. Our first example in Figure 4.12 shows an aerial image of size  $1500 \times 1500$ . The second set of images in Figure 4.13 are real singlelook SAR images provided by Sandia National Laboratories at [177]. To obtain first restoration results we have applied the nonlocal filter (4.28) with weights (4.30) adapted to multiplicative Rayleigh noise. For the improved results we performed an additional updating step. Again, our filter produces very good results, where most of the details are restored. Only extremely small details or details which lack a sufficient number of similar patches in their direct surrounding are not correctly restored.





mean(PSNR) = 30.40 (std. 0.02)

mean(MAE) = 42.87 (std. 0.06)

mean(PSNR) = 30.83 (std. 0.02)

mean(MAE) = 39.01 (std. 0.06)

Figure 4.12: *Top*: Original image ( $1500 \times 1500$ ), copyright [165], with values in  $[1, 2047]$  (left) and noisy version corrupted by multiplicative Rayleigh noise with  $\theta = 1$  (right). *Bottom*: Results by our nonlocal filter (4.28) using (4.30) with  $l = 7$ ,  $\omega = 21$ ,  $a = 2$ ,  $h = 0.4$ ,  $q = 0.6$  (left) and after an additional updating step with  $\tilde{l} = 5$ ,  $\tilde{a} = 1.5$ ,  $d = 0.05$  (right). Note that the displayed images have been subsampled to the size  $500 \times 500$  to better meet the standard screen and printer resolutions. The PSNR and MAE values have again been averaged over the results for 500 noisy realizations of the original image.

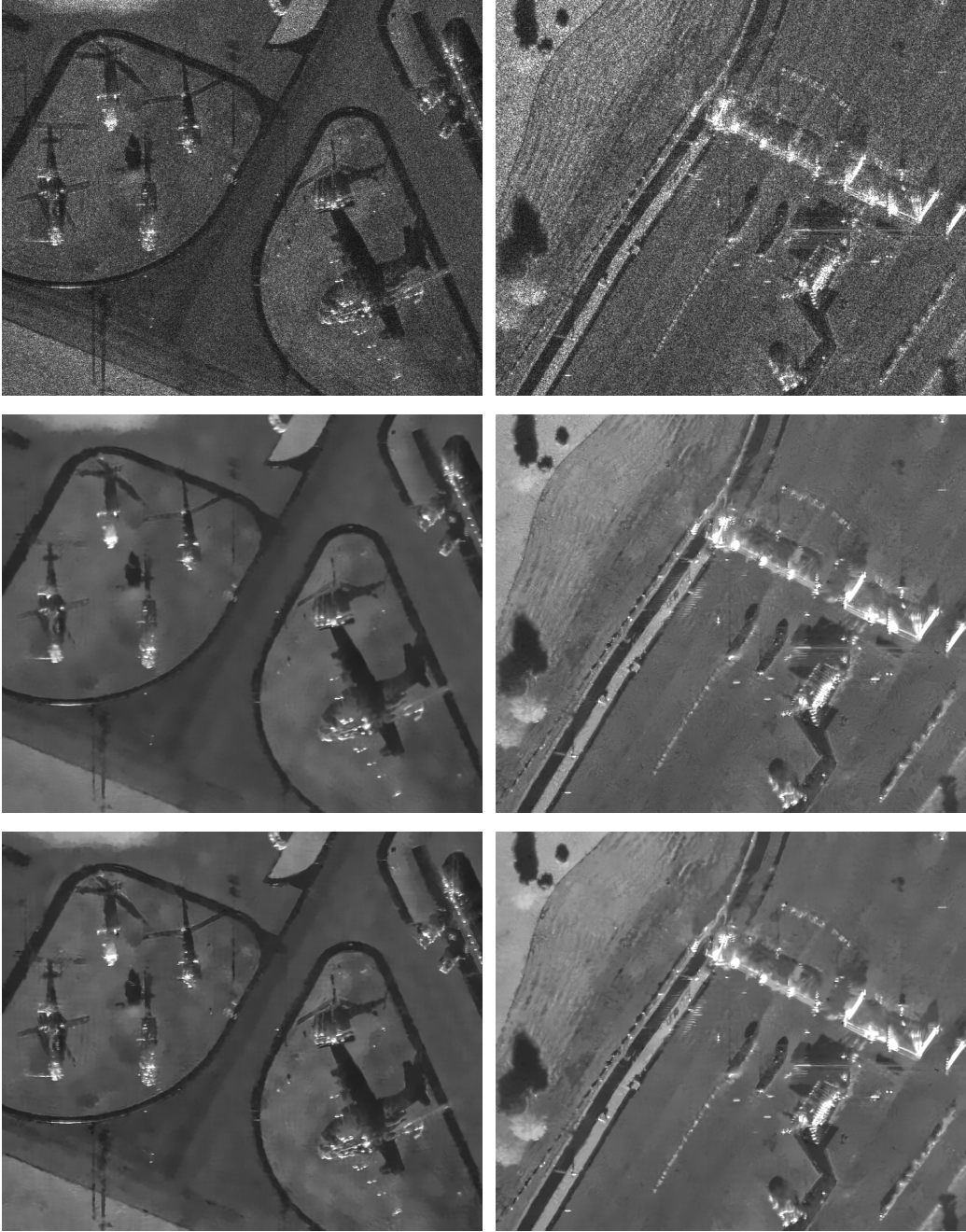


Figure 4.13: *Top*: Two real singlelook SAR images contaminated by multiplicative Rayleigh noise with  $\Theta = \frac{1}{\sqrt{2}}$ , copyright [177]. *Middle*: Results by our nonlocal filter (4.28) using (4.30) with  $l = 9$ ,  $\omega = 29$ ,  $a = 4$ ,  $h = 0.2$ ,  $q = 0.2$  (left) and  $h = 0.05$ ,  $q = 0$  (right). *Bottom*: Results after an additional updating step with  $\tilde{l} = 5$ ,  $\tilde{a} = 1$ ,  $d = 0.03$  (left) and  $d = 0.2$  (right). For contrast enhancement, cf. [177], the images are displayed after taking their square root and truncating the values outside of  $[0, \frac{1}{4} \max F]$  with  $F$  being the respective noisy image.

## CHAPTER 5

# Conclusions and Perspectives

Although a large variety of image restoration methods has been proposed in the last decades, there are still open problems and potential for improvements. One challenge faced in this thesis is the correct restoration of advanced image features like sharp corners and junctions in images. Another one is the development of new methods for restoring images corrupted by multiplicative noise.

In the first part of this thesis we focused on the restoration of images corrupted by strong additive Gaussian noise and presented different regularization and diffusion approaches to correctly restore in particular sharp corners and X junctions. For this purpose, we studied different orientation estimation tensors for locally determining up to two significant orientations per image pixel and incorporated the obtained orientation information into the restoration process.

As considered in [149] natural images may of course also contain image features like arbitrary junctions, where more than two significant orientations are required for a correct description. As a topic of future research it would be interesting to study also such structures and to examine appropriate extensions of the proposed methods which also take more than two significant orientations into account. To further improve the restoration results it would also be interesting to incorporate higher order derivatives into the restoration process as, e.g., done in [30, 131, 182, 189, SST11]. Hereby, it would for example be possible to overcome the staircasing effects observed in Figure 2.19.

In the very recent work [31] Bredies et al. proposed a new class of vertex-penalizing regularizers. Here, it might be worth exploring whether additional orientation information can also be incorporated in such approaches. Furthermore, since phase information obtained by approaches like, e.g., [106, 208] can provide a lot of information about an image, it would be interesting to also use such information for image restoration purposes.



In Chapter 3 we turned our attention to variational models for the removal of Poisson and multiplicative Gamma noise. Here, we compared different approaches from the literature for multiplicative Gamma noise and were able to prove the equivalence between one of these models and a standard maximum a posteriori (MAP) estimation based model for the removal of Poisson noise. This model for Poisson noise has not been considered for multiplicative Gamma noise before. Therefore, we studied in detail its properties for more general regularizers including also nonlocal ones. Besides, we proposed an efficient algorithm for determining the minimizer of the involved minimization problem. This algorithm was designed in such a way that it can also handle an additional linear transformation of the data, which can for example represent an additional blur operator.

The results in the first part of this chapter show that although the data fidelity term of variational models can often be determined by MAP estimation and a standard regularizer could be used, this need not lead to a good variational model. Instead, the regularizer needs to be appropriately chosen to fit to the determined data fidelity term. Hence, an open problem is still to find general rules for selecting data fidelity terms and appropriate regularizers for arbitrary types of noise. In addition, an automatic selection of the regularization parameter as done in [152, 220] would be desirable. Having the results in Chapter 2 in mind, it would be interesting to construct also for multiplicative noise regularizers which incorporate orientation information into the restoration process. However, in the presence of such noise the estimation of orientations in images is not straightforward due to the heavy distortions corrupting the data. To improve the results obtained by the nonlocal regularizers we could for example apply similar updating strategies as proposed for the nonlocal filters in Subsection 4.5.3.

