Westfälische
Wilhelms-Universität
Münster

(Nonlocal) Total Variation in Medical Imaging

Alex Sawatzky

- 2011 -

Westfälische Wilhelms-Universität Münster

(Nonlocal) Total Variation in Medical Imaging

Alex Sawatzky

- 2011 -

WESTFÄLISCHE
WILHELMS-UNIVERSITÄT
MÜNSTER

(Nonlocal) Total Variation in Medical Imaging

Fach: Mathematik

Inaugural-Dissertation zur Erlangung des
Doktorgrades der Naturwissenschaften
- Dr. rer. nat. -
im Fachbereich Mathematik und Informatik
der Mathematisch-Naturwissenschaftlichen Fakultät
der Westfälischen Wilhelms-Universität Münster

vorgelegt von
Alex Sawatzky
- 2011 -

wissen.leben
WWU Münster

Dekan:	Prof. Dr. Matthias Löwe
Erster Gutachter:	Prof. Dr. Martin Burger (Westfälische Wilhelms-Universität Münster)
Zweite Gutachterin:	Prof. Dr. Gabriele Steidl (Technische Universität Kaiserslautern)
Tag der mündlichen Prüfung:	
Tag der Promotion:	

Abstract

This thesis deals with the reconstruction of images where the measured data are corrupted by Poisson noise and a specific signal-dependent speckle noise, which occurs e.g. in medical ultrasound images. Since both noise types fundamentally differ from the frequently studied additive Gaussian noise in mathematical image processing, adapted variational models are required to handle these types of noise accurately.

The first part of this thesis introduces variational regularization frameworks for inverse problems with data corrupted by Poisson and ultrasound speckle noise. Due to the strong nonlinearity of both data fidelity terms, a forward-backward splitting approach is used to provide efficient numerical schemes allowing the use of arbitrary convex regularization energies, in particular singular ones. Moreover, analytical results such as the well-posedness of the variational problems as well as the positivity preservation and convergence of the proposed iteration methods are proved. Finally, an iterative extension of both frameworks is proposed in order to refine the systematic errors of variational regularization techniques, using inverse scale space methods and Bregman distance iterations.

The second part of this thesis considers the use of the (nonlocal) total variation functional as regularization energy in both previously developed frameworks. In particular, a modified version of the projected gradient descent algorithm of Chambolle and an augmented Lagrangian method are presented to solve the weighted (nonlocal) ROF model arising in both previously developed frameworks. In the case of the total variation regularization strategy, analytical results obtained previous in the general context of a convex regularization functional are carried over to the TV seminorm. In the case of the nonlocal regularization approach, a continuous framework of nonlocal derivative operators on directed graphs is introduced. This framework generalizes the nonlocal operators on undirected graphs in continuous and discrete setting and is consistent to the discrete local derivative operators.

Finally, the performance of the proposed algorithms is illustrated by 2D and 3D synthetic and real data reconstructions. To validate the method proposed for inverse problems with data corupted by Poisson noise, simulated PET data (2D) and real cardiac $H_2{}^{15}O$ (2D) and ^{18}F-FDG (3D) PET measurements with low count rates are used. Additionally, a denoising and reconstruction comparison between TV and nonlocal TV regularization is presented using 2D synthetic Poisson data. In the case of denoising problems in medical US imaging, results on real patient data (2D) are illustrated.

Key words: Inverse Problems, Image Processing, Poisson Noise, Ultrasound Speckle Noise, Variational Regularization Theory, Singular Regularization Energies, (Nonlocal) Total Variation Regularization, Forward-Backward Splitting, Expectation-Maximization Algorithm, Richardson-Lucy Algorithm, Bregman Distances, Inverse Scale Space Methods, Positron Emission Tomography, Medical Ultrasound Imaging

Acknowledgments

The process of writing a PhD thesis is a time and energy consuming task, which requires a lot of endurance, nerves made of steel and a lot of supporters and good friends who carry you all the way to your goal. I want to take the chance and thank the latter ones.

Martin Burger for giving me the opportunity to work on an interesting and challenging topic. For giving me a lot of support, for supervising this thesis, and for being the best mentor and 'Chefe' ever.

Gabriele Steidl for co-reviewing this thesis and for giving me the opportunity to present my work on the Imaging Science Conference in Chicago.

My colleagues in the 'Imaging' workgroup: Marzena Franek, Frank Wübbeling, Bärbel Schlake, Christoph Brune, Martin Benning, Jahn Müller, Michael Möller, Ralf Engbers, Marcisse Fouego, Thomas Kösters, Felix Lucka, Louise Reips, Lars Ruthotto; and in the Institute for Computational and Applied Mathematics for the wonderful years, a lot of fun and nonsense, and new best friendships.

My proof-readers Alexander Wenner, Daniel Tenbrinck, Jahn Müller and Ralf Engbers for all the many corrections and for preserving this thesis from uncountable more mistakes than the ones that are still contained.

Klaus Schäfers and Thomas Kösters for providing PET data and various variants of the EM algorithm.

Jörg Stypmann for providing medical ultrasound data and for interesting and helpful discussions.

The Federal Ministry of Education and Research BMBF for supporting me through the project *INVERS*.

Olga Schiklo, for being the sweetest anticipation when you are coming home. For always being insightful and patient with me for the last couple of months. And for giving me a lot of support and encouragement in the time of desperation when the things did not work as I wanted it.

My family, Rita, Walentina, Hedwig and Alfred Sawatzky, Jan, Sofia, Anastasia, Rita and Dieter Kotkas, for giving me an incredible support and for having solid faith in me and the things I am doing. Thanks a lot!

Contents

List of Algorithms

1 Introduction

This thesis deals with *variational regularization techniques* for *inverse problems* with *non-standard noise models*, with a particular focus on *image reconstruction problems*. The field of inverse problems is a wide and important area in applied mathematics and its related scientific disciplines, which arises in a wide variety of application areas, such as in medical imaging, biophysics, remote sensing, ocean acoustic tomography, geophysics, non-destructive testing, astronomy, and many others. The main part of this work concentrates on inverse problems arising in the field of *medical imaging*.

In the following we give an overview of the contents of this work. We start with the basic motivations of this thesis and give an outline of its contributions. Finally, we provide a sketch of how this thesis is organized.

1.1 Motivation

The main task of inverse problems consists in the reconstruction of desired parameters in mathematical models (in this thesis *images*) from indirectly observed data. Mathematically, this reconstruction process can be often modeled as computing a function u from an operator equation of the form

$$f \;=\; Ku \tag{1.1}$$

with measured data f. In this thesis, the operator K represents a model for the imaging device and is assumed to be linear. The difficulty of such a modeling is that the computation of u by direct inversion of K is not reasonable, since in practice the data f are corrupted by noise induced in the process of physical measurements and the problem (1.1) is usually ill-posed in the sense of Hadamard, i.e. it is in particular highly sensitive with respect to errors in f. Hence, some type of regularization is required to enforce stability during the inversion process and to compute useful reconstructions.

A frequently used class of regularization techniques are variational methods based on the minimization of functionals of the form

$$\frac{1}{p} \|Ku - f\|^p_{L^p(\Sigma)} + \alpha J(u) , \qquad \alpha > 0 , \qquad p \in [1, \infty) , \tag{1.2}$$

where the first term penalizes the deviation from the operator equation (1.1) and the second term introduces a-priori information about the expected solution. However, from the viewpoint of statistical modeling, the functionals in (1.2) result from the assumption that the raw data f are perturbed in the form $f = K\bar{u} + \eta$, where $\bar{u}$ denotes the desired exact image and η is an exponentially distributed random variable. Typical examples are that η is Laplace distributed or Gaussian and results in $p = 1$ and $p = 2$ in (1.2), respectively.

Most works in the area deal with the case of additive Gaussian noise so far. However, in real life there are several applications in which different types of noise occur. For instance, so-called *Poisson or photon counting noise* appears in positron emission tomography in medical imaging [148, 160, 167], fluorescence microscopy [87, 57], CCD cameras [152], or astronomical images [108, 102]. Other non-Gaussian noise models are *salt and pepper noise* or the different types of *multiplicative noise*, for example appearing in synthetic aperture radar (SAR) or *speckle noise* in medical ultrasound imaging [158, 92]. These types of noise fundamentally differ from the common studied exponentially distributed noise in (1.2) and, consequently, adapted variational regularization models are required in order to handle these kinds of noise accurately.

A commonly used idea to realize variational regularization techniques with statistical motivation is the *Bayesian model.* Using a *Gibbs a-priori density*, the *maximum a-posteriori probability (MAP) estimation* via the negative log-likelihood function leads to a more general form of optimization problems in image processing and inverse problems, namely

$$\min_u \; H_f(Ku - f) + \alpha J(u) , \tag{1.3}$$

where H_f and J are usually convex. The functional H_f denotes a general *data fidelity term* and has the task to penalize the deviation from the operator equality (1.1). The functional J is a *regularization functional* and penalizes the deviation from a certain ideal structure (smoothness) of the solution u. The *regularization parameter* α is a relative weight for both terms and controls the influence of the data fidelity and the regularization term on the solution.

The form of the data fidelity term H_f in (1.3) is solely dependent on the kind of noise, which occurs in the considered inverse problem, and should be adapted to the underlying statement of the problem. Another important aspect in the variational regularization technique is the specific choice of the regularization functional J in (1.3), since this choice decides on the way a-priori information about the expected solution is incorporated into the reconstruction process. In the past, smooth, in particular quadratic, regularizations have attracted most attention. However, such regularization approaches lead to blurring of reconstructions, in particular they cannot yield image reconstructions with sharp edges. Recently, singular regularization energies, in particular those of ℓ^1 - or L^1 -type, have attracted strong attention in inverse problems. However, such functionals are not differentiable in the common sense and other concepts are required to deal with these problems.

The wide variety of different types of noise in inverse problems combined with the stable inversion of such problems using variational techniques with singular regularization energies are the main motivation of this thesis. Based on the statements above we summarize the contributions of this work in the following section.

1.2 Contributions

In this thesis we address the task to reconstruct images in inverse problems in which the measured data are corrupted by *Poisson noise*, which occurs in various real life applications as e.g. in positron emission tomography in medical imaging or fluorescence microscopy. Another noise model considered in this thesis is a specific *speckle noise* which occurs in *medical ultrasound (US) imaging.* In both cases, the resulting variational regularization problems differ siginificantly from the usually studied models in (1.2) and have the following form using the MAP estimation model in (1.3),

$$\begin{aligned} &\min_{u \geq 0} \int_\Sigma (Ku - f \log Ku) \; d\mu + \alpha J(u) \,, && \textit{(Poisson)} \\ &\min_{u \geq 0} \int_\Sigma \frac{(f - Ku)^2}{Ku} \; d\mu + \alpha J(u) \,. && \textit{(US Speckle)} \end{aligned} \tag{1.4}$$

A particular complication of these minimization problems compared to (1.2) is the *strong nonlinearity in the data fidelity terms* which leads to issues in the computation of minimizers. A further challenge is the *positivity constraint* on the solution u in (1.4) which is absolutely necessary, since in typical applications these functions represent densities or intensity information. Hence, one of the main contributions of this thesis is the *development of efficient and stable numerical schemes* to compute the solution of minimization problems in (1.4), which may in addition guarantee the *positivity* of a solution and can also handle *singular regularization energies*. A future contribution of this work is to provide a *theoretical framework* for the variational problems (1.4) and the resulting numerical schemes using quite general convex regularization functional.

Besides the development of numerical schemes for variational problems in (1.4), we discuss the use of *total variation (TV) regularization* and its *nonlocal (NL) extension* in the Poisson and US speckle noise frameworks. In the case of the NL-TV functional, the high complexity of the nonlocal weighted graph requires suitable approaches to reduce this complexity. However, some strategies lead to a directed structure of the weighted graph such that the essential symmetry assumption of the nonlocal derivative operators proposed so far is violated. Hence, another contribution of this thesis is the *generalization of nonlocal derivative operators* on directed graphs in a continuous setting.

Finally, as a future contribution in this thesis, computational realization of the proposed algorithms is illustrated on 2D and 3D synthetic and real data reconstructions in positron emission tomography and medical ultrasound imaging.

1.3 Organization of this Work

In *Chapter 2* we provide a general introduction of variational regularization methods. In particular, we consider the mathematical modeling of inverse problems and image reconstruction problems. Moreover, we recall a commonly used idea to realize variational regularization techniques with statistical motivation via the Bayesian model.

In *Chapter 3* we provide an overview on the basic concepts of functional analysis, convex analysis, and calculus of variations, which will be needed in the course of this work.

In *Chapters 4* and *5* the main contribution of this thesis follows, namely the development of efficient and stable numerical schemes for inverse problems with data corrupted by Poisson and US speckle noise. Moreover, a theoretical framework for the corresponding variational regularization problems and computational methods is proposed. The results in *Chapter 4* are based on joint work with C. Brune and M. Burger [140, 141, 30, 31, 32].

In *Chapters 6* and *7* the use of the total variation functional and its nonlocal extension as regularization energies in both variational regularization frameworks of *Chapters 4* and *5* is discussed. Additionally, a further contribution of this work can be found in *Chapter 7*, namely the generalization of nonlocal derivative operators to directed graphs.

Finally, the performance of the numerical methods proposed in *Chapters 4 - 7* is presented in *Chapter 8* using 2D and 3D reconstructions on synthetic and real data in positron emission tomography and medical ultrasound imaging.

2 Basics : Variational Methods

2.1 Inverse Problems

The field of inverse problems is a wide and important area in applied mathematics and other sciences that has been rapidly growing over the last decades. The reason for the increasing interest is a wide variety of applications in sciences and engineering, such as in medical imaging, biophysics, geophysics, remote sensing, ocean acoustic tomography, nondestructive testing, astronomy, and many other areas.

The main task of inverse problems consists in the reconstruction of desired parameters in mathematical models, such as signals or images, from indirectly observed data. In this work, we will focus our attention in particular on *image reconstruction problems*. Hence, we first introduce a mathematical representation of an image in a discrete and continuous setting.

Definition 2.1.1 (*Continuous Image Representation [117, Def. 3.1]*). *Let $\Omega \subset \mathbb{R}^d$, $d \in \mathbb{N}$, be the image spatial domain. A function $u : \Omega \to \mathbb{R}$ is called a d-dimensional image if the following conditions are fulfilled,*

(1) *u has a compact support, if Ω is not bounded,*

(2) $0 \leq u(x) < \infty \quad$ *for all* $\quad x \in \Omega$, *(intensity boundedness)*

(3) $\displaystyle\int_\Omega u(x)\, dx < \infty$. *(energy boundedness)*

As we will see later, such a representation of images is an elegant way to deliver a simple basis for the analysis and construction of mathematical methods. However, this kind of image description is actually only an idealization, which cannot be realized on any computer and does not correspond to the reality of applications. Therefore, we are also interested in digital images, which arise as a naturally result from the discrete output of an imaging device.

Definition 2.1.2 (*Discrete Image Representation*). *Let* $\Omega := (0,1)^d \subset \mathbb{R}^d$, $d \in \mathbb{N}$, *be the image spatial domain with a regular grid of* $N_1 \times \cdots \times N_d$ *points. A grid matrix* $u \in \mathbb{R}^{N_1 \times \cdots \times N_d}$ *is called a d-dimensional image if the following conditions are fulfilled,*

(1) $0 \leq u_{i_1,\ldots,i_d} < \infty \quad \text{for} \quad 1 \leq i_k \leq N_k\,, \quad k = 1,\ldots,d\,,$

(2) $\displaystyle\sum_{i_1=1}^{N_1} \cdots \sum_{i_d=1}^{N_d} u_{i_1,\ldots,i_d} < \infty\,.$

Additionaly, we denote with $h_k = \frac{1}{N_k}$, $k = 1,\ldots,d$, *the stepsizes of the image grid in the* k*-th direction.*

Remark.

- If u is not only an one-channeled gray value image but also a color image, the *Definitions 2.1.1* and *2.1.2* extend for each channel in a straight-forward way.
- In this thesis, we work mainly with 2D ($d = 2$) or 3D ($d = 3$) images.
- The connection between the definition of the continuous and the discrete image representation is the usual interpretation of a grid matrix as locally constant gray values in a cell around respective grid point. In this manner, we construct to each discrete also a continuous description of an image using a piecewise constant approximation.

Next, in order to model a mathematical representation of image reconstruction problems, we have to consider a suitable mathematical description of a physical imaging system. In the following, we discuss only the relevant aspects of such a model construction and refer to the work of Bertero et al. [23] for a detailed discussion. An imaging system consists in general of two structural elements:

- A collection of different physical components generating signals that contain useful information about spatial properties of an object.
- A detector system that provides measurements of occurring signals, which also causes the undesirable sampling and noise effects in many cases.

Hence, we assume in the following that the raw data have the following properties:

- The data are discrete and the discretization is specified by the physical configuration of the detectors. In addition, we assume that the data are given in form of a vector $f \in \mathbb{R}^N$.

- The data are realizations of random variables, since the noise is a random process caused by the detector system or the forward process of an imaging device, as for instance the random decay of weak radioactively marked pharmaceuticals in positron emission tomography (PET). Hence, we consider the detected value f_i as a realization of a random variable F_i .

After the consideration of the raw data, it is necessary to model the imaging device which describes the generation and expansion of signals during the data acquisition process. Mathematically, the aim is to find a transformation that maps the spatial distribution of an object to the signals arriving at the detectors. In this work, we focus on problems where such a transformation is a linear operator and the data acquisition process can be described by a linear operator equation of the form

$$\bar{f} \;=\; \bar{K}\bar{u} \; . \tag{2.1}$$

Here, $\bar{K} : U(\Omega) \to V(\Sigma)$ denotes a linear and compact operator (thus with a nonclosed range), where $U(\Omega)$ and $V(\Sigma)$ are Banach spaces of functions on bounded and compact sets Ω respectively Σ . A typical example of (2.1) is a Fredholm integral equation of the first kind with

$$(\bar{K}u)(x) \;=\; \int_\Omega \bar{k}(x,y)\, u(y)\; dy \; , \qquad x \;\in\; \Sigma \;=\; \Omega \; ,$$

where $\bar{k}$ is the kernel of the operator $\bar{K}$. In (2.1), the function $\bar{u}$ describes the desired exact properties of an object and $\bar{f}$ denotes the exact signals before detection. Problems of the type above can be found in numerous real life applications, such as positron emission tomography in medical imaging [123, 148, 167], fluorescence microscopy [87, 57], astronomy [23], geophysics [103] or radar imaging [88, 123].

Consequently, the modeling of data acquisition in the manner described above, leads to problem of object reconstruction as the solution of a linear operator equation of the form (2.1). However, as mentioned above, in practice only noisy (and discrete) versions f and K of the exact data $\bar{f}$ and operator $\bar{K}$ are available so that only an approximate solution u of $\bar{u}$ can be computed from the equation

$$f \;=\; Ku \; . \tag{2.2}$$

The operator $K : U(\Omega) \to \mathbb{R}^N$ in (2.2) is a semi-discrete operator based on $\bar{K}$, which transforms the desired properties u , in contrary to $\bar{K}$, to the discrete raw data. Note that in the special case of the identity operator K , i.e. $K = I$, interpreted furthermore as a discretization (projection) operator, we call the image reconstruction problem (2.2) as an image denoising problem, where the observed image $f = u$ is a certain noisy version of the exact image property $\bar{u}$. Finally, for the completeness of the modeling of the image reconstruction problem, we additionally need due to the noise in the measurements f a model for the probability density of the noise. This aspect of the modeling process will be discussed in detail in the following *Section 2.2*.

Concluding, we see now in (2.1) and (2.2) that the inverse problems are often modeled as certain operator equations. However, the difficulty of such a modeling is that the computation of u by a direct inversion of K in (2.2) is not reasonable, since (2.1) is an ill-posed problem (due to the compactness of the forward operator $\bar{K}$) [67, 84]. But note that the problem (2.2) is not ill-posed in usual sense, because the operator K has a finite range. Nevertheless, the problem is highly ill-conditioned, since K approximates $\bar{K}$. For the sake of completeness, we recall that a problem

$$g = \bar{K}u \tag{2.3}$$

is called *ill-posed* (sometimes also called *incorrect*), if one of the following conditions is not fulfilled (cf. e.g. [83] or [103]),

- for each $g \in V(\Sigma)$ there is a solution $u \in U(\Omega)$ (*existence condition*),
- the solution u is unique in $U(\Omega)$ (*uniqueness condition*),
- the dependence of u upon g is continuous (*stability condition*).

In general, the stability condition above is responsible for the ill-posedness of the operator equation (2.3) and hence some type of regularization is required to enforce stability during the inversion process and to compute useful reconstructions, see *Section 2.2*. Some further examples of ill-posed operator equations can be found e.g. in [67, Chapt. 1] or [103, Chapt. 1].

2.2 Bayesian Modeling

As already suggested in *Section 2.1*, a complete mathematical modeling of an image reconstruction problem requires a model for the probability density of the noise that occurs in the measurements f. We denote this density with $p_F(f|u)$ and mean with this a conditional probability density of data f given an image u with respect to random variables F_i, $1 \leq i \leq N$. For the following observations, we make additionally the assumption that the random variables F_i are pairwise independent and identically distributed, i.e. we can write

$$p_F(f|u) \;=\; \prod_{i=1}^{N} p_F(f_i|u) \;.$$

This assumption is in general reasonable since each random variable can be assigned to a specific detector element which is usually independent of all the other detector elements.

Typical examples for probability densities $p_F(f|u)$ are exponentially distributed raw data f. A canonical choice in most works are data of the form $f = K\bar{u} + \eta$, i.e. perturbed by additive noise, where η denotes a vector valued Gaussian distributed random variable with expected value 0 and variance σ^2. In this case, (2.4) shows the corresponding probability density $p_F(f|u)$ [23]. In (2.4), we see also a relatively similar although more complicated structure of the data, where the measurements are corrupted with a signal-dependent noise of the form $f = K\bar{u} + \sqrt{K\bar{u}}\,\eta$ with a Gaussian distributed random variable η as just now. Such a type of noise can be found for instance in medical ultrasound (US) imaging [101] and corresponds to an experimental derived model of multiplicative speckle noise [158] in ultrasound images. Other typical models are also Poisson distributed data [23] or Gamma distributed data [5] that corresponds to multiplicative noise in f, i.e.

$$\begin{aligned}
p_F(f|u) \;&\sim\; e^{-\frac{1}{2\sigma^2}\|Ku - f\|^2_{L^2_\mu(\Sigma)}} \;, && \textit{(Gaussian)} \\
p_F(f|u) \;&\sim\; e^{-\frac{1}{2\sigma^2}\left\|\frac{Ku - f}{\sqrt{Ku}}\right\|^2_{L^2_\mu(\Sigma)}} \;, && \textit{(US Speckle)} \\
p_F(f|u) \;&=\; \prod_{i=1}^{N} \frac{(Ku)_i^{f_i}}{f_i\,!}\, e^{-(Ku)_i} \;, && \textit{(Poisson)} \\
p_F(f|u) \;&=\; \prod_{i=1}^{N} \frac{n^n}{(Ku)_i^n\,\Gamma(n)}\, f_i^{n-1}\, e^{-n\,\frac{f_i}{(Ku)_i}} \;. && \textit{(Gamma)}
\end{aligned} \tag{2.4}$$

Here, Γ denotes the Gamma function and we suppose for the multiplicative Gamma distributed noise that each f_i is the mean over n measurements. Additionally, in order to avoid the differentiation between the (semi-) discrete and continuous form of raw data and operators discussed above, we use in (2.4) and later in (2.5) a measure μ, which is a Lebesque measure in continuous setting and a point measure in discrete setting.

To illustrate the different characteristics of the noise forms in (2.4), we show in *Fig. 2.1* and *Fig. 2.2* a simple one dimensional (1D) noise free signal and a two dimensional (2D) noise free image corrupted with the corresponding noise types presented in (2.4). In this results, we can in particular observe that all noise forms are usually stronger than the classical additive Gaussian noise. Interesting is also the fact, that the Poisson noise in *Fig. 2.1d* is practically almost ever stronger than the additive Gaussian noise in *Fig. 2.1b*, excepting the results on the right-hand side of both figures. However, the issue of the similarity of both results on the right-hand side of *Fig. 2.1b* and *Fig. 2.1d* is not really surprising, since the Poisson distribution is asymptotically normal with mean λ and standard deviation $\sqrt{\lambda}$ (cf. e.g. [53, Sections 16.5 and 20.2]), where λ denotes the positive parameter of a Poisson distribution. Finally, notice also that due to the signal-dependent Gaussian degradation, the ultrasound speckle noisy objects in *Fig. 2.1c* can also be negative although the original object is always positive.

Now, as we already suggested in *Section 2.1*, a direct inversion of K in (2.2) is not reasonable for computing u, since the operator K is ill-conditioned. Therefore, if the probability density $p_F(f|u)$ of the noise is known, the desired object u appears in this formulation via the noise density function as a set of unknown parameters and the image reconstruction problem corresponds to the classical problem of parameter estimation. In such problems, a classical approach is the so-called *maximum likelihood (ML) estimator* that computes a solution by maximizing the likelihood respectively by minimizing the negative log-likelihood function, i.e.

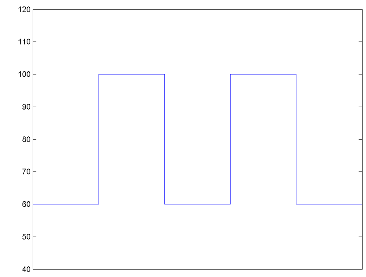
$$u_{ML} \in \arg\min_u \{ -\log p_F(f|u) \} .$$

In case of the probability densities given in (2.4), the maximum likelihood estimation leads to minimizing of functionals of the form [23, 5],

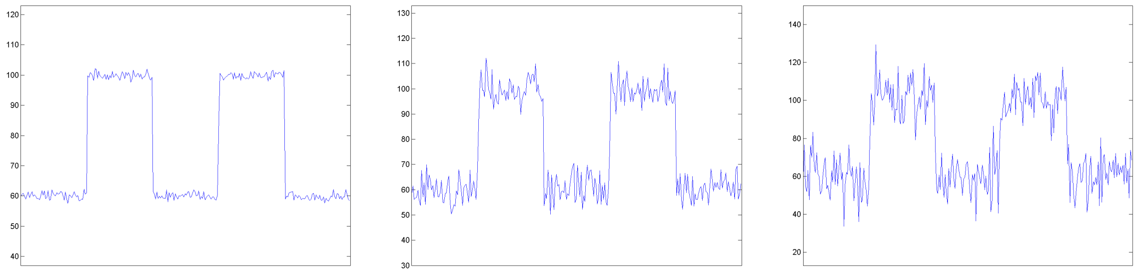
$$\begin{aligned}
u_{ML} &\in \arg\min_u \left\{ \frac{1}{2} \|Ku - f\|^2_{L^2_\mu(\Sigma)} \right\} , && \textit{(Gaussian)} \\
u_{ML} &\in \arg\min_u \left\{ \frac{1}{2} \left\| \frac{Ku - f}{\sqrt{Ku}} \right\|^2_{L^2_\mu(\Sigma)} \right\} , && \textit{(US Speckle)} \\
u_{ML} &\in \arg\min_u \left\{ \int_\Sigma (Ku - f \log Ku) \, d\mu \right\} , && \textit{(Poisson)} \\
u_{ML} &\in \arg\min_u \left\{ \int_\Sigma \left(\log Ku + \frac{f}{Ku} \right) d\mu \right\} , && \textit{(Gamma)}
\end{aligned} \tag{2.5}$$

in which we neglect additive terms independent of u. In case of Poisson distributed data, we discuss the corresponding functional in more detail in *Section 4.2*.

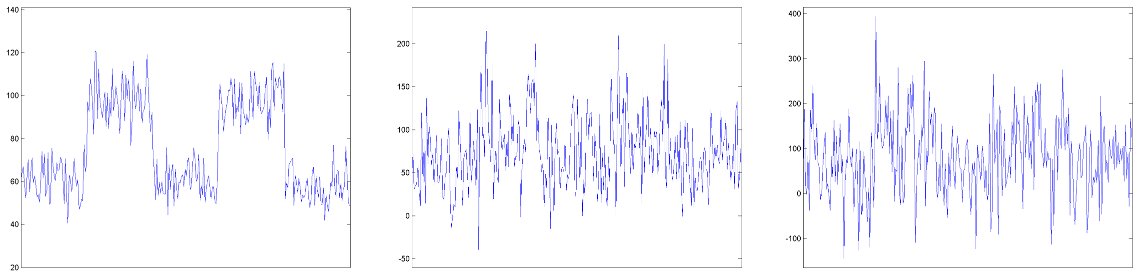
However, the image recontruction problems formulated as maximum likelihood (ML) estimation remain ill-posed respectively ill-conditioned, because the ML approach uses only information about the noise. It is generally well known that if the modeling of



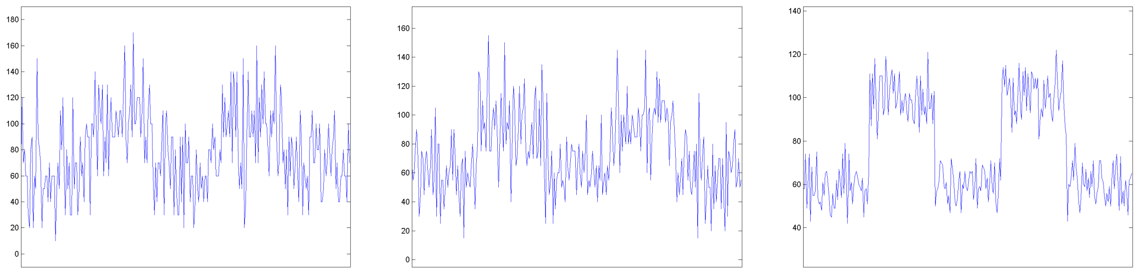
(a) *Noise free signal*



(b) *Additive Gaussian noise with* $\sigma = 1$ *(left),* $\sigma = 5$ *(middle) and* $\sigma = 10$ *(right)*

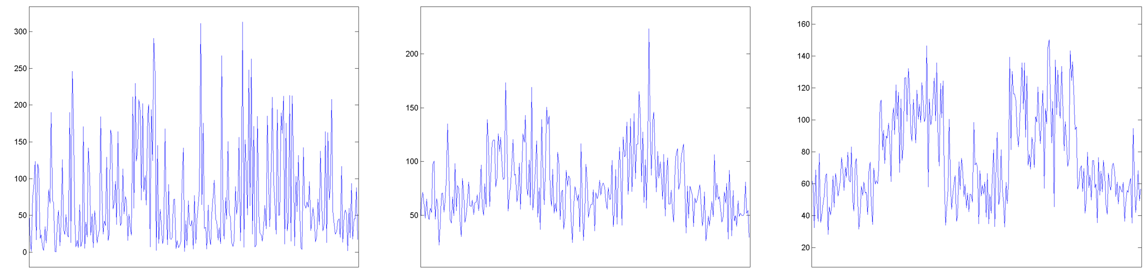


(c) *Ultrasound speckle noise with* $\sigma = 1$ *(left),* $\sigma = 5$ *(middle) and* $\sigma = 10$ *(right)*



(d) *Poisson noise with scaling factor* 10 *(left),* 5 *(middle) and* 1 *(right)*

Fig. 2.1. *Illustration of in* (2.4) *presented noise forms in one dimension (notice that the vertical scales are different in the signals). (a) Noise free 1D signal* $\bar{u}$ *. (b)* $\bar{u}$ *degraded by additive Gaussian noise with different standard deviations* σ *. (c)* $\bar{u}$ *degraded by ultrasound specific speckle noise with different standard deviations* σ *. (d)* $\bar{u}$ *degraded by Poisson noise in the form that* $\bar{u}$ *is first scaled down by a factor, subsequently degraded by Poisson noise and finally scaled back with the same factor.*



(e) *Multiplicative Gamma noise with mean* 1 *(left), mean* 10 *(middle) and mean* 15 *(right)*

Fig. 2.1. *Continued illustration of in* (2.4) *presented noise forms in one dimension (notice that the vertical scales are different in the signals). (e)* $\bar{u}$ *degraded by multiplicative Gamma noise with different assumptions of measurement means.*

the problem does not use some additional information about the desired object, the ill-posedness spreads from the operator equation (2.2) to the new model, what happens also in the case of ML approach. Therefore, a commonly used idea to realize stable inversion methods with statistical motivation is the *Bayesian model*, where the additional information are given in the form of statistical properties of desired object. In this model, one assumes that the desired solution u is a realization of a random variable U, i.e.

$$p_F(f|u) \;=\; p_F(f|U=u)$$

and we use simply $p(f|u)$. With this notation, the Bayes formula delivers the a-posteriori probability density of u for a given value f of F,

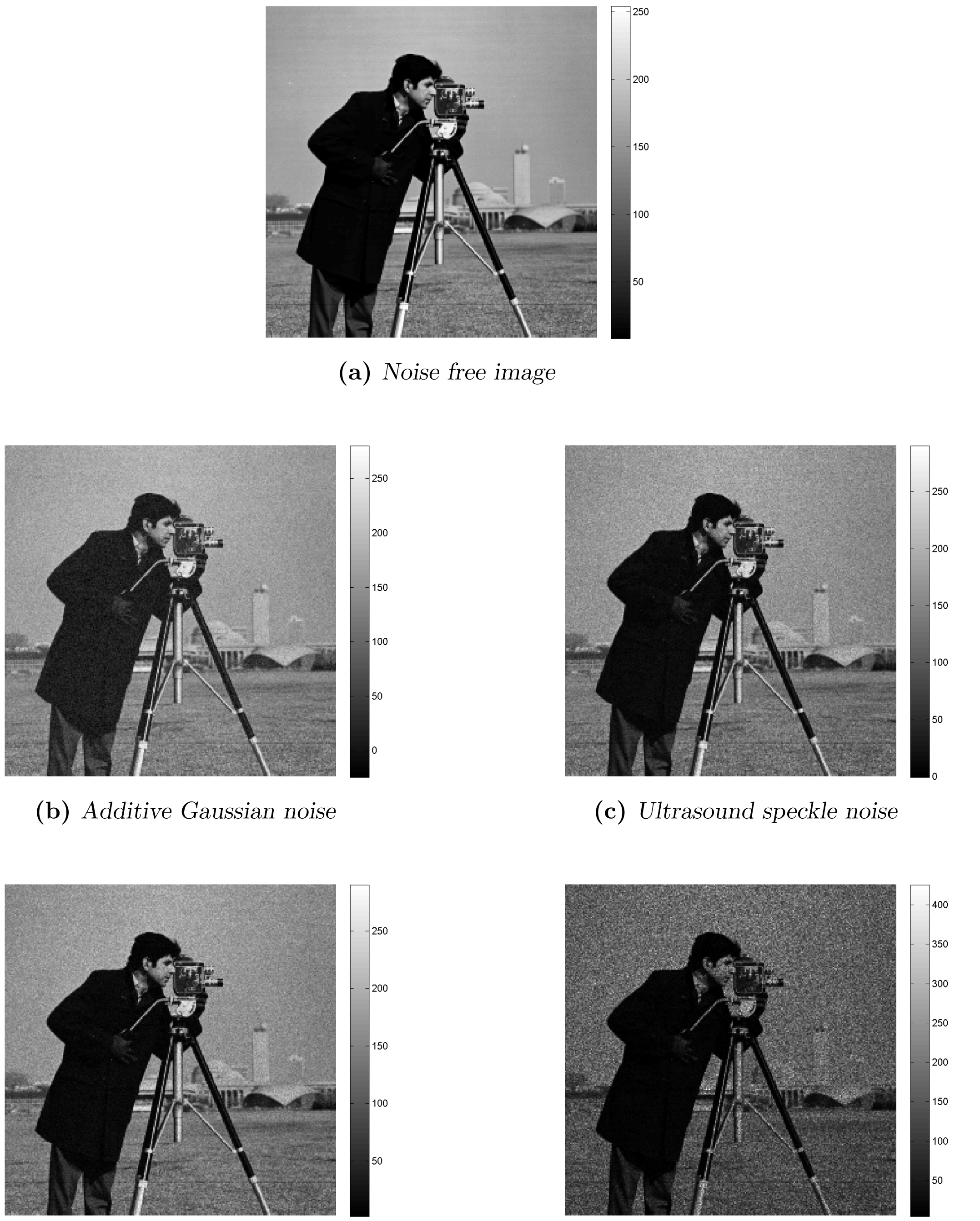
$$p(u|f) \;=\; \frac{p(f|u)\,p(u)}{p(f)}\;. \tag{2.6}$$

Inserting the given measurements f, the density $p(u|f)$ is denoted as the a-posteriori likelihood function, which depends on u only. To determine an approximation to the unknown object u, we use the *maximum a-posteriori probability (MAP) estimator* which maximizes the likelihood respectively minimizes the negative log-likelihood function, i.e.

$$u_{MAP} \;\in\; \arg\min_u \{\, -\log p(u|f)\,\} \;\stackrel{(2.6)}{=}\; \arg\min_u \{\, -\log p(f|u) \;-\; \log p(u)\,\}\;, \tag{2.7}$$

where we neglect the additive term $\log p(f)$, which is independent of u.

The main advantage of the Bayesian approach (2.6) is that it allows to incorporate additional prior information in form of statistical properties of the desired object u via the a-priori probability density $p(u)$ into the reconstruction process. The most frequently used a-priori densities are *Gibbs functions* [75, 76], in analogy to statistical



(a) *Noise free image*

(b) *Additive Gaussian noise*

(c) *Ultrasound speckle noise*

(d) *Poisson noise*

(e) *Multiplicative Gamma noise*

Fig. 2.2. *Illustration of in* (2.4) *presented noise forms in two dimensions (notice that the gray level values are different in the images). (a) Noise free 2D image* $\bar{u}$. *(b)* $\bar{u}$ *is degraded by additive Gaussian noise with standard deviation* $\sigma = 10$. *(c)* $\bar{u}$ *is degraded by ultrasound specific speckle noise with standard deviation* $\sigma = 1$. *(d)* $\bar{u}$ *is degraded by Poisson noise. (e)* $\bar{u}$ *is degraded by multiplicative Gamma noise with assumption of* 20 *measurement means.*

mechanics,

$$p(u) \ \sim \ e^{-\alpha J(u)} \ , \tag{2.8}$$

where α is a positive parameter and $J : W(\Omega) \to \mathbb{R}_{\geq 0}$ an energy functional, usually convex, on a Banach space $W(\Omega) \subset U(\Omega)$.

Now, if we plug the Gibbs a-priori density (2.8) in the maximum a-posteriori approch (2.7), we obtain subsequently minimization problems of the form

$$u_{MAP} \ \in \ \arg\min_{u} \{ \ -\log p(f|u) \ + \ \alpha \, J(u) \ \} \ , \tag{2.9}$$

where we can use suitable probability densities $p(f|u)$ depending on the model of the noise in the given data f. In the case of Gaussian, ultrasound speckle, Poisson, and multiplicative Gamma noise, the negative log-likelihood function $-\log p(f|u)$ has the forms as in (2.5). We note that in context of inverse problems, the functional J in the Gibbs a-priori density (2.8) is related to a regularization functional and the resulting functional from the negative log probability density of the noise $-\log p(f|u)$ to a data fidelity term, cf. (2.10) below.

2.3 General Form of Image Reconstruction Problems

In *Section 2.2* before, we have seen that the modeling of image reconstruction problems via the Bayesian approach and maximum a-posteriori probability estimation led us to minimization problems of the form (2.9). Such problems are related to the calculus of variations if the kind of the noise and thereby the probability density of the noise $p(f|u)$ in the given data f is known, as for example in case of Gaussian, ultrasound speckle, Poisson, or multiplicative Gamma noise given in (2.5). A more general form of all such optimization problems in image reconstruction and inverse problems is given by

$$\min_{u \, \in \, W(\Omega)} \ H_f(Ku - f) \ + \ \alpha \, J(u) \ , \tag{2.10}$$

where

$$H_f \ : \ V(\Sigma) \ \to \ \mathbb{R} \cup \{+\infty\}$$

denotes a general data fidelity term dependent on the given data f and a linear and compact operator

$$K \ : \ U(\Omega) \ \to \ V(\Sigma)$$

with Banach spaces of functions $U(\Omega)$ and $V(\Sigma)$ on bounded and compact sets Ω and Σ. The task of the functional H_f is to penalize the deviation from the operator equality $f = Ku$. In order to guarantee that the data fidelity is centered at zero, we use $Ku - f$ as argument, i.e.

$$H_f(Ku - f) = 0 \qquad \text{if} \qquad Ku = f \,. \tag{2.11}$$

The second term in (2.10) is defined in the same way as the Gibbs functions in (2.8), where

$$J : W(\Omega) \to \mathbb{R}_{\geq 0} \,, \qquad W(\Omega) \subset U(\Omega) \,,$$

denotes a general regularization functional and penalized the deviation from a certain ideal structure (smoothness) of the solution u. The regularization parameter α in (2.10) is a relative weight for both terms and controls the influence of the data fidelity and the regularization term on the solution. For the sake of completeness, note that if we choose K as the identity operator, i.e. $K = I$ and $V(\Sigma) = U(\Omega)$, the variational problem (2.10) becomes a *denoising model* where f is the known noisy image.

Finally, we discuss briefly the importance of the energy functional J in (2.8) respectively (2.9) and (2.10). The specific choice of the regularization functional J is important for the way as a-priori information about the expected solution u are incorporated into the reconstuction process. Smooth, in particular quadratic regularizations have attracted most attention in the past, mainly due to the simplicity in analysis and computation. However, such regularization approaches always lead to blurring of reconstructions, in particular they cannot yield image reconstructions with sharp edges, what is highly unnatural for the human eye.

Recently, singular regularization energies, in particular those of ℓ^1- or L^1-type, have attracted strong attention in inverse problems. In this work, we will concentrate mainly on the *total variation (TV) regularization* functional and its *nonlocal extension*, which we will discuss in *Chapters 6* and *7* in detail. But yet briefly, TV regularization has been derived as a denoising technique in [138] and has been generalized to various other imaging tasks subsequently. The exact definition of TV [2], also used in this work, is

$$|u|_{BV(\Omega)} = \sup_{\substack{g \in C_0^\infty(\Omega, \mathbb{R}^d) \\ \|g\|_\infty \leq 1}} \int_\Omega u \operatorname{div} g \,,$$

which is formally (true if u is sufficiently regular) given by $|u|_{BV(\Omega)} = \int_\Omega |\nabla u| \, dx$. The motivation for using TV is the effective suppression of noise and the realization of homogeneous regions with mostly sharp edges. These features are attractive for almost all image reconstruction problems where the goal is to identify object shapes that are separated by sharp edges and shall be analyzed quantitatively.

3

Basics : Mathematical Foundations

In the previous *Section 2.2*, we have seen that the mathematical modeling of the image reconstruction problems via statistical formulations lead to minimization problems of the form (2.9), where the precise form of the probability density $p(f|u)$ is dependent on the underlying imaging problem, in particular on the present noise in the given data, cf. (2.5). Such a minimization of functionals is a problem of variational calculus with a more general form of optimization problems shown in (2.10). In the following *Chapter 4*, we assume also that the regularization energies J in (2.9) are convex functionals. Hence, we give in this chapter an overview on the basic concepts of *functional analysis*, *convex analysis*, and *calculus of variations* that will be needed in the course of the work. The following results are mainly taken from [143, 64, 66, 114].

3.1 Functional Analysis

The functionals examined in this work will be functionals on Banach spaces associated with various topologies, which can be also weaker as the usually used norm topologies. Thus, we recall here some basic results and concepts of *topology* and *functional analysis*.

3.1.1 General Topology

In this section, we introduce shortly some topological notions and concepts, which will be essential in the course of the work. We recall the definition of a *topological space* and introduce the required expression in order to define a *Banach space* and to characterize a *locally convex space*, which represents a generalization of a normed linear space.

Definition 3.1.1 (*Topology and Topological Space*). *Let X be a set. A* topology τ *on X is a family of subsets of X, called* open sets*, such that*

(O1) The empty set and the whole space are open, i.e. $\emptyset \in \tau$ and $X \in \tau$.

(O2) If $O_1, O_2 \in \tau$, then $O_1 \cap O_2 \in \tau$, i.e. the intersection of two open sets is open.

(O3) If $\{O_i\}_{i \in I}$ is a family of sets $O_i \in \tau$, then $\bigcup_i O_i \in \tau$, i.e. the union of arbitrary many open sets is open.

A pair (X, τ) consisting of a set X and a topology τ is called a topological space.

Remark. Let (X, τ) be a topological space. A set $K \subset X$ is called *closed*, if its complement $X \setminus K$ is open, i.e. $X \setminus K \in \tau$. It is also easy to show that the family τ_C of closed sets fulfills the following properties dual to *(O1)-(O3)* [66, p. 13]:

(C1) The empty set and the whole space are closed, i.e. $\emptyset \in \tau_C$ and $X \in \tau_C$.

(C2) If $C_1, C_2 \in \tau_C$, then $C_1 \cup C_2 \in \tau_C$, i.e. the union of two closed sets is closed.

(C3) If $\{C_i\}_{i \in I}$ is a family of sets $C_i \in \tau_C$, then $\bigcap_i C_i \in \tau_C$, i.e. the intersection of arbitrary many closed sets is closed.

Definition 3.1.2 (*Interior and Closure of a Set*). *Let (X, τ) be a topological space and $A \subset X$. The* interior $\mathbf{int}(A)$ *of A is the largest open set contained in A or equivalently the union of all open sets contained in A, i.e.*

$$\mathbf{int}(A) = \bigcup_{O \subset A \text{ open}} O .$$

The closure $\mathbf{cl}(A)$ *of A is the smallest closed set containing A or equivalently the intersection of all closed sets containing A, i.e.*

$$\mathbf{cl}(A) = \bigcap_{C \supset A \text{ closed}} C .$$

Definition 3.1.3 (*Sequence*). *A* sequence *in a set X is a mapping $\phi : \mathbb{N} \to X$, where $\mathbb{N}$ denotes the set of natural numbers. We will write $x_n := \phi(n)$ for the elements in the sequence and denote the sequence as a whole by (x_n). A* subsequence *of a sequence (x_n) is itself a sequence (x_{n_j}) such that there exists a strictly increasing mapping $N : \mathbb{N} \to \mathbb{N}$ with $x_{n_j} = x_{N(n)}$.*

Definition 3.1.4 (*Convergent Sequence in Topological Space*). *A sequence (x_n) in a topological space (X, τ) is called convergent to some $x \in X$, if for every open set O containing x there exists $n_0 \in \mathbb{N}$ such that $x_n \in O$ for all $n \geq n_0$.*

Definition 3.1.5 (*Continuous Mapping in Topological Spaces, [114, Def. 2.1.7]*). *Let (X, τ_X) and (Y, τ_Y) be topological spaces. A mapping $M : (X, \tau_X) \to (Y, \tau_Y)$ is*

- continuous*, if $M^{-1}(O) \in \tau_X$ for any $O \in \tau_Y$, i.e. if the inverse image of any open subset of Y is open in X,*
- sequentially continuous*, if for every sequence (x_n) in X converging to $x \in X$, the sequence $\big(M(x_n)\big)$ in Y converges to $M(x) \in Y$.*

Remark. Continuity always implies the sequential continuity. However, the converse, i.e. that continuity is equivalent to sequential continuity, only holds if the topological space (X, τ_X) satisfies the *first countability axiom* [120, p. 190], in particular if X is a metric space [95, Thm. 4.11].

Definition 3.1.6 (*Metric Space*). *A* metric *on a set X is a function $d : X \times X \to \mathbb{R}_{\geq 0}$ satisfying*

(M1) $d(x, y) = 0$, *if and only if* $x = y$,

(M2) $d(x, y) = d(y, x)$ *for all* $x, y \in X$, *(symmetry)*

(M3) $d(x, z) \leq d(x, y) + d(y, z)$ *for all* $x, y, z \in X$. *(triangle inequality)*

A metric space *is a pair (X, d) consisting of a set X and a metric d, the number $d(x, y)$ is called the* distance *between x and y.*

Remark. A metric space (X, d) always defines a topological space (X, τ) in the following manner: A set $O \subset X$ is open, if and only if for every $x \in O$ there exists $\epsilon > 0$ such that $\{ y \in X : d(x, y) < \epsilon \} \subset O$. A proof that this generates actually a topology τ on the set X can be found e.g. in [66, pp. 248 - 249]. Hence, we always consider a metric space as a topological space equipped with the topology induced by the metric. Note also that various metrics can well induce the same topology.

Example 3.1.7 (*Metric Space*).

- *Let X be an arbitrary set, then the following distance for $x, y \in X$,*

$$d(x, y) = \begin{cases} 1, & \text{if } x \neq y, \\ 0, & \text{if } x = y, \end{cases}$$

defines a metric on X.

- *The set of real numbers $\mathbb{R}$ and the closed unit interval $[0, 1]$ are metric spaces with the distance between two points defined by the absolute value of their difference.*

Definition 3.1.8 (*Convergent Sequence in Metric Space*). *A sequence* (x_n) *in a metric space* (X, d) *is called convergent to some* $x \in X$ *, if for every* $\epsilon > 0$ *there exists* $n_0 \in \mathbb{N}$ *such that* $d(x_n, x) < \epsilon$ *for all* $n \geq n_0$ *or equivalently, if the sequence of real numbers* $\big(d(x_n, x)\big)$ *converges to zero.*

Definition 3.1.9 (*Cauchy Sequence and Complete Space*). *A sequence* (x_n) *in a metric space* (X, d) *is called* Cauchy sequence, *if for every* $\epsilon > 0$ *there exists* $n_0 \in \mathbb{N}$ *such that* $d(x_n, x_m) < \epsilon$ *for all* $n, m \geq n_0$ *. A metric space* (X, d) *is* complete, *if every Cauchy sequence* (x_n) *in* X *converges to some element* $x \in X$ *.*

Definition 3.1.10 (*Compact Set*). *Let* (X, τ) *be a topological space and* $K \subset X$ *. The set* K *is*

- compact, *if every family* $\{O_i \,:\, i \in I\}$ *of open sets* O_i *satisfying* $K \subset \bigcup_i O_i$ *has a finite subfamily* $O_1, \ldots, O_k$ *such that* $K \subset \bigcup_{i=1}^{k} O_i$ *,*
- precompact, *if its closure is compact,*
- sequentially compact, *if every sequence* $(x_n) \subset K$ *has a subsequence* (x_{n_j}) *converging to some* $x \in K$ *,*
- sequentially precompact, *if every sequence* $(x_n) \subset K$ *has a subsequence* (x_{n_j}) *converging to some* $x \in X$ *, but the limit needs not to be in* K *.*

Theorem 3.1.11. *Let* (X, d) *be a metric space and* $K \subset X$ *. Then* K *is compact, if and only if* K *is sequentially compact.*

Proof. See [95, Thm. 5.5]. □

Theorem 3.1.12. *Let* (X, d) *be a complete, metric space and* $K \subset X$ *. Then* K *is precompact, if and only if* K *is sequentially precompact.*

Proof. See [115, Cor. 4.10]. □

Definition 3.1.13 (*Semi-Norm and Normed Linear Space*). *Let* X *be a linear space (also called vector space) over the real numbers* $\mathbb{R}$ *. A* semi-norm *on* X *is a function* $p : X \to \mathbb{R}_{\geq 0}$ *such that*

(N1) $p(\lambda\, x) = |\lambda|\, p(x)$ *for all* $x \in X$ *and* $\lambda \in \mathbb{R}$ *,* *(positively homogeneous)*

(N2) $p(x + y) \leq p(x) + p(y)$ *for all* $x, y \in X$ *.* *(subadditivity)*

If p *additionally satisfies*

(N3) $p(x) = 0$ *, if and only if* $x = 0$ *,*

then p *is called a* norm *on* X *. In this case, the norm of* $x \in X$ *is denoted by* $\|x\|_X := p(x)$ *and a pair* $(X, \|\cdot\|_X)$ *consisting of a linear space* X *and a norm* $\|\cdot\|_X$ *is called a* normed linear space.

Remark. A normed linear space $(X, \|\cdot\|_X)$ is always a metric space with the distance $d(x,y) := \|x - y\|_X$.

Example 3.1.14 (*Normed Linear Space*).

- *For each* p *such that* $1 \leq p \leq \infty$ *, the Lebesque space* $L^p(\Omega)$ *on an open subset* Ω *of* $\mathbb{R}^d$ *is a normed linear space with the norm* $\|\cdot\|_p$ *given by*

$$\|g\|_p = \begin{cases} \left(\int_\Omega |g|^p \, d\lambda \right)^{1/p} , & \text{if} \quad 1 \leq p < \infty , \\ \inf\limits_{t \in \mathbb{R}} \{ \lambda (\{ x \in \Omega : g(x) > t \}) = 0 \} , & \text{if} \quad p = \infty , \end{cases}$$

where λ *denotes the Lebesque measure.*

- *Many other examples of normed linear spaces can be found e.g. in [114, pp. 11-13].*

Definition 3.1.15 (*Banach Space*). *A* Banach space *is a complete, normed linear space.*

Example 3.1.16 (*Banach Space*).

- *Spaces of continuous functions on compact sets equipped with the supremum norm.*
- *The space of Lebesque measurable functions* $L^p(\Omega)$ *for each* $1 \leq p \leq \infty$ *, as in Example 3.1.14 above.*

Finally, we introduce the notion of a *locally convex, linear topological space*, or shortly *locally convex space*, which generalizes the notion of a normed linear space. Here, we use only a special characterization of such spaces via a family of semi-norms given in [169, p. 26, Thm.]. For a more general definition of locally convex spaces, we refer to [114, Def. 2.2.1] or [169, Def. I.1.3].

Definition 3.1.17 (*Locally Convex Space*). *Let* X *be a linear space and* $(p_i)_{i \in I}$ *a (not necessarily countable) family of semi-norms on* X *. The family of semi-norms is called* separating *if*

$$x = 0 \qquad \text{if and only if} \qquad p_i(x) = 0 \quad \text{for all} \quad i \in I .$$

A pair $(X, (p_i)_{i \in I})$ *consisting of a linear space* X *and a separating family of semi-norms* $(p_i)_{i \in I}$ *is called a* locally convex space.

Remark 3.1.18.

(1) Locally convex spaces generalize the notion of normed linear spaces. Therefore, every normed linear space is locally convex (cf. [114, Thm. 2.2.3]), where the family of semi-norms consists of a single element, namely of the norm.

(2) A family of semi-norms $(p_i)_{i \in I}$ defines a (strong) topology on a locally convex space X in the following manner: A set $O \subset X$ is open, if for every $x \in O$ there exist $\epsilon > 0$ and a finite non-empty set $I' \subset I$ of indices such that

$$\bigcap_{i \in I'} \{ y \in X : p_i(y - x) < \epsilon \} \subset O \, .$$

Example 3.1.19 (*Locally Convex Space*).

- *Every normed linear space is locally convex, cf. Remark 3.1.18, Item (1).*
- *Every normed space equipped with a weaker topology than the strong norm topology, such as the weak or weak* topology (cf. Section 3.1.3), is locally convex.*

Definition 3.1.20 (*Convergent Sequence in Locally Convex Space*). *A sequence* (x_n) *in a locally convex space* $(X, (p_i)_{i \in I})$ *is called convergent to some* $x \in X$, *if and only if* $p_i(x_n - x) \to 0$ *for all* $i \in I$.

3.1.2 Bounded Linear Operators and Functionals

Definition 3.1.21 (*Linear Operator and Functional*). *Let* X *and* Y *be linear spaces. A mapping* $L : X \to Y$ *is a* linear operator, *if*

$$L(x + y) = Lx + Ly \, , \qquad L(\lambda x) = \lambda Lx \qquad \text{for all} \qquad x, y \in X \, , \, \lambda \in \mathbb{R} \, .$$

If $Y = \mathbb{R}$, *then we denote* L *as a* linear functional.

Theorem 3.1.22 (*Continuous and Bounded Linear Operator*). *Let* X *and* Y *be locally convex spaces and* $L : X \to Y$ *a linear operator. The operator* L *is* continuous, *if and only if* L *is* bounded, *i.e. if for every semi-norm* q *on* Y *there exist a semi-norm* p *on* X *and* $C \geq 0$ *such that*

$$q(Lx) \leq C \, p(x) \qquad \text{for all} \qquad x \in X \, . \tag{3.1}$$

If X *and* Y *are normed linear spaces, the condition* (3.1) *is equivalent to*

$$\|Lx\|_Y \leq C \, \|x\|_X \qquad \text{for all} \qquad x \in X \, .$$

Proof. See [169, Thm. I.6.1] in case of locally convex spaces and [169, Cor. I.6.2], [114, Thm. 1.4.2] or [133, Lem. 6.44] in case of normed linear spaces. □

Remark. Some examples with (bounded) linear operators can be found e.g. in [114, p. 26] or [133, Sect. 8.1.4].

The space of bounded linear operators from X to Y is denoted by $L(X,Y)$ and is a linear space with pointwise addition and scalar multiplication,

$$(L_1 + \lambda L_2)(x) = L_1 x + \lambda L_2 x \ , \qquad L_1, L_2 \in L(X,Y) \ , \ x \in X \ , \ \lambda \in \mathbb{R} \ .$$

If X and Y are normed linear spaces, a norm on $L(X,Y)$ is defined by

$$\|L\|_{L(X,Y)} := \sup \{ \|Lx\|_Y : \|x\|_X \leq 1 \} = \sup \left\{ \frac{\|Lx\|_Y}{\|x\|_X} : x \neq 0 \right\} .$$

If Y is a Banach space, then so is $L(X,Y)$ [114, Thm. 1.4.8], independent of the completeness of the normed linear space X.

Definition 3.1.23 (*Continuous Embedding*). *Let X and Y be locally convex spaces. We say X is* continuously embedded *in Y and write $X \hookrightarrow Y$, if $X \subset Y$ and for every semi-norm q on Y there exist a semi-norm p on X and a constant $C \geq 0$ such that*

$$q(x) \leq C\, p(x) \qquad \textit{for all} \qquad x \in X \ .$$

That means, there exists a well defined identity operator from X to Y which is bounded.

Definition 3.1.24 (*Compact Operator [114, Prop. 3.4.4]*). *A linear operator L between Banach spaces X and Y is* compact, *if and only if for every bounded sequence (x_n) in X, the sequence (Lx_n) is precompact in Y, i.e. (cf. Definition 3.1.10) there exists a subsequence (x_{n_j}) such that (Lx_{n_j}) converges in Y.*

Remark. Every compact linear operator from a Banach space into a Banach space is bounded [114, Prop. 3.4.2].

Definition 3.1.25 (*Compact Embedding [133, Def. 7.25]*). *Let X and Y be Banach spaces such that X is continuously embedded in Y (see Definition 3.1.23). We say that X is* compactly embedded *in Y and write $X \overset{c}{\hookrightarrow} Y$, if every bounded sequence in X has a subsequence which converges in Y, i.e. the identity operator from X to Y is compact.*

Definition 3.1.26 (*Dual Space*). *Let X be a locally convex space. The* dual space *$X^* := L(X,\mathbb{R})$ of X is the collection of all bounded linear functionals $x^* : X \to \mathbb{R}$. If $x^* \in X^*$, we write its evaluation at $x \in X$ as*

$$\langle x^*, x\rangle := \langle x^*, x\rangle_{X^*,X} := x^* x$$

and denote $\langle\cdot,\cdot\rangle$ as the standard duality product.

If X is a locally convex space, then its dual X^* is locally convex again with the strong topology on X^* induced by the family of semi-norms [169, Def. IV.7.2],

$$p(x^*) = p(x^*; B) = \sup_{x \in B} |\langle x^*, x\rangle_{X^*,X}| , \qquad B \subset X \text{ bounded}, \quad x^* \in X^* .$$

If X is a normed linear space, then its dual X^* is a Banach space, in particular locally convex space, with the norm [169, Thm. IV.7.1],

$$\|x^*\|_{X^*} := \sup_{\|x\|_X \le 1} |\langle x^*, x\rangle_{X^*,X}| = \sup\left\{ \frac{|\langle x^*, x\rangle_{X^*,X}|}{\|x\|_X} : x \neq 0 \right\} .$$

Hence, in both cases we can define the dual space $X^{**} := (X^*)^*$ of X^*, called the *bidual* of X. In this context, the space X is called *reflexive* if $X^{**} = X$.

Example 3.1.27 (*Dual and Reflexive Spaces*). *Let Ω be an open subset of $\mathbb{R}^d$. Then the space of Lebesgue measurable functions*

- *$X = L^p(\Omega)$ is reflexive for $1 < p < \infty$. The dual space of $L^p(\Omega)$ is $L^{p_*}(\Omega)$ for $1 \le p < \infty$ with $p^{-1} + p_*^{-1} = 1$, using the convention that $1_* = \infty$.*
- *$X = L^1(\Omega)$ is nonreflexive and $X^* = L^\infty(\Omega)$.*

Theorem 3.1.28 (*Adjoint Operator*). *Let X and Y be normed linear spaces. Then, for each $L \in L(X,Y)$ exists a unique bounded linear operator $L^* \in L(Y^*, X^*)$ given by the formula $L^* y^* := y^* \circ L$ with $y^* \in Y^*$, i.e. the operator is defined by*

$$\langle L^* y^*, x\rangle_{X^*,X} = \langle y^*, Lx\rangle_{Y^*,Y} , \qquad y^* \in Y^* , \quad x \in X .$$

The operator L^* is called the adjoint *of L and satisfies*

$$\|L^*\|_{L(Y^*,X^*)} = \|L\|_{L(X,Y)} .$$

Proof. See [139, Thm. 4.10]. □

3.1.3 Weak and Weak* Topologies

The most frequently used topology in functional analysis is the strong topology, which is induced by a norm. However, as the name already implies, such topologies are strong in the sense that they contain many open sets. This characteristic has of course some advantages, as for instance that functionals in such topological spaces appear simpler to fulfill the continuity property, but contains also some disadvantages and hence is not always suitable. The most crucial weakness of strong topologies is the fact that in an infinite dimensional normed space the topology contains so many open sets such that its closed unit ball cannot be compact. Therefore, many facts which are well known for finite dimensional normed spaces, as for instance that all closed bounded subsets are compact, cannot be generalized in a straight-forward way to the infinite dimensional case. Hence, in this section we introduce the *weak* and the *weak* topology* in order to allow such a transfer. Both topologies are in general weaker than the strong norm topology in the sense that they have fewer open sets, but are strong enough to allow for instance the compactness of closed unit balls in the infinite dimensional case.

In the following, we recall the definitions of the weak and weak* topologies only for a normed linear space X, since we will be more interested later in this case. However, we notice that the both notions can be also generalized to a locally convex space X, where the definitions of the respective families of semi-norms can be found for instance in [169, Sect. IV.7] and [143, Def. 8.46].

Throughout this section, let X be a normed linear space, X^* the dual space and $X^{**} = (X^*)^*$ the bidual space of X. The following results are taken from [169, Sect. V.1] and [6, Sect. 2.1.1].

Definition 3.1.29 (*Topologies on* X). *Let* (x_n) *be a sequence in* X. *Then*

- *the* strong topology, *denoted by* $x_n \to x$, *is defined by*

$$\|x_n - x\|_X \to 0 \ .$$

- *the* weak topology, *denoted by* $x_n \rightharpoonup x$, *is defined by*

$$\langle x^*, x_n \rangle_{X^*,X} \to \langle x^*, x \rangle_{X^*,X} \qquad \textit{for every} \qquad x^* \in X^* \ .$$

Remark.

- The weak topology on X is the smallest topology on X such that every member of the dual space X^* is continuous with respect to that topology, see [114, p. 212].

- One can simply show from the definition of the weak topology that the strong convergence implies weak convergence [169, Thm. V.1.1], but the converse is in general false, see e.g. [6, p. 25].

Definition 3.1.30 (*Topologies on* X^*). *Let* (x_n^*) *be a sequence in* X^*. *Then*

- *the* strong topology, *denoted by* $x_n^* \to x^*$, *is defined by*

$$\|x_n^* - x^*\|_{X^*} \to 0 .$$

- *the* weak topology, *denoted by* $x_n^* \rightharpoonup x^*$, *is defined by*

$$\langle x^{**}, x_n^* \rangle_{X^{**},X^*} \to \langle x^{**}, x^* \rangle_{X^{**},X^*} \qquad \text{for every} \qquad x^{**} \in X^{**} .$$

- *the* weak* topology, *denoted by* $x_n^* \rightharpoonup^* x^*$, *is defined by*

$$\langle x_n^*, x \rangle_{X^*,X} \to \langle x^*, x \rangle_{X^*,X} \qquad \text{for every} \qquad x \in X .$$

Remark.

- The weak* topology on X^* is the smallest topology on X^* such that, for each $x \in X$, the linear functional $x^* \mapsto \langle x^*, x \rangle_{X^*,X}$ on X^* is continuous with respect to that topology, see [114, p. 223].

- The weak* topology on X^* is in general weaker than the weak topology on X^* [114, Thm. 2.6.2], since it holds $X \subset X^{**}$.

- If X is reflexive, i.e. $X = X^{**}$, then the weak and weak* topology coincide, cf. e.g. [114, Thm. 2.6.2]. This easily follows from the definition of the weak topology on X^*, as it uses the same convergence as the weak* topology on X^*, and from the definition of the weak* topology on X^{**}, as it uses the same convergence as the weak topology on X.

The following *Theorem of Banach-Alaoglu* demonstrates now why the weak* topology on X^* is notably different from the strong and weak topology on X. As already mentioned at the beginning of this section, the main difficulty in the transfer of results from the finite dimensional normed space to the infinite dimensional case is the loss of the Heine-Borel property, which means for a finite dimensional X that all closed bounded subsets of the space are compact. Fortunately, such results can be obtained for the infinite dimensional case, if one uses weaker topologies than the strong one. Actually, due to the fact that the weak topology on X is a proper subtopology of the strong one allows a closed unit ball to be weakly compact than compact. However, this property is unfortunately true if and only if X is reflexive, see [114, Thm. 2.8.2]. For a general normed linear space X, we will see in the next theorem, that *all closed bounded subspaces of X^* are always compact if we use the weak* topology on X^*.*

Theorem 3.1.31 (*Banach-Alaoglu Theorem*). *The set*

$$\{ x^* \in X^* : \|x^*\|_{X^*} \leq C \} , \qquad C > 0 ,$$

is compact in the weak topology.*

Proof. See [55, Sect. 4.1, Thm. 1] or [114, Thm. 2.6.18]. □

3.2 Convex Analysis and Calculus of Variations

The minimization of functionals is a classical problem of variational calculus, which deals with general optimization problems in mathematics. As usual in optimization problems, we are primarily interested in optimality conditions based on derivatives. However, the classical notions of derivatives, as Gâteaux or Fréchet differentiability, will be insufficient in this work since we will study convex singular regularization energies, i.e. especially that they are non differentiable in the classical sense, and hence essential results from the convex analysis are necessary to handle such problems. Therefore, we review in the following the basic concepts of *convex analysis* and *calculus of variations*. Some standard reference works dealing with these topics are e.g. [64] and [137].

3.2.1 Convex and Lower Semicontinuous Functionals

In this thesis, we study convex functionals in the context of variational methods. Thus, we recall here some basic concepts of *convexity*, which are also crucial for the uniqueness of a solution in minimization problems. Furthermore, since the continuity of a convex functional is not so simple to characterize due to nonlinearity, we introduce an essentially weaker property, namely the *lower semicontinuity* of a functional.

Definition 3.2.1 (Convex Set). *Let U be a linear space and $C \subset U$. The set C is* convex, *if*

$$\lambda u + (1 - \lambda) v \in C \qquad \text{for all} \qquad u, v \in C, \quad \lambda \in (0,1) .$$

Definition 3.2.2. *Let U be a linear space and $F : U \to \mathbb{R} \cup \{+\infty\}$ a functional.*

- *The* effective domain *of F is the set*

$$\mathcal{D}(F) \;:=\; \{\; u \;\in\; U \;:\; F(u) \;<\; +\infty \;\} \;.$$

- *The functional F is* proper, *if $\mathcal{D}(F) \;\neq\; \emptyset$, i.e. F is not identically equal to $+\infty$.*

- *The functional F is* convex *if it satisfies*

$$F(\lambda\, u + (1-\lambda)\, v) \;\leq\; \lambda\, F(u) + (1-\lambda)\, F(v) \;, \qquad u,\, v \;\in\; U, \quad \lambda \;\in\; [0,1] \;. \tag{3.2}$$

- *The functional F is* strictly convex, *if the inequality in* (3.2) *is strict whenever $u \neq v \in \mathcal{D}(F)$ and $\lambda \in (0,1)$.*

Remark 3.2.3.

(1) It is easy to show that, if $F : U \to \mathbb{R} \cup \{+\infty\}$ is a convex functional on a linear space U, then the *sub-level sets of F* defined by

$$\{\; u \;\in\; U \;:\; F(u) \;\leq\; a \;\} \qquad \text{and} \qquad \{\; u \;\in\; U \;:\; F(u) \;<\; a \;\}$$

are convex for every $a \in \mathbb{R} \cup \{+\infty\}$.

(2) Note also that with the property in the first item, the effective domain of a convex functional is a convex set of U.

(3) We also shortly clarify why we allow the value $+\infty$. If U is a linear space and $F : A \subset U \to \mathbb{R}$ a functional, we can extend F to a functional $\widetilde{F}$ on the whole space U by

$$\widetilde{F}(u) \;=\; \begin{cases} F(u)\;, & \text{if} \quad u \;\in\; A\;, \\ +\infty\;, & \text{if} \quad u \;\in\; U \setminus A\;. \end{cases}$$

Thus, $\widetilde{F}$ is convex if and only if $A \subset U$ is convex and $F : A \to \mathbb{R}$ is convex. Therefore, the extension by $+\infty$ allows to deal only with such convex functionals, which are always defined everywhere, in the theory of convex analysis. In addition, in the context of minimization problems, we can extend convex functionals on larger function spaces without changing the stationary points of the minimization problem. In many cases, such an approach allows a simpler mathematical handling of the problem without manipulation of admissible sets.

Now, we pass over to the topological properties of general convex functionals, which are not as simple as in the case of linear functionals. In *Theorem 3.1.22*, the continuity of a linear functional is a simple consequence of boundedness. In the case of a convex functional, this property is in general not ensured such that we need another concepts from the convex analysis.

Definition 3.2.4 (*Lower Semicontinuous Functional [64, pp. 9 - 10]*). *Let U be a locally convex space and $F : U \to \mathbb{R} \cup \{+\infty\}$ a functional (not necessarily convex). Then, F is* lower semicontinuous*, if it satisfies the following equivalent conditions:*

(i) The sub-level sets

$$\{ \; u \; \in \; U \; : \; F(u) \; \leq \; a \; \}$$

are closed for every $a \in \mathbb{R}$.

(ii) For any $u \in U$ and for every converging sequence (u_n) with limit u it holds

$$F(u) \;\leq\; \liminf_{n \to \infty} F(u_n) \; .$$

This condition is also known as the sequential lower semicontinuity *of F.*

Remark. In particular, *Definition 3.2.4* is applicable for a Banach space U equipped with the strong or weak topology, since every normed space is locally convex (cf. *Remark 3.1.18*) and is naturally equipped with a strong norm topology. Moreover, we can also additionally introduce a weak topology on a locally convex space, cf. *Section 3.1.3*.

Lemma 3.2.5 (*Continuous Convex Functional*). *Every lower semicontinuous convex functional over a Banach space is continuous over the interior of its effective domain.*

Proof. See [64, p. 13, Cor. 2.5]. □

3.2.2 Subdifferentiability

The singular regularization energies that we will study in the course of the work are non-differentiable functionals in the classical sense, i.e. not Gâteaux or Fréchet differentiable, and thus other methods are needed. Fortunately, convex analysis delivers a powerful concept of generalized derivatives for convex functionals called *subdifferential*, which we introduce in this section.

Definition 3.2.6 (*Subdifferential*). *Let U be a locally convex space, U^* the dual space of U and $F : U \to \mathbb{R} \cup \{+\infty\}$ a functional (not necessarily convex). A* subgradient *of F at $u \in U$ is an element $u^* \in U^*$ such that*

$$F(v) - F(u) - \langle u^*, v - u \rangle \geq 0 \qquad \textit{for all} \qquad v \in U \,. \tag{3.3}$$

For each $u \in U$, the set of all subgradients of F at u is called the subdifferential *at u and is denoted by $\partial F(u)$, i.e.*

$$\partial F(u) = \{ u^* \in U^* \;:\; F(v) - F(u) - \langle u^*, v - u \rangle \geq 0 \quad \textit{for all} \quad v \in U \} \,.$$

If $\partial F(u) \neq \emptyset$, F is said to be subdifferentiable *at u.*

Remark 3.2.7.

(1) A subgradient $u^* \in \partial F(u) \subset U^*$ can be identified with the slope of a hyperplane in $U \times \mathbb{R}$ through $(u, F(u))$, that lies under the graph of F, as illustrated in *Fig. 3.1*.

(2) The subdifferential of F is a multivalued mapping ∂F which assigns the set $\partial F(u) \subset U^*$ (possibly empty or singleton) to each $u \in U$, cf. e.g. *Example 3.2.8* or *Fig. 3.1*.

(3) In *Lemma 3.2.17*, we will see at least in the context of convex functionals that the subdifferentiability is actually a generalization of the classical notion of Gâteaux differentiability. I.e., if F is convex and Gâteaux differentiable at u then $\partial F(u)$ is singleton and coincides with the Gâteaux derivative at u, cf. also *Fig. 3.1*.