In Chapter 4 we finally presented new nonlocal filters for the removal of multiplicative noise. A central point here was the definition of a suitable similarity measure for determining whether two given noisy pixels had the same initial gray value or not. For the deduction of this measure we studied in detail the similarity measure proposed for general noise models by Deledalle et al. in [60]. Furthermore, we proved different advantageous properties of our new measure and showed close relations to several of the measures studied recently in [59, 62]. Additionally, we could prove that several of the similarity measures compared in [59, 62] coincide for the considered multiplicative noise model.

As a topic of future research it would be interesting to see whether such relations can also be deduced for other types of noise. Besides, the use of our measure is of course not restricted to nonlocal filtering. It could be incorporated into appropriate variational methods or used for other applications such as inpainting and segmentation, see, e.g., [90, 91, 222]. Beyond that it could also be applied for block matching in registration problems relevant, e.g., for stereo vision or motion estimation, cf. [59, 143, 221].



# APPENDIX A

## Densities of Transformed Random Variables

To compute the probability density functions occurring in the former chapters we have used the following results from probability theory. Here, all random variables are supposed to be real-valued and continuous.

**Theorem A.1** (*Jacobi's Transformation Formula*, cf. [58, p.331], [111, p.92f])  
 Let  $X := (X_1, \dots, X_n)$  have the joint probability density function  $p_X$  and let  $T_1 \subseteq \mathbb{R}^n$  be an open set with  $P(X \in T_1) = 1$ . For  $T_2 \subseteq \mathbb{R}^n$  let  $g : T_1 \rightarrow T_2$  be an injective function which has a continuously differentiable inverse  $g^{-1}$  on  $T_2$  with non-vanishing Jacobian. Then, the density of  $Y = g(X)$  is given by

$$p_Y(y) = \begin{cases} p_X(g^{-1}(y)) |\det J_{g^{-1}}(y)| & \text{if } y \in T_2, \\ 0 & \text{otherwise.} \end{cases}$$

This theorem immediately implies the next three corollaries:

**Corollary A.2** Let  $X$  be a random variable with probability density function  $p_X$ , where  $p_X$  vanishes outside the interval  $[0, +\infty)$ . Then, the probability density function of  $Y = \sqrt{X}$  is given by

$$p_Y(y) = 2y p_X(y^2) 1_{\mathbb{R}_{\geq 0}}(y).$$

**Corollary A.3** Let  $X$  be a random variable with probability density function  $p_X$ , where  $p_X$  vanishes outside the interval  $[0, +\infty)$ . Then, the probability density function of  $Y = \log(X)$  is given by

$$p_Y(y) = \exp(y) p_X(\exp(y)).$$

**Corollary A.4** (Cf. [58, p. 336] and [101, p. 109])

For constants  $c, c_1, c_2 \neq 0$  and random variables  $X, X_1, X_2$ , where  $X_1$  and  $X_2$  are independent, we obtain the following probability density functions:

i)

$$p_{c_1 X_1 + c_2 X_2}(y) = \frac{1}{|c_1 c_2|} \int_{-\infty}^{\infty} p_{X_1} \left( \frac{t}{c_1} \right) p_{X_2} \left( \frac{y-t}{c_2} \right) dt,$$

ii)

$$p_{c X_1 X_2}(y) = \int_{-\infty}^{\infty} \frac{1}{|ct|} p_{X_1} \left( \frac{t}{c} \right) p_{X_2} \left( \frac{y}{t} \right) dt,$$

iii)

$$p_{c \frac{X_1}{X_2}}(y) = \int_{-\infty}^{\infty} \left| \frac{t}{c} \right| p_{X_1} \left( \frac{yt}{c} \right) p_{X_2}(t) dt$$

iv)

$$p_{cX}(y) = \frac{1}{|c|} p_X \left( \frac{y}{c} \right),$$

v)

$$p_{\frac{1}{X}}(y) = \frac{1}{y^2} p_X \left( \frac{1}{y} \right) \quad \text{for } y \neq 0.$$

Setting  $c_1 = -c_2 = 1$  in Corollary A.4 i) we obtain the following lemma:

**Lemma A.5** For independent random variables  $X_1, X_2$ , the probability density function  $p_{X_1 - X_2}$  has the following properties:

i)  $p_{X_1 - X_2}(y) = p_{X_2 - X_1}(-y)$ ,

ii) for identically distributed random variables  $X_1, X_2$  we have

$$p_{X_1 - X_2}(y) \leq p_{X_1 - X_2}(0) \quad \forall y \in \mathbb{R}.$$

---

**Proof:** Assertion i) follows directly from Corollary A.4 i). To prove ii) we use in addition the Cauchy-Schwarz inequality:

$$\begin{aligned} p_{X_1-X_2}(y) &= \int_{-\infty}^{\infty} p_{X_1}(t) p_{X_2}(t-y) dt \leq \sqrt{\int_{-\infty}^{\infty} (p_{X_1}(t))^2 dt \int_{-\infty}^{\infty} (p_{X_2}(t-y))^2 dt} \\ &= \int_{-\infty}^{\infty} (p_{X_1}(t))^2 dt = p_{X_1-X_2}(0). \end{aligned}$$

□

For the special case of exponentially distributed random variables, Corollary A.4 leads to the following lemma, see, e.g., [58, p. 190]:

**Lemma A.6** *Let  $X_i, i = 1, \dots, L$ , be independent exponentially distributed random variables with expectation value  $\lambda = 1$ . Then,  $X := \frac{1}{L} \sum_{i=1}^L X_i$  is Gamma distributed with density*

$$p_X(x) = \frac{L^L}{\Gamma(L)} x^{L-1} \exp(-Lx) 1_{\mathbb{R}_{\geq 0}}(x). \quad (\text{A.1})$$

**Proof:** To prove the assertion we use mathematical induction. For  $L = 1$  and a exponentially distributed random variable  $X_1$  with  $\lambda = 1$  it holds that

$$p_{X_1}(x) = \exp(-x) 1_{\mathbb{R}_{\geq 0}}(x) = \frac{1^1}{\Gamma(1)} x^0 \exp(-x) 1_{\mathbb{R}_{\geq 0}}(x),$$

i.e.,  $X := X_1$  is Gamma distributed with  $L = 1$ .

Now, let us assume that for some fixed  $L \in \mathbb{N} \setminus \{0\}$  the assertion holds true. Then, for  $L + 1$  exponentially distributed random variables  $X_i$  with  $\lambda = 1$  and  $\tilde{X} := \frac{1}{L} \sum_{i=1}^L X_i$  we have that

$$X := \frac{1}{L+1} \sum_{i=1}^{L+1} X_i = \frac{1}{L+1} \sum_{i=1}^L X_i + \frac{1}{L+1} X_{L+1} = \frac{L}{L+1} \tilde{X} + \frac{1}{L+1} X_{L+1}.$$

Hence, Corollary A.4 i) and the fact that  $\tilde{X}$  is Gamma distributed with density (A.1) imply that

$$p_X(x) = \frac{(L+1)^2}{L} \int_{-\infty}^{\infty} p_{\tilde{X}}\left(\frac{L+1}{L} t\right) p_{X_{L+1}}((L+1)(x-t)) dt$$

$$\begin{aligned}
 &= \frac{(L+1)^2}{L} 1_{\mathbb{R}_{\geq 0}}(x) \int_0^x \frac{L^L}{\Gamma(L)} \left( \frac{L+1}{L} t \right)^{L-1} \exp(-(L+1)t) \exp(-(L+1)(x-t)) dt \\
 &= \frac{(L+1)^{L+1}}{\Gamma(L)} \exp(-(L+1)x) 1_{\mathbb{R}_{\geq 0}}(x) \int_0^x t^{L-1} dt.
 \end{aligned}$$

Since

$$\int_0^x t^{L-1} dt = \left[ \frac{t^L}{L} \right]_0^x = \frac{x^L}{L}$$

and  $L \cdot \Gamma(L) = L(L-1)! = L! = \Gamma(L+1)$  we finally see that

$$p_X(x) = \frac{(L+1)^{L+1}}{\Gamma(L+1)} x^L \exp(-(L+1)x) 1_{\mathbb{R}_{\geq 0}}(x).$$

Thus,  $X$  is Gamma distributed with parameter  $L+1$ , which finishes the proof.  $\square$

Theorem A.1 and the definition of the conditional density yield the following result for the sum, respectively the product of two independent random variables:

**Proposition A.7** *Let  $X, Y$  be independent random variables.*

- i) *If  $Z := X + Y$ , then for any  $x \in \mathbb{R}$  with  $p_X(x) > 0$  the conditional density function of  $Z$  given  $X = x$  is*

$$p_{Z|X}(z|x) = p_Y(z-x).$$

- ii) *If  $Z := XY$ , then for any  $x \neq 0$  with  $p_X(x) > 0$  the conditional density function of  $Z$  given  $X = x$  is*

$$p_{Z|X}(z|x) = \frac{1}{|x|} p_Y\left(\frac{z}{x}\right).$$

# Bibliography

- [1] The MOSEK optimization toolbox. <http://www.mosek.com>. 39
- [2] T. Aach, C. Mota, I. Stuke, M. Mühlich, and E. Barth. Analysis of superimposed oriented patterns. *IEEE Transactions on Image Processing*, 15(12):3690–3700, 2006. 3, 16, 24, 27
- [3] R. Acar and C. R. Vogel. Analysis of bounded variation penalty methods for ill-posed problems. *Inverse Problems*, 10(6):1217–1229, 1994. 32, 68, 77
- [4] S. Aja-Fernández and C. Alberola-López. On the estimation of the coefficient of variation for anisotropic diffusion speckle filtering. *IEEE Transactions on Image Processing*, 15(9):2694–2701, 2006. 69
- [5] C. D. Aliprantis and K. C. Border. *Infinite Dimensional Analysis: A Hitchhiker’s Guide*. Springer, third edition, 2006. 81
- [6] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Mathematical Programming*, 95:3–51, 2003. 39
- [7] F. Alter, V. Caselles, and A. Chambolle. Evolution of characteristic functions of convex sets in the plane by the minimizing total variation flow. *Interfaces and Free Boundaries*, 7(1):29–53, 2005. 2, 14
- [8] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Mathematical Monographs. Oxford University Press, 2000. 68
- [9] F. Andreu-Vaillio, V. Caselles, and J. M. Mazón. *Parabolic Quasilinear Equations Minimizing Linear Growth Functionals*, volume 223 of *Progress in Mathematics*. Birkhäuser, 2004. 70, 78, 84
- [10] G. Aubert and J.-F. Aujol. A variational approach to removing multiplicative noise. *SIAM Journal on Applied Mathematics*, 68(4):925–946, 2008. 70, 71, 72, 83, 101, 112, 127
- [11] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, volume 147 of *Applied Mathematical Sciences*. Springer, New York, second edition, 2006. 9, 70, 83

- [12] A. Auslender and M. Teboulle. *Asymptotic Cones and Functions in Optimization and Variational Inequalities*. Springer Monographs in Mathematics. Springer, New York, 2003. [82](#), [87](#)
- [13] S. P. Awate and R. T. Whitaker. Higher-order image statistics for unsupervised, information-theoretic, adaptive, image filtering. In *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, volume 2, pages 44–51, 2005. [106](#)
- [14] S. P. Awate and R. T. Whitaker. Unsupervised, information-theoretic, adaptive image filtering for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3):364–376, 2006. [106](#)
- [15] J. M. Bardsley and A. Luttmann. Total variation-penalized Poisson likelihood estimation for ill-posed problems. *Advances in Computational Mathematics*, 31(1–3):35–59, 2009. [4](#), [68](#)
- [16] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer, New York, 2011. [45](#)
- [17] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing*, 18(11):2419–2434, 2009. [39](#), [93](#)
- [18] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009. [39](#), [93](#)
- [19] F. Benvenuto, A. La Camera, C. Theys, A. Ferrari, H. Lantéri, and M. Bertero. The study of an iterative method for the reconstruction of images corrupted by Poisson and Gaussian noise. *Inverse Problems*, 24(3), 2008. [64](#)
- [20] B. Berkels, M. Burger, M. Droske, O. Nemitz, and M. Rumpf. Cartoon extraction based on anisotropic image classification. In *Vision, Modeling, and Visualization Proceedings*, pages 293–300. 2006. [3](#), [15](#), [29](#)
- [21] M. Bertero and P. Boccacci. *Introduction to Inverse Problems in Imaging*. IOP Publishing, Bristol, 1998. [63](#)
- [22] M. Bertero, P. Boccacci, Desiderà, and G. Vicidomini. Image deblurring with Poisson data: from cells to galaxies. *Inverse Problems*, 25(12), 2009. [62](#), [63](#), [64](#), [67](#)
- [23] M. Bertero, H. Lantéri, and L. Zanni. Iterative image reconstruction: a point of view. In Y. Censor, M. Jiang, and A. K. Luis, editors, *Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation Therapy*, volume 7 of *Publications of the Scuola Normale Superiore, CRM Series*, pages 37–63. Springer, 2008. [67](#)
- [24] J. Bigun. *Vision with Direction*. Springer, Berlin Heidelberg, 2006. [17](#)

- 
- [25] J. M. Bioucas-Dias and M. A. T. Figueiredo. Multiplicative noise removal using variable splitting and constrained optimization. *IEEE Transactions on Image Processing*, 19(7):1720–1730, 2010. [64](#), [65](#)
- [26] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011. [4](#), [50](#), [86](#), [90](#)
- [27] A. Braides.  *$\Gamma$ -Convergence for Beginners*, volume 22 of *Oxford Lecture Series in Mathematics and Its Applications*. Oxford University Press, Oxford, 2002. [80](#), [81](#)
- [28] E. Bratsolis and M. Sigelle. A spatial regularization method preserving local photometry for Richardson-Lucy restoration. *Astronomy & Astrophysics*, 375(3):1120–1128, 2001. [63](#)
- [29] E. Bratsolis and M. Sigelle. Fast SAR image restoration, segmentation, and detection of high-reflectance regions. *IEEE Transactions on Geoscience and Remote Sensing*, 41(12):2890–2899, 2003. [64](#)
- [30] K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, 2010. [1](#), [139](#)
- [31] K. Bredies, T. Pock, and B. Wirth. Convex relaxation of a class of vertex penalizing functionals. SFB-Report 2012-001, Graz, 2012. [139](#)
- [32] L. Bregman. The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7:200–217, 1967. [67](#)
- [33] T. Brox, J. Weickert, B. Burgeth, and P. Mrázek. Nonlinear structure tensors. *Image and Vision Computing*, 24(1):41–55, 2006. [17](#)
- [34] T. Brox, M. Welk, G. Steidl, and J. Weickert. Equivalence results for TV diffusion and TV regularisation. In L. D. Griffin and M. Lillholm, editors, *Scale Space Methods in Computer Vision, 4th International Conference*, volume 2695 of *LNCS*, pages 86–100. Springer, 2003. [1](#), [11](#)
- [35] C. Brune. *4D Imaging in Tomography and Optical Nanoscopy*. PhD thesis, Institute for Computational and Applied Mathematics, University of Münster, 2010. [63](#), [67](#), [77](#), [78](#), [79](#), [94](#)
- [36] C. Brune, M. Burger, A. Sawatzky, T. Kösters, and F. Wübbeling. Forward-backward EM-TV methods for inverse problems with Poisson noise. *Inverse Problems*, 2011. Submitted. [94](#)
- [37] C. Brune, A. Sawatzky, and M. Burger. Primal and dual bregman methods with application to optical nanoscopy. *International Journal of Computer Vision*, 92(2):211–229, 2011. [95](#)

- [38] A. Buades, B. Coll, and J.-M. Morel. Online demo: Non-local means denoising. Available at [http://www.ipol.im/pub/algo/bcm\\_non\\_local\\_means\\_denoising](http://www.ipol.im/pub/algo/bcm_non_local_means_denoising). 129
- [39] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 60–65, 2005. 1, 2, 5, 15, 52, 68, 105, 106
- [40] A. Buades, B. Coll, and J.-M. Morel. Image denoising methods. A new nonlocal principle. *SIAM Review*, 52(1):113–147, 2010. 106
- [41] C. B. Burckhardt. Speckle in ultrasound B-mode scans. *IEEE Transactions on Sonics and Ultrasonics*, 25(1):1–6, 1978. 64
- [42] M. Burger, K. Frick, S. Osher, and O. Scherzer. Inverse total variation flow. *Multiscale Modeling & Simulation*, 6(2):366–395, 2007. 14
- [43] M. Burger, G. Gilboa, S. Osher, and J. Xu. Nonlinear inverse scale space methods. *Communications in Mathematical Sciences*, 4(1):179–212, 2006. 14
- [44] E. Candès and D. Donoho. Curvelets - a surprisingly effective nonadaptive representation for objects with edges. In A. Cohen, C. Rabut, and L. Schumaker, editors, *Curves and Surface fitting: Saint-Malo 1999*, pages 105–120, 2000. 16
- [45] F. J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986. 4, 52
- [46] F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM Journal on Numerical Analysis*, 29(1):182–193, 1992. 14, 41
- [47] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1–2):89–97, 2004. 81, 93, 94, 95
- [48] A. Chambolle. Total variation minimization and a class of binary MRF models. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 3757 of *LNCS*, pages 136–152. Springer, 2005. 9, 39, 95
- [49] A. Chambolle, V. Caselles, M. Novaga, D. Cremers, and T. Pock. An introduction to total variation for image analysis. In M. Fornasier, editor, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, pages 263–340. De Gruyter, 2010. 2, 14
- [50] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(2):167–188, 1997. 9, 13, 44, 70
- [51] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011. 39, 93