Example 3.2.8 (*Subdifferential of the Absolute Value Function*). *Let $U = \mathbb{R}$ and $F : U \to \mathbb{R}_{\geq 0}$, $u \mapsto |u|$, be the absolute value function. For $u < 0$ the subgradient is unique, namely $\partial F(u) = \{-1\}$. It is also similar for $u > 0$, we have $\partial F(u) = \{1\}$. For $u = 0$, the subdifferential is defined by the inequality*

$$|v| \geq u^* v \qquad \textit{for all} \qquad v \in U \,,$$

which is satisfied if and only if $u^* \in [-1, 1]$. I.e., the subdifferential of F at u is given by

$$\partial F(u) = \begin{cases} \{-1\} \,, & \textit{for} \quad u < 0 \,, \\ [-1, 1] \,, & \textit{for} \quad u = 0 \,, \\ \{1\} \,, & \textit{for} \quad u > 0 \,. \end{cases}$$

In general, it is not easy to decide whether a functional is subdifferentiable or not. However, in case of a convex functional there is a simple criterion for the subdifferentiability.

Lemma 3.2.9. *Let U be a locally convex space and $F : U \to \mathbb{R} \cup \{+\infty\}$ a convex functional. Then $\partial F(v) \neq \emptyset$ for all $v \in \mathbf{int}\big(\mathcal{D}(F)\big)$, where $\mathbf{int}\big(\mathcal{D}(F)\big)$ denoted the interior of the effective domain $\mathcal{D}(F)$ of F. In particular, $\partial F(u) \neq \emptyset$ if F is finite and continuous at $u \in U$.*

Proof. See [64, p. 22, Prop. 5.2]. □

Remark 3.2.10. In [64, p. 23, Remark 5.1], the authors remark that a proper lower semicontinuous convex functional F defined on a complete normed linear space is subdifferentiable "almost everywhere" (more precisely, over a dense subset) inside the effective domain $\mathcal{D}(F)$ of F, cf. to *Lemma 3.2.5.*

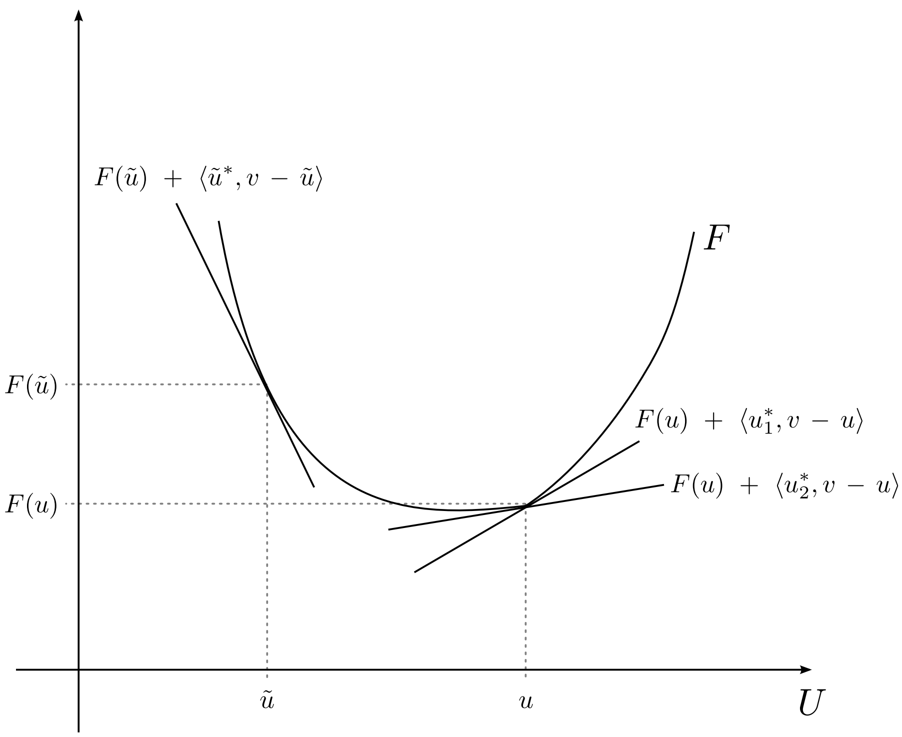


Fig. 3.1. *The convex functional F is differentiable at $\tilde{u} \in U$ and has therefore an unique subgradient at $\tilde{u}$ (see Lemma 3.2.17), namely the Gâteaux derivative $\tilde{u}^* \in U^*$. At point $u \in U$, the functional F is not differentiable and has hence at this point multiple subgradients, where we shown here only $u_1^* \in U^*$ and $u_2^* \in U^*$.*

Next, we introduce now a characterization of the subdifferentials in the special case of one-homogeneous functionals, which will be required during the theoretical analysis of numerical schemes proposed in *Section 4.6.*

Lemma 3.2.11 (*Characterization of Subdifferentials for one-homogeneous Functional*). *Let U be a locally convex space and $F : U \to \mathbb{R} \cup \{+\infty\}$ a convex one-homogeneous* *functional, i.e.* F *satisfies* $F(\lambda u) = \lambda F(u)$ *for all* $\lambda > 0$. *Then, the subdifferential of* F *at* $u \in U$ *is given by*

$$\partial F(u) = \{ u^* \in U^* : \langle u^*, u \rangle = F(u) \text{ and } \langle u^*, v \rangle \leq F(v) \text{ for all } v \in U \} .$$

Proof. Let u^* be a subgradient of F at $u \in U$. Then, the definition of subgradient in (3.3) yields

$$\langle u^*, v - u \rangle \leq F(v) - F(u) \quad \text{for all} \quad v \in U . \tag{3.4}$$

Using the one-homogeneity of F and the fact that U is a linear space, we obtain consequently for $v = 0$ the inequality

$$\langle u^*, u \rangle \geq F(u)$$

and for $v = 2u$ the upper bound

$$\langle u^*, u \rangle \leq F(2u) - F(u) = 2F(u) - F(u) = F(u) .$$

Hence, we have altogether that $\langle u^*, u \rangle = F(u)$ and (3.4) delivers then $\langle u^*, v \rangle \leq F(v)$ for all $v \in U$. □

As already noticed repeatedly, we deal in this work with minimization of functionals and hence we are also interested particularly in the role of the subdifferential in optimization problems.

Lemma 3.2.12 (*Subdifferential and Optimality*). *Let* U *be a locally convex space and* $F : U \to \mathbb{R} \cup \{+\infty\}$ *a convex functional. Then,* $u \in U$ *is a minimizer of* F *if and only if* F *is subdifferentiable at* u *and*

$$0 \in \partial F(u) . \tag{3.5}$$

Due to the convexity of F, *this condition is not only necessary, but also sufficient.*

Proof. Let $0 \in \partial F(u)$, then the definition of $\partial F(u)$ (3.3) implies that $F(u) \leq F(v)$ for all $v \in U$, which is equivalent to stating that u minimizes F. Vice versa, let u be a minimizer of F, then the inequality

$$0 \leq F(v) - F(u) = F(v) - F(u) - \langle 0, v - u \rangle \quad \text{for all} \quad v \in U$$

holds, which implies that $0 \in \partial F(u)$. □

Finally, the following lemma provides an important property of the subdifferential, which will be necessary later.

Lemma 3.2.13 (*Subdifferential Calculus*). *Let* $F, S : U \to \mathbb{R} \cup \{+\infty\}$ *be convex and assume that there exists* $v \in \mathcal{D}(F) \cap \mathcal{D}(S)$ *such that* F *is continuous in* v. *Then*

$$\partial(F + S)(u) = \partial F(u) + \partial S(u) \qquad \text{for all} \qquad u \in U .$$

Proof. See [64, p. 26, Prop. 5.6]. □

3.2.3 Differentiability of Functionals

Finally, we recall the definition of directional derivatives of functionals and see, in the context of convex functionals, that subdifferentiability is actually a generalization of classical differentiability. The definitions here are taken from [64, p. 23, Def. 5.2].

Definition 3.2.14 (*Directional Derivative*). *Let* $F : U \to \mathbb{R} \cup \{+\infty\}$ *be a functional (not necessarily convex) on a locally convex space* U. *The* directional derivative *of* F *at* $u \in U$ *in the direction* $v \in U$, *denoted by* $F'(u; v)$, *is defined as*

$$F'(u; v) = \lim_{t \to 0^+} \frac{F(u + t\,v) - F(u)}{t} , \tag{3.6}$$

if that limit exists.

Remark 3.2.15.

(1) If the functional F is convex, then the expression (3.6) always has a limit, which however can be $\pm\infty$.

(2) In practice, a useful strategy to compute a directional derivative is to define a function $\phi_v : \mathbb{R}_{\geq 0} \to \mathbb{R} \cup \{\pm\infty\}$, $t \mapsto F(u + t\,v)$, for a direction $v \in U$. Then, the expression (3.6) for $F'(u; v)$ can be computed by

$$F'(u; v) = \lim_{t \to 0^+} \frac{\phi_v(t) - \phi_v(0)}{t} = \phi_v'(t)\,|_{t=0} .$$

Definition 3.2.16 (*Gâteaux Differential*). *Let* $F : U \to \mathbb{R} \cup \{+\infty\}$ *be a functional on a locally convex space* U *and assume that the directional derivatives* $F'(u; v)$ *of* F *at* $u \in U$ *exist for all directions* $v \in U$. *Then, if there exists a bounded linear operator* $F'(u) \in U^*$ *such that*

$$F'(u; v) = \langle F'(u), v \rangle_{U^*, U} = F'(u)\,v \qquad \text{for all} \qquad v \in U ,$$

we say that F *is* Gâteaux differentiable *at* u *and call* $F'(u)$ *the* Gâteaux derivative *of* F *at* u.

Finally, we see that for convex functionals subdifferentiability is actually a generalization of Gâteaux differentiability, i.e. Gâteaux differentiability is the same as the uniqueness of the subgradient.

Lemma 3.2.17. *Let U be a locally convex space, $F : U \to \mathbb{R} \cup \{+\infty\}$ a convex functional and $u \in U$. If F is Gâteaux differentiable at u, then the subdifferential $\partial F(u)$ of F at u is singleton and it holds $\partial F(u) = \{F'(u)\}$. Conversely, if F is continuous and finite at u and the subdifferential $\partial F(u)$ of F at u is singleton, then F is Gâteaux differentiable at u and it holds $\partial F(u) = \{F'(u)\}$.*

Proof. See [64, p. 23, Prop. 5.3]. □

Remark. Let $F : U \to \mathbb{R} \cup \{+\infty\}$ be a functional on a locally convex space U. If $u \in U$ is a minimizer of F, then by definition

$$F(u + t\,v) - F(u) \geq 0 \qquad \text{for all} \quad v \in U \quad \text{and} \quad t > 0 \,.$$

Consequently, it holds for the directional derivatives $F'(u;\, v)$ that

$$F'(u;\, v) \geq 0 \qquad \text{for all} \qquad v \in U \,. \tag{3.7}$$

Additionally, if F is Gâteaux differentiable at u, then (3.7) is equivalent to

$$F'(u) = 0 \,, \tag{3.8}$$

since $F'(u;\, -v) = -F'(u;\, v)$. The conditions (3.7) and (3.8) are called *(first order) optimality conditions* for a minimizer of F. Here, we see also that the optimality condition (3.5) in *Lemma 3.2.12* is actually a generalization of (3.8) for non differentiable convex functionals.

4 Imaging : Poisson Framework

In this chapter, we introduce a *variational regularization framework* for inverse problems with data corrupted by *Poisson noise*. This kind of noise, also called *photon counting noise*, occurs in several real life applications, such as medical imaging, light microscopy, or astronomy, and differentiates crucial from the common studied additive Gaussian noise in image processing. Consequently, an adapted variational model is required in order to handle this type of noise accurate. Hence, we study in the following the regularized Poisson likelihood estimation problem obtaining via a *Bayesian model* and *maximum a-posteriori probability estimation*. However, the Poisson likelihood reconstruction model leads to issues in the computation of minimizers, caused by the strong nonlinearity in the data fidelity term. Hence, we propose a robust *forward-backward splitting approach* to provide an efficient numerical scheme, based on the *expectation-maximization algorithm*.

4.1 Introduction

Image reconstruction is a fundamental problem in several areas of applied sciences, such as medical imaging, biophysics, geophysics, or astronomy, with an enormous number of applications. Mathematically, all these image reconstruction problems can in general be formulated as an inverse and ill-posed problem, cf. the discussion in *Section 2.1*,

$$\bar{f} \;=\; \bar{K}\bar{u} \;,$$

with a linear and compact operator $\bar{K} : U(\Omega) \to V(\Sigma)$, exact data $\bar{f}$ and the desired exact image $\bar{u}$. Unfortunately, in practice only noisy versions f and K of $\bar{f}$ and $\bar{K}$ are available and an approximation u of $\bar{u}$ from the operator equation (2.2),

$$f \;=\; Ku \tag{4.1}$$

is wanted, where the operator $K : U(\Omega) \to \mathbb{R}^N$ here is a semi-discrete operator based on $\bar{K}$, which transforms the desired properties of u to the discrete raw data f. As we already discussed at the end of *Section 2.1*, the computation of u by a direct inversion of K is not reasonable since the problem (4.1) is highly ill-conditioned due to the compactness of the operator $\bar{K}$. Hence, some type of regularization is required to enforce stability during the inversion process and to compute useful reconstructions.

A frequently used class of regularization techniques are variational methods based on the minimization of functionals of the form

$$\frac{1}{p} \|Ku - f\|_{L^p(\Sigma)}^p + \alpha J(u) , \qquad \alpha > 0 , \qquad p \in [1, \infty) . \tag{4.2}$$

The first term, so-called *data fidelity term*, penalizes the deviation from the operator equality (4.1). The second term is a *regularization functional*, typically convex, which introduces a-priori information about the expected solution. Finally, the *regularization parameter* α acts as a relative weight for both terms. However, from the viewpoint of statistical modeling in *Section 2.2*, it is important to notice that the functionals in (4.2) result from the additive assumption of exponentially distributed raw data, i.e. we have $f = K\bar{u} + \eta$, where η is a vector valued random variable with statistically independent and identically distributed components. Typical examples are that η is Laplace distributed $(p = 1)$ or Gaussian distributed $(p = 2)$, cf. *Section 2.2.* We notice also that in case of $p = 2$, the minimization strategy (4.2) corresponds to the classical Tikhonov regularization methods, cf. e.g. [83].

Most works in the area deal with the case of additive Gaussian noise so far. However, in real life there are several applications in which different types of noise occur. For instance, so-called *Poisson or photon counting noise* appears in positron emission tomography in medical imaging [148, 160, 167], fluorescence microscopy [87, 57], CCD cameras [152] or astronomical images [108, 102]. Other non-Gaussian noise models are *salt and pepper noise* or the different types of *multiplicative noise*, for example appearing in synthetic aperture radar (SAR) or *speckle noise* in medical ultrasonic imaging [158, 92]. For such cases, different variational models, i.e. different data fidelity terms related to the negative log-likelihood of the noise distribution, can be derived in the framework of maximum a-posteriori probability (MAP) estimation, see *Section 2.2.* Consequently, dependent on the chosen variational model, we need an analysis different from that in the case of additive exponentially distributed noise as in (4.2).

In this thesis, we mainly address the task to reconstruct images where the data are corrupted by Poisson noise, which is important in various applications. An example for such an application is positron emission tomography (PET), a biomedical imaging technique in nuclear medicine that generates images of living organisms by visualizing weak radioactively marked pharmaceuticals, so-called tracers. Due to the crucial possibility of measuring temporal tracer uptake from list-mode data, this imaging modality is particularly suitable for investigating biochemical and physiological processes, such as glucose metabolism, blood flow or receptor concentrations, see e.g. [167, 12]. Another application of image reconstruction problems with Poisson distributed data is fluorescence microscopy [87]. It represents an important technique for investigating living biological cells at nanoscales. In this type of applications, image reconstruction arises in terms of deconvolution problems, where the undesired blurring effects are caused by diffraction of light.

In the both applications that we just mentioned, the measured data are of stochastic nature due to the radioactive decay of tracers in PET and laser scanning techniques in fluorescence microscopy. Furthermore, the random variables of the measured data are not Gaussian but rather Poisson distributed [148, 160]. As already briefly discussed in *Section 2.2* and will deduced in more detail in *Section 4.2*, the MAP estimation via the negative log-likelihood function leads in this kind of noise to the following variational problem,

$$\min_{u \geq 0} \int_{\Sigma} (Ku \; - \; f \log Ku) \; d\mu \; + \; \alpha \, J(u) \; . \tag{4.3}$$

Up to additive terms independent of u, the data fidelity term here is the so-called *Kullback-Leibler divergence* (also known as cross entropy or I-divergence) between two probability measures f and Ku. A particular complication of (4.3) compared to (4.2) is the *strong nonlinearity in the data fidelity term*, leading to issues in the computation of minimizers. A further challenge in this work is the *positivity constraint* on the solution u in (4.3), which is absolutely necessary since in typical applications these functions represent densities or intensity information. In the absence of regularization, i.e. $J \equiv 0$ in (4.3), the *expectation-maximization (EM)* [56, 148], or Richardson-Lucy [136, 111], algorithm has become a standard reconstruction scheme, which is however difficult to generalize to the regularized case. The robust and accurate solution of this problem for appropriate models of J, as well as its analysis are the main contributions of this thesis.

As usual in regularization theory of ill-posed problems, the specific choice of the regularization functional J in (4.3) is crucial to the way a-priori information about the expected solution is incorporated into the reconstruction process. Smooth, in particular quadratic regularizations, have attracted most attention in the past, mainly due to the simplicity in analysis and computation. However, such regularization approaches always lead to blurring in the reconstructions, in particular, they cannot yield reconstructions with sharp edges. Hence, singular regularization energies, especially those of ℓ^1- or L^1-

type, have recently attracted strong attention in variational problems, but are difficult to handle due to their non differentiable nature. Nevertheless, we will focus our attention on the *total variation (TV) regularization* functional in *Chapter 6* and its *nonlocal extension* in *Chapter 7*. As anticipated here, TV regularization has been derived as a denoising technique in [138] and has been generalized to various other imaging tasks subsequently. The exact definition of TV [2], used in this work, is

$$|u|_{BV(\Omega)} \;=\; \sup_{\substack{g \,\in\, C_0^\infty(\Omega,\mathbb{R}^d)\\ \|g\|_\infty \,\leq\, 1}} \int_\Omega u \,\mathrm{div} g \;, \tag{4.4}$$

which is formally (true if u is sufficiently regular)

$$|u|_{BV(\Omega)} \;=\; \int_\Omega |\nabla u| \; dx \;. \tag{4.5}$$

The motivation for using TV is the effective suppression of noise and the realization of homogeneous regions with mostly sharp edges. These features are attractive for almost all image reconstruction problems where the goal is to identify object shapes that are separated by sharp edges and shall be analyzed quantitatively.

In the past, various methods have been suggested for the regularized Poisson likelihood estimation problem (4.3), for instance in case of Tikhonov regularization [17], diffusion regularization [18] or L^1 regularization functional [112]. However, most works deal with TV regularization functionals (4.4) or (4.5), e.g. in application to PET [93, 127], deconvolution problems [19, 57, 72, 147], or denoising problems with identity operator K [104], but still with some restrictions. These limitations can be traced back to the strong computational difficulties in the minimization of (4.3) with TV regularization and can be separated in two main problems:

- The methods use the exact definition of total variation in (4.4), but require an inversion of the operator K^*K, where K^* is the adjoint operator of K. For example, the authors in [72] and [147] proposed two algorithms, called *PIDAL* and *PIDSplit+*, using an augmented Lagrangian approach and the equivalent split Bregman method, respectively. Both algorithms require an inversion of the operator $I + K^*K$, where I is the identity operator, with the result that such methods are efficient only if K^*K is diagonalizable and can be inverted efficiently, as for instance a convolution operator via fast Fourier transform (FFT) or discrete cosine transform (DCT). Additionally, in contrast to the PIDSplit+ algorithm in [147], the PIDAL algorithm in [72] ensures that Ku is nonnegative and not that the final solution u is nonnegative, which however is essential in (4.3).

- To overcome the non differentiability of the TV regularization functional, the other class of methods uses an approximation of (4.5) by differentiable functionals,

$$|u|^{\epsilon}_{BV(\Omega)} \;=\; \int_{\Omega} \sqrt{|\nabla u|^2 \;+\; \epsilon}\; dx \;, \qquad \epsilon \;>\; 0 \;, \tag{4.6}$$

  and creates blurring effects in reconstructed images. In [14], Bardsley proposed an efficient computational method based on gradient projection and lagged-diffusivity, where the nonnegativity constraint is guaranteed via a simple projection on the feasible set. On the other hand, the schemes in [57], [93] and [127] are realized as elementary modifications of the EM algorithm (see *Section 4.3*) with a fully explicit or semi-implicit treatment of TV in the iteration. A major disadvantage of these approaches is that the regularization parameter α need to be chosen very small, since otherwise the positivity of solutions is not guaranteed and the EM based algorithm cannot be continued. Due to the additional parameter dependence on ϵ these algorithms are even less robust.

In this chapter, we propose a robust framework for the regularized Poisson likelihood estimation problem (4.3) using arbitrary convex regularization energy J, which can be even singular (i.e. non differentiable in the usual sense and not strictly convex). In *Section 4.4*, we use a *forward-backward (FB) splitting* approach [107, 128, 157] for this framework, that can be realized by alternating a classical EM reconstruction step and solving a convex variational problem which is just a simple modification of the well-known version of the regularized L^2 problem with a weight in the data fidelity term. The first advantage of this approach is that the EM algorithm does not needed any inversion of the operator K and is hence applicable for an arbitrary operator, as e.g. the Radon or X-ray transform [123]. Additionally, if the implementation of the EM algorithm already exists, one can use it without any modifications. Furthermore, due to the decoupling of reconstruction and smoothing done by the FB splitting approach, one can use standard numerical schemes known from the regularized L^2 problem in the smoothing step, depending on the chosen regularization functional J, with some simple modifications caused by the weight in the data fidelity term, e.g. even for the exact TV regularization functional (4.4). Hence, this approach enables a high flexibility, can be performed equally well for large regularization parameters and is also favourably applicable for problems with a low signal-to-noise ratio.

In *Section 4.5*, we then propose a regularized variational model to denoise an image corrupted by Poisson noise. Of course, this model corresponds to the reconstruction problem (4.3) with the identity operator K and we can reduce the FB splitting approach for $K = I$. This procedure leads to an iterative strategy where in each step a weighted

L^2 regularization model is to be solved. Due to the iterative nature of this approach, we investigate additionally a second order Taylor approximation of the Poisson data fidelity term in (4.3) with $K = I$, which yields a more straight-forward numerical solution.

In *Section 4.6*, we additionally study the FB splitting model from an analytical point of view. We investigate the well-posedness of the minimization problem (4.3) with a convex regularization functional J in terms of existence, uniqueness and stability, prove the positivity preservation of the FB splitting algorithm and provide a convergence analysis for a damped FB splitting strategy.

Finally, in image processing it is well known that regularization techniques lead in almost all cases to so-called systematic errors in the reconstructions. Hence, we present in *Section 4.7* an iterative refinement strategy in order to improve the reconstruction results obtained with the FB splitting approach using inverse scale space methods based on Bregman distance regularization. We shall also repeat the main aspects of the (inverse) scale space methods and clarify the connection to the Bregman distance iteration.

4.2 Statistical Modeling

In *Section 2.2*, we already discussed the basic concepts of the statistical problem formulation of image reconstruction, which we would like to consider here more exactly in case of Poisson distributed raw data. To recall, the problem of image reconstruction can be formulated as a solution of a linear and highly ill-conditioned operator equation

$$f = Ku , \tag{4.7}$$

where f are given raw data in form of a vector $f \in \mathbb{R}^N$ and $K : U(\Omega) \to \mathbb{R}^N$ is an operator that transforms the spatial distribution of the desired object into sampled signals f on the detectors. Since in practice the given raw data contain noise caused by the detector system with the noise being a random process, we consider the detected value f_i as a realization of a random variable F_i. Hence, in order to obtain a complete mathematical modeling of the image reconstruction problem, we need a model for the probability density of the noise.

In this work, we concentrate on a specific non-Gaussian noise, namely so-called *Poisson or photon counting noise*. This type of noise appears for example in positron emission tomography (PET) due to the radioactive decay of tracers and counting of photon coincidences [148, 167], or in optical nanoscopy due to photon counts by laser sampling of an object [122, 57]. For this kind of noise, every F_i corresponds to a Poisson random variable with an expectation value given by $(Ku)_i$, i.e.

$$F_i \text{ is Poisson distributed with parameter } (Ku)_i \, .$$

Hence, we obtain the following conditional probability density $p(f|u)$ of noisy data f given an image u with respect to the random variables F_i (cf. (2.4)),

$$p(f|u) \;=\; \prod_{i=1}^{N} \frac{(Ku)_i^{f_i}}{f_i\,!} \, e^{-(Ku)_i} \; . \tag{4.8}$$

Now, to determine an approximation to the unknown object u, the *Bayesian model* and the *maximum a-posteriori probability (MAP) estimation* via the negative log-likelihood function $-\log p(u|f)$ and a *Gibbs a-priori density* $p(u)$ of the form (2.8) lead to a minimization problem of the form (cf. (2.7) and (2.9)),

$$u_{MAP} \;\in\; \mathop{\arg\min}_{\substack{u \,\in\, W(\Omega) \\ u \,\geq\, 0 \;\; a.e.}} \left\{ \, -\log p(u|f) \, \right\} \;=\; \mathop{\arg\min}_{\substack{u \,\in\, W(\Omega) \\ u \,\geq\, 0 \;\; a.e.}} \left\{ \, -\log p(f|u) \;+\; \alpha \, J(u) \, \right\} \; , \tag{4.9}$$

where $J : W(\Omega) \;\rightarrow\; \mathbb{R}_{\geq 0}$ is an energy functional, usually convex, on a Banach space $W(\Omega) \;\subset\; U(\Omega)$. Note that an additional positivity constraint on the solution is taken into account, which is important in typical applications with Poisson distributed raw data, since in such cases the functions in general represent densities or intensity information.

Now, for a detailed specification of the likelihood function $p(u|f)$ in (4.9), we take the probability density $p(f|u)$ of the Poisson noise (4.8) and obtain the following negative log-likelihood function,

$$-\log p(u|f) \;=\; \sum_{i=1}^{N} \Big((Ku)_i \;-\; f_i \log (Ku)_i \Big) \;+\; \alpha \, J(u) \; , \tag{4.10}$$

in which the additive terms independent of u are neglected. At this point, in order to have a simpler basis for the construction and the later analysis of the methods, we will pass over from a discrete to a continuous representation of the data. For this sake, we assume that any element g in the discrete data space $\mathbb{R}^N$ can be interpreted as samples of a function in $V(\Sigma)$, which we denote for the sake of convenience with g again. Then, with the indicator function

$$\chi_{M_i}(x) \;=\; \begin{cases} 1 \, , & \text{if} \quad x \;\in\; M_i \; , \\ 0 \, , & \text{else} \; , \end{cases} \tag{4.11}$$

where M_i is the region of the i-th detector, we can interpret the discrete data as mean values,

$$f_i \;=\; \int_{M_i} f \; dx \;=\; \int_{\Sigma} \chi_{M_i} f \; dx \;.$$

Thus, with the negative log-likelihood function in (4.10), we can rewrite the MAP estimate in (4.9) as following continuous variational problem,

$$u_{MAP} \;\in\; \operatorname*{arg\,min}_{\substack{u \,\in\, W(\Omega) \\ u \,\geq\, 0 \;\; a.e.}} \left\{ \; \int_{\Sigma} (Ku \;-\; f \log Ku) \; d\mu \;+\; \alpha \, J(u) \; \right\} \;, \tag{4.12}$$

with $d\mu \;=\; \sum_{i=1}^{N} \chi_{M_i} \; d\lambda$, where λ denotes the Lebesque measure.

4.3 Reconstruction Method : EM Algorithm

In the literature there are in general two classes of reconstruction methods that are used. On the one hand analytical (direct) methods and on the other hand algebraic (iterative) strategies. A classical representative for a direct method is the Fourier-based filtered backprojection (FBP). Although FBP is well understood and can be computed efficiently, iterative type methods obtain more and more attention in practice. The major reason is the high noise level, i.e. low signal-to-noise ratio, and the special type of statistics found in measurements of various applications, such as positron emission tomography or fluorescence microscopy, which cannot be taken into account by direct methods. Thus, we shall use in this work the *expectation-maximization (EM) algorithm* [56], which is a popular iterative method to maximize the likelihood function $p(f|u)$ in problems with incomplete data.

In the previous section, we presented a statistical problem formulation for inverse problems with measured data drawn from Poisson statistics and could observe that the Bayesian MAP approach leads to a constrained minimization problem (4.12). In this section, we will give a review on a currently standard iterative reconstruction method for this problem, namely the so-called *expectation-maximization (EM) algorithm* [56, 123, 148], which finds numerous applications, for instance in medical imaging, microscopy or astronomy. In the two latter ones, the algorithm is also known as *Richardson-Lucy algorithm* [136, 111]. The EM algorithm is an iterative procedure to maximize the likelihood function $p(f|u)$ in problems with incomplete data and will form a basis for our algorithms introduced later. Here, we neglect the prior knowledge and assume that any object u has the same relevance, i.e. the Gibbs a-priori density $p(u)$ in (2.8) is constant. For simplicity, we normalize $p(u)$ such that $J \;\equiv\; 0$. Hence, the problem in

(4.12) reduces to the following variational problem with a positivity constraint,

$$\min_{\substack{u \in U(\Omega) \\ u \geq 0 \ a.e.}} \int_\Sigma (Ku \; - \; f \log Ku) \; d\mu \; . \tag{4.13}$$

To derive the algorithm, we consider the first order optimality condition of the constrained minimization problem (4.13). Formally, the Karush-Kuhn-Tucker (KKT) conditions [89, Thm. 2.1.4] provide the existence of a Lagrange multiplier $\lambda \geq 0$, such that the stationary points of the functional in (4.13) need to fulfill the equations,

$$\begin{aligned} 0 \; &= \; K^* \mathbf{1}_\Sigma \; - \; K^* \left(\frac{f}{Ku} \right) \; - \; \lambda \; , \\ 0 \; &= \; \lambda \, u \; , \end{aligned} \tag{4.14}$$

where K^* is the adjoint operator of K and $\mathbf{1}_\Sigma \in (V(\Sigma))^*$ is the characteristic function on Σ. Since the optimization problem (4.13) is convex (cf. the comments to equation (4.17)), every function u fulfilling the equations in (4.14) is a global minimum of (4.13). Now, multiplying the first equation in (4.14) by u, the Lagrange multiplier λ can be eliminated by the second equation and the subsequent division by $K^* \mathbf{1}_\Sigma$ leads to a simple iteration scheme

$$u_{k+1} \; = \; \frac{u_k}{K^* \mathbf{1}_\Sigma} \, K^* \left(\frac{f}{K u_k} \right) \; , \tag{4.15}$$

which preserves positivity if the operator K preserves positivity and the initialization u_0 is positive. This iteration scheme is the well-known EM or Richardson-Lucy algorithm. In [148], Shepp und Vardi showed that this iteration is actually a closed example of the EM algorithm proposed by Dempster, Laird and Rubin in [56], who presented the algorithm in a more general setup.

In the case of noise-free data $f = \bar{f}$, several convergence proofs of the EM algorithm to the maximum likelihood estimate (4.9), i.e. the solution of (4.13), can be found in literature [123, 135, 160, 91]. Besides, it is known that the speed of convergence of iteration (4.15) is slow. A further property of the iteration is a lack of smoothing, whereby the so-called "checkerboard effect" arises, i.e. single pixels become visible in the iterates.

For noisy data f, the convergence issue requires a distinction between the discrete and continuous setting. In the fully discrete case, i.e. if K is a matrix und u a vector, the existence of a minimizer can be guaranteed since the smallest singular value is bounded away from zero by a positive value. Hence, the iterates are bounded during the iteration and convergence is ensured. However, if K is a general continuous operator, the convergence is not only difficult to prove, but even a divergence of the EM algorithm is possible due to the underlying ill-posedness of the image reconstruction problems. This aspect can be seen as a lack of additional a-priori knowledge about the unknown u resulting from $J \equiv 0$ and is not really surprising according to the discussion in *Section 2.2*. Since, the variational problem (4.13) corresponds to the clasical approach of maximum likelihood (ML) estimation, cf. (2.5). More precisely, the ML approach remains ill-posed because it uses only information about the noise and not additional information about the desired object, for what reason the ill-posedness spreads from the operator equation (4.7) to the ML approach (4.13).

As described in [135], the EM iterates show the following typical behavior for ill-posed problems. Namely, the (metric) distance between the iterates and the exact solution decreases initially before it increases as the noise is amplified during the iteration process. This issue might be regulated by using appropriate stopping rules to obtain reasonable results. In [135], the authors have shown that certain stopping rules indeed allow stable approximations. Another possibility to improve reconstruction results considerably are regularization techniques, which we shall discuss in the following sections.

4.4 Reconstruction Method : Regularized EM Algorithm

4.4.1 FB-EM-REG Algorithm

The EM or Richardson-Lucy algorithm, that we discussed in *Section 4.3*, is currently the standard iterative reconstruction method for most inverse problems with incomplete Poisson data based on the linear equation (4.7). However, by setting $J \equiv 0$, no a-priori knowledge about the expected solution is taken into account, i.e. different images have the same a-priori probability. Especially in case of measurements with low signal-to-noise ratio (SNR), as they occur in positron emission tomography examinations with lower dose rate or tracers with short radioactive half life, the multiplicative fixed point iteration (4.15) delivers unsatisfactory and noisy results even with early termination. Therefore, we propose to integrate nonlinear variational methods into the reconstruction process to make an efficient use of a-priori information and to obtain improved results.

An attractive approach to improve the reconstructions from the EM algorithm is to use a regularization approach. In the classical EM algorithm, the negative log-likelihood functional (4.13) is minimized. In the regularized EM approach, we modify the functional by adding a weighted regularization term $J(u)$,

$$\min_{\substack{u \in W(\Omega) \\ u \geq 0 \ a.e.}} \int_\Sigma (Ku - f \log Ku) \, d\mu + \alpha \, J(u) \, , \qquad \alpha > 0 \, . \tag{4.16}$$

The specific choice of $J(u)$ is important for the way a-priori information about the expected solution is incorporated into the reconstruction process. In the recent past, singular regularization energies, especially those of ℓ^1 - or L^1 -type, have attracted strong attention in variational problems, but are difficult to handle due to their non differentiable nature.

In the following, we consider a *general Poisson framework* where the regularization energy J is any convex functional that can be also singular, i.e. in particular non differentiable in the classical sense. Subsequently, we discuss especially in *Chapters 6* and *7* a popular regularization technique in mathematical image processing, namely the *TV regularization* and its *nonlocal extension*. For the solution of (4.16), we propose here a *forward-backward splitting algorithm*, which can be realized by alternating classical EM steps with almost standard denoising steps as encountered in image denoising. For designing the proposed algorithm, we consider the first order optimality condition of (4.16). However, due to the assumptions on the regularization functional J, this variational problem can be non differentiable in the usual sense and other concepts are required to deal with such problems.

Fortunately, there are powerful methods from convex analysis which are available for general convex variational problems, see *Section 3.2*. As described in [134], the data fidelity term in (4.16) can be extended to a convex functional without changing the stationary points of the minimization problem, namely

$$\min_{\substack{u \in W(\Omega) \\ u \geq 0 \;\; a.e.}} \; D_{KL}(f, Ku) \; + \; \alpha \, J(u) \tag{4.17}$$

with the *Kullback-Leibler (KL) functional* D_{KL} defined by

$$D_{KL}(v, u) \; = \; \int_\Sigma \left(v \log \frac{v}{u} \; - \; v \; + \; u \right) \, d\mu \; .$$

For further comments and properties regarding the KL functional see *Remark 4.6.2* and *Section 4.6.1*. Now, due to the assumption of convexity of the regularization functional J, the minimization problem (4.17) is convex and we can use the methods from convex analysis, which we summarized in *Section 3.2*.

In order to compute the first order optimality condition of the minimization problem (4.17), we use a generalized derivative for convex functionals called *subdifferential* (see *Definition 3.2.6*), that we denote with ∂. Generally, the subdifferential is a multi-valued set and a single element is called *subgradient*. For the use of the subdifferential calculus on the functional in (4.17), note that due to the definition of functional $D_{KL}(f, K\,\cdot\,)$

on $U(\Omega)$, its subgradients are elements in $(U(\Omega))^*$, while the subgradients of J are elements in a larger function space $(W(\Omega))^*$, since $W(\Omega) \subset U(\Omega)$. However, we can extend J to a convex functional on $U(\Omega)$ (see *Remark 3.2.3, Item (3)*) by setting

$$J(u) \;:=\; \infty \qquad \text{if} \qquad u \;\in\; U(\Omega) \setminus W(\Omega) \;, \tag{4.18}$$

such that during the minimization process in (4.17) solutions from the smaller space $W(\Omega)$ will still be preferred. Hence, due to the continuity of the KL functional and the subdifferential calculus in *Lemma 3.2.13*, we obtain the following identity

$$\partial \Big(D_{KL}(f, Ku) \;+\; \alpha\, J(u) \Big) \;=\; \partial_u\, D_{KL}(f, Ku) \;+\; \alpha\, \partial J(u)$$

in $(U(\Omega))^* \subset (W(\Omega))^*$ for any $f \in V(\Sigma)$. Finally, since the subdifferentials ∂_u of the KL functional D_{KL} are singletons, the first optimality condition of (4.17) for a positive solution u is given by

$$K^* \mathbf{1}_\Sigma \;-\; K^* \left(\frac{f}{Ku} \right) \;+\; \alpha\, p \;=\; 0 \;, \qquad p \;\in\; \partial J(u) \;, \tag{4.19}$$

where K^* denotes the adjoint operator of K and $\mathbf{1}_\Sigma$ the characteristic function on Σ.