- 
- [52] T. Chan, S. Esedoglu, F. Park, and A. Yip. Total variation image restoration: Overview and recent developments. In N. Paragios, Y. Chen, and O. Faugeras, editors, *Handbook of Mathematical Models in Computer Vision*, pages 17–32. Springer, 2006. 9, 13, 40, 69
- [53] D. Cireşan and D. Damian. Preserving topological information in the windowed Hough transform for rectangle extraction. In K. Franke, K.-R. Müller, B. Nickolay, and R. Schäfer, editors, *Pattern Recognition, 28th DAGM Symposium*, volume 4174 of *LNCS*, pages 172–181. Springer, 2006. 4, 53, 56
- [54] P. Coupé, P. Hellier, C. Kervrann, and C. Barillot. Nonlocal means-based speckle filtering for ultrasound images. *IEEE Transactions on Image Processing*, 18(10):2221–2229, 2009. 106
- [55] I. Csiszár. Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *The Annals of Statistics*, 19(4):2032–2066, 1991. 67
- [56] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007. 1
- [57] G. Dal Maso. *An Introduction to  $\Gamma$ -Convergence*. Birkhäuser, Boston, 1993. 79
- [58] A. DasGupta. *Fundamentals of Probability: A First Course*. Springer Texts in Statistics. Springer, first edition, 2010. 141, 142, 143
- [59] C.-A. Deledalle. *Image denoising beyond additive Gaussian noise - Patch-based estimators and their application to SAR imagery*. PhD thesis, Telecom ParisTech, 2011. <http://tel.archives-ouvertes.fr/tel-00662520>. 5, 6, 122, 123, 124, 126, 140
- [60] C.-A. Deledalle, L. Denis, and F. Tupin. Iterative weighted maximum likelihood denoising with probabilistic patch-based weights. *IEEE Transactions on Image Processing*, 18(12):2661–2672, 2009. 5, 6, 106, 107, 108, 110, 114, 126, 131, 140
- [61] C.-A. Deledalle, F. Tupin, and L. Denis. Poisson NL means: Unsupervised non local means for Poisson noise. In *Image Processing (ICIP), 17th IEEE International Conference on*, pages 801–804, 2010. 107
- [62] C.-A. Deledalle, F. Tupin, and L. Denis. Patch similarity under non Gaussian noise. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 1845–1848, 2011. 5, 6, 122, 123, 124, 126, 140
- [63] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977. 67, 93

- [64] N. Dey, L. Blanc-Feraud, C. Zimmer, P. Roux, Z. Kam, J.-C. Olivo-Marin, and J. Zerubia. Richardson-Lucy algorithm with total variation regularization for 3D confocal microscope deconvolution. *Microscopy Research and Technique*, 69(4):260–266, 2006. [63](#)
- [65] S. Didas. *Denoising and Enhancement of Digital Images - Variational Methods, Integro-differential Equations, and Wavelets*. PhD thesis, Saarland University, Saarbrücken, Germany, 2008. [41](#)
- [66] D. Donoho and J. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994. [1](#)
- [67] J. Douglas and H. H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American Mathematical Society*, 82(2):421–439, 1956. [86](#)
- [68] R. Duits and E. Franken. Left-invariant diffusions on the space of positions and orientations and their application to crossing-preserving smoothing of HARDI images. *International Journal of Computer Vision*, 92(3):231–264, 2011. [16](#)
- [69] S. Durand, J. Fadili, and M. Nikolova. Multiplicative noise cleaning via a variational method involving curvelet coefficients. In X.-C. Tai, K. Morken, M. Lysaker, and K.-A. Lie, editors, *Scale Space and Variational Methods in Computer Vision*, volume 5567 of *LNCS*, pages 282–294. Springer, 2009. [71](#)
- [70] S. Durand, J. Fadili, and M. Nikolova. Multiplicative noise removal using L1 fidelity on frame coefficients. *Journal of Mathematical Imaging and Vision*, 36(3):201–226, 2010. [ii](#), [71](#), [133](#), [135](#), [136](#)
- [71] J. Eckstein. *Splitting methods for monotone operators with applications to parallel optimization*. PhD thesis, Massachusetts Institute of Technology, 1989. <http://dspace.mit.edu/handle/1721.1/14356>. [86](#)
- [72] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(3):293–318, 1992. [87](#)
- [73] I. Ekeland and R. Témam. *Convex Analysis and Variational Problems*. Society for Industrial and Applied Mathematics, Philadelphia, USA, 1999. [31](#), [76](#), [80](#)
- [74] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, 2006. [1](#)
- [75] S. Esedoglu and S. Osher. Decomposition of images by the anisotropic Rudin-Osher-Fatemi model. *Communications in Pure and Applied Mathematics*, 57(12):1609–1626, 2004. [3](#), [9](#), [15](#), [29](#), [30](#), [32](#), [33](#)

- 
- [76] E. Esser. Applications of Lagrangian-based alternating direction methods and connections to split Bregman. CAM-Report 09-31, UCLA, Los Angeles, 2009. 50, 86, 88
- [77] E. Esser, X. Zhang, and T. F. Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM Journal on Imaging Sciences*, 3(4):1015–1046, 2010. 39, 93
- [78] European Space Agency (ESA). Multilook SAR image. 104
- [79] L. C. Evans and R. F. Gariepy. *Measure Theory and Fine Properties of Functions*. Studies in Advanced Mathematics. CRC Press, 1992. 77
- [80] J. Fadili, M. Nikolova, and S. Durand. Matlab toolbox: Multiplicative noise removal by L1-data fidelity on frame coefficients. Available at <http://www.greyc.ensicaen.fr/~jfadili/software.html>, 2009. 136
- [81] J. Fan, M. Farnen, and I. Gijbels. Local maximum likelihood estimation and inference. *Journal of the Royal Statistical Society, Series B, Statistical methodology*, pages 591–608, 1998. 126
- [82] M. A. T. Figueiredo and J. M. Bioucas-Dias. Deconvolution of Poissonian images using variable splitting and augmented Lagrangian optimization. In *IEEE Workshop on Statistical Signal Processing*, Cardiff, 2009. 92, 93
- [83] M. A. T. Figueiredo and J. M. Bioucas-Dias. Restoration of Poissonian images using alternating direction optimization. *IEEE Transactions on Image Processing*, 19(12):3133–3145, 2010. 91, 92, 93, 95
- [84] I. Fonseca and G. Leoni. *Modern Methods in the Calculus of Variations:  $L^p$  Spaces*. Springer Monographs in Mathematics. Springer, 2007. 78
- [85] W. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *Proc. ISPRS Intercommission Conf. on Fast Processing of Photogrammetric Data*, pages 281–305, 1987. 3, 15, 16
- [86] E. Franken and R. Duits. Crossing-preserving coherence-enhancing diffusion on invertible orientation scores. *International Journal of Computer Vision*, 85(3):253–278, 2009. 2, 16
- [87] D. Gabay. Applications of the method of multipliers to variational inequalities. In M. Fortin and R. Glowinski, editors, *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, volume 15 of *Studies in Mathematics and its Applications*, chapter 9, pages 299–331. Elsevier Science Publishers B.V., Amsterdam, 1983. 86, 87
- [88] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications*, 2(1):17–40, 1976. 50, 86