The simplest iteration scheme to compute a solution of the variational problem (4.16) is a gradient-type method. However, such an approach is not robust in case of singular regularization energies, such that severe step size restrictions are needed, since the subgradient p of J is treated explicitely. A better idea is to use an iteration scheme which evaluates the nonlocal term (including the operator K) in (4.19) at the last iterate u_k and the local term (including the subgradient of J) at the new iterate u_{k+1}, i.e.

$$\mathbf{1}_\Omega \;-\; \frac{1}{K^* \mathbf{1}_\Sigma}\, K^* \left(\frac{f}{Ku_k} \right) \;+\; \alpha\, \frac{1}{K^* \mathbf{1}_\Sigma}\, p_{k+1} \;=\; 0 \;, \qquad p_{k+1} \;\in\; \partial J(u_{k+1}) \;, \tag{4.20}$$

with an additional division of (4.19) by $K^* \mathbf{1}_\Sigma$. However, in this iteration the new iterate u_{k+1} appears only as a point of reference for the subdifferential of J. That is a considerable drawback, since u_{k+1} cannot be determined from (4.20) due to the missing of one-to-one relation between subgradients (dual variables) and the primal variable u. In addition, such iteration schemes cannot guarantee preservation of positivity. Hence, we obtain an improved method if we approximate the constant term $\mathbf{1}_\Omega$ in (4.20) by $\frac{u_{k+1}}{u_k}$, where such an approximation seems of course reasonable in case of convergence of iterates, so that u_{k+1} appears directly, i.e.

$$u_{k+1} \;-\; \frac{u_k}{K^* \mathbf{1}_\Sigma}\, K^* \left(\frac{f}{Ku_k} \right) \;+\; \alpha\, \frac{u_k}{K^* \mathbf{1}_\Sigma}\, p_{k+1} \;=\; 0 \;, \qquad p_{k+1} \;\in\; \partial J(u_{k+1}) \;. \tag{4.21}$$

In order to verify that this iteration scheme actually preserves positivity, we proceed in an analogous way to the EM algorithm in *Section 4.3*. Namely, due to the nonnegativity constraint in (4.16), the Karush-Kuhn-Tucker (KKT) conditions [89, Thm. 2.1.4] provide the existence of a Lagrange multiplier $\lambda \geq 0$, such that the stationary points of the functional in (4.17) need to fulfill

$$\begin{aligned} 0 \;&\in\; K^* \mathbf{1}_\Sigma \;-\; K^* \left(\frac{f}{Ku} \right) \;+\; \alpha\, \partial J(u) \;-\; \lambda \;, \\ 0 \;&=\; \lambda\, u \;. \end{aligned} \tag{4.22}$$

By multiplying the first equation in (4.22) by u, the Lagrange multiplier λ can be eliminated by the second equation and the subsequent division by $K^* \mathbf{1}_\Sigma$ leads to a fixed point equation of the form

$$u \;-\; \frac{u}{K^* \mathbf{1}_\Sigma} K^* \left(\frac{f}{Ku} \right) \;+\; \alpha\, \frac{u}{K^* \mathbf{1}_\Sigma}\, p \;=\; 0 \;, \qquad p \;\in\; \partial J(u) \;, \tag{4.23}$$

which is just the optimality condition (4.19) multiplied by u, i.e. this multiplication corresponds to the nonnegativity constraint in (4.16). Furthemore, the iteration (4.21) is just a semi-implicit approach to the fixed point equation (4.23). In *Section 4.6.4*, we will then prove that the iteration method (4.21) actually preserves positivity if the operator K preserves positivity and the initialization u_0 is positive.

It is remarkable that the second term in the iteration (4.21) is just a single EM step from (4.15). Therefore, the method (4.21) solving the variational problem (4.16) and (4.17), respectively, can be realized as a nested two step iteration of the form

$$\begin{aligned} u_{k+\frac{1}{2}} \;&=\; \frac{u_k}{K^* \mathbf{1}_\Sigma} K^* \left(\frac{f}{K u_k} \right) \;, && \textit{(EM step)} \\ u_{k+1} \;&=\; u_{k+\frac{1}{2}} \;-\; \alpha\, \frac{u_k}{K^* \mathbf{1}_\Sigma}\, p_{k+1} \;, \qquad p_{k+1} \;\in\; \partial J(u_{k+1}) \;. && \textit{(REG step)} \end{aligned} \tag{4.24}$$

Thus, we alternate an EM reconstruction step with a suitable regularization (REG) step to compute a solution of (4.16), respectively (4.17). In *Section 4.4.3*, we will especially see that this iteration scheme can be interpreted as a *modified forward-backward (FB) splitting strategy* and denote it thus as *FB-EM-REG algorithm*. The complex second half step from $u_{k+\frac{1}{2}}$ to u_{k+1} in (4.24) can be realized by solving the convex variational problem

$$u_{k+1} \;\in\; \operatorname*{arg\,min}_{u \,\in\, W(\Omega)} \left\{ \; \frac{1}{2} \int_\Omega \frac{K^* \mathbf{1}_\Sigma \left(u \;-\; u_{k+\frac{1}{2}} \right)^2}{u_k} \;+\; \alpha\, J(u) \; \right\} \;. \tag{4.25}$$

Inspecting the first order optimality condition of (4.25) confirms the equivalence of this minimization problem with the regularization step in (4.24) due to convexity. The solution approach of (4.25) is now dependent on the corresponding choice of the regularization functional J. It is however interesting to note that the problem (4.25) is just a modified version of the frequently used variational model in image denoising problems based on Gaussian assumed noise in the data, i.e. usual L^2 data deviation, with the difference of a weight $\frac{K^* \mathbf{1}_\Sigma}{u_k}$ in the data fidelity term. Hence, dependent on J, one can use standard numerical schemes known for these regularization strategies with some simple modifications, cf. e.g. the numerical realizations of (4.25) in the case of (nonlocal) total variation regularization in *Sections 6.3.3*, *6.3.4*, *7.4.1* and *7.4.2*.

4.4.2 Damped FB-EM-REG Algorithm

The alternating structure of the proposed iteration (4.24) has a particular advantage that we might control the interaction between reconstruction and regularization via a simple adaption of the second half step. A possibility is a damped version of the regularization step, namely

$$u_{k+1} \;=\; (1 - \omega_k)\, u_k \;+\; \omega_k\, u_{k+\frac{1}{2}} \;-\; \omega_k\, \alpha\, \frac{u_k}{K^* \mathbf{1}_\Sigma}\, p_{k+1} \;, \qquad \omega_k \;\in\; (0, 1] \;, \tag{4.26}$$

which relates the current EM iterate $u_{k+\frac{1}{2}}$ to the previous regularized iterate u_k in a way of a convex combination by using a damping parameter ω_k. The damped half step (4.26) can also be realized in an analogous way to (4.25), namely by minimizing the variational problem

$$u_{k+1} \in \operatorname*{arg\,min}_{u \,\in\, W(\Omega)} \left\{ \frac{1}{2} \int_\Omega \frac{K^* \mathbf{1}_\Sigma \left(u \;-\; \left(\omega_k\, u_{k+\frac{1}{2}} \;+\; (1 - \omega_k)\, u_k \right) \right)^2}{u_k} \;+\; \omega_k\, \alpha\, J(u) \right\} . \tag{4.27}$$

This aspect is required to attain a monotone descent of the objective functional in (4.16) and (4.17) during the minimization process (see (4.65)). Finally, for $\omega_k = 1$, the iteration (4.26) simplifies to the original regularization step in (4.24). For a small ω_k, the iterations stay close to regularized solutions u_k. For an adequate choice of $\omega_k \in (0, 1]$, we can prove in *Theorem 4.6.14* the convergence of the proposed two step iteration with respect to the damped regularization step (4.26), additionally with an explicit bound on ω_k in the special case of denoising problems (see *Corollary 4.6.15*), i.e. K being the identity operator.

4.4.3 Forward-Backward Splitting

In the previous *Sections 4.4.1* and *4.4.2*, we introduced the FB-EM-REG reconstruction method as a two step algorithm (4.24) with an additional damping modification (4.26). This two step strategy can be interpreted as an operator splitting algorithm. In convex optimization such splitting methods arise in the context of decomposition problems. Recently, some works in literature picked up these splitting ideas, providing efficient algorithms in image processing, see e.g. [146, 27, 51, 52]. Most of the papers dealing with convex splitting strategies go back to early works of Douglas and Rachford [61] and other authors in [107] and [157].

The optimality condition (4.19) of our underlying variational problem (4.16) can be interpreted as a decomposition problem $(C = A + B)$, regarding the convex Kullback-Leibler functional and the convex regularization term, respectively their subdifferentials. Hence, we consider the stationary equation

$$0 \;\in\; C(u) \;:=\; \underbrace{K^*\mathbf{1}_\Sigma \;-\; K^*\left(\frac{f}{Ku}\right)}_{=:\;A(u)} \;+\; \underbrace{\alpha\,\partial J(u)}_{=:\;B(u)}\;,$$

with two maximal monotone operators A and B. Hence, the damped two step iteration (4.24) with the modified regularization step (4.26) and $\omega_k \in (0,1]$ reads as follows,

$$\begin{aligned} \frac{K^*\mathbf{1}_\Sigma\left(u_{k+\frac{1}{2}} \;-\; u_k\right)}{u_k} \;+\; A\,u_k \qquad &= \; 0 \\ \frac{K^*\mathbf{1}_\Sigma\left(u_{k+1} \;-\; \omega_k\,u_{k+\frac{1}{2}} \;-\; (1-\omega_k)\,u_k\right)}{u_k} \;+\; \omega_k\,B\,u_{k+1} \;&=\; 0 \end{aligned} \tag{4.28}$$

and can easily be formulated as a *forward-backward splitting algorithm* of the form

$$\begin{aligned} \frac{K^*\mathbf{1}_\Sigma\left(\tilde{u}_{k+\frac{1}{2}} \;-\; u_k\right)}{\omega_k\,u_k} \;+\; A\,u_k \quad &= \; 0 \qquad\qquad \text{(forward step on } A\text{)} \\ \frac{K^*\mathbf{1}_\Sigma\left(u_{k+1} \;-\; \tilde{u}_{k+\frac{1}{2}}\right)}{\omega_k\,u_k} \;+\; B\,u_{k+1} \;&=\; 0 \qquad\qquad \text{(backward step on } B\text{)} \end{aligned}$$

with

$$\tilde{u}_{k+\frac{1}{2}} \;=\; \omega_k\,u_{k+\frac{1}{2}} \;+\; (1-\omega_k)\,u_k\;.$$

Compared to the undamped iteration strategy (4.24), in case of damped iteration scheme (4.26), the artificial time step size is not only given by u_k, but can also be controlled via the additional damping parameter ω_k. In a more compact form, the whole iteration can be formulated as

$$u_{k+1} \;=\; \left(I \;+\; \frac{\omega_k\, u_k}{K^*\mathbf{1}_\Sigma}\, B \right)^{-1} \left(I \;-\; \frac{\omega_k\, u_k}{K^*\mathbf{1}_\Sigma}\, A \right) u_k \tag{4.29}$$

$$\;=\; (L_k \;+\; B)^{-1}\, (L_k \;-\; A)\, u_k$$

with a multiplication operator L_k defined by $\frac{K^*\mathbf{1}_\Sigma}{\omega_k\, u_k}$.

The forward-backward splitting approach for maximal monotone operators has been suggested independently by Lions and Mercier [107] and Passty [128]. In our case, we will see in *Theorem 4.6.14* that the key to proving the convergence of the FB-EM-REG splitting algorithm lies in the incorporation of damping parameters ω_k .

Finally, notice that there are alternatives to the forward-backward splitting strategy such as the Peaceman-Rachford or Douglas-Rachford splitting schemes, see e.g. [107] or [146], which are indeed unconditionally stable. However, these approaches have the numerical drawback that also an additional backward step on A has to be performed, which would mean an inversion of the operator K^*K , cf. e.g. [72] or [147]. There, the authors use an augmented Lagrangian approach and a split Bregman method in case of total variation regularization, which are equivalent [69] and correspond to the Douglas-Rachford splitting strategy applied to the dual problem of (4.16) [146]. However, these methods require an inversion of the operator K^*K and hence are only efficient if K^*K is diagonalizable and can be inverted efficiently, for example in the case of the convolution operator K using the fast Fourier transform or the discrete cosine transform.

4.4.4 Stopping Rules

After the description of the damped FB-EM-REG algorithm in *Sections 4.4.1* and *4.4.2*, it is useful to define appropriate stopping rules in order to guarantee the accuracy of the proposed algorithm. In addition to a maximum number of iterations, the error in the optimality condition (4.19) can be taken as a basic stopping criterion in a suitable norm. For this purpose, we define a weighted norm deduced from a weighted scalar product,

$$\langle u, v \rangle_{\mathsf{w}} \;:=\; \int_\Omega u\, v\, \mathsf{w}\; d\lambda \qquad \text{and} \qquad \|u\|_{2,\mathsf{w}} \;:=\; \sqrt{\langle u, u \rangle_{\mathsf{w}}}\;, \tag{4.30}$$

with a positive weight function w and the standard Lebesque measure λ on Ω . Hence, the error in the optimality condition can be measured reasonably in the following norm,

$$opt_{k+1} \;:=\; \left\| \, K^*\mathbf{1}_\Sigma \;-\; K^* \left(\frac{f}{K u_{k+1}} \right) \;+\; \alpha\, p_{k+1} \, \right\|_{2,u_{k+1}}^2 . \tag{4.31}$$

Furthermore, due to the fact that we use a damped two step iteration, we are not only interested in the improvement of the whole optimality condition (4.19), but also in the convergence of the sequence of primal functions (u_k) and the sequence of subgradients (p_k) with $p_k \in \partial J(u_k)$. Hence, in order to establish appropriate stopping rules for these sequences, we consider the damped regularization step (4.26) with the EM reconstruction step in (4.24),

$$u_{k+1} \;-\; \omega_k \, \frac{u_k}{K^* \mathbf{1}_\Sigma} \, K^* \left(\frac{f}{K u_k} \right) \;-\; (1 - \omega_k) \, u_k \;+\; \omega_k \, \alpha \, \frac{u_k}{K^* \mathbf{1}_\Sigma} \, p_{k+1} \;=\; 0 \; .$$

By combining this iteration scheme with the optimality condition (4.23) evaluated at u_k, which must be fulfilled in the case of convergence, we obtain an optimality statement for the sequences (p_k) and (u_k),

$$\alpha \, (p_{k+1} \;-\; p_k) \;+\; \frac{K^* \mathbf{1}_\Sigma \, (u_{k+1} \;-\; u_k)}{\omega_k \, u_k} \;=\; 0 \; . \tag{4.32}$$

With the aid of the weighted norm (4.30), we now have additional stopping criteria for the FB-EM-REG algorithm, which guarantee the accuracy of the primal functions (u_k) and the subgradients (p_k), namely

$$\begin{aligned} u_{opt_{k+1}} \; &:= \; \left\| \frac{K^* \mathbf{1}_\Sigma \, (u_{k+1} \;-\; u_k)}{\omega_k \, u_k} \right\|^2_{2, u_{k+1}} \; , \\ p_{opt_{k+1}} \; &:= \; \| \, \alpha \, (p_{k+1} \;-\; p_k) \, \|^2_{2, u_{k+1}} \; . \end{aligned} \tag{4.33}$$

We finally mention that the stopping criteria (4.31) and (4.33) are well defined, since we can prove that each iterate u_k of the damped FB-EM-REG splitting strategy is strictly positive, see *Lemma 4.6.12*.

4.4.5 Pseudocode and Some Remarks

If we summarize the observations in the *Sections 4.4.1*, *4.4.2* and *4.4.4*, we can now use *Algorithm 4.1* to solve the regularized Poisson likelihood estimation problem (4.16).

Remark. Selecting a reasonable regularization parameter α in our model is a common problem. In the case of additive Gaussian noise, there exist several works in literature dealing with this problem in the case of the total variation regularization, see e.g. [106, 155, 162]. Finding an "optimal" regularization parameter is, in general, more complicated for non-Gaussian noise models. Nevertheless, there exist a few works in literature addressing this issue, see e.g. [16] and the references therein or [22].

Algorithm 4.1 *(Damped) FB-EM-REG Algorithm*

1. ***Parameters:*** f, $\alpha > 0$, $\omega \in (0,1]$, $maxEMIts \in \mathbb{N}$, $tol > 0$
2. ***Initialization:*** $k = 0$, $u_0 := c > 0$
3. ***Iteration:***

   while $((k < maxEMIts)$ **and**

   $(opt_k \geq tol$ **or** $u_{opt_k} \geq tol$ **or** $p_{opt_k} \geq tol))$ **do** $\triangleright$ (4.31), (4.33)

   i) Compute $u_{k+\frac{1}{2}}$ via EM step in (4.24).

   ii) Set $\omega_k = \omega$.

   iii) Compute u_{k+1} via convex variational problem (4.27).

   iv) $k \leftarrow k+1$

   end while
4. ***Return*** u_k

4.5 Image Denoising

In this section, we are interested in the problem of image denoising, which is a widely studied problem in applied mathematics and has a wide application in fields ranging from computer vision to medical imaging. For an overview of the subject and various methods, we refer e.g. to [6, 49] and references therein. However, most works deal with additive white Gaussian noise, i.e. given an original image $\bar{u}$, it is assumed that the observed image f is corrupted by some additive white Gaussian noise η. The denoising problem is then to recover $\bar{u}$ from the data $f = \bar{u} + \eta$. In the literature, there are many effective methods to tackle this problem, like wavelet approaches [60], [44], stochastic approaches [75] and variational approaches [138]. In the case of Gaussian noise, the variational methods can be written as minimization problems for an energy functional of the form [49, 143],

$$\min_{u \in W(\Omega)} \frac{1}{2} \int_\Omega (u - f)^2 \, d\mu + \alpha J(u) ,$$

in order to obtain a denoised version u of a given image f, which is certainly only an approximation to the original $\bar{u}$. One of the most popular approaches is the Rudin-Osher-Fatemi (ROF) model (6.6) [138], which used total variation as regularization functional J (cf. *Section 6.1*) and realized results preserving edges. However, the Poisson noise in the images has not yet obtained wide attention in the literature, some methods are to find in [104, 46] and references therein.

In [104], Le, Chartrand and Asaki proposed a total variation based variational model to denoise an image corrupted by Poisson noise, which can be generalized to the following minimization problem with an arbitrary convex regularization functional J,

$$\min_{\substack{u \in W(\Omega) \\ u \geq 0 \ a.e.}} \int_\Omega (u - f \log u) \, d\mu + \alpha J(u) . \tag{4.34}$$

The optimality condition of this problem is given via the Karush-Kuhn-Tucker (KKT) conditions [89, Thm. 2.1.4], similar to (4.22) with identity operator K,

$$u \left(1 - \frac{f}{u} + \alpha p \right) = 0 , \qquad p \in \partial J(u) . \tag{4.35}$$

To solve the minimization problem (4.34) with the total variation regularization J, the authors in [104] suggest to use a gradient descent algorithm based on the Euler-Lagrange equation. However, such an approach requires always an approximation of TV by differentiable functionals (4.6) and needs a severe step size restriction. Here, we propose an approach which is based on the FB-EM-REG strategy introduced in the previous *Sections 4.4.1* and *4.4.2*.

4.5.1 Exact Denoising Model

To establish the problem formulation (4.34), the authors in [104] use Bayes' theorem and the maximum a-posteriori probability estimation via the negative log-likelihood function such that it is not surprising that the problem (4.34) coincides in the case of identity operator K with the Poisson likelihood reconstruction model (4.16). Hence, to propose a numerical iteration scheme for the Poisson denoising problem, we can use the FB-EM-REG splitting strategy (4.24) with the damped modification (4.26), which simply results in the following iteration scheme (note that the EM reconstruction step vanishes in the denoising case, i.e. it holds $u_{k+\frac{1}{2}} = f$),

$$u_{k+1} = (1 - \omega_k) \, u_k + \omega_k \, f - \omega_k \, \alpha \, u_k \, p_{k+1} , \quad p_{k+1} \in \partial J(u_{k+1}) , \tag{4.36}$$

with $\omega_k \in (0, 1]$. Analogous to the FB-EM-REG algorithm, this iteration step can be realized by solving the convex variational problem of the form (cf. (4.25) and (4.27)),

$$u_{k+1} \in \arg\min_{u \in W(\Omega)} \left\{ \frac{1}{2} \int_\Omega \frac{\big(u - \big(\omega_k \, f + (1 - \omega_k) \, u_k \big) \big)^2}{u_k} + \omega_k \, \alpha \, J(u) \right\} . \tag{4.37}$$

Note in particular that in the undamped case (i.e. $\omega_k = 1$), the algorithm (4.36) is just a semi-implicit iteration scheme with respect to the optimality condition (4.35) and thus actually computes a denoised image in the Poisson case.

4.5.2 Approximated Denoising Model

As we have seen, the iteration scheme (4.36) solves the Poisson denoising problem (4.34) by a sequence of modified L^2 variational models of the form (4.37). The advantage is that, in this way, we obtain a maximum a-posteriori probability (MAP) estimate, but at the price of high computational effort, which is comparable to the incorporated FB-EM-REG reconstruction strategy (4.24).

Hence, to reduce the computational complexity in the case of Poisson image denoising, we introduce an approximation of the denoising problem, based on the second order Taylor approximation of the data fidelity term in (4.34). To justify the usage of the Taylor approximation, note that as usual for variational methods, we expect that the solution u fulfills a certain regularity or smoothness (caused by the regularization functional J) but yet approximates moderately well the given image f (caused by the data fidelity term). Hence, we can expect a certain closeness of u to f such that it is formally possible to perform the Taylor expansion of the data fidelity term in (4.34) at f. For this purpose, we define an auxiliary functional G_f as the data fidelity term in (4.34),

$$G_f(v) \;:=\; \int_\Omega (v \,-\, f \log v) \; d\mu \;,$$

and obtain the following linearization

$$G_f(u) \,=\, G_f(f) \,+\, G'_f(f;\, u \,-\, f) \,+\, \frac{1}{2}\, G''_f(f;\, u \,-\, f,\, u \,-\, f) \,+\, \mathcal{O}\left((u \,-\, f)^3\right) \;, \tag{4.38}$$

using the terminology of directional derivatives in *Definition 3.2.14*. Now, to compute these derivatives, we use the strategy proposed in *Remark 3.2.15*, *Item (2)*, and consider first the directional derivative G'_f. To do this, let $\phi_{w_1} \,:\, \mathbb{R}_{\geq 0} \,\to\, \mathbb{R} \,\cup\, \{\pm\infty\}$ be a function defined by $\phi_{w_1}(t) \,:=\, G_f(v \,+\, t\, w_1)$ for any $w_1 \,\in\, U(\Omega)$, then the directional derivative $G'_f(v;\, w_1)$ of G_f at v in the direction w_1 is given by

$$\begin{aligned} G'_f(v;\, w_1) \;=\; \phi'_{w_1}(t)\big|_{t\,=\,0} \;&=\; \int_\Omega \frac{\partial}{\partial t} \big((v \,+\, t\, w_1) \,-\, f \log(v \,+\, t\, w_1) \big) \; d\mu \,\bigg|_{t\,=\,0} \\ &=\; \int_\Omega \left(1 \,-\, \frac{f}{v} \right) \, w_1 \; d\mu \;. \end{aligned} \tag{4.39}$$

Next, let ϕ_{w_2} be defined by $\phi_{w_2}(t) \,:=\, G'_f(v \,+\, t\, w_2;\, w_1)$ for any $w_2 \,\in\, U(\Omega)$, then the second directional derivative $G''_f(v;\, w_1,\, w_2)$ of G_f at v in directions w_1 and w_2

$$\begin{aligned} G''_f(v;\, w_1,\, w_2) \;=\; \phi'_{w_2}(t)\big|_{t\,=\,0} \;&=\; \int_\Omega \frac{\partial}{\partial t} \left(\left(1 \,-\, \frac{f}{v \,+\, t\, w_2} \right) w_1 \right) \, d\mu \,\bigg|_{t\,=\,0} \\ &=\; \int_\Omega \left(\frac{f\, w_2\, w_1}{v^2} \right) \; d\mu \;. \end{aligned} \tag{4.40}$$

Now, combining the results from (4.39) and (4.40), the Taylor expansion (4.38) delivers the following linearization of the data fidelity term in (4.34),

$$\int_\Omega (u \,-\, f \log u)\; d\mu \;=\; \int_\Omega (f \,-\, f \log f)\; d\mu \;+\; \frac{1}{2} \int_\Omega \frac{(u \,-\, f)^2}{f}\; d\mu \;+\; \mathcal{O}\left((u \,-\, f)^3\right) .$$

Thus, neglecting terms of higher order and additive terms independent of u, we can approximate the denoising problem (4.34) by

$$\min_{\substack{u \,\in\, W(\Omega) \\ u \,\geq\, 0 \;\; a.e.}} \; \frac{1}{2} \int_\Omega \frac{(u \,-\, f)^2}{f}\; d\mu \;+\; \alpha\, J(u)\ . \tag{4.41}$$

The efficiency of this linearization in comparison to the scheme in (4.36), where we have to solve a sequence of variational models (4.37), is that we can compute a denoised image by solving just a single modified L^2 variational problem (4.41).

4.6 Analysis

In this section, we carry out a mathematical analysis of the regularized Poisson based variational model (4.16) of the form

$$\min_{\substack{u \,\in\, W(\Omega) \\ u \,\geq\, 0 \;\; a.e.}} \; \int_\Sigma (Ku \,-\, f \log Ku)\; d\mu \;+\; \alpha\, J(u)\ , \qquad \alpha \;>\; 0\ . \tag{4.42}$$

In the recent past, some authors proposed already several theoretical frameworks for the regularized Poisson likelihood estimation problem (4.42), using different regularization energies $J(u)$. For instance, Bardsley proposed in [15] a theoretical framework for the Tikhonov regularization, regularization by diffusion, and total variation regularization functional, referring to earlier joint works with Laobeul [17], [18] and Luttman [19]. Moreover, Resmerita and Anderssen studied in [134] the Kullback-Leibler functional as regularization entropy $J(u)$ in order to approximate solutions of inverse problems with Poisson distributed data. In this thesis, we concentrate on the following aspects in the mathematical analysis of (4.42):

- In contrast to the mentioned works of Bardsley [15] and Resmerita [134], we propose here a theoretical framework for an arbitrary convex regularization functional $J(u)$, in particular for those which can be also non differentiable in the classical sense, as the total variation or general ℓ^1- or L^1-type functionals. For such energies, we prove the well-posedness, i.e. the existence, uniqueness, and stability of a solution, with respect to the variational regularization methods in (4.42).

- Moreover, we study the damped FB-EM-REG algorithm proposed in *Sections 4.4.1* and *4.4.2* with respect to preservation of positivity of a solution and a stable convergence behavior of this iteration scheme.

4.6.1 Kullback-Leibler Functional

In order to simplify the analysis of the regularized Poisson likelihood estimation problem (4.42), we give in this section the definition of the *Kullback-Leibler (KL) functional* and recall from [134] a collection of basic results about the KL functional, which will be necessary in the following analysis. For further information on the KL functional, we refer to [63] and [134].

Definition 4.6.1 (*Kullback-Leibler Functional*). *The Kullback-Leibler (KL) functional is a function* $D_{KL} : L^1(\Sigma) \times L^1(\Sigma) \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$ *given by*

$$D_{KL}(\varphi, \psi) \;=\; \int_\Sigma \left(\varphi \log \left(\frac{\varphi}{\psi} \right) \;-\; \varphi \;+\; \psi \right) d\nu \qquad \textit{for all} \qquad \varphi,\, \psi \;\geq\; 0 \;\; \textit{a.e.}\;, \tag{4.43}$$

where ν *is a measure. Note that, using the convention* $0 \log 0 \;=\; 0$, *the integrand in* (4.43) *is nonnegative and vanishes if and only if* $\varphi \;=\; \psi$ *.*

Remark 4.6.2. In the literature, there are further notations for the KL functional, like cross-entropy, information for discrimination or Kullback's I-divergence (cf. e.g. [54], [63] or [134]). The functional (4.43) generalizes the well known *Kullback-Leibler entropy*,

$$E_{KL}(\varphi, \psi) \;\;=\;\; \int_\Sigma \varphi \log \left(\frac{\varphi}{\psi} \right) \, d\nu \;,$$

for functions which are not necessarily probability densities. In the definition above, you get the extension by adding (linear) terms, which are chosen so that (4.43) is a Bregman distance or divergence (see *Definition 4.7.1*) with respect to the *Boltzman-Shannon entropy* [134, Sect. 3.1] given by

$$E_{BS}(\vartheta) \;\;=\;\; \begin{cases} \int_\Sigma \vartheta \log \vartheta \; d\nu \;, & \text{if} \quad \vartheta \;\geq\; 0 \text{ a.e.} \quad \text{and} \quad \vartheta \log \vartheta \;\in\; L^1(\Sigma) \;, \\ +\infty \;, & \text{else} \;. \end{cases}$$

Lemma 4.6.3 (*Properties of KL Functional*). *Let* $A \;:\; U(\Omega) \;\to\; V(\Sigma)$ *be a linear operator between locally convex spaces* $U(\Omega)$ *and* $V(\Sigma)$ *(see* Definition 3.1.17*), such that* $U(\Omega)$ *and* $V(\Sigma)$ *are associated with topologies* τ_U *and* τ_V *. Additionally, we assume* *that the operator A is sequentially continuous (see* Definition 3.1.5*) with respect to the topologies τ_U and τ_V and that $V(\Sigma)$ is continuously embedded in $L^1(\Sigma)$ (see* Definition 3.1.23*). Moreover, we suppose that the operator A preserves positivity, i.e. it satisfies $Au \geq 0$ a.e. for any $u \geq 0$ a.e. . Then, the following statements hold:*

(i) The function $(\varphi, \psi) \mapsto D_{KL}(\varphi, \psi)$ is convex and thus, due to the linearity of the operator A, the function $(\varphi, u) \mapsto D_{KL}(\varphi, Au)$ is also convex.

(ii) For any fixed nonnegative $\varphi \in L^1(\Sigma)$, the function $u \mapsto D_{KL}(\varphi, Au)$ is lower semicontinuous with respect to the topology τ_U.

(iii) For any nonnegative functions φ and ψ in $L^1(\Sigma)$, one has

$$\|\varphi - \psi\|^2_{L^1(\Sigma)} \leq \left(\frac{2}{3} \|\varphi\|_{L^1(\Sigma)} + \frac{4}{3} \|\psi\|_{L^1(\Sigma)} \right) D_{KL}(\varphi, \psi) .$$

Proof. ***(i)*** See [134, Lemma 3.4]. ***(iii)*** See [134, Lemma 3.3] and [24, Lemma 2.2]. ***(ii)*** The proof is almost identical with the one in [134, Lemma 3.4 (iii)], only a few modifications are incorporated. Fix a nonnegative function $\varphi \in L^1(\Sigma)$. Let (u_n) be a sequence in the domain of the function $v \mapsto D_{KL}(\varphi, Av)$, which converges in the topology τ_U to some $u \in \{ v \in U(\Omega) : v \geq 0 \ a.e. \}$. Then, due to the sequential continuity of the operator A with respect to the topologies τ_U and τ_V, as well as the continuous embedding of $V(\Sigma)$ in $L^1(\Sigma)$, we also obtain the convergence of the sequence (Au_n) to Au in the norm topology on $L^1(\Sigma)$, as well as the pointwise convergence almost everywhere on Σ. Thus, the sequence $\big(\varphi \log(\varphi \,/\, Au_n) - \varphi + Au_n \big)$ converges almost everywhere to $\varphi \log(\varphi \,/\, Au) - \varphi + Au$ and by applying Fatou's Lemma, we obtain

$$\int_\Sigma \left(\varphi \log \left(\frac{\varphi}{Au} \right) - \varphi + Au \right) d\nu \leq \liminf_{n \to \infty} \int_\Sigma \left(\varphi \log \left(\frac{\varphi}{Au_n} \right) - \varphi + Au_n \right) d\nu .$$

Now, this inequality means that the function $v \mapsto D_{KL}(\varphi, Av)$ is lower semicontinuous with respect to the topology τ_U. □

Corollary 4.6.4. *If (φ_n) and (ψ_n) are bounded sequences in $L^1(\Sigma)$, then*

$$\lim_{n \to \infty} D_{KL}(\varphi_n, \psi_n) = 0 \qquad \Rightarrow \qquad \lim_{n \to \infty} \|\varphi_n - \psi_n\|_{L^1(\Sigma)} = 0 .$$

Proof. The statement follows directly from *Lemma 4.6.3 (iii).* □

4.6.2 Assumptions

In this section, we introduce the necessary foundations for the following analysis of the regularized Poisson likelihood estimation problem (4.42), in particular we state the required assumptions on the forward operator K and the regularization functional J.

However, to simplify the analysis of (4.42), we begin here with an extension of the data fidelity term in (4.42), such that we can take advantage of already known results and properties of the Kullback-Leibler (KL) functional introduced in *Section 4.6.1.* We see that the data fidelity term in (4.42) has almost the form of the KL functional presented in *Definition 4.6.1.* Hence, if we add the expression $f \log f - f$ to the data fidelity term, which is independent from the desired function u, the stationary points of the minimization problem are not influenced (if they exist) and (4.42) is equivalent to

$$\min_{\substack{u \in W(\Omega) \\ u \geq 0 \ a.e.}} D_{KL}(f, Ku) \ + \ \alpha \, J(u) \ , \qquad \alpha \ > \ 0 \ , \tag{4.44}$$

where D_{KL} is the Kullback-Leibler functional as in *Definition 4.6.1.* Subsequently, we can also extend the regularization functional J. To recall, we assumed in *Sections 2.2* and *4.4* that the regularization functional

$$J \, : \, W(\Omega) \ \to \ \mathbb{R}_{\geq 0}$$

is convex on a Banach space $W(\Omega) \subset U(\Omega)$, where $U(\Omega)$ is a Banach space itself and denotes the domain of the forward operator K. Hence, to have a simpler basis for the analysis of (4.44), we can use a helpful extension of J to the whole space $U(\Omega)$ by

$$J(u) \ := \ +\infty \qquad \text{if} \qquad u \ \in \ U(\Omega) \setminus W(\Omega) \ . \tag{4.45}$$

Despite this extension, notice in the following that the extended functional J is convex on $U(\Omega)$ (see *Remark 3.2.3, Item (3)*) and, furthermore, that solutions from the smaller space $W(\Omega)$ are preferred during the minimization of J. Consequently, we can also extend the admissible solution set of the minimization problem (4.44) from $W(\Omega)$ to $U(\Omega)$ and denote for the following analysis the objective functional with $F(u)$,

$$\min_{\substack{u \in U(\Omega) \\ u \geq 0 \ a.e.}} F(u) \ := \ D_{KL}(f, Ku) \ + \ \alpha \, J(u) \ , \qquad \alpha \ > \ 0 \ . \tag{4.46}$$

Next, for the following assumptions, we recall the basic characteristics of the forward operator K. As introduced in *Section 2.1,* we consider in this work the operator K as a semi-discrete operator based on

$$\bar{K} \, : \, U(\Omega) \ \to \ V(\Sigma) \ ,$$

where $U(\Omega)$ and $V(\Sigma)$ are Banach spaces of functions on bounded and compact sets Ω and Σ. The operator $\bar{K}$ was claimed to be linear and compact. The difference between both operators K and $\bar{K}$ is that K transforms the desired functions from $U(\Omega)$ to the discrete data space $\mathbb{R}^N$. Nevertheless, to be able to present a unified theory with respect to the continuous problem formulation (4.42), we passed over from a discrete to a continuous representation of the raw data using a point measure μ in *Section 4.2*. There, we assumed that any element in the discrete data space $\mathbb{R}^N$ can be interpreted as samples of a function in $V_\mu(\Sigma)$, where $V_\mu(\Sigma)$ denotes a Banach space of functions with respect to the measure μ. Note that in the case of a fully continuous formulation of the forward operator, the measure μ has to be set to the Lebesque measure and we have $V_\mu(\Sigma) = V(\Sigma)$. Hence, we consider $K : U(\Omega) \to V_\mu(\Sigma)$ below.

Now, based on the observations in this section, we make the following assumptions.

Assumption 4.6.5.

(i) *The Banach spaces $U(\Omega)$ and $V_\mu(\Sigma)$ are associated with topologies τ_U and τ_V, where τ_U can also be weaker than the strong norm topology.*

(ii) *The Banach space $V_\mu(\Sigma)$ is continuously embedded in $L^1_\mu(\Sigma)$ (Definition 3.1.23), where $L^1_\mu(\Sigma)$ is the Lebesque space $L^1(\Sigma)$ with respect to the measure μ.*

(iii) *The operator $K : U(\Omega) \to V_\mu(\Sigma)$ is linear and sequentially continuous with respect to the topologies τ_U and τ_V (Definition 3.1.5).*

(iv) *The operator K preserves positivity, i.e. it satisfies $Ku \geq 0$ a.e. for any $u \geq 0$ a.e. and the equality is fulfilled if and only if $u = 0$.*

(v) *If $u \in U(\Omega)$ satisfies $c_1 \leq u \leq c_2$ a.e. for some positive constants c_1, c_2, then there exist $c_3, c_4 > 0$ such that $c_3 \leq Ku \leq c_4$ a.e. on Σ.*

(vi) *The functional $J : U(\Omega) \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$ is convex, lower semicontinuous with respect to the topology τ_U (Definition 3.2.4) and can also be singular, i.e. it is not differentiable in the classical sense.*

(vii) *For any nonnegative function $f \in V_\mu(\Sigma)$,*

$$\mathcal{D}(F) \;:=\; \mathcal{D}(D_{KL}(f, K\,\cdot\,)) \;\cap\; \mathcal{D}(J) \;\neq\; \emptyset$$

holds, where $\mathcal{D}$ denotes the effective domain of a functional (see Definition 3.2.2). In particular, this implies that the functional J is proper (see Definition 3.2.2).

(viii) For every $a > 0$, *the sub-level sets* $\mathcal{S}_J(a)$ *of the functional* J, *defined by*

$$\mathcal{S}_J(a) \;:=\; \{\; u \;\in\; U(\Omega) \;:\; J(u) \;\leq\; a \;\} \;,$$

are sequentially precompact with respect to the topology τ_U *(Definition 3.1.10).*

Remark 4.6.6.

(1) The *Assumptions 4.6.5 (ii)* and *(iv)* are mainly necessary due to the fact that we use a positivity constraint on the solution u in the minimization problem (4.46) and that the domain of the KL functional as data fidelity term in (4.44) is strictly included in the set $\{\, \varphi \in L^1(\Sigma) \;:\; \varphi \geq 0 \; a.e. \,\} \times \{\, \psi \in L^1(\Sigma) \;:\; \psi \geq 0 \; a.e. \,\}$, see [134, Lemma 3.3].

(2) The *Assumption 4.6.5 (v)* is more of technical nature, but nevertheless it is not significantly restrictive in most practical situations, since there are many classes of linear ill-posed problems for which the required condition is fulfilled. An example are integral equations of the first kind, which have smooth, bounded and positive kernels. Such integral equations appear in numerous fields of applications, e.g. in geophysics and potential theory or in deconvolution problems such as fluorescence microscopy or astronomy. Another interesting example of operators, which fulfill the *Assumption 4.6.5 (v)*, is the X-ray transform which assigns the integral values along all straight lines to a function. This transform coincides in two dimensions with the well-known Radon transform and is strongly applied in medical imaging. The *Assumption 4.6.5 (v)* is fulfilled in this example, if the length of the lines is bounded and bounded away from zero, a condition which is obviously satisfied in practice.

(3) The *Assumptions 4.6.5 (vi)-(viii)* are standard assumptions on the regularization functional J, which are necessary to ensure the existence of a regularized solution and to obtain the stability and convergence of the regularization method.

(4) Due to the definition of the effective domain (*Definition 3.2.2*) and the extension property (4.45), it is easy to see that *Assumption 4.6.5 (vii)* implies $\mathcal{D}(F) \subset W(\Omega)$. Of course, this means that solutions from the smaller space $W(\Omega) \subset U(\Omega)$ are preferred during the minimization of the functional F in (4.46).

(5) On order to verify the validity of the *Assumption 4.6.5 (viii)* for a specific regularization energy J, it suffices in general to show that the functional J on a Banach space $U(\Omega)$ is $W(\Omega)$-coercive (see *Definition 4.6.7*), where the Banach space $W(\Omega)$ is compactly embedded in $U(\Omega)$ (see *Definition 3.1.25*).

Definition 4.6.7 (*Coercivity*). *Let $U(\Omega)$ and $W(\Omega)$ be Banach spaces such that $W(\Omega) \subset U(\Omega)$. A functional G defined on $U(\Omega)$ is called $W(\Omega)$-coercive (cf. [89, Def. IV.3.2.6]), if the sub-level sets of G are bounded in the $\|\cdot\|_{W(\Omega)}$ norm, i.e. for all $a \in \mathbb{R}_{\geq 0}$ the set $\{ u \in U(\Omega) : G(u) \leq a \}$ is uniformly bounded in the $W(\Omega)$-norm; or equivalently*

$$G(u) \to +\infty \qquad \textit{whenever} \qquad \|u\|_{W(\Omega)} \to +\infty .$$

4.6.3 Well-Posedness of Minimization Problem

In this section, we verify the existence, uniqueness, and stability of the regularized Poisson likelihood estimation problem consisting in the minimization of (4.46).

Theorem 4.6.8 (*Existence of Minimizers*). *Let $U(\Omega)$, $V_\mu(\Sigma)$, K, J, and F satisfy Assumption 4.6.5. Assume that $\alpha > 0$ and $f \in V_\mu(\Sigma)$ is nonnegative. Then, the functional F defined in* (4.46) *has a minimizer.*

Proof. We use the direct method of the calculus of variations, see e.g. [6, Sect. 2.1.2]: Since $\mathcal{D}(F) \neq \emptyset$, there exists at least one $v \in U(\Omega)$ such that $F(v) < \infty$. Thus, let $(u_n) \subset \mathcal{D}(F)$, $u_n \geq 0$ a.e., be a minimizing sequence of the functional F, i.e.

$$\lim_{n \to \infty} F(u_n) = \inf_{u \in \mathcal{D}(F)} F(u) =: F_{min} < \infty . \tag{4.47}$$

Then, we obtain that for every $\epsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$,

$$a := F_{min} + \epsilon \geq F(u_n) \overset{(4.46)}{=} D_{KL}(f, K u_n) + \alpha J(u_n) \geq J(u_n) , \tag{4.48}$$

due to the positivity of the KL functional D_{KL} and $\alpha > 0$. Thus $(u_n)_{n \geq n_0} \subset \mathcal{S}_J(a)$ and it follows from *Assumption 4.6.5 (viii)* that (u_n) has a τ_U-convergent subsequence (u_{n_j}), which converges to some $\tilde{u} \in U(\Omega)$. Because J is lower semicontinuous with respect to the topology τ_U, we have

$$J(\tilde{u}) \overset{Def.\ 3.2.4}{\leq} \liminf_{j \to \infty} J(u_{n_j}) \overset{(4.48)}{\leq} a$$

and with it that $\tilde{u} \in \mathcal{S}_J(a)$. Simultaneously, caused by *Lemma 4.6.3 (ii)*, also the objective functional F in (4.46) is lower semicontinuous with respect to the topology τ_U and implies the inequality

$$F(\tilde{u}) \overset{Def.\ 3.2.4}{\leq} \liminf_{j \to \infty} F(u_{n_j}) \overset{(4.47)}{=} F_{min} ,$$

which means that $\tilde{u}$ is a minimizer of F. □

Next, we consider the uniqueness of the minimizers, for which it suffices to verify the strict convexity of the objective functional F. For this purpose, it is straight-forward to see that the negative logarithm function is strictly convex and consequently also the function $u \mapsto D_{KL}(f, Ku)$, if $f \in V_\mu(\Sigma)$ fulfills $\inf_\Sigma f > 0$ and the operator K is injective, i.e. the null space of K is trivial since K is linear (cf. *Assumption 4.6.5 (iii)*). Therefore, we can immediately conclude the following result.

Theorem 4.6.9 (*Uniqueness of Minimizers*). *Let $U(\Omega)$, $V_\mu(\Sigma)$, K, J, and F satisfy Assumption 4.6.5. Assume that K is an injective operator and $f \in V_\mu(\Sigma)$ fulfills $\inf_\Sigma f > 0$. Then, the function $u \mapsto D_{KL}(f, Ku)$ and also the functional F from* (4.46) *is strictly convex. In particular, the minimizer of F is unique.*

After existence and uniqueness of minimizers, we show in the following the stability of the regularized Poisson estimation problem (4.46) with respect to a certain kind of data perturbations. In *Section 4.2*, we already described that the given measurements in practice are discrete and can be interpreted in our framework as averages of a function $f \in V(\Sigma)$. The open question is certainly the suitable choice of the function f. Moreover, the physically limited discrete construction of the detector leads to a natural loss of information, because not all signals can be acquired. Consequently, a stability result is required guaranteeing that the regularized approximations converge to a solution u, if e.g. the approximated data converge to a preferably smooth function f. Because the measurements are still a realization of Poisson distributed random variables, it is natural to assess the convergence in terms of the KL functional, as shown in (4.49).

Theorem 4.6.10 (*Stability with Respect to Perturbations in Measurements*). *Let $U(\Omega)$, $V_\mu(\Sigma)$, K, J, and F satisfy Assumption 4.6.5. Fix $\alpha > 0$ and assume that the functions $f_n \in V_\mu(\Sigma)$, $n \in \mathbb{N}$, are nonnegative approximations of a data function $f \in V_\mu(\Sigma)$ in the form that*

$$\lim_{n \to \infty} D_{KL}(f_n, f) \;=\; 0 \; . \tag{4.49}$$

Moreover, let

$$u_n \;\in\; \operatorname*{arg\,min}_{\substack{v \,\in\, U(\Omega) \\ v \,\geq\, 0 \;\; a.e.}} \left\{ F_n(v) \;:=\; D_{KL}(f_n, Kv) \;+\; \alpha\, J(v) \right\} , \qquad n \;\in\; \mathbb{N} \; , \tag{4.50}$$

and u a solution of the regularized problem (4.46) *corresponding to the data function f. Additionally, we assume that $\log f$ and $\log Ku$ belong to the function space $L^\infty_\mu(\Sigma)$ and there exist positive constants $c_1, \, \dots \, , c_4$ such that*

$$0 \;<\; c_1 \;\leq\; f \;\leq\; c_2 \qquad \textit{and} \qquad 0 \;<\; c_3 \;\leq\; Ku \;\leq\; c_4 \qquad \textit{a.e. on } \Sigma \; . \tag{4.51}$$

Moreover, we suppose that the sequence (f_n) *is uniformly bounded in the* $V_\mu(\Sigma)$ *-norm, i.e. it exists a positive constant* c_5 *such that*

$$\|f_n\|_{V_\mu(\Sigma)} \;\le\; c_5 \;, \qquad \forall n \;\in\; \mathbb{N} \;. \tag{4.52}$$

Then, the problem (4.46) *is stable with respect to perturbations in the data, i.e. the sequence* (u_n) *has a* τ_U *-convergent subsequence and every* τ_U *-convergent subsequence converges to a minimizer of the functional* F *in* (4.46).

Proof. For the existence of a τ_U -convergent subsequence of (u_n), we will use the precompactness property of the sub-level sets $\mathcal{S}_J$ from *Assumption 4.6.5 (viii)*. Hence, we have to show the uniform boundedness of the sequence $\big(J(u_n)\big)$. Let $\alpha > 0$ be a fixed regularization parameter. For any $n \in \mathbb{N}$, the positivity of the KL functional and the definition of u_n as a minimizer of the objective functional F_n in (4.50) implies that

$$J(u_n) \;\le\; \underbrace{D_{KL}(f_n, Ku_n) \;+\; \alpha\, J(u_n)}_{=\;F_n(u_n)} \;\le\; \underbrace{D_{KL}(f_n, Ku) \;+\; \alpha\, J(u)}_{=\;F_n(u)} \;. \tag{4.53}$$

Hence, the sequence $\big(J(u_n)\big)$ is bounded, if the sequence $\big(D_{KL}(f_n, Ku)\big)$ on the right-hand side of (4.53) is bounded. To show this, we use condition (4.52) and obtain the uniform boundedness of sequence (f_n) in the $L^1_\mu(\Sigma)$ -norm, due to continuous embedding of $V_\mu(\Sigma)$ in *Assumption 4.6.5 (ii)*. Hence, condition (4.49) and the result in *Corollary 4.6.4* yield the strong convergence of (f_n) to f in $L^1_\mu(\Sigma)$, i.e. we have

$$\lim_{n\,\to\,\infty} \|f \;-\; f_n\|_{L^1_\mu(\Sigma)} \;=\; 0 \;. \tag{4.54}$$

Thus, the condition (4.51) implies together with the inequality

$$\begin{aligned}
&\Big| D_{KL}(f_n, Ku) \;-\; D_{KL}(f, Ku) \;-\; D_{KL}(f_n, f) \Big| \\
&\qquad\qquad = \; \left| \int_\Sigma (\log Ku \;-\; \log f)\,(f \;-\; f_n)\, d\mu \right| \\
&\qquad\qquad \le \; \underbrace{\|\log Ku \;-\; \log f\|_{L^\infty_\mu(\Sigma)}}_{<\,\infty} \; \underbrace{\|f \;-\; f_n\|_{L^1_\mu(\Sigma)}}_{\overset{(4.54)}{\to}\;0} \;,
\end{aligned}$$

the following convergence,

$$\lim_{n\,\to\,\infty} D_{KL}(f_n, Ku) \;=\; D_{KL}(f, Ku) \;. \tag{4.55}$$

Because u is a minimizer of the regularized problem (4.46) corresponding to the data function f, the expressions $D_{KL}(f, Ku)$ and $J(u)$ are bounded and thus also the sequence $\big(D_{KL}(f_n, Ku)\big)$ is bounded, since convergent to $D_{KL}(f, Ku)$. This fact delivers, together with the boundedness of $J(u)$ and the property (4.53), the uniform boundedness of the sequence $\big(J(u_n)\big)$.

The uniform boundedness of the sequence $\big(J(u_n)\big)$ means that there exists $a \in \mathbb{R}_{\geq 0}$ such that $\big(J(u_n)\big)$ is contained in the sub-level set $\mathcal{S}_J(a)$. Thus, the precompactness *Assumption 4.6.5 (viii)* ensures the existence of a τ_U-convergent subsequence (u_{n_j}), which converges to some $\tilde{u} \in U(\Omega)$. Actually, the function $\tilde{u}$ lies in $\mathcal{S}_J(a)$, since J is lower semicontinuous with respect to the topology τ_U and therefore $\mathcal{S}_J(a)$ is τ_U-closed (see *Definition 3.2.4*).

Now, let (u_{n_j}) be an arbitrary subsequence of (u_n), which converges to some $\tilde{u} \in U(\Omega)$ with respect to the topology τ_U. Due to the sequential continuity of the operator K with respect to the topologies τ_U and τ_V, as well as the continuous embedding of $V_\mu(\Sigma)$ in $L^1_\mu(\Sigma)$ (see *Assumptions 4.6.5 (ii)-(iii)*), we have also the convergence of (Ku_{n_j}) to $K\tilde{u}$ in the strong norm topology on $L^1_\mu(\Sigma)$, as well as the pointwise convergence almost everywhere on Σ. Additionally, a similar behavior holds also for the sequence (f_n), which converges strongly to f in $L^1_\mu(\Sigma)$ (4.54). Thus, since the functions f_n and u_n are nonnegative for all $n \in \mathbb{N}$ and K is an operator that preserves positivity (see *Assumption 4.6.5 (iv)*), we can apply Fatou's Lemma to the sequence $\big(f_{n_j} \log\big(f_{n_j} / Ku_{n_j}\big) - f_{n_j} + Ku_{n_j} \big)$ and obtain

$$D_{KL}(f, K\tilde{u}) \;\leq\; \liminf_{j \to \infty} D_{KL}(f_{n_j}, Ku_{n_j}) \; . \tag{4.56}$$

Due to the lower semicontinuity of the regularization energy J (see *Assumption 4.6.5 (vi)*) and due to (4.53), (4.55) and (4.56), we obtain now the following inequality,

$$\begin{aligned}
D_{KL}(f, K\tilde{u}) \;+\; \alpha\, J(\tilde{u}) \quad &\overset{(4.56)}{\leq} \quad \liminf_{j \to \infty} D_{KL}(f_{n_j}, Ku_{n_j}) \;+\; \alpha \liminf_{j \to \infty} J(u_{n_j}) \\
&\leq \quad \liminf_{j \to \infty} \Big(D_{KL}(f_{n_j}, Ku_{n_j}) \;+\; \alpha\, J(u_{n_j}) \Big) \\
&\leq \quad \limsup_{j \to \infty} \Big(D_{KL}(f_{n_j}, Ku_{n_j}) \;+\; \alpha\, J(u_{n_j}) \Big) \\
&\overset{(4.53)}{\leq} \quad \limsup_{j \to \infty} \Big(D_{KL}(f_{n_j}, Ku) \;+\; \alpha\, J(u) \Big) \\
&\overset{(4.55)}{=} \quad D_{KL}(f, Ku) \;+\; \alpha\, J(u) \; ,
\end{aligned}$$

which means that $\tilde{u}$ is a minimizer of the functional F in (4.46). $\square$

Remark.

- For the proof of stability, condition (4.51) is required, which assumes that the functions $\log f$ and $\log Ku$ belong to $L^\infty_\mu(\Sigma)$, where u is a regularized solution of the minimization problem (4.46). In the case of the data function f, this assumption is not significantly restrictive in most practical situations. The boundedness from above is fulfilled naturally due to the finite acquisition time of the data. The almost everywhere boundedness on Σ away from zero is reasonable, when a sufficient amount of measurements has been collected. In addition, in most practical applications a certain level of background noise is present, which causes the positivity of the data. In the case of the function Ku, condition (4.51) is not simple to justify and requires a more precise analysis of the variational problem (4.46). Due to *Assumption 4.6.5 (v)*, it suffices to prove that u is bounded and bounded away from zero. For instance, the authors in [134] show that this condition on u is available, if we use the KL functional $D_{KL}(\,\cdot\,, u^*)$ as regularization energy J, where u^* denotes the a-priori estimation of the solution and satifies the boundedness condition from above and away from zero. Roughly speaking, this is possible because, during the minimization, the linear part of the KL functional in u tries to keep the function bounded and the log part tries to push the function away from zero. However, every other choice of the regularization energy J needs a particular study of this property depending on the specific form of this functional. Nevertheless, note that in *Section 4.6.4* we can show at least that the iterate sequence (u_k) of the FB-EM-REG splitting algorithm (4.24) has the boundedness and the boundedness away from zero property, assuming that the data function f belongs to $L^\infty_\mu(\Sigma)$ with $\inf_\Sigma f > 0$ and the initialization function u_0 is strictly positive. For this reason, we think that condition (4.51) is an acceptable assumption.

- As in *Theorem 4.6.10*, it is also possible to consider perturbations of the operator K. The proof is similar to the one above and only slight modifications are necessary. However, several assumptions on the perturbed operators K_n are needed, like the boundedness of operators for each $n \in \mathbb{N}$ and pointwise convergence to K. Unfortunately, it is also essential that the operators K_n fulfill the assumption (4.51), i.e. that $K_n u$ is bounded and bounded away from zero for any $n \in \mathbb{N}$, where u is a solution of the minimization problem (4.46). Therefore, this condition is severely restrictive for the possible perturbations of the operator K.

- We finally mention that some stability estimates for the regularized Poisson likelihood estimation problem (4.46) have been also derived in [15] and [112] in case of Tikhonov, diffusion, total variation, and L^1 regularization functionals, but in a different setting. There, the assumptions on the possible data perturbations are more restrictive (convergence in the supremum norm), while the assumptions on the operator perturbations are relaxed.

4.6.4 Positivity Preservation of FB-EM-REG Algorithm

In this section, we consider a particular property of the iteration sequence (u_k) obtained by the FB-EM-REG splitting approach (4.24) and its damped modification (4.26), namely the positivity preservation of this iteration scheme. Given a strictly positive $u_k \in U(\Omega)$ for some $k \geq 0$, it is straight-forward to see that the result $u_{k+\frac{1}{2}}$ of the reconstruction half step is well defined and strictly positive due to the form of the EM iteration step in (4.24), if the data function f is strictly positive and the operator K fulfills the positivity preservation property in *Assumption 4.6.5 (v)*. Consequently, an existence and uniqueness proof for the regularization half step (4.25) and its damped variant (4.27), analogous to the classical results for the unweighted L^2 model depending on the special choice of the regularization functional J, delivers also the existence of $u_{k+1} \in U(\Omega)$. Now, in order to show inductively the well-definedness of the complete iteration sequence (u_k), it remains to verify that u_{k+1} is again strictly positive.

However, note that if any u_k is negative during the iteration, the objective functional in the regularization half step (4.25) and its damped modification (4.27) is in general not convex anymore. Moreover, the minimization problem becomes a maximization problem and the existence and uniqueness of u_{k+1} cannot be guaranteed. Thus, the non negativity of a solution in the regularization half step, and with it also the positivity of the whole iteration sequence, is strongly desired, in particular since in typical applications the functions represent densities or intensity information. The latter aspect is considered explicitly by using the positivity constraint in the Poisson based log-likelihood optimization problem (4.16).

Now, to clarify the positivity preservation of the (damped) FB-EM-REG iteration scheme, we present a maximum principle for the following weighted L^2 regularization problem,

$$\min_{u \,\in\, U(\Omega)} S(u) \;:=\; \frac{1}{2} \int_\Omega \frac{(u \,-\, q)^2}{h} \, dx \;+\; \beta \, J(u) \;, \qquad \beta \;>\; 0 \;, \tag{4.57}$$

which represents the more general form of the regularization half step (4.25) and its damped modification (4.27) in the forward-backward splitting strategy.

Lemma 4.6.11 (*Maximum Principle for the Weighted L^2 Regularization Problem*). *Let $\tilde{u} \in U(\Omega)$ be a minimizer of the variational problem* (4.57)*, where the function q belongs to $L^\infty(\Omega)$ with $\inf_\Omega q > 0$ and the weighting function h is strictly positive. Additionally, we assume that for any positive constants a and b with $a < b$, the regularization functional J fulfills*

$$J(v) \;\leq\; J(\tilde{u}) \qquad \textit{for} \qquad v \;=\; \min\{\max\{\tilde{u},\, a\},\, b\} \;. \tag{4.58}$$

Then, the following maximum principle holds

$$0 \;<\; \inf_\Omega q \;\le\; \inf_\Omega \tilde{u} \;\le\; \sup_\Omega \tilde{u} \;\le\; \sup_\Omega q\;. \tag{4.59}$$

Proof. Let $\tilde{u}$ be a minimizer of the functional S defined in (4.57). For the proof of the maximum principle, we show that there exists a function v with

$$0 \;<\; \inf_\Omega q \;\le\; \inf_\Omega v \;\le\; \sup_\Omega v \;\le\; \sup_\Omega q \tag{4.60}$$

and

$$S(v) \;\le\; S(\tilde{u})\;. \tag{4.61}$$

Then, the desired boundedness property (4.59) follows directly from the strict convexity of the functional S in (4.57), i.e. from the uniqueness of the solution.

Now, we define the function v as a version of $\tilde{u}$ cut off at $\inf_\Omega q$ and $\sup_\Omega q$, i.e.

$$v \;:=\; \min\{\max\{\tilde{u}, \inf q\}, \sup q\}\;.$$

With this definition, property (4.60) is directly guaranteed and we observe that, due to assumption (4.58), also $J(v) \le J(\tilde{u})$ holds. To show (4.61), it remains now to estimate the weighted L^2 fidelity term in (4.57). To do this, we use the set

$$M \;:=\; \{\, x \,\in\, \Omega \;:\; v(x) \;=\; \tilde{u}(x)\,\} \;\subseteq\; \Omega$$

and see that the data fidelity terms with respect to v and $\tilde{u}$ agree on M, due to the definition of the function v. In case of $x \in \Omega \setminus M$, we distinguish two cases:

Case 1: If $\tilde{u}(x) \ge \sup q$ then $v(x) \;=\; \sup q$ and

$$\begin{aligned} 0 \;\le\; v(x) \;-\; q(x) \;=\; \sup q \;-\; q(x) \;\le\; \tilde{u}(x) \;-\; q(x) \\ \Rightarrow\quad (v(x) \;-\; q(x))^2 \;\le\; (\tilde{u}(x) \;-\; q(x))^2\;. \end{aligned}$$

Case 2: If $\tilde{u}(x) \le \inf q$ then $v(x) \;=\; \inf q$ and

$$\begin{aligned} 0 \;\le\; -\,v(x) \;+\; q(x) \;=\; -\,\inf q \;+\; q(x) \;\le\; -\,\tilde{u}(x) \;+\; q(x) \\ \Rightarrow\quad (v(x) \;-\; q(x))^2 \;\le\; (\tilde{u}(x) \;-\; q(x))^2\;. \end{aligned}$$

Finally, we obtain

$$(v - q)^2 \leq (\tilde{u} - q)^2 , \qquad \forall x \in \Omega ,$$

and property (4.61) is fulfilled due to the strict positivity of the weighting function h and assumption (4.58). □

Lemma 4.6.12 (*Positivity of (damped) FB-EM-REG Algorithm*). *Let (ω_k) be a given sequence of damping parameters with $\omega_k \in (0,1]$ for all $k \geq 0$ and the initialization function u_0 be strictly positive. Additionally, we assume that the data function f lies in $L^\infty_\mu(\Sigma)$ with $\inf_\Sigma f > 0$, the operator K satisfies the positivity preservation property in Assumption 4.6.5 (v) and that the regularization functional J fulfills the condition* (4.58) *for the maximum principle above. Then, each half step of the (damped) FB-EM-REG splitting method and therewith also the solution is strictly positive.*

Proof. Since $u_0 > 0$, $f > 0$ and the operator K and therewith also the adjoint operator K^* does not affect the strict positivity, the first EM reconstruction step $u_{\frac{1}{2}}$ in (4.24) is strictly positive. Because the regularization step in (4.24) can be realized via the convex variational problem (4.25), the maximum principle in *Lemma 4.6.11* using $q := u_{\frac{1}{2}} > 0$ and $h := \frac{u_0}{K^* \mathbf{1}_\Sigma} > 0$ yields $u_1 > 0$. With the same argument, we also obtain $u_1 > 0$, if we take the damped regularization step (4.26) via the variational problem (4.27), using the maximum principle with $q := \omega_0 \, u_{\frac{1}{2}} + (1 - \omega_0) \, u_0 > 0$ for $\omega_0 \in (0,1]$ and $h := \frac{u_0}{K^* \mathbf{1}_\Sigma} > 0$. Inductively, the strict positivity of the whole nested iteration sequence (u_k) and with it the strict positivity of the solution is obtained by the same arguments using *Lemma 4.6.11*. □

Finally, we consider also the positivity preservation of the Poisson denoising strategy, which we proposed in *Sections 4.5.1* and *4.5.2*. Although the denoising iteration in *Section 4.5.1* is a special case of the damped FB-EM-REG algorithm with identity operator K, we study its properties here explicitly, because it will later simplify the convergence criteria of the Poisson denoising method. In the following lemma, we study directly the damped form of the Poisson denoising scheme (4.37) with damping parameters $\omega_k \in (0,1]$. However, note that we obtain the original denoising strategy, if we choose $\omega_k = 1$ for all $k \geq 0$.

Lemma 4.6.13 (*Maximum Principle and Positivity of the Poisson Denoising Scheme*). *Let (ω_k) be a sequence of damping parameters with $\omega_k \in (0,1]$ for all $k \geq 0$, the data function f lie in $L^\infty_\mu(\Omega)$ with $\inf_\Omega f > 0$ and the initialization function u_0 fulfill*

$$0 < \inf_\Omega f \leq \inf_\Omega u_0 \leq \sup_\Omega u_0 \leq \sup_\Omega f . \tag{4.62}$$

Moreover, let (u_k) *be a sequence of iterates generated by the damped Poisson denoising scheme* (4.37) *and the regularization functional* J *satisfy the condition* (4.58) *for the maximum principle in Lemma 4.6.11. Then, the following maximum principle holds,*

$$0 \;<\; \inf_\Omega f \;\leq\; \inf_\Omega u_k \;\leq\; \sup_\Omega u_k \;\leq\; \sup_\Omega f \;, \qquad \forall k \;\geq\; 0 \;. \tag{4.63}$$

Simultaneously, this result guarantees also that each step of the damped Poisson denoising method (4.36) *and with it also the solution is strictly positive.*

Proof. We prove the assertion by induction. For $k = 0$, the condition (4.63) is fulfilled due to (4.62). For a general $k \geq 0$, *Lemma 4.6.11* offers a maximum principle for the Poisson denoising model (4.37) using $q := \omega_k f + (1 - \omega_k) u_k$ and $h := u_k$, i.e. we have

$$\begin{aligned} 0 \;<\; \inf_\Omega \{ \omega_k f \;+\; (1-\omega_k)\, u_k \} \;\leq\; \inf_\Omega u_{k+1} & \\ \leq\; \sup_\Omega u_{k+1} \;\leq\; \sup_\Omega \{ \omega_k f \;+\; (1-\omega_k)\, u_k \} \;. & \end{aligned} \tag{4.64}$$

Due to the fact that $\omega_k \in (0,1]$ for all $k \geq 0$ and the inequalities

$$\inf_\Omega \{ \omega_k f \;+\; (1-\omega_k)\, u_k \} \;\geq\; \omega_k \inf_\Omega f \;+\; (1-\omega_k) \inf_\Omega u_k$$

and

$$\sup_\Omega \{ \omega_k f \;+\; (1-\omega_k)\, u_k \} \;\leq\; \omega_k \sup_\Omega f \;+\; (1-\omega_k) \sup_\Omega u_k \;,$$

we obtain from (4.64) and the induction hypothesis the desired maximum principle (4.63). □

Remark.

- The assumption (4.62) on the initialization function u_0 is in general fulfilled, since u_0 will be usually chosen as a positive and constant function or as the given noisy image f itself.
- In *Section 4.5.2*, we also proposed an approximation of the denoising problem (4.34) in the form that we reduced the sequence of the variational models (4.37) to a single modified L^2 variational problem (4.41). Therefore, this approximated denoising model also preserves positivity according to the maximum principle in *Lemma 4.6.11* with $q = h = f$, if the given noisy image $f \in L^\infty_\mu(\Omega)$ is strictly positive.

4.6.5 Convergence of Damped FB-EM-REG Algorithm

In *Section 4.4.3*, we interpreted the (damped) FB-EM-REG reconstruction method, proposed in *Sections 4.4.1* and *4.4.2*, as a forward-backward operator splitting algorithm. In the past, several works in convex analysis have been proposed dealing with the convergence of such splitting strategies for solving decomposition problems, see e.g. Tseng [157] and Gabay [74]. For the proposed algorithm (4.29),

$$u_{k+1} \;=\; \left(I \;+\; \frac{\omega_k \, u_k}{K^* \mathbf{1}_\Sigma} \, B \right)^{-1} \left(I \;-\; \frac{\omega_k \, u_k}{K^* \mathbf{1}_\Sigma} \, A \right) u_k \;,$$

Gabay provided in [74] a proof of a weak convergence of the forward-backward splitting approach under the assumption of a fixed damping parameter ω strictly less than twice the modulus of A^{-1}. On the other hand, Tseng gave later in [157] a convergence proof, where in our case, the damping values $\frac{\omega_k \, u_k}{K^* \mathbf{1}_\Sigma}$ need to be bounded in the following way,

$$\epsilon \;\leq\; \frac{\omega_k \, u_k}{K^* \mathbf{1}_\Sigma} \;\leq\; 4m \,-\, \epsilon \;, \qquad \epsilon \;\in\; (0, 2m] \;,$$

where the Kullback-Leibler data fidelity functional needs to be strictly convex with modulus m. Unfortunately, the results above cannot be used in our case, since we cannot verify the modulus assumption on the data fidelity and in particular, we cannot provide the upper bounds for the iterates u_k.

For these reasons, we prove the necessity of a damping strategy manually, in order to guarantee a monotone descent of the objective functional F in (4.46) with respect to the iterates u_k of the FB-EM-REG algorithm. In the following theorem, we will establish the convergence of the damped FB-EM-REG splitting algorithm under appropriate assumptions on the damping parameters ω_k.

Theorem 4.6.14 (*Convergence of Damped FB-EM-REG Algorithm*). *Let* $U(\Omega)$, $V_\mu(\Sigma)$, K , J , *and* F *satisfy Assumption 4.6.5. Moreover, let* (u_k) *be a sequence of iterates obtained by the damped FB-EM-REG algorithm* (4.28), *i.e. with the iteration scheme* (4.24) *and the damped regularization step* (4.26). *Regarding this sequence of iterates, we make additional assumptions:*

- *The data function* f *lies in* $L^\infty_\mu(\Sigma)$ *and fulfills* $\inf_\Sigma f \,>\, 0$.
- *The regularization functional* J *fulfills the condition* (4.58) *for the maximum principle in Lemma 4.6.11.*

Now, if there exists a sequence of corresponding damping parameters (ω_k) , $\omega_k \in (0,1]$, *satisfying the inequality*

$$\omega_k \;\leq\; \frac{\displaystyle\int_\Omega \frac{K^*\mathbf{1}_\Sigma\,(u_{k+1} - u_k)^2}{u_k}\,d\lambda}{\displaystyle\sup_{v\,\in\,[u_k,u_{k+1}]} \frac{1}{2}\int_\Sigma \frac{f\,(Ku_{k+1} - Ku_k)^2}{(Kv)^2}\,d\mu}\,(1-\epsilon)\;, \qquad \epsilon \in (0,1)\;, \tag{4.65}$$

then the objective functional F *defined in* (4.46) *is decreasing during the iteration. If, in addition, we assume that the following assumptions are fulfilled,*

- *the function* $K^*\mathbf{1}_\Sigma$ *, the damping parameters and the iterates are bounded away from zero by positive constants* c_1 *,* c_2 *and* c_3 *such that for all* $k \geq 0$ *,*

$$0 < c_1 \leq K^*\mathbf{1}_\Sigma\;, \qquad 0 < c_2 \leq \omega_k\;, \qquad 0 < c_3 \leq u_k\;, \tag{4.66}$$

- *the regularization functional* J *is homogeneous of degree one, i.e. it satisfies* $J(\lambda\,u) = \lambda\,J(u)$ *for all* $\lambda > 0$ *, and there exists a constant* $c_4 > 0$ *such that*

$$\sup_{\|v\|_{W(\Omega)} \,\leq\, 1} J(v) \;\leq\; c_4\;, \tag{4.67}$$

- *the functional* F *defined in* (4.46) *is* $U(\Omega)$ *-coercive (see Definition 4.6.7) and the Banach space* $U(\Omega)$ *is continuously embedded in* $L^1(\Omega)$ *(see Definition 3.1.23),*

- *there exists a locally convex space* (X,τ_X) *such that the topology spaces* $(U(\Omega),\tau_U)$ *and* $(L^1(\Omega),\|\cdot\|_{L^1(\Omega)})$ *are continuously embedded in* (X,τ_X) *(see Definition 3.1.23), i.e. we have*

$$(U(\Omega),\tau_U) \hookrightarrow (X,\tau_X) \qquad \text{and} \qquad (L^1(\Omega),\|\cdot\|_{L^1(\Omega)}) \hookrightarrow (X,\tau_X)\;, \tag{4.68}$$

then the sequence of iterates (u_k) *has a* τ_U *-convergent subsequence and every* τ_U *-convergent subsequence converges to a minimizer of the functional* F *defined in* (4.46).

Proof. This proof is divided into several steps. First, we show the monotone descent of the objective functional F . In the following steps, we prove the existence of convergent subsequences of the primal iterates (u_k) and of the subgradients (p_k) corresponding to (u_k) . Subsequently, we verify that the limit p of the dual iterates (p_k) is actually a subgradient of the regularization functional J at the limit u of the primal iterates (u_k) , i.e. $p \in \partial J(u)$. In the last step, we show that the found limit u is a minimizer of the objective functional F .