- [89] G. Gilboa, J. Darbon, S. Osher, and T. F. Chan. Nonlocal convex functionals for image regularization. CAM-Report 06-57, UCLA, Los Angeles, 2006. [68](#)
- [90] G. Gilboa and S. Osher. Nonlocal linear image regularization and supervised segmentation. *Multiscale Modeling & Simulation*, 6(2):595–630, 2007. [68](#), [140](#)
- [91] G. Gilboa and S. Osher. Nonlocal operators with applications to image processing. *Multiscale Modeling & Simulation*, 7(3):1005–1028, 2008. [4](#), [68](#), [81](#), [100](#), [140](#)
- [92] E. Giusti. *Minimal Surfaces and Functions of Bounded Variation*, volume 80 of *Mono-graphs in Mathematics*. Birkhäuser, Basel, Boston, Stuttgart, 1984. [68](#)
- [93] R. Glowinski and A. Marroco. Sur l’approximation, par éléments finis d’ordre un, et la résolution, par pénalisation-dualité d’une classe de problèmes de Dirichlet non linéaires. *Revue française d’automatique, informatique, recherche opérationnelle. Analyse numérique*, 9(2):41–76, 1975. [50](#), [86](#)
- [94] M. Goldbach. Image ”Durchgang”. Available 2008 at <http://www.photocase.com/de/upload/02/rdva4ha8/photocasefknsx7u23.jpg>. [55](#)
- [95] D. Goldfarb, Z. Wen, and W. Yin. A curvilinear search method for  $p$ -harmonic flows on spheres. *SIAM Journal on Imaging Sciences*, 2(1):84–109, 2009. [16](#)
- [96] T. Goldstein and S. Osher. The split Bregman method for L1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009. [50](#), [86](#), [93](#)
- [97] J. W. Goodman. Statistical properties of laser speckle patterns. In J. C. Dainty, editor, *Laser Speckle and Related Phenomena*, volume 9 of *Topics in Applied Physics*, pages 9–75. Springer, 1975. [64](#)
- [98] J. W. Goodman. *Speckle Phenomena in Optics*. Roberts & Company, 2007. [64](#), [65](#)
- [99] M. Grasmair. A coarea formula for anisotropic total variation regularisation. Technical Report 103, FSP 092: Joint Research Program of Industrial Geometry, 2010. [i](#), [4](#), [33](#), [78](#)
- [100] M. Grasmair and F. Lenzen. Anisotropic total variation filtering. *Applied Mathematics & Optimization*, 62(3):323–339, 2010. [38](#)
- [101] G. R. Grimmett and D. R. Stirzaker. *Probability and Random Processes*. Oxford University Press, third edition, 2001. [62](#), [108](#), [142](#)
- [102] K. Guo, G. Kutyniok, and D. Labate. Sparse multidimensional representations using anisotropic dilation and shear operators. In G. Chen and M.-J. Lai, editors, *Wavelets and splines: Athens 2005*, Modern Methods in Mathematics, pages 189–201. Nashboro Press, 2006. [16](#)

- 
- [103] A. B. Hamza and H. Krim. A variational approach to maximum a posteriori estimation for image denoising. In M. A. T. Figueiredo, J. Zerubia, and A. K. Jain, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition, Third International Workshop*, volume 2134 of *LNCS*, pages 19–34, 2001. [66](#)
- [104] P. C. Hansen, J. G. Nagy, and D. P. O’Leary. *Deblurring Images: Matrices, Spectra, and Filtering*. Fundamentals of Algorithms. SIAM, 2006. [92](#), [95](#)
- [105] B. S. He, H. Yang, and S. L. Wang. Alternating direction method with self-adaptive penalty parameters for monotone variational inequalities. *Journal of Optimization Theory and Applications*, 106(2):337–356, 2000. [4](#), [90](#)
- [106] S. Held, M. Storath, P. Massopust, and B. Forster. Steerable Wavelet frames based on the Riesz transform. *IEEE Transactions on Image Processing*, 19(3):653–667, 2010. [139](#)
- [107] M. Hintermüller and K. Kunisch. Total bounded variation regularization as a bilaterally constrained optimization problem. *SIAM Journal on Applied Mathematics*, 4(64):1311–1333, 2004. [9](#)
- [108] E. J. Hoffman, S.-C. Huang, M. E. Phelps, and D. E. Kuhl. Quantitation in positron emission computed tomography: 4. Effect of accidental coincidences. *Journal of Computer Assisted Tomography*, 5(3):391–400, 1981. [63](#)
- [109] T. J. Holmes, D. L. Snyder, and D. C. Ficke. The effect of accidental coincidences in time-of-flight positron emission tomography. *IEEE Transactions on Medical Imaging*, 3(2):68–79, 1984. [63](#)
- [110] Y.-M. Huang, M. K. Ng, and Y.-W. Wen. A new total variation method for multiplicative noise removal. *SIAM Journal on Imaging Sciences*, 2(1):20–40, 2009. [73](#)
- [111] J. Jacod and P. Protter. *Probability Essentials*. Springer, second edition, 2004. [64](#), [141](#)
- [112] A. Jeffrey and H.-H. Dai. *Handbook of Mathematical Formulas and Integrals*. Academic Press, fourth edition, 2008. [71](#), [122](#)
- [113] Z. Jin and X. Yang. Analysis of a new variational model for multiplicative noise removal. *Journal of Mathematical Analysis and Applications*, 362(2):415–426, 2010. [74](#)
- [114] E. Jonsson, S.-C. Huang, and T. Chan. Total variation regularization in positron emission tomography. CAM-Report 98-48, UCLA, Los Angeles, 1998. [4](#), [68](#)
- [115] J. P. Kaipio and K. Somersalo. *Statistical and Computational Inverse Problems*, volume 160 of *Applied Mathematical Sciences*. Springer, 2005. [66](#)
- [116] R. W. Keener. *Theoretical Statistics*. Springer Texts in Statistics. Springer, 2010. [125](#)

- [117] C. Kervrann, J. Boulanger, and P. Coupé. Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal. In F. Sgallari, A. Murli, and N. Paragios, editors, *Scale Space and Variational Methods in Computer Vision*, volume 4485 of *LNCIS*, pages 520–532. Springer, 2007. [106](#)
- [118] S. Kichenassamy. The Perona-Malik paradox. *SIAM Journal on Applied Mathematics*, 57(5):1328–1342, 1997. [14](#)
- [119] R. Kimmel, R. Malladi, and N. Sochen. Images as embedded maps and minimal surfaces: Movies, color, texture, and volumetric medical images. *International Journal of Computer Vision*, 39(2):111–129, 2000. [1](#)
- [120] R. Kimmel and N. Sochen. Orientation diffusion or how to comb a porcupine? *Journal of Visual Communication and Image Representation*, 13(1–2):238–248, 2002. [16](#)
- [121] S. Kindermann, S. Osher, and P. W. Jones. Deblurring and denoising of images by nonlocal functionals. *Multiscale Modeling & Simulation*, 4(4):1091–1115, 2005. [1](#), [68](#)
- [122] P. Kornprobst, R. Deriche, and G. Aubert. Image sequence analysis via partial differential equations. *Journal of Mathematical Imaging and Vision*, 11(1):5–26, 1999. [70](#)
- [123] K. Krissian, C.-F. Westin, R. Kikinis, and K. G. Vosburgh. Oriented speckle reducing anisotropic diffusion. *IEEE Transactions on Image Processing*, 16(5):1412–1424, 2007. [64](#), [69](#)
- [124] A. Kryvanos, J. Hesser, and G. Steidl. Nonlinear image restoration methods for marker extraction in 3D fluorescent microscopy. In C. A. Bouman and E. L. Miller, editors, *Computational Imaging III, Proc. SPIE*, volume 5674, pages 432–443, 2005. [63](#)
- [125] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel. Adaptive noise smoothing filter for images with signal-dependent noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2):165–177, 1985. [69](#)
- [126] D. Labate, W.-Q. Lim, G. Kutyniok, and G. Weiss. Sparse multidimensional representation using shearlets. In M. Papadakis, A. F. Laine, and M. A. Unser, editors, *Wavelets XI*, volume 5914 of *Proceedings of SPIE*, pages 254–262, 2005. [16](#)
- [127] S. Lang. *Calculus of Several Variables*. Addison-Wesley series in Mathematics. Addison-Wesley Publishing Company, 1973. [13](#)
- [128] H. Lantéri and C. Theys. Restoration of astrophysical images - the case of Poisson data with additive Gaussian noise. *EURASIP Journal on Applied Signal Processing*, 15:2500–2513, 2005. [63](#), [64](#)
- [129] T. Le, R. Chartrand, and T. J. Asaki. A variational approach to reconstructing images corrupted by Poisson noise. *Journal of Mathematical Imaging and Vision*, 27(3):257–263, 2007. [4](#), [68](#)