First step: Monotone descent of the objective functional

To get a descent of the objective functional F using an adequate damping strategy, we look for a condition on the damping parameters (ω_k), which guarantees for all $k \geq 0$ a descent of the form

$$F(u_{k+1}) \; + \; \underbrace{\frac{\epsilon}{\omega_k} \int_\Omega \frac{K^* \mathbf{1}_\Sigma \, (u_{k+1} \; - \; u_k)^2}{u_k} \, d\lambda}_{\geq \, 0} \; \leq \; F(u_k) \; , \qquad \epsilon \; > \; 0 \; . \tag{4.69}$$

This condition ensures actually a descent of F, since the second term on the left-hand side of (4.69) is positive due to $\omega_k > 0$ and due to the strict positivity of the iterates u_k (see *Lemma 4.6.12*). To show (4.69), we start with the damped regularization step (4.26), multiply it with $u_{k+1} - u_k$ and integrate the result over the domain Ω. Thus, for $p_{k+1} \in \partial J(u_{k+1}) \subset (U(\Omega))^*$, we obtain

$$\begin{aligned} 0 \;\; &= \;\; \int_\Omega \frac{K^* \mathbf{1}_\Sigma \, (u_{k+1} \; - \omega_k \, u_{k+\frac{1}{2}} \; - \; (1 \; - \; \omega_k) \, u_k) \, (u_{k+1} \; - \; u_k)}{u_k} \, d\lambda \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad + \; \omega_k \, \alpha \, \langle p_{k+1}, u_{k+1} \; - \; u_k \rangle \\ &= \;\; \int_\Omega \frac{K^* \mathbf{1}_\Sigma \, (u_{k+1} \; - \; u_k)^2}{u_k} \; + \; \omega_k \, \frac{K^* \mathbf{1}_\Sigma \, (u_k \; - \; u_{k+\frac{1}{2}}) \, (u_{k+1} \; - \; u_k)}{u_k} \, d\lambda \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad + \; \omega_k \, \alpha \, \langle p_{k+1}, u_{k+1} \; - \; u_k \rangle \; . \end{aligned}$$

Due to the definition of subgradients (see *Definition 3.2.6*), we now have that

$$\langle p_{k+1}, u_{k+1} \; - \; u_k \rangle \; \geq \; J(u_{k+1}) \; - \; J(u_k)$$

and thus

$$\begin{aligned} \alpha \, J(u_{k+1}) \; - \; \alpha \, J(u_k) + \frac{1}{\omega_k} \int_\Omega \frac{K^* \mathbf{1}_\Sigma \, (u_{k+1} \; - \; u_k)^2}{u_k} \, d\lambda & \\ \leq \; - \int_\Omega \frac{K^* \mathbf{1}_\Sigma \, (u_k \; - \; u_{k+\frac{1}{2}}) \, (u_{k+1} \; - \; u_k)}{u_k} \, d\lambda \; . & \end{aligned} \tag{4.70}$$

Adding the difference $D_{KL}(f, K u_{k+1}) - D_{KL}(f, K u_k)$ on both sides of this inequality and considering the definitions of the KL functional in (4.43) and the objective functional

F in (4.46) yields

$$\begin{aligned} F(u_{k+1}) \;-\; F(u_k) \;+\; \frac{1}{\omega_k} \int_\Omega \frac{K^*\mathbf{1}_\Sigma\,(u_{k+1} \;-\; u_k)^2}{u_k}\, d\lambda & \\ \leq \int_\Sigma \left(f \log\left(\frac{f}{Ku_{k+1}}\right) \;+\; Ku_{k+1} \;-\; f \log\left(\frac{f}{Ku_k}\right) \;-\; Ku_k \right) d\mu & \\ - \int_\Omega \left(K^*\mathbf{1}_\Sigma\,(u_{k+1} \;-\; u_k) \;-\; \frac{K^*\mathbf{1}_\Sigma\, u_{k+\frac{1}{2}}}{u_k}\,(u_{k+1} \;-\; u_k) \right) d\lambda & \qquad (4.71) \\ = \int_\Sigma \left(f \log\left(\frac{f}{Ku_{k+1}}\right) \;-\; f \log\left(\frac{f}{Ku_k}\right) \right) d\mu & \\ + \int_\Omega \left(K^*\left(\frac{f}{Ku_k}\right)(u_{k+1} \;-\; u_k) \right) d\lambda \;. & \end{aligned}$$

The last equality in (4.71) holds, since $u_{k+\frac{1}{2}}$ is given by the EM reconstruction step in (4.24) and K^* is the adjoint operator of K, i.e. we have, according to *Theorem 3.1.28*,

$$\int_\Omega K^*\mathbf{1}_\Sigma\, u\, d\lambda \;=\; \langle K^*\mathbf{1}_\Sigma, u\rangle_{(U(\Omega))^*,\, U(\Omega)} \;=\; \langle \mathbf{1}_\Sigma, Ku\rangle_{(V_\mu(\Sigma))^*,\, V_\mu(\Sigma)} \;=\; \int_\Sigma Ku\, d\mu \;. \quad (4.72)$$

Now, we try to characterize the right-hand side of inequality (4.71). For this sake, we define an auxiliary functional $G : U(\Omega) \to \mathbb{R} \cup \{+\infty\}$ as

$$G(u) \;:=\; \int_\Sigma f \log\left(\frac{f}{Ku}\right)\, d\mu$$

and consider the directional derivatives of G using the proposed strategy in *Remark 3.2.15, Item (2)*. That is, we define for $u \in U(\Omega)$ and any $w_1 \in U(\Omega)$ a function $\phi_{w_1}(t) := G(u + t\,w_1)$ and see that the directional derivative $G'(u; w_1)$ of G at u in the direction w_1 is given by

$$\begin{aligned} G'(u;\, w_1) \;=\; \phi'_{w_1}(t)\,\Big|_{\,t\,=\,0} \quad &= \quad \int_\Sigma \frac{\partial}{\partial t}\left(f \log\left(\frac{f}{Ku \;+\; t\,Kw_1}\right)\right) d\mu\,\Bigg|_{\,t\,=\,0} \\ &= \quad \left\langle -\,\frac{f}{Ku},\, Kw_1 \right\rangle_{(V_\mu(\Sigma))^*,\, V_\mu(\Sigma)} \\ &\overset{Thm.\ 3.1.28}{=} \quad \left\langle -\,K^*\left(\frac{f}{Ku}\right),\, w_1 \right\rangle_{(U(\Omega))^*,\, U(\Omega)} \;. \end{aligned}$$

Interpreting the right-hand side of inequality (4.71) formally as a Taylor linearization of G yields

$$\begin{aligned}
F(u_{k+1}) \;-\; F(u_k) \;+\; \frac{1}{\omega_k} \int_\Omega \frac{K^* \mathbf{1}_\Sigma \, (u_{k+1} \;-\; u_k)^2}{u_k} \, d\lambda & \\
\leq\; G(u_{k+1}) \;-\; G(u_k) \;-\; G'(u_k;\, u_{k+1} \;-\; u_k) & \\
=\; \frac{1}{2} G''(v;\, u_{k+1} \;-\; u_k,\, u_{k+1} \;-\; u_k) \;, \qquad v \;\in\; [u_k, u_{k+1}] \;, & \\
\leq\; \sup_{v \,\in\, [u_k, u_{k+1}]} \; \frac{1}{2} G''(v;\, u_{k+1} \;-\; u_k,\, u_{k+1} \;-\; u_k) \;. &
\end{aligned} \tag{4.73}$$

Now, analogous to the first directional derivative $G'(u;\, w_1)$, we can compute the second directional derivative $G''(u;\, w_1,\, w_2)$ using the function $\phi_{w_2}(t) \;:=\; G'(u \,+\, t\, w_2;\, w_1)$ for any $w_2 \,\in\, U(\Omega)$,

$$\begin{aligned}
G''(u;\, w_1,\, w_2) \;=\; \phi'_{w_2}(t) \,\Big|_{\,t\,=\,0} \;&=\; - \int_\Sigma \frac{\partial}{\partial t} \left(\frac{f}{Ku \;+\; t\, K w_2} \, K w_1 \right) d\mu \,\Bigg|_{\,t\,=\,0} \\
&=\; \int_\Sigma \frac{f\, K w_2 \, K w_1}{(Ku)^2} \, d\mu \;.
\end{aligned}$$

Plugging the computed derivative $G''(u;\, w_1,\, w_2)$ in the inequality (4.73), we obtain

$$\begin{aligned}
F(u_{k+1}) \;-\; F(u_k) \;+\; \frac{1}{\omega_k} \int_\Omega \frac{K^* \mathbf{1}_\Sigma \, (u_{k+1} \;-\; u_k)^2}{u_k} \, d\lambda & \\
\leq\; \sup_{v \,\in\, [u_k, u_{k+1}]} \; \frac{1}{2} \int_\Sigma \frac{f\, (K u_{k+1} \;-\; K u_k)^2}{(Kv)^2} \, d\mu \;. &
\end{aligned} \tag{4.74}$$

Finally, we split the third term on the left-hand side of (4.74) with $\epsilon \,\in\, (0, 1)$,

$$\begin{aligned}
F(u_{k+1}) \;+\; \frac{\epsilon}{\omega_k} \int_\Omega \frac{K^* \mathbf{1}_\Sigma \, (u_{k+1} \;-\; u_k)^2}{u_k} \, d\lambda \;+\; \frac{1 \,-\, \epsilon}{\omega_k} \int_\Omega \frac{K^* \mathbf{1}_\Sigma \, (u_{k+1} \;-\; u_k)^2}{u_k} \, d\lambda & \\
\leq\; \sup_{v \,\in\, [u_k, u_{k+1}]} \; \frac{1}{2} \int_\Sigma \frac{f\, (K u_{k+1} \;-\; K u_k)^2}{(Kv)^2} \, d\mu \;+\; F(u_k) \;, &
\end{aligned}$$

and obtain the desired condition (4.69), and therewith the descent of the objective functional F, if

$$\sup_{v \,\in\, [u_k, u_{k+1}]} \; \frac{1}{2} \int_\Sigma \frac{f\, (K u_{k+1} \;-\; K u_k)^2}{(Kv)^2} \, d\mu \;\leq\; \frac{1 \,-\, \epsilon}{\omega_k} \int_\Omega \frac{K^* \mathbf{1}_\Sigma \, (u_{k+1} \;-\; u_k)^2}{u_k} \, d\lambda \;. \tag{4.75}$$

By solving inequality (4.75) for ω_k, we obtain the required condition (4.65) for the damping parameters (ω_k) in order to have a descent of the objective functional F.

Additionally, by a suitable choice of ϵ in (4.65), we can guarantee that $\omega_k \leq 1$ for all $k \geq 0$.

Second step: Convergence of the primal iterates

Next, to show that the iteration method converges to a minimizer of F, we need a convergent subsequence of the primal iterates (u_k). For this purpose, we can use the precompactness requirement on the sub-level sets $\mathcal{S}_J$ of the regularization function J in *Assumption 4.6.5 (viii).* Since the functional F is positive, we set $a := F(u_0) \geq 0$ and can always obtain $a < \infty$ for a suitable choice of the nonnegative initialization function u_0. Due to the monotone decrease of the sequence $\big(F(u_k)\big)$ with the corresponding choice of the damping parameters (ω_k) in (4.65), it is now simple to see that for all $k \geq 0$,

$$a \;:=\; F(u_0) \;\geq\; F(u_k) \;=\; D_{KL}(f, K u_k) \;+\; \alpha\, J(u_k) \;\geq\; J(u_k)\;, \qquad (4.76)$$

due to the positivity of the KL functional D_{KL} and $\alpha > 0$. Thus, $u_k \in \mathcal{S}_J(a)$ for all $k \geq 0$ and it follows from *Assumption 4.6.5 (viii)* that (u_k) has a τ_U-convergent subsequence (u_{k_l}), which converges to some $u \in U(\Omega)$.

Subsequently, we can also consider the sequence of the primal iterates (u_{k+1}) and obtain with the same argumentation that there exists a τ_U-convergent subsequence (u_{k_l+1}) which converges to some $\tilde{u} \in U(\Omega)$. Now, we show that the limits of the subsequences (u_{k_l}) and (u_{k_l+1}) coincide, i.e. that it holds $u = \tilde{u}$. For this purpose, we apply inequality (4.69) recursively and obtain the following estimate,

$$F(u_{k+1}) \;+\; \epsilon \sum_{j\,=\,0}^{k} \int_\Omega \frac{K^* \mathbf{1}_\Sigma \,(u_{j+1} \,-\, u_j)^2}{\omega_j\, u_j}\, d\lambda \;\leq\; F(u_0) \;<\; \infty\;, \quad \forall k \;\geq\; 0\;. \qquad (4.77)$$

Thus, the series of functional descent values on the left-hand side of (4.77) is summable and the Cauchy criterion for convergence delivers

$$\lim_{k\,\to\,\infty} \int_\Omega \frac{K^* \mathbf{1}_\Sigma \,(u_{k+1} \,-\, u_k)^2}{\omega_k\, u_k}\, d\lambda \;=\; 0\;. \qquad (4.78)$$

Additionally, the Cauchy-Schwarz inequality yields the following estimate,

$$\| u_{k+1} \,-\, u_k \|_{L^1(\Omega)}^2 \;\leq\; \underbrace{\int_\Omega \frac{\omega_k\, u_k}{K^* \mathbf{1}_\Sigma}\, d\lambda}_{\overset{(4.66)}{\leq}\; c_1\, \|u_k\|_{L^1(\Omega)}} \; \underbrace{\int_\Omega \frac{K^* \mathbf{1}_\Sigma \,(u_{k+1} \,-\, u_k)^2}{\omega_k\, u_k}\, d\lambda}_{\overset{(4.78)}{\to}\; 0}\;. \qquad (4.79)$$

The first term on the right-hand side of (4.79) is uniformly bounded for all $k \geq 0$, since $\omega_k \in (0,1]$, the function $K^* \mathbf{1}_\Sigma$ is bounded away from zero (4.66) and the sequence (u_k) is uniformly bounded in the $L^1(\Omega)$-norm. The latter holds, because according to (4.76) we have that $u_k \in \{ v \in U(\Omega) : F(v) \leq F(u_0) \}$ for all $k \geq 0$, we additionally assumed that the functional F is $U(\Omega)$-coercive (see *Definition 4.6.7*) and that the space $U(\Omega)$ is continuously embedded in $L^1(\Omega)$. To conclude, since the second term on the right-hand side of (4.79) converges to zero (cf. (4.78)), we obtain from (4.79) that

$$u_{k+1} - u_k \to 0 \qquad \text{in } L^1(\Omega) . \tag{4.80}$$

Hence, due to the assumption (4.68) that the topology spaces $(L^1(\Omega), \|\cdot\|_{L^1(\Omega)})$ and $(U(\Omega), \tau_U)$ are continuously embedded in a locally convex space (X, τ_X), the uniqueness of the limit in the locally convex space (X, τ_X) implies $u = \tilde{u}$.

Third step: Convergence of the dual iterates

In addition to the second step, we also need a convergent subsequence of the subgradients (p_k) corresponding to the sequence (u_k), i.e. $p_k \in \partial J(u_k)$ for all $k \geq 0$. Since the subgradients exist only in the effective domain of a functional (see *Lemma 3.2.9* and *Remark 3.2.10*), we obtain that $p_k \in (W(\Omega))^*$ for all $k \geq 0$, due to extension property (4.45). Moreover, we assumed that the regularization functional J is convex and one-homogeneous. Hence, we can use the general property from *Lemma 3.2.11*, in order to characterize the subdifferentials of such a functional by

$$\partial J(u) = \{ p \in (W(\Omega))^* : \langle p, u \rangle = J(u) \text{ and } \langle p, v \rangle \leq J(v) \text{ for all } v \in W(\Omega) \} .$$

Then, we can see with assumption (4.67) that for each subgradient p_k the dual norm is bounded by

$$\|p_k\|_{(W(\Omega))^*} = \sup_{\|v\|_{W(\Omega)} \leq 1} \langle p_k, v \rangle \leq \sup_{\|v\|_{W(\Omega)} \leq 1} J(v) \overset{(4.67)}{\leq} c_4 .$$

Hence, the sequence (p_k) is uniformly bounded in the $(W(\Omega))^*$-norm and the Banach-Alaoglu *Theorem 3.1.31* delivers the compactness in the weak* topology on $(W(\Omega))^*$, which implies the existence of a subsequence, again denoted by index k_l, such that

$$p_{k_l+1} \rightharpoonup^* p \qquad \text{in } (W(\Omega))^* .$$

Fourth step: Show that $p \in \partial J(u)$
We have now the τ_U-convergence of sequences (u_{k_l}) and (u_{k_l+1}) in $U(\Omega)$ and the weak* convergence of (p_{k_l+1}) in $(W(\Omega))^*$. Next, we will show that the limit p of the dual iterates is a subgradient of J at the limit u of the primal iterates, i.e. $p \in \partial J(u)$. Hence, we have to prove that (see the definition of the subgradient in *Definition 3.2.6*),

$$J(u) \;+\; \langle p, v \,-\, u \rangle \;\leq\; J(v) \;, \qquad \forall v \;\in\; W(\Omega) \;. \tag{4.81}$$

For this purpose, let $p_{k_l+1} \in \partial J(u_{k_l+1})$, then the definition of the subgradient of J yields

$$J(u_{k_l+1}) \;+\; \langle p_{k_l+1}, v \,-\, u_{k_l+1} \rangle \;\leq\; J(v) \;, \qquad \forall v \;\in\; W(\Omega) \;. \tag{4.82}$$

Since we assumed that J is lower semicontinuous with respect to the topology τ_U in *Assumption 4.6.5 (vi)*, we can estimate the functional J at u from above,

$$J(u) \;\leq\; \liminf_{l \to \infty} J(u_{k_l+1}) \;\leq\; J(u_{k_l+1}) \;,$$

and (4.82) delivers

$$J(u) \;+\; \langle p_{k_l+1}, v \,-\, u_{k_l+1} \rangle \;\leq\; J(v) \;, \qquad \forall v \;\in\; W(\Omega) \;. \tag{4.83}$$

In addition, in the third step we verified the weak* convergence of (p_{k_l+1}) in $(W(\Omega))^*$, i.e. it holds corresponding to *Definition 3.1.30* that

$$\langle p_{k_l+1}, v \rangle \;\to\; \langle p, v \rangle \;, \qquad \forall v \;\in\; W(\Omega) \;.$$

Hence, to prove $p \in \partial J(u)$, it remains to show with respect to (4.83) and (4.81) that

$$\langle p_{k_l+1}, u_{k_l+1} \rangle \;\to\; \langle p, u \rangle \;. \tag{4.84}$$

For this purpose, we consider the complete iteration scheme of the damped FB-EM-REG algorithm (4.26) with $u_{k+\frac{1}{2}}$ in (4.24),

$$u_{k_l+1} \;-\; (1 - \omega_{k_l})\, u_{k_l} \;-\; \omega_{k_l} \left(\frac{u_{k_l}}{K^* \mathbf{1}_\Sigma} \, K^* \left(\frac{f}{K u_{k_l}} \right) \right) \;+\; \omega_{k_l}\, \alpha \, \frac{u_{k_l}}{K^* \mathbf{1}_\Sigma} \, p_{k_l+1} \;=\; 0 \;, \tag{4.85}$$

which is equivalent to

$$-\, \alpha \, p_{k_l+1} \;=\; \frac{K^* \mathbf{1}_\Sigma \, (u_{k_l+1} \,-\, u_{k_l})}{\omega_{k_l}\, u_{k_l}} \;+\; K^* \mathbf{1}_\Sigma \;-\; K^* \left(\frac{f}{K u_{k_l}} \right) \;.$$

Multiplying this formulation of the iteration scheme with u_{k_l+1} and integrating over the domain Ω yields

$$
\begin{aligned}
-\alpha \langle p_{k_l+1}, u_{k_l+1} \rangle \;&=\; \int_\Omega \frac{K^* \mathbf{1}_\Sigma \,(u_{k_l+1} \,-\, u_{k_l})\, u_{k_l+1}}{\omega_{k_l}\, u_{k_l}} \, d\lambda \\
&\qquad\qquad + \left\langle \mathbf{1}_\Sigma \,-\, \frac{f}{K u_{k_l}}, K u_{k_l+1} \right\rangle \\
&=\; \underbrace{\int_\Omega \frac{K^* \mathbf{1}_\Sigma \,(u_{k_l+1} \,-\, u_{k_l})^2}{\omega_{k_l}\, u_{k_l}} \, d\lambda}_{\overset{(4.78)}{\to}\, 0} \\
&\qquad + \underbrace{\int_\Omega \frac{K^* \mathbf{1}_\Sigma \,(u_{k_l+1} \,-\, u_{k_l})\, u_{k_l}}{\omega_{k_l}\, u_{k_l}} \, d\lambda}_{\overset{(4.80)}{\to}\, 0} \\
&\qquad\qquad + \left\langle \mathbf{1}_\Sigma \,-\, \frac{f}{K u_{k_l}}, K u_{k_l+1} \right\rangle \; .
\end{aligned}
\tag{4.86}
$$

The second term on the right-hand side of (4.86) vanishes in the limit, since the term $\frac{K^* \mathbf{1}_\Sigma}{\omega_{k_l}}$ is uniformly bounded in the supremum norm (caused by the boundedness away from zero of the damping parameters ω_k (4.66) and the boundedness preservation of the operator K in *Assumption 4.6.5 (v)*) and due to the $L^1(\Omega)$-norm convergence of the primal iterates in (4.80). Now, using the sequentially continuity of the operator K in *Assumption 4.6.5 (iii)*, we obtain

$$
\left\langle \mathbf{1}_\Sigma \,-\, \frac{f}{K u_{k_l}}, K u_{k_l+1} \right\rangle \;\to\; \left\langle \mathbf{1}_\Sigma \,-\, \frac{f}{K u}, K u \right\rangle
$$

and thus can deduce from (4.86) that

$$
-\alpha \langle p_{k_l+1}, u_{k_l+1} \rangle \;\to\; \int_\Omega \left(K^* \mathbf{1}_\Sigma \,-\, K^* \left(\frac{f}{K u} \right) \right) u \, d\lambda \;\overset{(4.88)}{=}\; -\alpha \langle p, u \rangle \; .
$$

Hence, we can conclude (4.84) and therewith $p \in \partial J(u)$.

Fifth step: Convergence to a minimizer of the objective functional
Now, let (u_{k_l}) and (u_{k_l+1}) be arbitrary τ_U-convergent subsequences of the primal iteration sequence (u_k), which converge to some $u \in U(\Omega)$. Then, as seen in the third and fourth step, there exists a weak* convergent subsequence (p_{k_l+1}) of the dual iteration sequence (p_k), which converges to some $p \in (W(\Omega))^*$ such that $p \in \partial J(u)$. To verify the convergence of the damped FB-EM-REG splitting algorithm, it remains to show that u is a minimizer of the functional F defined in (4.46). For this purpose, we consider the complete iteration scheme of the damped FB-EM-REG algorithm (4.85)

with reference to the convergent subsequences and show their weak* convergence to the optimality condition (4.19) of the variational problem (4.17). Note that it actually suffices to prove only the convergence to (4.19) and not to (4.23), since the function u is positive due to the strict positivity assumption on the iterates u_k for all $k \geq 0$ in (4.66).

An equivalent formulation of equation (4.85) reads as follows

$$\frac{u_{k_l+1} \; - \; u_{k_l}}{\omega_{k_l} \, u_{k_l}} \; + \; \mathbf{1}_\Omega \; - \; \frac{1}{K^* \mathbf{1}_\Sigma} \, K^* \left(\frac{f}{K u_{k_l}} \right) \; + \; \frac{\alpha}{K^* \mathbf{1}_\Sigma} \, p_{k_l+1} \; = \; 0 \; . \tag{4.87}$$

The convergence can be verified in the following way. Due to the boundedness away from zero assumptions in (4.66), we can use the result (4.78) in order to deduce the following convergence,

$$c_1 \, c_2 \, c_3 \int_\Omega \frac{(u_{k+1} \; - \; u_k)^2}{\omega_k^2 \, u_k^2} \, d\lambda \; \leq \; \int_\Omega \frac{K^* \mathbf{1}_\Sigma \, (u_{k+1} \; - \; u_k)^2}{\omega_k^2 \, u_k^2} \, \omega_k \, u_k \, d\lambda \; \overset{(4.78)}{\to} \; 0 \; .$$

Since the integrand on the left-hand side is positive, we obtain with the uniqueness of the limit, that

$$\lim_{l \to \infty} \frac{u_{k_l+1} \; - \; u_{k_l}}{\omega_{k_l} \, u_{k_l}} \; = \; 0 \; .$$

Therefore, if we pass over to the weak* limit of the subsequences in (4.87) using the sequential continuity of the operator K in *Assumption 4.6.5 (iii)* for the convergence of the term $\frac{f}{K u_{k_l}}$, we obtain that both limit functions u and p of the subsequences (u_{k_l}) and (p_{k_l+1}) fulfill the optimality condition (4.19) of the variational problem (4.17),

$$\mathbf{1}_\Omega \; - \; \frac{1}{K^* \mathbf{1}_\Sigma} \, K^* \left(\frac{f}{K u} \right) \; + \; \frac{\alpha}{K^* \mathbf{1}_\Sigma} \, p \; = \; 0 \; . \tag{4.88}$$

This means that the subsequence (u_{k_l}) converges in the topology τ_U to a minimizer of the functional F defined in (4.46). □

Remark.

- We note here that inequality (4.75) in the proof above motivates at the same time the mathematical necessity of a damping in the FB-EM-REG splitting strategy. In the undamped case, i.e. $\omega_k = 1$, the term on the right-hand side of (4.75) is maximal for $\epsilon \to 0^+$, due to the strict positivity of $K^* \mathbf{1}_\Sigma$ and u_k for all $k \geq 0$ in (4.66). In general, one cannot say whether this term is greater than the supremum on the left-hand side of (4.66) or not and with it whether the objective functional F is decreasing during the iteration or not. Hence, we need a parameter $\omega_k \in (0, 1)$, which increases the term on the right-hand side of (4.75) in order to guarantee a descent of the objective functional F.

- The assumptions on boundedness away from zero in (4.66) are reasonable from our point of view. In case of the function $K^*\mathbf{1}_\Sigma$, the assumption is practical since if there exists a point $x \in \Omega$ with $(K^*\mathbf{1}_\Sigma)(x) = 0$, then it is a-priori impossible to reconstruct the information in this point. Moreover, the assertion on the damping parameters ω_k makes sense because a strong damping is certainly undesirable. The boundedness away from zero of the iterates u_k can be absolutely guaranteed, if the regularization functional J fulfills the condition (4.58), such that from *Lemma 4.6.12* each half step of the (damped) FB-EM-REG method is strictly positive.

- Inspired by the relaxed EM reconstruction step in [123, Chap. 5.3.2], another possibility to influence convergence arises in the FB-EM-REG strategy by adding a relaxation parameter $\nu > 0$ to the EM fixed point iteration in the form,

$$u_{k+\frac{1}{2}} \;=\; u_k \left(\frac{1}{K^*\mathbf{1}_\Sigma} \, K^* \left(\frac{f}{K u_k} \right) \right)^{\nu} \qquad \text{(relaxed EM step)} \; .$$

  Correspondingly, one can obtain a reasonable regularization step in the FB-EM-REG splitting idea via

$$u_{k+1} \;=\; \left(u_{k+\frac{1}{2}}^{\frac{1}{\nu}} \;-\; \alpha \, u_k^{\frac{1}{\nu}} \, p_{k+1} \right)^{\nu} \; , \quad p_{k+1} \;\in\; \partial J(u_{k+1}) \; , \qquad \text{(relaxed REG step)} \; ,$$

  with the relaxed EM step $u_{k+\frac{1}{2}}$ above. The relaxed terms in the regularization step are necessary to fit the basic variational problem (4.16) and its optimality condition (4.19). Due to the computational challenge of the relaxed regularization step, which would require novel methods, a comparison of this strategy with our damping strategy proposed in *Section 4.4.2* would go beyond the scope of this thesis.

In practice, determining the damping parameters ω_k via the general condition in (4.65) is not straight-forward and one would be interested in an explicit bound for all damping parameters ω_k. Unfortunately, this is not possible in the case of a general operator K, but we can provide such an explicit bound on ω_k in the case of the Poisson denoising strategy (4.36), i.e. for the identity operator K.

Corollary 4.6.15 (*Convergence of the Damped Poisson Denoising Scheme*). *Let* (u_k) *be a sequence of iterates generated by the damped Poisson denoising scheme* (4.37) *and the given noisy function* $f \in L^\infty_\mu(\Omega)$ *satisfy* $\inf_\Omega f > 0$. *In order to guarantee the convergence in the case of the identity operator* K, *the condition* (4.65) *in Theorem 4.6.14 on the damping parameters simplifies to*

$$\omega_k \;\leq\; \frac{2\,(\inf_\Omega f)^2}{(\sup_\Omega f)^2}\,(1 \,-\, \epsilon) \; , \qquad \epsilon \;\in\; (0,1) \; . \tag{4.89}$$

Proof. In the special case of the identity operator K, the maximum principle of the damped Poisson denoising scheme from *Lemma 4.6.13* is the main idea for simplifying the desired condition (4.65). For this sake, we consider inequality (4.75), which guarantees a monotone descent of the objective functional if

$$\frac{1}{2} \int_\Omega \frac{f\, u_k}{v^2} \, \frac{(u_{k+1} \,-\, u_k)^2}{u_k} \, d\lambda \;\leq\; \frac{1 \,-\, \epsilon}{\omega_k} \int_\Omega \frac{(u_{k+1} \,-\, u_k)^2}{u_k} \, d\lambda \;, \qquad \forall v \,\in\, [u_k, u_{k+1}] \;.$$

Our goal is now to find an estimate for the coefficients $\frac{f\, u_k}{2\, v^2}$. Due to the fact that $v \in [u_k, u_{k+1}]$ and that (u_k) are iterates generated by the damped Poisson denoising scheme (4.37), we can apply the maximum principle from *Lemma 4.6.13* and obtain an estimate for the coefficients,

$$\frac{f\, u_k}{2\, v^2} \;\leq\; \frac{(\sup_\Omega\, f)\, (\sup_\Omega\, u_k)}{2\, (\inf_\Omega \{u_k,\, u_{k+1}\})^2} \;\leq\; \frac{(\sup_\Omega\, f)^2}{2\, (\inf_\Omega\, f)^2} \;, \qquad \forall k \,\geq\, 0 \;,$$

which should be less or equal $\frac{1-\epsilon}{\omega_k}$. Thus, choosing ω_k according to the estimate (4.89) guarantees a monotone descent of the objective functional. □

4.7 Iterative Refinement via Bregman Distance Iteration

In *Sections 4.4.1* and *4.4.2*, we presented the (damped) FB-EM-REG algorithm in order to solve the regularized Poisson likelihood estimation problem (4.16). However, in image processing it is well known that variational regularization techniques lead in almost all cases to so-called *systematic errors*, causing an oversmoothing effect in the reconstructions. One of the typical examples is the loss of contrast, which has been analyzed in the case of total variation minimization by Meyer [116]. Therefore, we propose in this section to extend the regularized Poisson likelihood estimation problem (4.16), and with it also the (damped) FB-EM-REG algorithm, by introducing an iterative regularization strategy in order to refine the reconstruction results. More precisely, we perform a refinement approach using *inverse scale space methods* based on *Bregman distance iteration*. These techniques have been derived by Osher et al. in [125], with a detailed analysis for Gaussian-type problems (4.2) $(p \,=\, 2)$, and have been generalized to time-continuity [38], L^p-norm data fitting terms [37] and nonlinear inverse problems [11]. Following these methods, an iterative refinement in our case is realized by a sequence of modified Poisson likelihood estimation problems based on (4.16).

4.7.1 Bregman Distances

Since the just mentioned iterative refinement approach is based on a Bregman distance iteration, we begin here with a short review of (generalized) Bregman distances related to a convex functional F. The Bregman distance is named after L. M. Bregman, who introduced this concept in [28] using Gâteaux differentiable functionals F. In [100], Kiwiel generalized the concept of Bregman distances to nonsmooth functionals which are strictly convex. However, since in image processing singular regularization functionals, i.e. neither continuously differentiable nor strictly convex, play an important role, Burger and Osher introduced in [39] a further generalization of Bregman distances based on subgradients. In this case, the subdifferential of a functional can also be multi-valued and one obtains a family of distances.

Definition 4.7.1 (*Generalized Bregman Distance [143, Def. 3.15]*). *Let U be a Banach space and $F : U \to \mathbb{R} \cup \{+\infty\}$ a convex and proper functional (see Definition 3.2.2) with subdifferential ∂F (see Definition 3.2.6). Then, the Bregman distance of F at $u \in U$ and $p \in \partial F(u) \subset U^*$ is defined by*

$$D_F^p(v,u) \;:=\; F(v) \;-\; F(u) \;-\; \langle p, v-u \rangle_{U^*,U} \;, \qquad v \;\in\; U \;, \tag{4.90}$$

where $\langle \cdot, \cdot \rangle_{U^*,U}$ denotes the standard duality product between the dual space U^* and U (see Definition 3.1.26).

Remark.

- For a Gâteaux differentiable functional F (*Definition 3.2.16*), the subdifferential ∂F of F contains a unique element p (see *Lemma 3.2.17*) and consequently a unique Bregman distance. In this case, one actually obtains $D_F^p(u,u) = 0$, and the convexity of F implies that D_F^p is really a distance, i.e. it holds $D_F^p(v,u) \geq 0$ for any v, $u \in U$.

- If the functional F is not Gâteaux differentiable, then the subdifferential ∂F of F is in general multivalued (cf. *Remark 3.2.7, Item (2)*), such that each element $d \in D_F^p(v,u)$ represents a distance between the elements v and u. Additionally, for not strictly convex functionals, it is also possible that $0 \in D_F^p(v,u)$ for $v \neq u$. Moreover, it is not guaranteed that $D_F^p(v,u)$ is nonempty, since $\partial F(u)$ can also be nonempty.

- The Bregman distance can be visualized as the difference between the tangent and the convex functional, cf. *Fig. 4.1*, or to be more precise, as the difference between the value of F at v and the value of the tangent at u evaluated at v (cf. here also the interpretation of the subgradients in *Remark 3.2.7, Item (1)*). In other words, the Bregman distance can be seen as a part of the Taylor linearization.

- Some examples of Bregman distances and their underlying functions are presented in *Table 4.1*.

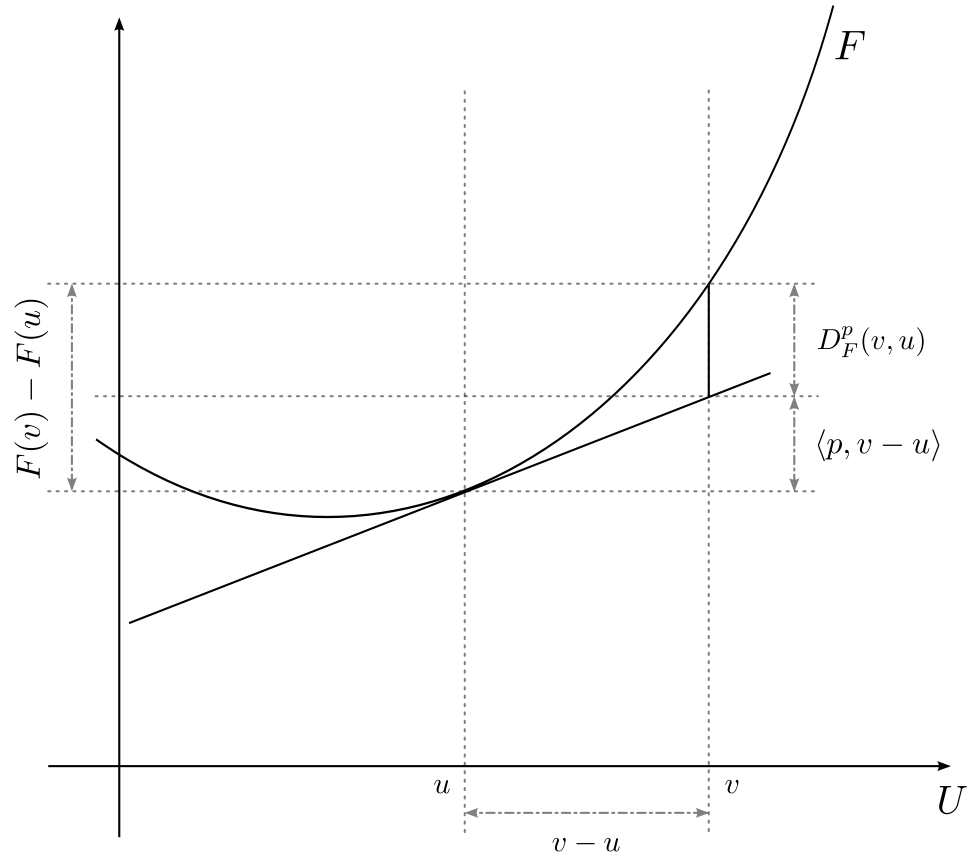


Fig. 4.1. *Bregman distance.*

The following lemma shows that the Bregman distance can in fact be viewed as a measure of similarity of two elements of U, since it is nonnegative. The proof of these properties is a direct consequence of the definitions of the Bregman distance and the subgradients in *Definition 3.2.6*.

Lemma 4.7.2 (*Properties of the Bregman Distance*). *Let $F : U \to \mathbb{R} \cup \{+\infty\}$ be a convex and proper functional on a Banach space U. Then, for $u \in U$ with a non-empty subdifferential $\partial F(u)$ and $p \in \partial F(u)$, the mapping $v \mapsto D_F^p(v,u)$ is convex and nonnegative due to the convexity of F and hence defines a distance in the sense that*

$$D_F^p(v,u) \begin{cases} = 0 \;, & \text{if } v = u \;, \\ \geq 0 \;, & \text{else} \;. \end{cases}$$

Additionally, if F is strictly convex, then $D_F^p(v,u) = 0$ if and only if $v = u$.