- 
- [130] J.-S. Lee. Digital image enhancement and noise filtering by use of local statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(2):165–168, 1980. 69
- [131] S. Lefkimmatis, A. Bourquard, and M. Unser. Hessian-based norm regularization for image restoration with biomedical applications. *IEEE Transactions on Image Processing*, 21(3):983–995, 2012. 139
- [132] F. Lenzen, F. Becker, J. Lellmann, S. Petra, and C. Schnörr. Variational image denoising with adaptive constraint sets. In A. M. Bruckstein, B. M. ter Haar Romeny, A. M. Bronstein, and M. M. Bronstein, editors, *Scale Space and Variational Methods in Computer Vision, Third International Conference, SSVM 2011*, volume 6667 of *LNCIS*, pages 206–217. Springer, 2012. 38
- [133] B. C. Levy. *Principles of Signal Detection and Parameter Estimation*. Springer, 2008. 125
- [134] P. L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979. 86
- [135] Y. Lou, X. Zhang, S. Osher, and A. Bertozzi. Image recovery via nonlocal operators. *Journal of Scientific Computing*, 42(2):185–197, 2010. 68
- [136] L. B. Lucy. An iterative technique for the rectification of observed distributions,. *The Astronomical Journal*, 79(6):745–754, 1974. 68
- [137] O. Lysaker, S. Osher, and X.-C. Tai. Noise removal using smoothed normals and surface fitting. *IEEE Transactions on Image Processing*, 13(10):1345–1357, 2004. 16
- [138] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2272–2279, 2009. 1
- [139] H. Maître. *Processing of Synthetic Aperture Radar Images*. ISTE Ltd and John Wiley & Sons, 2008. 64, 65
- [140] M. Mäkitalo and A. Foi. Optimal inversion of the Anscombe transformation in low-count Poisson image denoising. *IEEE Transactions on Image Processing*, 20(1):99–109, 2011. 106
- [141] M. Mäkitalo, A. Foi, D. Fevrale, and V. Lukin. Denoising of single-look SAR images based on variance stabilization and nonlocal filters. In *Mathematical Methods in Electromagnetic Theory (MMET), 2010 International Conference on*, pages 1–4, 2010. 106
- [142] J. V. Manjon-Herrera and A. Buades. Matlab implementation: Non-local means filter. Available at <http://www.mathworks.com/matlabcentral/fileexchange/13176-non-local-means-filter>, 2006. 51, 52, 57, 58



- [143] Y. Matsushita and S. Lin. A probabilistic intensity similarity measure based on noise distributions. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR '07*, 2007. 108, 140
- [144] H. Mayer. Automatic object extraction from aerial imagery - a survey focusing on buildings. *Computer Vision and Image Understanding*, 74(2):138–149, 1999. 53
- [145] Y. Meyer. *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, volume 22 of *University Lecture Series*. American Mathematical Society, Providence, 2001. 68
- [146] T. P. Minka. Distance measures as prior probabilities. Technical report, 2000. <http://research.microsoft.com/en-us/um/people/minka/papers/metric/minka-metric.pdf>. 123
- [147] T. P. Minka. Bayesian inference, entropy, and the multinomial distribution. Technical report, 2003. <http://research.microsoft.com/en-us/um/people/minka/papers/minka-multinomial.pdf>. 39, 123
- [148] C. Mota, I. Stuke, T. Aach, and E. Barth. Estimation of multiple orientations at corners and junctions. In C. E. Rasmussen, H. H. Bülthoff, B. Schölkopf, and M. A. Giese, editors, *Pattern Recognition, 26th DAGM Symposium, Proceedings*, volume 3175 of *LNCIS*, pages 163–170. Springer, 2004. 3, 16, 24
- [149] M. Mühlich and T. Aach. A theory for multiple orientation estimation. In H. Bischof and A. Leonardis, editors, *Proceedings European Conference on Computer Vision 2006, Part II*, volume 3952 of *LNCIS*, pages 69–82. Springer, 2006. 24, 139
- [150] F. Natterer and F. Wübbeling. *Mathematical Methods in Image Reconstruction*. SIAM Monographs on Mathematical Modeling and Computation. SIAM, Philadelphia, 2001. 67, 93
- [151] Y. Nesterov. Smooth minimization of non-smooth functions. *Mathematical Programming*, 103(1):127–152, 2005. 39, 93
- [152] M. K. 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):2710–2736, 2010. 140
- [153] M. Nikolova. Model distortions in Bayesian MAP reconstruction. *Inverse Problems and Imaging*, 1(2):399–422, 2007. 66
- [154] C. Olsson, M. Byröd, N. C. Overgaard, and F. Kahl. Extending continuous cuts: Anisotropic metrics and expansion moves. In *International Conference on Computer Vision*, page 4, 2009. 33
- [155] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation based image restoration. *Multiscale Modeling and Simulation*, 4:460–489, 2005. 1, 14



- 
- [156] V. Y. Panin, G. L. Zeng, and G. T. Gullberg. Total variation regulated EM algorithm. *IEEE Transactions on Nuclear Science*, 46(6):2202–2210, 1999. [63](#), [68](#)
- [157] S. Paris and F. Durand. A fast approximation of the bilateral filter using a signal processing approach. *MIT Technical Report*, MIT-CSAIL-TR-2006-073, 2006. [15](#)
- [158] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand. Bilateral filtering: Theory and applications. *Foundations and Trends in Computer Graphics and Vision*, 4(1):1–73, 2009. [1](#)
- [159] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990. [1](#), [10](#), [14](#), [41](#)
- [160] G. Peyré, S. Bogleux, and L. Cohen. Non-local regularization of inverse problems. In D. Forsyth, P. Torr, , and A. Zisserman, editors, *Computer Vision - ECCV 2008, Proceedings, Part III*, volume 5304 of *LNCS*, pages 57–68. Springer, 2008. [68](#)
- [161] G. Peyré, S. Bogleux, and L. Cohen. Non-local regularization of inverse problems. *Inverse Problems and Imaging*, 5(2):511–530, 2011. [68](#)
- [162] D. G. Politte and D. L. Snyder. Corrections for accidental coincidences and attenuation in maximum-likelihood image reconstruction for positron-emission tomography. *IEEE Transactions on Medical Imaging*, 10(1):82–89, 1991. [63](#)
- [163] J. Polzehl and V. Spokoiny. Propagation-separation approach for local likelihood estimation. *Probability Theory and Related Fields*, 135(3):335–362, 2006. [126](#), [131](#)
- [164] D. Potts and G. Steidl. Optimal trigonometric preconditioners for nonsymmetric Toeplitz systems. *Linear Algebra and its Applications*, 281(1–3):265–292, 1998. [92](#)
- [165] Quickbird sensor, Digital Globe. Optical image. [137](#)
- [166] T. Rahman, X.-C. Tai, and S. Osher. A TV-Stokes denoising algorithm. In F. Sgallari and A. M. und N. Paragios, editors, *Scale Space and Variational Methods in Computer Vision, First International Conference*, volume 4485 of *LNCS*, pages 473–483. Springer, 2007. [16](#)
- [167] S. Remmele, M. Seeland, and J. Hesser. Fluorescence microscopy deconvolution based on Bregman iteration and Richardson-Lucy algorithm with TV regularization. In T. Tolxdorff, J. Braun, T. M. Deserno, H. Handels, A. Horsch, and H.-P. Meinzer, editors, *Bildverarbeitung für die Medizin 2008*, Informatik aktuell, pages 72–76, Berlin, 2008. Springer. [63](#)
- [168] E. Resmerita. Regularization of ill-posed problems in Banach spaces: convergence rates. *Inverse Problems*, 21(4):1303–1314, 2005. [76](#)

- [169] E. Resmerita and R. S. Anderssen. Joint additive Kullback-Leibler residual minimization and regularization for linear inverse problems. *Mathematical Methods in the Applied Sciences*, 30(13):1527–1544, 2007. [67](#)
- [170] E. Resmerita, H. W. Engl, and A. N. Iusem. The expectation-maximization algorithm for ill-posed integral equations: a convergence analysis. *Inverse Problems*, 23(6):2575–2588, 2007. [67](#), [93](#)
- [171] W. H. Richardson. Bayesian-based iterative method of image restoration. *Journal of the Optical Society of America*, 62(1):55–59, 1972. [68](#)
- [172] C. P. Robert. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer Texts in Statistics. Springer, second edition, 2007. [117](#), [123](#), [125](#), [126](#)
- [173] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970. [45](#), [46](#)
- [174] R. T. Rockafellar. Integral functions, normal integrands and measurable selections. In J. P. Gossez, E. J. Lami Dozo, J. Mawhin, and L. Waelbroeck, editors, *Nonlinear Operators and the Calculus of Variations*, volume 543 of *Lecture Notes in Mathematics*, pages 157–207. Springer, 1976. [76](#), [77](#), [80](#), [129](#)
- [175] L. I. Rudin, P.-L. Lions, and S. Osher. Multiplicative denoising and deblurring: Theory and algorithms. In S. Osher and N. Paragios, editors, *Geometric Level Set Methods in Imaging, Vision, and Graphics*, pages 103–119. Springer, 2003. [73](#)
- [176] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992. [1](#), [2](#), [9](#), [13](#), [69](#)
- [177] Sandia National Laboratories. Sandia SAR data. Available online at <http://www.sandia.gov/radar/sar-data.html>. [136](#), [138](#)
- [178] A. Sawatzky. *(Nonlocal) Total Variation in Medical Imaging*. PhD thesis, Institute for Computational and Applied Mathematics, University of Münster, 2011. [67](#), [68](#), [78](#)
- [179] A. Sawatzky, C. Brune, J. Müller, and M. Burger. Total variation processing of images with Poisson statistics. In X. Jiang and N. Petkov, editors, *Computer Analysis of Images and Patterns, 13th International Conference, Proceedings*, volume 5702 of *LNCS*, pages 533–540. Springer, 2009. [63](#), [93](#), [94](#)
- [180] A. Sawatzky, C. Brune, F. Wübbeling, T. Kösters, K. Schäfers, and M. Burger. Accurate EM-TV algorithm in PET with low SNR. In *IEEE Nuclear Science Symposium Conference Record*, pages 5133–5137, 2008. [4](#), [63](#), [93](#), [94](#)
- [181] H. Schar. *Optimal Operators in Digital Image Processing*. PhD thesis, Interdisciplinary Center for Scientific Computing, Ruprecht-Karls-Universität Heidelberg, 2000. [42](#), [43](#)

- 
- [182] H. Scharr. Diffusion-like reconstruction schemes from linear data models. In K. Franke, K.-R. Müller, B. Nickolay, and R. Schäfer, editors, *Pattern Recognition, 28th DAGM Symposium*, volume 4174 of *LNCIS*, pages 51–60. Springer, 2006. 16, 54, 57, 59, 139
- [183] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging*, volume 167 of *Applied Mathematical Sciences*. Springer, 2009. 68, 77, 78
- [184] O. Scherzer and J. Weickert. Relations between regularization and diffusion filtering. *Journal of Mathematical Imaging and Vision*, 12(1):43–63, 2000. 1, 11
- [185] M. Seeger. Covariance kernels from Bayesian generative models. In T. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14, Proceedings of the 2001 Neural Information Processing Systems Conference*, volume 2, pages 905–912. MIT Press, 2002. 123
- [186] S. Setzer. Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage. In X.-C. Tai, K. Morken, M. Lysaker, and K.-A. Lie, editors, *Scale Space Methods and Variational Methods in Computer Vision, Second International Conference, Proceedings*, volume 5567 of *LNCIS*, pages 464–476. Springer, 2009. 86, 92
- [187] S. Setzer. *Splitting Methods in Image Processing*. PhD thesis, University of Mannheim, 2009. <http://ub-madoc.bib.uni-mannheim.de/2924>. 85, 87
- [188] S. Setzer. Operator splittings, Bregman methods and frame shrinkage in image processing. *International Journal of Computer Vision*, 92(3):265–280, 2011. 39, 86, 87
- [189] S. Setzer and G. Steidl. Variational methods with higher-order derivatives in image processing. In M. Neamtu and L. L. Schumaker, editors, *Approximation Theory XII: San Antonio 2007*, pages 360–385. Nashboro Press, 2008. 139
- [190] T. A. Severini. *Likelihood Methods in Statistics*, volume 22 of *Oxford statistical science series*. Oxford University Press, 2000. 126
- [191] L. A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging*, 1(2):113–122, 1982. 63, 67, 93
- [192] J. Shi and S. Osher. A nonlinear inverse scale space method for a convex multiplicative noise model. *SIAM Journal on Imaging Sciences*, 1(3):294–321, 2008. 4, 6, 73, 100, 101
- [193] D. L. Snyder, A. M. Hammoud, and R. L. White. Image recovery from data acquired with a charge-coupled-device camera. *Journal of the Optical Society of America A*, 10(5):1014–1023, 1993. 63, 64
- [194] D. L. Snyder, C. W. Helstrom, A. D. Lanterman, M. Faisal, and R. L. White. Compensation for read-out noise in HST image restoration. In R. J. Hanisch and R. L. White, editors, *The Restoration of HST Images and Spectra II, Proceedings of a workshop held at the Space Telescope Science Institute, Baltimore, Maryland*, pages 139–154, 1994. 63, 64

- [195] D. L. Snyder, C. W. Helstrom, A. D. Lanterman, M. Faisal, and R. L. White. Compensation for readout noise in CCD images. *Journal of the Optical Society of America A*, 12(2):272–283, 1995. [63](#), [64](#)
- [196] N. Sochen, R. Kimmel, and R. Malladi. From high energy physics to low level vision. In B. M. ter Haar Romeny, L. Florack, J. J. Koenderink, and M. A. Viergever, editors, *Scale-Space Theory in Computer Vision, First International Conference*, volume 1252 of *LNCS*, pages 236–247, 1997. [1](#)
- [197] N. Sochen, R. Kimmel, and R. Malladi. A general framework for low level vision. *IEEE Transactions on Image Processing*, 7(3):310–318, 1998. [1](#)
- [198] G. Steidl. A note on the dual treatment of higher-order regularization functionals. *Computing*, 76:135–148, 2006. [81](#)
- [199] G. Steidl and J. Weickert. Relations between soft wavelet shrinkage and total variation denoising. In L. Van Gool, editor, *Pattern Recognition, 24th DAGM Symposium*, volume 2449 of *LNCS*, pages 198–205. Springer, 2002. [13](#)
- [200] G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Welk. On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDes. *SIAM Journal on Numerical Analysis*, 42(2):686–713, 2004. [13](#)
- [201] T. Strömberg. *The operation of infimal convolution*. PhD thesis. Polska Akademia Nauk, Warszawa, Poland, 1996. [45](#)
- [202] Z. Tao, H. D. Tagare, and J. D. Beaty. Evaluation of four probability distribution models for speckle in clinical cardiac ultrasound images. *IEEE Transactions on Medical Imaging*, 25(11):1483–1491, 2006. [64](#)
- [203] T. Teuber. Anisotropic diffusion and regularization methods for restoring rotated shapes. Diploma thesis, University of Mannheim, 2008. [1](#), [11](#), [14](#), [15](#), [35](#)
- [204] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proc. Sixth Int. Conf. on Computer Vision*, pages 839–846. Narosa Publishing House, 1998. [15](#)
- [205] D. Tschumperlé. The CImg library. <http://cimg.sourceforge.net>. C++ Template Image Processing Library. [15](#)
- [206] D. Tschumperlé. Fast anisotropic smoothing of multivalued images using curvature preserving PDEs. *International Journal of Computer Vision*, 68(1):65–82, 2006. [1](#), [9](#), [15](#)
- [207] M. Tur, K. C. Chin, and J. W. Goodman. When is speckle noise multiplicative? *Applied Optics*, 21(7):1157–1159, 1982. [64](#)

- 
- [208] M. Unser, D. Sage, and D. Van De Ville. Multiresolution monogenic signal analysis using the Riesz-Laplace wavelet transform. *IEEE Transactions on Image Processing*, 18(11):2402–2418, 2009. [16](#), [139](#)
- [209] L. Vese and S. Osher. Numerical methods for  $p$ -harmonic flows and applications to image processing. *SIAM Journal on Numerical Analysis*, 40(6):2085–2104, 2002. [16](#)
- [210] C. R. Vogel. *Computational Methods for Inverse Problems*. SIAM, Philadelphia, 2002. [68](#)
- [211] R. F. Wagner, S. W. Smith, J. M. Sandrik, and H. Lopez. Statistics of speckle in ultrasound B-scans. *IEEE Transactions on Sonics and Ultrasonics*, 30(3):156–163, 1983. [64](#)
- [212] J. Weickert. Theoretical foundations of anisotropic diffusion in image processing. In W. Kropatsch, R. Klette, and F. Solina, editors, *Theoretical Foundations of Computer Vision*, Computing Supplement 11, pages 221–236. Springer, 1996. [1](#), [10](#)
- [213] J. Weickert. *Anisotropic Diffusion in Image Processing*. Teubner, Stuttgart, 1998. [1](#), [9](#), [10](#), [12](#), [41](#), [42](#)
- [214] J. Weickert. Coherence-enhancing diffusion filtering. *International Journal of Computer Vision*, 31(2/3):111–127, 1999. [1](#), [41](#)
- [215] J. Weickert and H. Scharr. A scheme for coherence-enhancing diffusion filtering with optimized rotation invariance. *Journal of Visual Communication and Image Representation*, 13(1/2):103–118, 2002. [42](#), [43](#)
- [216] J. Weickert and C. Schnörr. A theoretical framework for convex regularizers in PDE-based computation of image motion. *International Journal of Computer Vision*, 45(3):245–264, 2001. [1](#), [11](#)
- [217] J. Weickert, G. Steidl, P. Mrázek, M. Welk, and T. Brox. Diffusion filters and wavelets: What can they learn from each other? In N. Paragios, Y. Chen, and O. Faugeras, editors, *Handbook of Mathematical Models in Computer Vision*, pages 3–16. Springer, 2006. [13](#)
- [218] M. Welk, G. Steidl, and J. Weickert. Locally analytic schemes: A link between diffusion filtering and wavelet shrinkage. *International Journal of Computer Vision*, 24(2):195–224, 2008. [13](#)
- [219] M. Welk, J. Weickert, and G. Steidl. From tensor-driven diffusion to anisotropic wavelet shrinkage. In A. Leonardis, H. Bischof, and A. Pinz, editors, *Computer Vision – ECCV 2006*, volume 3951 of *LNCS*, pages 391–403, 2006. [13](#)
- [220] Y.-W. Wen and R. H. Chan. Parameter selection for total-variation-based image restoration using discrepancy principle. *IEEE Transactions on Image Processing*, 21(4):1770–1781, 2012. [140](#)