In general, the Bregman distance is not a metric (see *Definition 3.1.6*), since neither the triangular inequality nor the symmetry property is fulfilled. However, the symmetry can be achieved by introducing the so-called *symmetric Bregman distance*.

Function name	$F(u)$	$D_F(v,u)$
Squared norm	$\frac{1}{2}\,u^2$	$\frac{1}{2}\,(v\,-\,u)^2$
Shannon entropy	$u\,\log u\,-\,u$	$v\,\log\dfrac{v}{u}\,-\,v\,+\,u$
Bit entropy	$u\,\log u\,+\,(1\,-\,u)\,\log(1\,-\,u)$	$v\,\log\dfrac{v}{u}\,+\,(1\,-\,v)\,\log\dfrac{1\,-\,v}{1\,-\,u}$
Burg entropy	$-\,\log u$	$\dfrac{v}{u}\,-\,\log\dfrac{v}{u}\,-\,1$
Hellinger	$-\sqrt{1\,-\,u^2}$	$\dfrac{1\,-\,v\,u}{\sqrt{1\,-\,u^2}}\,-\,\sqrt{1\,-\,v^2}$
ℓ^p *quasi-norm*	$-\,u^p \qquad (0\,<\,p\,<\,1)$	$-\,v^p\,+\,p\,v\,u^{p-1}\,-\,(p\,-\,1)\,u^p$
ℓ^p *norm*	$\lvert u\rvert^p \qquad (1\,<\,p\,<\,\infty)$	$\lvert v\rvert^p\,-\,p\,v\,\mathrm{sgn}\,u\,\lvert u\rvert^{p-1}\,+\,(p\,-\,1)\,\lvert u\rvert^p$
Exponential	e^u	$e^v\,-\,(v\,-\,u\,+\,1)\,e^u$
Inverse	$\dfrac{1}{u}$	$\dfrac{1}{v}\,+\,\dfrac{v}{u^2}\,-\,\dfrac{2}{u}$

Table 4.1. *Overview of Bregman distances and their underlying functions, see [58].*

Definition 4.7.3 (*Symmetric Bregman Distance*). *Let* $F : U \to \mathbb{R} \cup \{+\infty\}$ *be a convex and proper functional with subdifferential* ∂F *on a Banach space* U. *The symmetric Bregman distance is defined by*

$$\begin{aligned} D_F^{symm}(v,u) \;&:=\; D_F^{p_u}(v,u) \;+\; D_F^{p_v}(u,v) \\ &=\; \langle p_v - p_u, v - u\rangle_{U^*,\,U} \end{aligned}$$

with $p_v \in \partial F(v)$ *and* $p_u \in \partial F(u)$.

4.7.2 (Inverse) Scale Space Methods and Bregman Iteration

As already mentioned, we will present in *Section 4.7.3* an iterative refinement strategy in order to improve the reconstruction results obtained with the FB-EM-REG algorithm, using inverse scale space methods based on Bregman distance iteration. Hence, we briefly repeat here the main aspects of the (inverse) scale space methods and clarify the connection to the iterative Bregman distance regularization technique. The following statements are based on [38, 37, 40, 85, 125, 144, 143].

In mathematical processing of noisy images there are at least two evolutionary concepts based on partial differential equations, namely the *scale space* and *inverse scale space methods*. Both approaches base on the concept of multiscale analysis caused by the assumption that the noise in images is usually expected to be a small scale feature. Therefore, a separation of small scales from larger ones allows to compute a smooth approximation of the noisy image.

Scale Space Methods

The main characteristics of scale space methods are that they use parabolic partial differential equations and smooth small scale features of an image faster than larger scale ones. The most important scale space methods in image processing are nonlinear diffusion filters, see e.g. [129] or [165], of the form

$$\begin{aligned} \frac{\partial u}{\partial t} &= \operatorname{div}\big(g\,(\,|\nabla u|^2\,)\,\nabla u \big) \\ u(0) &= f \end{aligned} \qquad \text{in} \qquad \Omega \times \mathbb{R}_{\geq 0}\,, \tag{4.91}$$

where f denotes the observed noisy image and the diffusion coefficient involves a positive and monotone function g. For given $t_0 > 0$, $u(t_0)$ is considered to be an approximation and filtered version of f and t_0 controls the amount of the filtering. Especially, it can be shown that such methods smooth small scales faster than large ones. Hence, if we stop the method at a suitable time, we can expect that the noise is smoothed, while large scale features are preserved to some degree.

However, in context of inverse problems and variational regularization theory, we are more interested in minimization problems of the form (cf. (2.10)),

$$\min_{u \in W(\Omega)} \; H_f(Ku - f) \; + \; \alpha\, J(u) \;. \tag{4.92}$$

In this setup, the scale space methods are only applicable for denoising problems, i.e. for the identity operator K. For instance, in the case of additive Gaussian noise in the images, the data fidelity term is given by

$$H_f(Ku - f) \;=\; H_f(u - f) \;=\; \frac{1}{2}\,\|u \,-\, f\|^2_{L^2(\Omega)}$$

and the scale space methods base on the following *gradient flow equation*,

$$\begin{aligned} \frac{\partial u}{\partial t} &= -p(t) \;\in\; \partial J(u(t)) \\ u(0) &= f \end{aligned} \qquad \text{in} \qquad \Omega \times \mathbb{R}_{\geq 0}\,, \tag{4.93}$$

where ∂J denotes the subdifferential of the regularization functional J (see *Definition 3.2.6*). The smoothing effect happens by stopping the flow after a certain timespan, which should be short since the long time limit corresponds to a minimizer of J. In particular, the total variation regularization functional J has attracted strong attention in image processing, due to the possibility to realize discontinuous solutions (cf. *Section 6.2*), where the resulting scale space flow (4.93) is called *TV flow equation*.

Inverse Scale Space Methods

Inverse scale space methods have been introduced by Groetsch and Scherzer in [85] and [144], and are based on a reverse characteristic comparing to the scale space methods. The naming of the inverse scale space methods is motivated by the fact that those have the following evolution properties in case of denoising problems,

$$\lim_{t \to \infty} u(t) \; = \; f \qquad \text{and} \qquad \lim_{t \to 0^+} u(t) \; = \; u_0 \; , \tag{4.94}$$

where f is the observed noisy image and u_0 denotes an arbitrary initialization function for the filtering process. Thus, the properties in (4.94) mean that the inverse scale space methods "invert" the axiom of fidelity in the scale space theory, which asserts that (cf. (4.91) and (4.93)),

$$\lim_{t \to 0^+} u(t) \; = \; f \; . \tag{4.95}$$

The difference between (4.94) and (4.95) is that instead of starting the smoothing process with the noisy image in (4.95) and gradually smoothing it, inverse scale space methods (4.94) start with an arbitrary image u_0 and adapt the noisy image f with increasing time, where the large scale features convergence faster than the small ones. Hence, if we stop the method at a suitable time, we can expect that large scale features are already incorporated into the reconstruction, while small scale features (including the noise part) are still missing.

Iterative Regularization with Bregman Distances

The inverse scale space methods have a close relation to regularization theory, in particular to iterative Tikhonov regularization as it has been introduced in [85] and [144]. However, the construction of such methods in [144] worked only well for quadratic regularization functionals, which lead to interesting but linear evolution equations. Hence, this approach is less applicable for other important functionals, such as total variation or other singular regularization energies of ℓ^1- or L^1-type (cf. e.g. [37]).

In [38] and [37], Burger et al. proposed a novel approach to construct inverse scale space methods, namely as the limit of an iterated refinement procedure previously introduced by Osher et al. in [125], which can also be used for singular regularization energies like total variation. In the general setup of variational regularization problems (4.92), the iterative refinement procedure starts with $u_0 = 0$ and $p_0 = 0 \in \partial J(u_0)$, and delivers a sequence of reconstructions via

$$u^{l+1} \in \operatorname*{arg\,min}_{u \in W(\Omega)} \left\{ H_f(Ku - f) + \alpha \left(J(u) - \left\langle p^l, u \right\rangle \right) \right\} , \qquad p^l \in \partial J(u^l) , \quad (4.96)$$

for $l = 0, 1, \ldots$, where $\langle \cdot, \cdot \rangle$ is the standard duality product and ∂J the subdifferential of the functional J. Clearly, this procedure can be generalized to an arbitrary initial value u_0 as long as there exists a subgradient $p_0 \in \partial J(u_0) \cap \mathcal{R}(K^*)$, where $\mathcal{R}(K^*)$ is the range of the operator K^*. By adding the term $\alpha \left(-J(u^l) + \left\langle p^l, u^l \right\rangle \right)$, which is independent of the minimization variable u, the iterative procedure (4.96) can equivalently be rewritten as

$$\begin{aligned} u^{l+1} &\in \operatorname*{arg\,min}_{u \in W(\Omega)} \left\{ H_f(Ku - f) + \alpha\, D_J^{p^l}(u, u^l) \right\} , \\ p^{l+1} &= p^l - \frac{1}{\alpha} K^* \left(\partial H_f(Ku^{l+1} - f) \right) \in \partial J(u^{l+1}) , \end{aligned} \quad (4.97)$$

with the *generalized Bregman distance* $D_J^{p^l}(u, u^l)$ defined in *Definition 4.7.1.* In the case of K being the identity operator and using data fidelity term

$$H_f(Ku - f) = H_f(u - f) = \frac{1}{p} \| u - f \|_{L^p(\Omega)}^p , \qquad p \in (1, \infty) ,$$

the authors in [37] have shown that the procedure (4.97) fulfills the *discrete inverse fidelity property*, which means that the sequence u^{l+1} approaches the original given (noisy) image f as $l \to \infty$.

Now, in order to connect the iterated refinement procedure (4.96) to the inverse scale space methods, we consider in (4.97) the limit $\frac{1}{\alpha} \to 0^+$. In this case, the first order optimality condition of (4.96), which corresponds to the update rule for p^l in (4.97), can be interpreted as a forward Euler discretization of the *flow equation*

$$\frac{\partial p}{\partial t} = -K^* \left(\partial H_f(Ku(t) - f) \right) , \qquad p(t) \in \partial J(u(t)) ,$$

which has been termed *nonlinear inverse scale space method* (cf. [38] and [37]) in analogy to the previous work on inverse scale space methods by Scherzer and Groetsch [144].

Finally, we consider a further interpretation of the iterated refinement strategy (4.96), which motivates us in particular to use this approach in the context of the regularized

Poisson likelihood estimation problem (4.16). Namely, in image processing it is well known that the variational regularization techniques lead in almost all cases to the so-called systematic errors in the reconstructions. This effect can be easily illustrated by the fact that for noise free data $f = \bar{f} = K\bar{u}$, which are generated by the exact image $\bar{u}$, a minimizer u of the general variational regularization problem (4.92) satisfies

$$H_f(Ku - \bar{f}) \; + \; \alpha \, J(u) \; \leq \; \underbrace{H_f(K\bar{u} - \bar{f})}_{\overset{(2.11)}{=} \; 0} \; + \alpha \, J(\bar{u}) \; .$$

This implies the inequality,

$$J(u) \; - \; J(\bar{u}) \; \leq \; -\frac{1}{\alpha} \, H_f(Ku - \bar{f}) \; < \; 0 \; ,$$

which means that the regularization functional J is smaller for the minimizer u than for the exact image $\bar{u}$. This is the so-called *systematic error* of a regularization strategy and causes an oversmoothing (also in case of noisy data f) of the reconstructions. One of the typical examples of systematic errors is the loss of contrast, which has been analyzed intensively in the case of the Rudin, Osher and Fatemi model (6.6) [138] by Meyer [116] (cf. *Section 6.2*). This loss of contrast can be interpreted as a deficiency of the scheme at larger scales, since usually in denoising one likes to eliminate only noise modeled as small scale features. Therefore, motivated by the systematic error of total variation regularization, Osher et al. proposed in [125] a new iterative regularization procedure for inverse problems based on Bregman distances, which has in general case the form (4.96) (4.97), respectively. Simultaneously, in the case of an L^2 data fidelity term the authors provided another equivalent formulation of the strategy (4.97), which enables a nice interpretation, why the iterative regularization procedure (4.96) actually leads to a refinement of systematic errors. In [32], Brune et al. generalized this equivalent formulation for arbitrary convex data fidelity terms and obtained the following form,

$$\begin{aligned} u^{l+1} \; &\in \; \operatorname*{arg\,min}_{u \, \in \, W(\Omega)} \left\{ \, H_f(Ku \, + \, r^l \, - \, f) \; + \; \alpha \, J(u) \, \right\} \; , \\ r^{l+1} \; &= \; r^l \; + \; Ku^{l+1} \; - f \; \in \; \partial H_f^*(- \, \alpha \, (K^*)^{-1} \, p^{l+1}) \; , \end{aligned} \tag{4.98}$$

where H_f^* denotes the convex conjugate of the data fidelity term H_f. The recursion formula with respect to the "noise" function r delivers now an interesting decomposition of the data function f involving "noise" at levels l and $l+1$ and signal at level $l+1$. Remarkably, the formulation (4.98) leads to the concept that the original given noisy data f should be modified by the current "noise" function r^l and the sum of both should then be processed by the regularization procedure. Therefore, the iterative regularization strategy (4.96) actually leads to a stepwise refinement of reconstructions compensating the systematic errors of regularization methods.

4.7.3 Bregman-FB-EM-REG Algorithm

To perform an iterative refinement of the reconstructions resulting from the (damped) FB-EM-REG algorithm, proposed in *Sections 4.4.1* and *4.4.2*, we use the approach of inverse scale space methods based on Bregman distance iteration, discussed in *Section 4.7.2*. There, following the procedure (4.96), an iterative refinement is realized by a sequence of modified FB-EM-REG problems based on (4.16). In our case, the desired method initially starts with a simple FB-EM-REG algorithm, i.e. it consists of computing a minimizer u^1 from (4.16) and (4.17), respectively. Subsequently, the results will be refined step by step by considering variational problems with a shifted regularization term, namely

$$u^{l+1} \in \operatorname*{arg\,min}_{\substack{u \in W(\Omega) \\ u \geq 0 \;\; a.e.}} \left\{ D_{KL}(f, Ku) \; + \; \alpha \left(J(u) \; - \; \langle p^l, u \rangle \right) \right\} , \qquad p^l \in \partial J(u^l) \; . \tag{4.99}$$

Using the Bregman distance considered in *Definition 4.7.1*, the shifted regularization term can be extended to a convex functional (see *Lemma 4.7.2*) without changing the stationary points of the minimization problem, such that we obtain

$$u^{l+1} \in \operatorname*{arg\,min}_{\substack{u \in W(\Omega) \\ u \geq 0 \;\; a.e.}} \left\{ D_{KL}(f, Ku) \; + \; \alpha \, D_J^{p^l}(u, u^l) \right\} , \qquad p^l \in \partial J(u^l) \; . \tag{4.100}$$

Note that the first iterate u^1 can also be realized by the variational problem (4.99) respectively (4.100), if the initialization function u^0 will be chosen constant and we set $p^0 := 0 \in \partial J(u^0)$.

From the point of view of the statistical problem formulation in *Section 2.2*, the Bregman distance regularized variational problem (4.100) uses an adapted a-priori probability density $p(u)$ (2.8) in the Bayesian model formulation (2.6). Namely, instead of a zero centered a-priori probability $J(u)$ as in the case of the FB-EM-REG algorithm (4.16), we consider in every Bregman refinement step a new a-priori probability, which is related to a shifted regularization functional, i.e. we use the following Gibbs function (2.8),

$$p(u) \; \sim \; e^{- \alpha \, D_J^{p^l}(u, u^l)} \; .$$

This a-priori probability density means that images with smaller regularity, where the type of regularity is depends on the chosen functional J, and with a close distance to the maximum likelihood estimator u^l of the previous FB-EM-REG problem are preferred in the minimization of (4.100).

To design a suitable two step iteration strategy analogous to the FB-EM-REG algorithm in *Section 4.4.1*, we consider the first order optimality condition for the variational problem (4.100). Due to the convexity of the Bregman distance in the first argument (see *Lemma 4.7.2*), we can determine the subdifferential of (4.100). To this aspect, due to the continuity of the Kullback-Leibler (KL) functional and the subdifferential calculus in *Lemma 3.2.13*, we obtain the following identity

$$\partial\Big(D_{KL}(f, Ku) \;+\; \alpha\,\big(\,J(u) \;-\; \langle p^l, u\rangle\,\big)\Big) \\ = \partial_u\, D_{KL}(f, Ku) \;+\; \alpha\,\Big(\partial J(u) \;-\; \partial\,\big(\,\langle p^l, u\rangle\,\big)\Big)\;.$$

Now, the subdifferentials of the KL functional and of the shift in the regularization functional are singletons of the form

$$\partial_u\, D_{KL}(f, Ku) \;=\; \left\{K^*\mathbf{1}_\Sigma \;-\; K^*\left(\frac{f}{Ku}\right)\right\} \qquad \text{and} \qquad \partial\,\big(\,\langle p^l, u\rangle\,\big) \;=\; \{p^l\}\;.$$

Therefore, due to the nonnegativity constraint in (4.99), the Karush-Kuhn-Tucker (KKT) conditions [89, Thm. 2.1.4] provide the existence of a Lagrange multiplier $\lambda \geq 0$, such that the stationary points of the functional in (4.100) need to fulfill

$$\begin{aligned} 0 \;&\in\; K^*\mathbf{1}_\Sigma \;-\; K^*\left(\frac{f}{Ku^{l+1}}\right) \;+\; \alpha\,\big(\,\partial J(u^{l+1}) \;-\; p^l\,\big) \;-\; \lambda\;, \\ 0 \;&=\; \lambda\, u^{l+1}\;, \end{aligned} \tag{4.101}$$

with $p^l \in \partial J(u^l)$. By multiplying the first equation in (4.101) by u^{l+1}, the Lagrange multiplier λ can be eliminated by the second equation and a subsequent division by $K^*\mathbf{1}_\Sigma$ leads to a fixed point equation of the form

$$0 \;\in\; u^{l+1} \;-\; \frac{u^{l+1}}{K^*\mathbf{1}_\Sigma}\, K^*\left(\frac{f}{Ku^{l+1}}\right) \;+\; \alpha\,\frac{u^{l+1}}{K^*\mathbf{1}_\Sigma}\,\big(\,\partial J(u^{l+1}) \;-\; p^l\,\big) \tag{4.102}$$

with $p^l \in \partial J(u^l)$. Particular advantageous is that this condition delivers a well defined update formula for the iterates p^l, namely

$$p^{l+1} \;:=\; p^l \;-\; \frac{1}{\alpha}\left(K^*\mathbf{1}_\Sigma \;-\; K^*\left(\frac{f}{Ku^{l+1}}\right)\right) \;\in\; \partial J(u^{l+1}) \tag{4.103}$$

with u^0 constant and $p^0 := 0 \in \partial J(u^0)$. Now, analogous to the fixed point equation (4.23) and (4.21) in the case of FB-EM-REG algorithm, we can apply the idea of the nested two step iteration (4.24) in every refinement step $l = 0, 1, \ldots$. Then, the condition (4.102) yields a strategy consisting of an EM step for the solution of (4.100),

$$u_{k+\frac{1}{2}}^{l+1} \;=\; \frac{u_k^{l+1}}{K^* \mathbf{1}_\Sigma} \, K^* \left(\frac{f}{K u_k^{l+1}} \right) \,, \tag{4.104}$$

followed by solving an adapted variational regularization problem

$$u_{k+1}^{l+1} \;\in\; \operatorname*{arg\,min}_{u \,\in\, W(\Omega)} \left\{ \frac{1}{2} \int_\Omega \frac{K^* \mathbf{1}_\Sigma \left(u \;-\; u_{k+\frac{1}{2}}^{l+1} \right)^2}{u_k^{l+1}} \;+\; \alpha \, \Big(J(u) \;-\; \left\langle p^l, u \right\rangle \Big) \right\} \,. \tag{4.105}$$

Following [125], we transfer now the shift term $\langle p^l, u \rangle$ to the data fidelity term. This approach facilitates the implementation of the iterated refinement with the Bregman distance via a slight modification of the FB-EM-REG algorithm. For this purpose, we use the scaling

$$v^l \;:=\; \frac{\alpha}{K^* \mathbf{1}_\Sigma} \, p^l \tag{4.106}$$

and obtain from (4.103) the following update formula for the iterates v^l,

$$v^{l+1} \;=\; v^l \;-\; \left(\mathbf{1}_\Omega \;-\; \frac{1}{K^* \mathbf{1}_\Sigma} \, K^* \left(\frac{f}{K u^{l+1}} \right) \right) \,, \qquad v^0 \;=\; 0 \,. \tag{4.107}$$

Using this scaled update, we can rewrite the second step (4.105) to

$$u_{k+1}^{l+1} \;\in\; \operatorname*{arg\,min}_{u \,\in\, W(\Omega)} \left\{ \frac{1}{2} \int_\Omega \frac{K^* \mathbf{1}_\Sigma \left(\left(u \;-\; u_{k+\frac{1}{2}}^{l+1} \right)^2 \;-\; 2 \, u \, u_k^{l+1} \, v^l \right)}{u_k^{l+1}} \;+\; \alpha \, J(u) \right\} \,.$$

Note that in the equation

$$\left(u - u_{k+\frac{1}{2}}^{l+1} \right)^2 - 2 \, u \, u_k^{l+1} \, v^l \;=\; \left(u \;-\; \left(u_{k+\frac{1}{2}}^{l+1} \;+\; u_k^{l+1} \, v^l \right) \right)^2 - 2 \, u_{k+\frac{1}{2}}^{l+1} \, u_k^{l+1} \, v^l + (u_k^{l+1})^2 \, (v^l)^2 \,,$$

the last two terms on the right-hand side are independent of u and hence (4.105) simplifies to

$$u_{k+1}^{l+1} \;\in\; \operatorname*{arg\,min}_{u \,\in\, W(\Omega)} \left\{ \frac{1}{2} \int_\Omega \frac{K^* \mathbf{1}_\Sigma \left(u \;-\; \left(u_{k+\frac{1}{2}}^{l+1} \;+\; u_k^{l+1} \, v^l \right) \right)^2}{u_k^{l+1}} \;+\; \alpha \, J(u) \right\} \,, \tag{4.108}$$

i.e. that the second half step (4.105) can be realized by a minor modification of the regularization step introduced in (4.25).

In *Section 4.4.2*, we additionally introduced a damped variant of the FB-EM-REG algorithm. This damping strategy can be also realized in each Bregman refinement step, namely the reconstruction step (4.108) simply needs to be adapted to

$$u_{k+1}^{l+1} \in \operatorname*{arg\,min}_{u \in W(\Omega)} \left\{ \frac{1}{2} \int_\Omega \frac{K^* \mathbf{1}_\Sigma \left(u - \tilde{u}_{k+\frac{1}{2}}^{l+1} \right)^2}{u_k^{l+1}} + \omega_k^{l+1} \alpha J(u) \right\} \tag{4.109}$$

with

$$\tilde{u}_{k+\frac{1}{2}}^{l+1} = \omega_k^{l+1} u_{k+\frac{1}{2}}^{l+1} + \omega_k^{l+1} u_k^{l+1} v^l + (1 - \omega_k^{l+1}) u_k^{l+1} .$$

4.7.4 Stopping Rules

As usual for iterative methods, the refinement strategy via the iterative Bregman distance regularization, described in *Section 4.7.3*, needs a suitable stopping criterion. Optimally, this rule should stop the method at an iteration, which offers a solution that approximates the desired true image as good as possible. This is essential in order to prevent that too many small scales, in particular the noise, are incorporated into the reconstruction results (cf. *Section 4.7.2*). In the case of Gaussian noise, the authors in [125] and [38] suggested to use the so-called *generalized discrepancy principle* (cf. [67] and [131] for a detailed discussion). This strategy consists in stopping the iteration at the index $l_* = l_*(\delta, f)$, where the residual $\| K u^{l_*} - f \|_{L^2(\Sigma)}$ reaches the noise level δ or an estimate of the noise level, i.e.

$$l_* = \max\{ l \in \mathbb{N} : \| K u^l - f \|_{L^2(\Sigma)} \geq \tau\delta \} , \qquad \tau > 1 .$$

However, in the case of raw data corrupted by Poisson noise, it makes more sense to stop the Bregman iteration, when the Kullback-Leibler (KL) distance between the given data f and the signal Ku^l reaches the noise level. For synthetic data, the noise level δ is naturally given by the KL distance between f and $K\bar{u}$, i.e.

$$\delta = D_{KL}(f, K\bar{u}) ,$$

where $\bar{u}$ denotes the exact and noise free image. For experimental data, it is necessary to find a suitable estimate of the noise level δ from the data counts.

In addition to a stopping criterion for the outer Bregman iteration, we also need suitable stopping rules for the inner FB-EM-REG iteration loop. For this purpose, we can proceed analogous to the discussion of stopping rules in the case of the FB-EM-REG algorithm in *Section 4.4.4*. In addition to a maximum number of iterations, we consider the error in the optimality condition (4.102) as a basic stopping criterion using the weighted norm $\| \cdot \|_{2,\mathrm{w}}$ (4.30), i.e.

$$\begin{aligned} opt_{k+1}^{l+1} \; &:= \; \left\| \, K^* \mathbf{1}_\Sigma \; - \; K^* \left(\frac{f}{K u_{k+1}^{l+1}} \right) \; + \; \alpha \, p_{k+1}^{l+1} \; - \; \alpha \, p^l \, \right\|_{2, u_{k+1}^{l+1}}^2 \\ &\overset{(4.106)}{=} \; \left\| \, K^* \mathbf{1}_\Sigma \; - \; K^* \left(\frac{f}{K u_{k+1}^{l+1}} \right) \; - K^* \mathbf{1}_\Sigma \, v^l \; + \; \alpha \, p_{k+1}^{l+1} \, \right\|_{2, u_{k+1}^{l+1}}^2 \; . \end{aligned} \tag{4.110}$$

Furthermore, we are also interested in the convergence of the primal function sequence (u_k^{l+1}) and subgradient sequence (p_k^{l+1}) with $p_k^{l+1} \in \partial J(u_k^{l+1})$. To establish appropriate rules for these sequences, we consider the damped regularization step (4.109) with the EM reconstruction step (4.104),

$$\begin{aligned} u_{k+1}^{l+1} \; - \; \omega_k^{l+1} \, \frac{u_k^{l+1}}{K^* \mathbf{1}_\Sigma} K^* \left(\frac{f}{K u_k^{l+1}} \right) \; - \; \omega_k^{l+1} \, u_k^{l+1} \, v^l \; - \; (1 - \omega_k^{l+1}) \, u_k^{l+1} & \\ + \; \omega_k^{l+1} \, \alpha \, \frac{u_k^{l+1}}{K^* \mathbf{1}_\Sigma} \, p_{k+1}^{l+1} \; &= \; 0 \; . \end{aligned}$$

By combining this iteration scheme with the optimality condition (4.102) evaluated at u_k^{l+1}, which must be fulfilled in the case of convergence, and applying the scaling (4.106), we obtain the following optimality statement for the sequences (u_k^{l+1}) and (p_k^{l+1}),

$$\alpha \, (p_{k+1}^{l+1} \; - \; p_k^{l+1}) \; + \; \frac{K^* \mathbf{1}_\Sigma \, (u_{k+1}^{l+1} \; - \; u_k^{l+1})}{\omega_k^{l+1} \, u_k^{l+1}} \; = \; 0 \; .$$

With the aid of the weighted norm (4.30), we now have additional stopping criteria for the inner FB-EM-REG iteration loop, which guarantee the accuracy of the primal functions (u_k^{l+1}) and the subgradients (p_k^{l+1}), namely

$$\begin{aligned} u_{opt_{k+1}^{l+1}} \; &:= \; \left\| \, \frac{K^* \mathbf{1}_\Sigma \, (u_{k+1}^{l+1} \; - \; u_k^{l+1})}{\omega_k^{l+1} \, u_k^{l+1}} \, \right\|_{2, u_{k+1}^{l+1}}^2 \; , \\ p_{opt_{k+1}^{l+1}} \; &:= \; \left\| \, \alpha \, (p_{k+1}^{l+1} \; - \; p_k^{l+1}) \, \right\|_{2, u_{k+1}^{l+1}}^2 \; . \end{aligned} \tag{4.111}$$

We finally mention that the stopping criteria (4.110) and (4.111) are well defined, since we proved in *Lemma 4.6.12* that each iterate u_k^{l+1} of the inner FB-EM-REG iteration loop is strictly positive.

4.7.5 Pseudocode and Some Remarks

Summarizing the observations in *Sections 4.7.3* and *4.7.4*, we now can make use of *Algorithm 4.2* to solve the stepwise refinement (4.99) of the regularized Poisson likelihood estimation problem (4.16).

Algorithm 4.2 *(Damped) Bregman-FB-EM-REG Algorithm*

1. ***Parameters:*** f , $\alpha > 0$, $\omega \in (0,1]$, $maxBregIts \in \mathbb{N}$, $\delta > 0$, $\tau > 1$, $maxEMIts \in \mathbb{N}$, $tol > 0$
2. ***Initialization:*** $l = 0$, $u_0^1 = u_0 := c > 0$, $v^0 := 0$
3. ***Iteration:***

while ($D_{KL}(f, Ku_0^{l+1}) \geq \tau\delta$ **and** $l < maxBregIts$) **do**

$a)$ Set $k = 0$.

while (($k < maxEMIts$) **and**
($opt_k^{l+1} \geq tol$ **or** $u_{opt_k^{l+1}} \geq tol$ **or**
$p_{opt_k^{l+1}} \geq tol$)) **do** $\triangleright$ (4.110), (4.111)

$i)$ Compute $u_{k+\frac{1}{2}}^{l+1}$ via EM step in (4.104).

$ii)$ Set $\omega_k^{l+1} = \omega$.

$iii)$ Compute u_{k+1}^{l+1} via convex variational problem (4.109).

$iv)$ $k \leftarrow k+1$

end while

$b)$ Compute update v^{l+1} via (4.107).

$c)$ Set $u_0^{l+2} = u_k^{l+1}$.

$d)$ $l \leftarrow l+1$

end while

4. ***Return*** u_0^{l+1}

Remark.

- Note that the update variable v in (4.107) has an interesting interpretation as an error function with reference to the optimality condition of the unregularized Poisson log-likelihood functional (4.13). That means that in every refinement step of the Bregman iteration the function v^{l+1} differs from v^l by the current error in the optimality condition of (4.13),

$$K^* \mathbf{1}_\Sigma \; - \; K^* \left(\frac{f}{Ku} \right) \; = \; 0 \; .$$

Hence, caused by the regularization step (4.108), we can expect that the iterative regularization based on the Bregman distance leads to a stepwise refinement. The reason is, that instead of fitting the new regularized solution u_{k+1}^{l+1} to the EM result $u_{k+\frac{1}{2}}^{l+1}$ in the weighted L^2 norm, as it occurs in the case of the FB-EM-REG step (4.25), the Bregman refinement strategy (4.108) uses an adapted "noisy" function in the data fidelity term, where the intensities of the EM solution $u_{k+\frac{1}{2}}^{l+1}$ are increased by a weighted error function v^l.

- Motivated by the work in [125], one may also consider the modeling of an iterative reconstruction refinement inside the FB-EM-REG algorithm. On the basis of the two step iteration proposed in (4.24) and (4.25), this strategy would lead to a regularization step which can be realized by a sequence of modified variational problems based on (4.25). More precisely, for any fixed index k, the iterate u_{k+1} is determined via a sequence of the following minimization problem,

$$u_{k+1}^{l+1} \;\in\; \operatorname*{arg\,min}_{u \,\in\, W(\Omega)} \left\{ \frac{1}{2} \int_\Omega \frac{K^* \mathbf{1}_\Sigma \left(u \;-\; u_{k+\frac{1}{2}} \right)^2}{u_k} \;+\; \alpha \left(J(u) \;-\; \langle p^l, u \rangle \right) \right\} \qquad (4.112)$$

with $p^l \in \partial J(u_{k+1}^l)$, a constant initialization u_{k+1}^0 and $p^0 := 0 \in \partial J(u_{k+1}^0)$. Analogous to *Section 4.7.3*, the scaling $K^* \mathbf{1}_\Sigma v^l := \alpha\, u_k\, p^l$ transfers the shift term $\langle p^l, u \rangle$ to the data fidelity term in such a way that (4.112) can be rewritten similar to (4.108), namely

$$u_{k+1}^{l+1} \;\in\; \operatorname*{arg\,min}_{u \,\in\, W(\Omega)} \left\{ \frac{1}{2} \int_\Omega \frac{K^* \mathbf{1}_\Sigma \left(u \;-\; \left(u_{k+\frac{1}{2}} \;+\; v^l \right) \right)^2}{u_k} \;+\; \alpha\, J(u) \right\}$$

with the update formula

$$v^{l+1} \;=\; v^l \;+\; \left(u_{k+\frac{1}{2}} \;-\; u_{k+1}^{l+1} \right) , \qquad v^0 \;=\; 0 \,.$$

However, this iteration scheme with the update formula seems rather related to an additive than a multiplicative setting and hence is less promising for our Poisson framework. Additionally, computational experiments indeed confirm that the inner refinement leads to worse reconstructions than the outer one, presented in *Section 4.7.3*.

4.7.6 A Further Refinement Approach

Finally, we briefly present an alternative refinement strategy, which iteratively improves the reconstruction results obtained by the (damped) FB-EM-REG algorithm, namely a *dual inverse scale space strategy* introduced by Brune et al. in [32]. As the name of the approach already implies, this strategy is based on a dual representation of the variational problem (4.16) and represent the dual counterpart of the primal inverse scale space strategy presented in *Section 4.7.3*. Analogous to the primal case, we apply the approach of the iterative Bregman distance regularization, here, however, to the dual formulation of (4.16). Subsequently, a further dual formulation of the shifted regularization strategy is computed such that we again obtain a simple primal (equal to bidual) iterative regularization technique, namely

$$u^{l+1} \in \operatorname*{arg\,min}_{\substack{u \in W(\Omega) \\ u \geq 0 \ a.e.}} \left\{ \int_\Sigma \big(Ku + r^l - f \log(Ku + r^l) \big) \, d\mu + \alpha \, J(u) \right\} , \tag{4.113}$$

with the following update formula for the residual function r^l ,

$$r^{l+1} = r^l + Ku^{l+1} - f , \qquad r^0 = 0 .$$

Fortunately, the minimization problem (4.113) has a simple structure, which is very similar to the initial variational problem (4.16) and (4.17). The crucial difference is that the objective functional in (4.113) has $D_{KL}(f, Ku + r^l)$ as data fidelity term and not $D_{KL}(f, Ku)$ as in the case of (4.16) and (4.17). Nevertheless, we can use the idea of the FB-EM-REG splitting strategy in *Section 4.4.1* and obtain in each refinement step $l = 0, 1, \ldots$ a strategy consisting of an modified EM step

$$u^{l+1}_{k+\frac{1}{2}} = \frac{u^{l+1}_k}{K^* \mathbf{1}_\Sigma} \, K^* \left(\frac{f}{K u^{l+1}_k + r^l} \right) ,$$

which is followed by solving a variational problem

$$u^{l+1}_{k+1} \in \operatorname*{arg\,min}_{u \in W(\Omega)} \left\{ \frac{1}{2} \int_\Omega \frac{K^* \mathbf{1}_\Sigma \big(u - u^{l+1}_{k+\frac{1}{2}} \big)^2}{u^{l+1}_k} + \alpha \, J(u) \right\} .$$

In comparison to the primal inverse scale space strategy proposed in *Section 4.7.3*, the dual approach (4.113) has an interesting interpretation with respect to a dynamical change of the background model based on the residual function r^l , which ensures the stepwise refinement of the reconstruction results.

However, the numerical results in [32] show that both inverse scale space methods compute very similar iterates and we could not recognize a difference in the performance so far. But in the case of the dual approach, we can provide error estimates and convergence rates for exact and noisy data (see [32]), which are not possible for the primal approach so far.

5 Imaging : Ultrasound (US) Speckle Framework

Based on ideas in *Chapter 4*, we introduce in this chapter a *variational regularization framework* for restoration problems in *medical ultrasound (US) imaging*. In this type of application, the measurements are corrupted with a signal-dependent noise, called *speckle noise*, which is fundamentally different as the commonly used additive Gaussian noise and the Poisson noise studied in the previous chapter. Hence, we consider here an adapted variational regularization model, which we derive using a *Bayesian model* and *maximum a-posteriori probability estimation*. Based on this model, which is highly nonlinear in the data fidelity term, we propose a robust and efficient numerical realization of this variational regularization problem, using a *forward-backward splitting approach*. Moreover, we also present a mathematical analysis of the US speckle based variational model and the proposed numerical algorithm.