- [221] M. Werlberger, T. Pock, and H. Bischof. Motion estimation with non-local total variation regularization. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2464–2471, 2010. [140](#)
- [222] M. Werlberger, M. Unger, T. Pock, and H. Bischof. Efficient minimization of the non-local Potts model. In A. M. Bruckstein, B. M. ter Haar Romeny, A. M. Bronstein, and M. M. Bronstein, editors, *Scale Space and Variational Methods in Computer Vision, Third International Conference, SSVM 2011*, volume 6667 of *LNCS*, pages 314–325. Springer, 2012. [140](#)
- [223] N. Wiest-Daesslé, S. Prima, P. Coupé, S. P. Morrissey, and C. Barillot. Rician noise removal by non-local means filtering for low signal-to-noise ratio MRI: applications to DT-MRI. In D. Metaxas, L. Axel, G. Fichtinger, and G. Szekely, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2008, 11th International Conference, Proceedings, Part II*, volume 5242 of *LNCS*, pages 171–179. Springer, 2008. [106](#)
- [224] M. Yavuz. *Statistical Tomographic Image Reconstruction Methods for Randoms-Precorrected PET Measurements*. PhD thesis, University of Michigan, 2000. [63](#)
- [225] S. Yi, D. Labate, G. R. Easley, and H. Krim. A shearlet approach to edge analysis and detection. *IEEE Transactions on Image Processing*, 18(5):929–941, 2009. [16](#)
- [226] Y.-L. You, W. Xu, A. Tannenbaum, and M. Kaveh. Behavioral analysis of anisotropic diffusion in image processing. *IEEE Transactions on Image Processing*, 5(11):1539–1553, 1996. [14](#), [41](#)
- [227] Y. Yu and S. T. Acton. Speckle reducing anisotropic diffusion. *IEEE Transactions on Image Processing*, 11(11):1260–1270, 2002. [69](#)
- [228] J. Yuan, C. Schnörr, and G. Steidl. Convex Hodge decomposition and regularization of image flows. *Journal of Mathematical Imaging and Vision*, 33(2):169–177, 2009. [16](#)
- [229] X. Zhang, M. Burger, X. Bresson, and S. Osher. Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. *SIAM Journal on Imaging Sciences*, 3(3):253–276, 2010. [68](#)
- [230] D. Zhou and B. Schölkopf. Regularization on discrete spaces. In W. Kropatsch, R. Sablatnig, and A. Hanbury, editors, *Pattern Recognition, Proceedings of the 27th DAGM Symposium*, volume 3663 of *LNCS*, pages 361–368. Springer, 2005. [68](#)
- [231] M. Zhu and T. Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. CAM-Report 08-34, UCLA, Los Angeles, 2008. [39](#), [93](#)
- [232] W. P. Ziemer. *Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation*, volume 120 of *Graduate Texts in Mathematics*. Springer, 1989. [68](#)

# Index

- additive noise, [109](#)
  - Gaussian distributed, [8](#), [109](#)
- ADMM, *see* alternating direction method of multipliers
- alternating direction method of multipliers, [50](#), [86–89](#)
- alternating split Bregman algorithm, [50](#), [86](#), [93](#)
- anisotropic wavelet shrinkage, [13](#)
- Bayesian likelihood ratio, [123](#)
- characteristic function, [32](#)
- conditional density function, [62](#), [144](#)
- conditional expectation, [129](#)
- conditional mass function, [62](#)
- conjugate function, [31](#)
- coupled shrinkage, [92](#)
- data fidelity term, [2](#)
- diffusion model, [9](#), [40](#)
  - anisotropic, [10](#), [12–13](#)
  - isotropic, [10](#), [12–13](#)
- diffusion tensor, [9](#), [40–41](#)
- diffusivity function, [10](#), [15](#), [40](#)
- domain of a function, [44](#)
- double orientation tensors, [29](#)
- Douglas-Rachford splitting algorithm, [86](#)
- dual problem, [82](#)
- duality gap, [82](#)
- edge-enhancing diffusion, [1](#), [10–11](#), [15](#)
- EM-TV algorithm, [94](#)
- equi-coercive sequence, [79](#)
- equi-mildly coercive sequence, [81](#)
- essential infimum, [69](#)
- essential supremum, [69](#)
- fast iterative shrinkage thresholding algorithm, [39](#), [93](#)
- FISTA, *see* fast iterative shrinkage thresholding algorithm
- forward differences, [33](#)
- forward-backward splitting, [39](#)
- generalized likelihood ratio, [124](#)
- I-divergence, [67](#)
- IDIV-NL model, [82](#)
- IDIV-TV model, [68](#), [82](#)
- indicator function, [31](#)
- infimal convolution, [44–45](#)
- integrand, [76](#)
  - normal, [76](#)
- interior of a set, [87](#)
- Jacobi's transformation formula, [141](#)
- Jeffreys' prior, [117](#), [123](#)
- Kullback-Leibler divergence
  - generalized, *see* I-divergence
  - symmetric, [131](#)
- maximum a posterior estimation, [66](#)
- maximum a posteriori estimation, [66](#)
- mean absolute error (MAE), [133](#)
- method of multipliers, [86](#)
- multilook data, [65](#)
- multiplicative noise, [64](#), [110](#)
  - exponentially distributed, [65](#)
  - Gamma distributed, [65](#), [113](#)
  - Nakagami distributed, [65](#)
  - Rayleigh distributed, [64](#), [115](#)



- uniformly distributed, [112](#)
- mutual information kernel, [123](#)
- Nesterov's algorithm, [39](#), [93](#)
- non-informative prior, [117](#)
- nonlocal means filter, [1](#), [15](#), [52](#), [54](#), [106](#)
- occlusion model, [24](#)
- patch, [106](#)
- PDHG, *see* primal-dual hybrid gradient algorithm
- peak signal to noise ratio (PSNR), [133](#)
- PIDAL algorithm, [92](#)
- PIDSplit+ algorithm, [91](#)
- Poisson noise, [62](#)
- primal-dual hybrid gradient algorithm, [39](#), [93](#)
- probability density function, [62](#), [141–143](#)
- probability mass function, [62](#)
- probability space, [62](#), [107](#)
- recession function, [44](#)
- regularization parameter, [8](#)
- regularizer, [2](#), [8](#)
  - anisotropic, [8–9](#), [12–13](#)
  - isotropic, [8–9](#), [12–13](#)
  - nonlocal, [68](#)
- Rudin-Osher-Fatemi model, [9](#), [13](#), [52](#), [54](#)
- search window, [131](#)
- second-order cone programming, [38–39](#), [50](#)
- space of functions of bounded variation, [68](#)
- structure tensor, [15–21](#)
- support function, [46](#)
- total variation, [68](#)
  - anisotropic, [9](#)
- transparent model, [24](#)
- Wulff shape, [9](#)



# Scientific Career

4/1984	<i>Born in Speyer, Germany</i>
3/2003	<i>University entrance qualification (Abitur)</i> Gymnasium, Paul-von-Denis-Schulzentrum Schifferstadt
10/2003 - 1/2008	<i>Undergraduate studies in Mathematics and Computer Science</i> University of Mannheim
1/2007 - 6/2007	<i>Semester abroad</i> Swansea University, United Kingdom
1/2008	<i>Diploma in Mathematics</i> Diploma thesis: Anisotropic Diffusion and Regularization Methods for Restoring Rotated Shapes
2/2008 - 1/2011	<i>Research assistant</i> Department of Mathematics and Computer Science, University of Mannheim
2/2011 - today	<i>Research assistant</i> Department of Mathematics, University of Kaiserslautern



# Wissenschaftlicher Werdegang

4/1984	<i>Geboren in Speyer</i>
3/2003	<i>Abitur</i> Gymnasium, Paul-von-Denis-Schulzentrum Schifferstadt
10/2003 - 1/2008	<i>Studium im integrierten Studiengang Mathematik und Informatik</i> Universität Mannheim
1/2007 - 6/2007	<i>Auslandssemester</i> Swansea University, Großbritannien
1/2008	<i>Diplom in Mathematik</i> Titel der Diplomarbeit: Anisotropic Diffusion and Regularization Methods for Restoring Rotated Shapes
2/2008 - 1/2011	<i>Wissenschaftliche Mitarbeiterin</i> Fakultät für Mathematik und Informatik, Universität Mannheim
2/2011 - heute	<i>Wissenschaftliche Mitarbeiterin</i> Fachbereich Mathematik, TU Kaiserslautern