5.1 Introduction

Over the last decades, *ultrasound imaging* has developed to one of the most important techniques in the field of medical diagnostic. This success can be traced back to its ability to be a non-invasive, low cost, and real-time application, which can be used in almost all medical fields. In particular, since ultrasound imaging is a low-risk and painless application, it can be used in sensible areas, such as in prenatal care or examination of kidneys and heart. However, the ultrasound images suffer from an acoustic noise called *speckle* [36], which causes the main degradation of the image quality and with it possibly the success of an ultrasound examination. These degradations are unavoidable *interference effects*, which will be caused by scattering of the ultrasound beam from tissue inhomogeneities [1]. These effects yield certain granular patterns, which distort the actual tissue structure. Hence, the interpretation of ultrasound images requires highly trained knowledge to derive essential information for diagnosis from an image.

In addition, the degradation by speckle noise reduces the success of an automatic image postprocessing, such as a segmentation of different anatomical structures in an image.

Due to the random scattering of the ultrasound beam, the measured data are of stochastic nature. In [163], the authors showed that the US speckle noise follows in many cases a Rayleigh distribution with mean proportional to the standard deviation. Hence, the speckle noise in this application can be modeled as *multiplicative noise* [158]. However, the noise in the displayed ultrasound images does not follow the Rayleigh distribution [159, 110] due to the modification of the original signal statistic by the ultrasound device, e.g. by logarithmic compression [94], low-pass filtering or interpolation of the signal. In [110], Loupas et al. showed that the typical linear relation between mean and standard deviation for Rayleigh distribution no longer holds for displayed ultrasound images and derived experimentally that the process of image degradation can be modeled as a signal-dependent noise of the form

$$f \;=\; \bar{u} \;+\; \sqrt{\bar{u}}\,\eta \;, \tag{5.1}$$

where $\bar{u}$ is the original signal and f the observed signal as introduced in *Section 2.1*, as well as η is a vector valued Gaussian distributed random variable with expectation value 0 and variance σ^2. However, deconvolution of ultrasound images (cf. e.g. [149] and the references therein) is also an relevant approach in this area, for instance to estimate the speed of sound in a media. Hence, we can generalize the data model (5.1) to a reconstruction model

$$f \;=\; K\bar{u} \;+\; \sqrt{K\bar{u}}\,\eta \;, \tag{5.2}$$

where K is a forward operator (e.g. convolution operator) as introduced in *Section 2.1.*

We can see in (5.2) that the model of the raw data in ultrasound imaging fundamentally differs from the commonly studied case of additive exponentially distributed raw data of the form $f \,=\, K\bar{u} + \eta$, where η is a Gaussian distributed random variable as in (5.1). Hence, from the viewpoint of statistical modeling of image reconstruction problems in *Section 2.2*, the ultrasound data model (5.2) leads to a more complicated form of the variational regularization problem. To derive an adapted minimization problem to (5.2), we use the general form of the negative log-likelihood function (2.9), obtained with the *Bayesian model* and the *maximum a-posteriori probability estimation*, as well as using a *Gibbs a-priori density* of the form (2.8). In this case, we have to characterize the probability $p(f|u)$ for the ultrasound data f in (5.2). Although the noise η in (5.2) is unknown, the distribution of η is assumed to be normal with mean 0 and variance σ^2 and we obtain

$$p(f|u) \;\sim\; e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\eta_i^2} \;=\; e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\frac{(f_i - (Ku)_i)^2}{(Ku)_i}} \;.$$

With this characterization, we obtain now the following negative log probability density of the noise,

$$- \log p(f|u) \;=\; \sum_{i=1}^{N} \frac{(f_i \;-\; (Ku)_i)^2}{(Ku)_i} \;, \tag{5.3}$$

where additive terms independent of u are neglected. At this point, we proceed analogously to *Section 4.2* and pass over from a discrete to a continuous representation of the data in order to have a simpler basis for the construction and analysis of the methods. In *Section 4.2*, we assumed that any element in the discrete data space $\mathbb{R}^N$ can be interpreted as samples of a function in $V(\Sigma)$. Then, using the negative log probability density of the noise in (5.3) and the Gibbs a-priori density in (2.8), the MAP estimate in (2.7) can be rewritten as the following continuous variational problem,

$$u_{MAP} \;\in\; \arg\min_{\substack{u \,\in\, W(\Omega) \\ u \,\geq\, 0 \;\; a.e.}} \left\{ \int_\Sigma \frac{(f \;-\; Ku)^2}{Ku} \, d\mu \;+\; \alpha \, J(u) \right\} \;, \tag{5.4}$$

where μ is a point measure and $d\mu \;=\; \sum_{i=1}^N \chi_{M_i} \; d\lambda$ with Lebesque measure λ and indicator function χ_{M_i} as defined in (4.11).

5.2 Reconstruction Method

In this section, we consider a *variational regularization framework* adapted to the *speckle noise* occuring in *medical ultrasound (US) imaging*. In this framework, we use a convex regularization functional J, which can be singular as in the Poisson framework discussed in *Section 4.4*. The specific choice of the total variation regularization and its nonlocal extension will then be discussed in *Sections 6.3* and *7.4*. Due to the nondifferentiability of singular regularization energies in the common sense and the strong nonlinearity of the data fidelity term, we propose a robust numerical scheme to solve the US speckle noise based variational regularization problem (5.4). The algorithm is based on a *forward-backward splitting approach* and can be realized by alternating a reconstruction step with an almost standard denoising step as encountered in image processing.

5.2.1 (Damped) US-FB-REC-REG Algorithm

In the following, we develop a numerical algorithm for the US speckle noise adapted variational regularization problem (5.4),

$$\min_{\substack{u \,\in\, W(\Omega) \\ u \,\geq\, 0 \;\; a.e.}} \underbrace{\int_\Sigma \frac{(f \;-\; Ku)^2}{Ku} \, d\mu}_{=:\, D_{US}(f, Ku)} \;+\; \alpha \, J(u) \;, \qquad \alpha \;>\; 0 \;, \tag{5.5}$$

where J is a convex regularization functional. We proceed analogously to *Section 4.4.1*, where we proposed an alternating iteration strategy to solve the Poisson likelihood estimation problem (4.16). To compute the first order optimality condition of the minimization problem (5.5), we extend the functional $J : W(\Omega) \to \mathbb{R}_{\geq 0}$ to a convex functional on $U(\Omega)$, $W(\Omega) \subset U(\Omega)$, analogous to (4.18) and compute the subdifferential of the objective functional in (5.5) (see *Definition 3.2.6*), denoted by ∂. Due to the continuity and convexity (cf. *Lemma 5.4.1*) of the data fidelity term D_{US} in (5.5), the subdifferential calculus in *Lemma 3.2.13* yields the following identity

$$\partial \Big(D_{US}(f, Ku) \; + \; \alpha \, J(u) \Big) \;=\; \partial_u \, D_{US}(f, Ku) \; + \; \alpha \, \partial J(u) \ ,$$

where the subdifferentials $\partial_u \, D_{US}(f, Ku)$ of D_{US} are singletons given by

$$\partial_u \, D_{US}(f, Ku) \;=\; \left\{ K^* \mathbf{1}_\Sigma \; - \; K^* \left(\frac{f}{Ku} \right)^2 \right\} . \tag{5.6}$$

Hence, the first order optimality condition of (5.5) for a positive solution u is given by

$$K^* \mathbf{1}_\Sigma \; - \; K^* \left(\frac{f}{Ku} \right)^2 \; + \; \alpha \, p \;=\; 0 \ , \qquad p \;\in\; \partial J(u) \ . \tag{5.7}$$

However, we additionally have to regard the positivity constraint in (5.5). Hence, the Karush-Kuhn-Tucker conditions [89, Thm. 2.1.4] provide the existence of a Lagrange multiplier $\lambda \geq 0$, such that the stationary points of the functional in (5.5) need to fulfill

$$\begin{aligned} 0 \;&\in\; K^* \mathbf{1}_\Sigma \; - \; K^* \left(\frac{f}{Ku} \right)^2 \; + \; \alpha \, \partial J(u) \; - \; \lambda \ , \\ 0 \;&=\; \lambda \, u \ . \end{aligned} \tag{5.8}$$

By multiplying the first equation in (5.8) by u, the Lagrange multiplier λ can be eliminated by the second equation and the subsequent division by $K^* \mathbf{1}_\Sigma$ leads to a fixed point equation of the form

$$u \; - \; \frac{u}{K^* \mathbf{1}_\Sigma} \, K^* \left(\frac{f}{Ku} \right)^2 \; + \; \alpha \, \frac{u}{K^* \mathbf{1}_\Sigma} \, p \;=\; 0 \ , \qquad p \;\in\; \partial J(u) \ . \tag{5.9}$$

To design an iteration scheme, we use a semi-implicit approach based on (5.9) and obtain a sequence of iteration steps of the form

$$u_{k+1} \; - \; \frac{u_k}{K^* \mathbf{1}_\Sigma} \, K^* \left(\frac{f}{Ku_k} \right)^2 \; + \; \alpha \, \frac{u_k}{K^* \mathbf{1}_\Sigma} \, p_{k+1} \;=\; 0 \ , \qquad p_{k+1} \;\in\; \partial J(u_{k+1}) \ . \tag{5.10}$$

This iteration approach has almost the same form as the numerical scheme proposed in the case of Poisson noisy data in (4.21), so that we can utilize the iteration strategy introduced in (4.24). According to this, the method (5.10) solving the variational problem (5.5) can be realized as a nested two step iteration of the form

$$\begin{aligned} u_{k+\frac{1}{2}} \; &= \; \frac{u_k}{K^* \mathbf{1}_\Sigma} \, K^* \left(\frac{f}{K u_k} \right)^2 \, , && \textit{(REC step)} \\ u_{k+1} \; &= \; u_{k+\frac{1}{2}} \; - \; \alpha \, \frac{u_k}{K^* \mathbf{1}_\Sigma} \, p_{k+1} \, , \qquad p_{k+1} \; \in \; \partial J(u_{k+1}) \, , && \textit{(REG step)} \end{aligned} \tag{5.11}$$

in which we alternate a reconstruction (REC) step with a suitable regularization (REG) step. In this iteration scheme, we can observe that the reconstruction half step in (5.11) has a light modified form of the EM algorithm presented in (4.15), where only a modification of the backprojected function $\frac{f}{Ku_k}$ is required. Moreover, the nesting in the present form leads to the fact that the regularization step in (5.11) coincides with the regularization step of the FB-EM-REG algorithm (4.24). Hence, the second half step from $u_{k+\frac{1}{2}}$ to u_{k+1} in (5.11) can be realized by solving the convex variational problem (4.25). In addition, a damping strategy in this regularization step can be introduced analogously to (4.26), which can be solved by minimizing the variational problem (4.27).

Finally, we note that the two step strategy (5.11) with its damping variant in (4.26) can be interpreted as an operator splitting algorithm. This is not surprising due to the analogy of this iteration scheme to the (damped) FB-EM-REG algorithm proposed in *Sections 4.4.1* and *4.4.2*. Hence, the nested iteration sequence (5.11) with the modified regularization step (4.26) can be formulated as a *forward-backward (FB) splitting algorithm* (see *Section 4.4.3*) using the following decomposition of the optimality condition (5.7),

$$0 \; \in \; C(u) \; := \; \underbrace{K^* \mathbf{1}_\Sigma \; - \; K^* \left(\frac{f}{K u} \right)^2}_{=: \; A(u)} \; + \; \underbrace{\alpha \, \partial J(u)}_{=: \; B(u)} \; .$$

For that reason, we denote the iteration scheme (5.11), which solved the regularized US speckle likelihood estimation problem (5.5), as *US-FB-REC-REG algorithm*.

5.2.2 Stopping Rules and Pseudocode

In order to derive appropriate stopping rules to guarantee the accuracy of the proposed US-FB-REC-REG algorithm, we again proceed analogously to the FB-EM-REG algorithm in *Section 4.4.4*. According to this, we consider the maximum number of iterations and the error in the optimality condition (5.7) as basic stopping criteria. The latter one will be measured in a weighted norm $\| \cdot \|_{2,\mathrm{w}}$ defined in (4.30) and has the form (cf. (4.31))

$$opt_{k+1} \; := \; \left\| K^* \mathbf{1}_\Sigma \; - \; K^* \left(\frac{f}{K u_{k+1}} \right)^2 \; + \; \alpha \, p_{k+1} \right\|^2_{2, u_{k+1}} \; . \tag{5.12}$$

In addition, we are also interested in the convergence of the sequence of primal functions (u_k) and the sequence of subgradients (p_k) with $p_k \in \partial J(u_k)$. Hence, we consider the damped regularization step (4.26) with the reconstruction step in (5.11),

$$u_{k+1} \; - \; \omega_k \, \frac{u_k}{K^* \mathbf{1}_\Sigma} \, K^* \left(\frac{f}{K u_k} \right)^2 \; - \; (1 - \omega_k) \, u_k \; + \; \omega_k \, \alpha \, \frac{u_k}{K^* \mathbf{1}_\Sigma} \, p_{k+1} \; = \; 0 \; .$$

By combining this iteration scheme with the optimality condition (5.9) evaluated at u_k, which must be fulfilled in the case of convergence, we obtain the optimality statement (4.32) for the sequences (u_k) and (p_k). Hence, the stopping criteria in (4.33) can be used to guarantee the accuracy of these sequences.

Following the results in *Section 5.2.1* and the stopping rules above, we can use *Algorithm 5.1* to solve the regularized US speckle noise based likelihood estimation problem (5.5).

Algorithm 5.1 *(Damped) US-FB-REC-REG Algorithm*

---

1. ***Parameters:*** f, $\alpha > 0$, $\omega \in (0,1]$, $maxRECIts \in \mathbb{N}$, $tol > 0$
2. ***Initialization:*** $k = 0$, $u_0 := c > 0$
3. ***Iteration:***

   while $(\, (\, k < maxRECIts \,)$ **and**
   $(\, opt_k \geq tol$ **or** $u_{opt_k} \geq tol$ **or** $p_{opt_k} \geq tol \,) \,)$ **do** ▷ (5.12), (4.33)

   i) Compute $u_{k+\frac{1}{2}}$ via reconstruction step in (5.11).

   ii) Set $\omega_k = \omega$.

   iii) Compute u_{k+1} via convex variational problem (4.27).

   iv) $k \leftarrow k + 1$

   end while
4. ***Return*** u_k

5.3 Image Denoising

In this section, we are interested in the problem of image denoising, which is a relevant issue in medical ultrasound imaging (cf. [110, 101, 92]). In [101], Krissian et al. proposed a total variation based model to denoise an ultrasound image corrupted by speckle noise, which can be generalized to the following minimization problem with an arbitrary convex regularization functional J,

$$\min_{\substack{u \in W(\Omega) \\ u \geq 0 \;\, a.e.}} \int_\Omega \frac{(f \,-\, u)^2}{u} \, d\mu \;+\; \alpha \, J(u) \; . \tag{5.13}$$

To solve this minimization problem with the total variation regularization, the authors in [101, 92] suggest a gradient descent algorithm based on the Euler-Lagrange equation. However, such an approach requires always an approximation of TV by differentiable functionals (4.6) and needs a severe step size restriction. Here, we propose a strategy based on the damped US-FB-REC-REG algorithm introduced in *Section 5.2.1* to solve the denoising problem (5.13).

We can see that the denoising problem (5.13) coincides with the US speckle noise based likelihood reconstruction model (5.5) in the case of identity operator K. Hence, we can use the US-FB-REC-REG splitting strategy (5.11) with the damped modification in (4.26) in order to obtain a numerical iteration scheme for the US denoising problem (5.13). Since the reconstruction step in (5.11) simplifies in the case of identity operator K to $u_{k+\frac{1}{2}} = \frac{f^2}{u_k}$, the damped regularization step in (4.26) results in the following iteration sequence,

$$u_{k+1} \;=\; (1 - \omega_k) \, u_k \;+\; \omega_k \, \frac{f^2}{u_k} \;-\; \omega_k \, \alpha \, u_k \, p_{k+1} \; , \qquad p_{k+1} \;\in\; \partial J(u_{k+1}) \; , \tag{5.14}$$

with $\omega_k \in (0, 1]$. This step can be realized by solving a convex variational problem of the form (cf. (4.27))

$$u_{k+1} \;\in\; \operatorname*{arg\,min}_{u \,\in\, W(\Omega)} \left\{ \frac{1}{2} \int_\Omega \frac{\Big(u \;-\; \big(\omega_k \, \frac{f^2}{u_k} \;+\; (1 - \omega_k) \, u_k\big)\Big)^2}{u_k} \;+\; \omega_k \, \alpha \, J(u) \right\} \; . \tag{5.15}$$

In the undamped case (i.e. $\omega_k = 1$), the algorithm (5.14) represents a semi-implicit iteration scheme with respect to the optimality condition of (5.13), which is given by

$$u \, (\, 1 \;-\; \frac{f^2}{u^2} \;+\; \alpha \, p \,) \;=\; 0 \; , \qquad p \;\in\; \partial J(u) \; , \tag{5.16}$$

and thus actually realized a denoised image in medical ultrasound imaging. Note that the optimality condition (5.16) is obtained via the Karush-Kuhn-Tucker conditions, analogous to (5.9) with identity operator K.

Remark. We have seen that the iteration scheme (5.14) solves the US denoising problem (5.13) by a sequence of modified L^2 variational models of the form (5.15). In this way, we obtain a maximum a-posteriori probability estimate, but unfortunately for the price of high computational efforte, which are comparable to the incorporated US-FB-REC-REG reconstruction strategy (5.11). Hence, motivated by the approximated denoising model in the case of Poisson noisy images in *Section 4.5.2*, one can consider the same approach to reduce the computational complexity of the computational sequence (5.15). Actually, using the Taylor linearization (4.38) with the function G_f setting as the data fidelity term in (5.13), we obtain an approximation of the form

$$\min_{\substack{u \in W(\Omega) \\ u \geq 0 \;\; a.e.}} \int_\Omega \frac{(f - u)^2}{f} \, d\mu + \alpha \, J(u) . \tag{5.17}$$

The use of this approximation only makes sense if the given noisy image f is strictly positive, since in the case of negative values the minimization problem becomes a maximization problem. Formally, the given noisy image can be negative due to the signal dependent perturbation of the form (5.1), however in practice the displayed ultrasound images are positive so that the approximation (5.17) can be used to reduce the complexity of the sequence (5.15).

5.4 Analysis

In this section, we carry out a mathematical analysis of the US speckle noise based variational model (5.5). In the case of identity operator K and total variation regularization, an existence and uniqueness proof as well as the positivity preservation of a solution was given in [92]. In this work, we concentrate on the mathematical analysis of the general reconstruction problem (5.5) with respect to the following aspects:

- We propose a theoretical framework for an arbitrary convex regularization functional J, also including singular energies such as the total variation or general ℓ^1- or L^1-type functionals. For such energies we prove the well-posedness, i.e. the existence, uniqueness, and stability of a solution with respect to the variational regularization model (5.5).

- We study the damped US-FB-REC-REG algorithm proposed in *Section 5.2.1* with respect to preservation of positivity of a solution and a stable convergence behavior of this iteration scheme.

5.4.1 Properties of Data Fidelity Term

For the analysis of the regularized US speckle noise based likelihood estimation problem (5.5), we first study the data fidelity term $D_{US} : L^2(\Sigma) \times L^2(\Sigma) \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$ of (5.5) defined by

$$D_{US}(\varphi, \psi) := \int_\Sigma \frac{(\varphi - \psi)^2}{\psi} \, d\nu \qquad \text{for all} \qquad \psi \geq 0 \ \ a.e. \,, \tag{5.18}$$

where ν is a measure. Note that the integrand is nonnegative and vanishes if and only if $\varphi = \psi$. In the following, we will study some analytical results of this functional, which will be necessary in the following analysis.

Lemma 5.4.1 (*Properties of Data Fidelity Term*). *Let $A : U(\Omega) \to V(\Sigma)$ be a linear operator between locally convex spaces $U(\Omega)$ and $V(\Sigma)$ (see Definition 3.1.17), such that $U(\Omega)$ and $V(\Sigma)$ are associated with topologies τ_U and τ_V. Additionally, we assume that the operator A is sequentially continuous (see Definition 3.1.5) with respect to the topologies τ_U and τ_V and that $V(\Sigma)$ is continuously embedded in $L^2(\Sigma)$ (see Definition 3.1.23). Moreover, we suppose that the operator A preserves positivity, i.e. it satisfies $Au \geq 0$ a.e. for any $u \geq 0$ a.e.. Then, the following statements hold:*

(i) For any fixed $\varphi \in L^2(\Sigma)$, the function $\psi \mapsto D_{US}(\varphi, \psi)$ is convex and thus, due to the linearity of the operator A, the function $u \mapsto D_{US}(\varphi, Au)$ is also convex.

(ii) For any fixed $\varphi \in L^2(\Sigma)$, the function $u \mapsto D_{US}(\varphi, Au)$ is lower semicontinuous with respect to the topology τ_U.

Proof. ***(i)*** The convexity of the function $\psi \mapsto D_{US}(\varphi, \psi)$ follows directly from the form of the second directional derivative of $D_{US}(\varphi, \cdot)$, namely $(D_{US}(\varphi, \cdot))''(\psi) = \frac{2\varphi^2}{\psi^3}$, which is positive for any $\psi \geq 0$ a.e. ***(ii)*** Fix a function $\varphi \in L^2(\Sigma)$. Let (u_n) be a sequence in the domain of the function $v \mapsto D_{US}(\varphi, Av)$ which converges in the topology τ_U to some $u \in \{ v \in U(\Omega) : v \geq 0 \ a.e. \}$. Then, due to the sequential continuity of the operator A with respect to the topologies τ_U and τ_V as well as the continuous embedding of $V(\Sigma)$ in $L^2(\Sigma)$, we obtain the convergence of the sequence (Au_n) to Au in the norm topology on $L^2(\Sigma)$ as well as the pointwise convergence almost everywhere on Σ. Thus, the sequence $\big((\varphi - Au_n)^2 / Au_n \big)$ converges almost everywhere to $(\varphi - Au)^2 / Au$ and by applying Fatou's Lemma we obtain

$$\int_\Sigma \frac{(\varphi - Au)^2}{Au} \, d\nu \leq \liminf_{n \to \infty} \int_\Sigma \frac{(\varphi - Au_n)^2}{Au_n} \, d\nu \,.$$

This inequality means that the function $v \mapsto D_{US}(\varphi, Av)$ is lower semicontinuous with respect to the topology τ_U. $\square$

5.4.2 Assumptions

In this section, we discuss the necessary foundations for the analysis of the regularized US speckle noise based likelihood estimation problem (5.5). Since almost all required assumptions on the forward operator K and the regularization functional J are identical with the ones described in *Assumptions 4.6.5*, we will only point out the few necessary changes in the following.

At the beginning, we extend the convex regularization funtional $J : W(\Omega) \to \mathbb{R}_{\geq 0}$ on a Banach space $W(\Omega) \subset U(\Omega)$ to a convex functional on Banach space $U(\Omega)$ using the extension (4.45) (cf. *Section 4.6.2*). Then, we can also extend the admissible solution of the minimization problem (5.5) from $W(\Omega)$ to $U(\Omega)$ and denote for the following analysis the objective functional with $F(u)$,

$$\min_{\substack{u \in U(\Omega) \\ u \geq 0 \ a.e.}} F(u) \ := \ D_{US}(f, Ku) \ + \ \alpha \, J(u) \ , \qquad \alpha \ > \ 0 \ , \tag{5.19}$$

where the functional D_{US} is defined in (5.18). Now, based on the discussion in *Section 4.6.2*, we make the following assumptions.

Assumption 5.4.2. *We demand the conditions given in Assumption 4.6.5, excluding the item (ii), which we replace in the case of the data fidelity term* (5.18) *by*

(ii) The Banach space $V_\mu(\Sigma)$ is continuously embedded in $L^2_\mu(\Sigma)$ (Definition 3.1.23), where $L^2_\mu(\Sigma)$ is the Lebesque space $L^2(\Sigma)$ with respect to the measure μ .

5.4.3 Well-Posedness of Minimization Problem

In the following, we verify the existence, uniqueness, and stability of the regularized US speckle noise based likelihood estimation problem represented by the minimization of (5.19).

Theorem 5.4.3 (Existence of Minimizers). *Let $U(\Omega)$, $V_\mu(\Sigma)$, K , J , and F satisfy Assumption 5.4.2. Assume that $\alpha > 0$ and $f \in V_\mu(\Sigma)$. Then, the functional F defined in* (5.19) *has a minimizer.*

Proof. Due to the positivity of the data fidelity term D_{US} and the lower semicontinuity of the function $v \mapsto D_{US}(f, Kv)$, $f \in V_\mu(\Sigma)$ fix, with respect to the topology τ_U in *Lemma 5.4.1*, the proof works analogously to the proof of *Theorem 4.6.8*. □

Next, we consider the uniqueness of the minimizers, for which it is enough to verify the strict convexity of the objective functional F defined in (5.19). For this purpose, it is straight-forward to see that the function $u \mapsto D_{US}(f, Ku)$ is strictly convex for functions with $\inf_\Omega u > 0$, if the operator K is injective, i.e. the null space is trivial since K is linear (cf. *Assumption 4.6.5*). Therefore, we can immediately conclude the following result.

Theorem 5.4.4 (*Uniqueness of Minimizers*). *Let $U(\Omega)$, $V_\mu(\Sigma)$, K, J, and F satisfy Assumption 5.4.2. Assume that K is an injective operator and $f \in V_\mu(\Sigma)$. Then, the function $u \mapsto D_{KL}(f, Ku)$ and also the functional F from* (4.46) *is strictly convex for functions u fulfilling $\inf_\Omega u > 0$. In particular, the minimizer of F is unique.*

Finally, we show the stability of the regularized US estimation problem (5.19) with respect to a certain kind of data perturbations. For the necessity of a stability result compare the discussion above *Theorem 4.6.10*.

Theorem 5.4.5 (*Stability with Respect to Perturbations in Measurements*). *Let $U(\Omega)$, $V_\mu(\Sigma)$, K, J, and F satisfy Assumption 5.4.2. Fix $\alpha > 0$ and assume that the functions $f_n \in V_\mu(\Sigma)$, $n \in \mathbb{N}$, are approximations of a data function $f \in V_\mu(\Sigma)$ in the $L^2_\mu(\Sigma)$-norm, i.e.*

$$\lim_{n \to \infty} \| f_n - f \|_{L^2_\mu(\Sigma)} = 0 . \tag{5.20}$$

Moreover, let

$$u_n \in \operatorname*{arg\,min}_{\substack{v \in U(\Omega) \\ v \geq 0 \text{ a.e.}}} \left\{ F_n(v) := D_{US}(f_n, Kv) + \alpha J(v) \right\} , \qquad n \in \mathbb{N} , \tag{5.21}$$

and u a solution of the regularized problem (5.19) *corresponding to the data function f. Additionally, we assume that f belongs to the function space $L^\infty_\mu(\Sigma)$ and there exists a positive constant c such that*

$$0 < c \leq Ku \qquad \text{a.e. on } \Sigma . \tag{5.22}$$

Then, the problem (5.19) *is stable with respect to perturbations in the data, i.e. the sequence (u_n) has a τ_U-convergent subsequence and every τ_U-convergent subsequence converges to a minimizer of the functional F in* (5.19).

Proof. The proof is similar to the one of *Theorem 4.6.10*, where we proved the stability of the Poisson likelihood estimation problem (4.46), and only modifications at the beginning of the proof are required. According to the latter one, we first show the uniform boundedness of the sequence $\big(J(u_n)\big)$ for the existence of a τ_U-convergent subsequence of (u_n). Let $\alpha > 0$ fix, then the positivity of the functional D_{US} and the definition of u_n as a minimizer of the objective functional F_n in (5.21) imply that

$$J(u_n) \;\leq\; D_{US}(f_n, Ku_n) \;+\; \alpha\, J(u_n) \;\leq\; D_{US}(f_n, Ku) \;+\; \alpha\, J(u) \;, \quad \forall n \in \mathbb{N} \;. \tag{5.23}$$

Hence, the sequence $\big(J(u_n)\big)$ is bounded, if the sequence $\big(D_{US}(f_n, Ku)\big)$ on the right-hand side of (5.23) is bounded. To show this, we use condition (5.20) and obtain with the Cauchy-Schwarz inequality the strong convergence of the sequence (f_n) to f in the $L^1_\mu(\Sigma)$-norm,

$$\lim_{n \to \infty} \| f \;-\; f_n \|_{L^1_\mu(\Sigma)} \;=\; 0 \;. \tag{5.24}$$

Thus, the condition (5.20) implies together with the inequality

$$\begin{aligned}
&\big|\, D_{US}(f_n, Ku) \;-\; D_{US}(f, Ku) \,\big| \\
&\quad= \big|\, D_{US}(f_n \;-\; f \;+\; f, Ku) \;-\; D_{US}(f, Ku) \,\big| \\
&\quad= \left| \int_\Sigma \frac{(f_n \;-\; f)^2}{Ku} \, d\mu \;+\; 2 \int_\Sigma \left(\frac{f}{Ku} \;-\; 1 \right) (f_n \;-\; f) \, d\mu \right| \\
&\quad\leq \underbrace{\left\| \frac{1}{Ku} \right\|_{L^\infty_\mu(\Sigma)}}_{<\,\infty} \underbrace{\| f \;-\; f_n \|^2_{L^2_\mu(\Sigma)}}_{\overset{(5.20)}{\to}\, 0} \;+\; 2\, \underbrace{\left\| \frac{f}{Ku} \;-\; 1 \right\|_{L^\infty_\mu(\Sigma)}}_{<\,\infty} \underbrace{\| f \;-\; f_n \|_{L^1_\mu(\Sigma)}}_{\overset{(5.24)}{\to}\, 0} \;,
\end{aligned}$$

the following convergence

$$\lim_{n \to \infty} D_{US}(f_n, Ku) \;=\; D_{US}(f, Ku) \;. \tag{5.25}$$

Note that the boundedness of $\frac{1}{Ku}$ follows from assumption (5.22) and the boundedness of $\frac{f}{Ku} - 1$ is derived from the assumption that f belongs to the function space $L^\infty_\mu(\Sigma)$.

With result (5.25), we can now proceed analogously to the proof of *Theorem 4.6.10*, replacing the Kullback-Leibler functional D_{KL} by D_{US} and noting that the space $V_\mu(\Sigma)$ is now continuously embedded in $L^2_\mu(\Sigma)$ (see *Assumption 5.4.2 (ii)*). □

5.4.4 Positivity Preservation of US-FB-REC-REG Algorithm

In the following, we analyze the positivity preservation of the US-FB-REC-REG splitting approach (5.11) and its damped modification (4.26). Since the regularization steps in this strategy coincides with the regularization steps of the damped FB-EM-REG algorithm, we can use the results obtained in *Section 4.6.4*, particularly the maximum principle for the weighted L^2 regularization problem (4.57) in *Lemma 4.6.11*.

Lemma 5.4.6 (*Positivity of (damped) US-FB-REC-REG Algorithm*). *Let (ω_k) be a given sequence of damping parameters with $\omega_k \in (0,1]$ for all $k \geq 0$, the initialization function u_0 be strictly positive and let the data function f lie in $L^\infty_\mu(\Sigma)$. Additionally, we assume that the operator K satisfies the positivity preservation in Assumption 4.6.5 (v) and that the adjoint operator K^* fulfills*

$$\begin{aligned} K^* g \ > \ 0 \ , \quad \forall g \ \in \ V_\mu(\Sigma) \qquad\qquad\qquad\qquad\qquad\qquad \\ \textit{such that} \quad g \ \geq \ 0 \quad \textit{and} \quad \exists x \quad \textit{with} \quad g(x) \ > \ 0 \ . \end{aligned} \tag{5.26}$$

Moreover, we suppose that the regularization functional J fulfills the condition (4.58) *for the maximum principle in Lemma 4.6.11. Then, each half step of the (damped) US-FB-REC-REG splitting method and therewith also the solution is strictly positive.*

Proof. Since $u_0 > 0$ and the operator K does not effect the strict positivity, we have $K u_0 > 0$ and thus at least a point x such that $(f \,/\, K u_0)^2(x) > 0$. With assumption (5.26), we obtain that the first reconstruction step $u_{\frac{1}{2}}$ in (5.11) is strictly positive. Because the regularization step in (5.11) can be realized via the convex variational problem (4.25), the maximum principle in *Lemma 4.6.11* using $q := u_{\frac{1}{2}} > 0$ and $h := \frac{u_0}{K^* \mathbf{1}_\Sigma} > 0$ yields $u_1 > 0$. With the same argument, we also obtain $u_1 > 0$ if we take the damped regularization step (4.26) via the variational problem (4.27), using the maximum principle with $q := \omega_0 \, u_{\frac{1}{2}} + (1 - \omega_0) \, u_0 > 0$ for $\omega_0 \in (0,1]$ and $h := \frac{u_0}{K^* \mathbf{1}_\Sigma} > 0$. Inductively, the strict positivity of the whole nested iteration sequence (u_k) and with it the strict positivity of the solution is obtained by the same arguments using *Lemma 4.6.11.* ☐

In the context of Poisson distributed data, we analyzed the positivity preservation of the (damped) FB-EM-REG splitting algorithm in *Section 4.6.4*. In this context, we could also prove a maximum principle and positivity preservation of the Poisson denoising strategy explicitely in *Lemma 4.6.13*. In the case of the US framework, it is not possible to provide the same procedure. The difficulty lies in the form of the “noisy” image $u_{k+\frac{1}{2}} = \omega_k \frac{f^2}{u_k} + (1 - \omega_k) \, u_k$, which will be restored during the denoising iteration sequence in (5.15). Since this term contains the function $\frac{f^2}{u_k}$, we could not find a suitable estimation so far such that we can proceed analogously to *Lemma 4.6.13* in order to obtain a maximum principle for the US denoising problem similar to (4.63) using *Lemma 4.6.11*. However, *Lemma 5.4.6* gives us also a positivity preservation result for the US denoising strategy (5.15) if we restrict the consideration of the noisy image f to the points, where the absolute values of f are strictly positive.

Corollary 5.4.7 (*Positivity of US Denoising Strategy*). *Let* (ω_k) *be a given sequence of damping parameters with* $\omega_k \in (0,1]$ *for all* $k \geq 0$ *and the initialization function* u_0 *be strictly positive. Furthermore, we assume that the data function* f *lies in* $L^\infty_\mu(\Omega)$ *and we define a set*

$$M \; := \; \{ \, x \, \in \, \Omega \; : \; |f(x)| \; > \; 0 \, \} \; .$$

Moreover, we suppose that the regularization functional J *fulfills the condition* (4.58) *for the maximum principle in Lemma 4.6.11. Then, each half step of the (damped) US denoising method* (5.15) *and with it also the solution is strictly positive on* M *.*

Proof. The result follows direct from *Lemma 5.4.6*, since the identity operator K certainly fulfills the condition (5.26) on the point set M . □

5.4.5 Convergence of Damped US-FB-REC-REG Algorithm

In this section, we show analogously to *Theorem 4.6.14* the convergence of the damped US-FB-REC-REG splitting algorithm under appropriate assumptions on the damping parameters ω_k .

Theorem 5.4.8 (*Convergence of Damped US-FB-REC-REG Algorithm*). *Let* $U(\Omega)$ *,* $V_\mu(\Sigma)$ *,* K *,* J *, and* F *satisfy Assumption 5.4.2. Moreover, let* (u_k) *be a sequence of iterates obtained by the damped US-FB-REC-REG algorithm* (5.11) *with the damped regularization step* (4.26)*. Regarding this sequence of iterates, we make additional assumptions:*

- *The data function* f *lies in* $L^\infty_\mu(\Sigma)$ *and the adjoint operator* K^* *fulfills the condition* (5.26)*.*
- *The regularization functional* J *fulfills the condition* (4.58) *for the maximum principle in Lemma 4.6.11.*

Now, if there exists a sequence of corresponding damping parameters (ω_k) *,* $\omega_k \in (0,1]$ *, satisfying the inequality*

$$\omega_k \; \leq \; \frac{\displaystyle\int_\Omega \frac{K^* \mathbf{1}_\Sigma \, (u_{k+1} \, - \, u_k)^2}{u_k} \, d\lambda}{\displaystyle\sup_{v \, \in \, [u_k, u_{k+1}]} \int_\Sigma \frac{f^2 \, (K u_{k+1} \, - \, K u_k)^2}{(Kv)^3} \, d\mu} \; (1 \, - \, \epsilon) \; , \qquad \epsilon \; \in \; (0,1) \; , \tag{5.27}$$

then the objective functional F *defined in* (5.19) *is decreasing during the iteration. If we additionally suppose the four assumptions introduced in Theorem 4.6.14, then the* *sequence of iterates* (u_k) *has a* τ_U *-convergent subsequence and every* τ_U *-convergent subsequence converges to a minimizer of the functional* F *defined in* (5.19).

Proof. The proof is build up analogously to the one of the damped FB-EM-REG algorithm in *Theorem 4.6.14*. Thus, we distinguish only the steps where a change is required.

First step: Monotone descent of the objective functional
To get a descent of the objective functional F, we look for a condition on the damping parameters (ω_k), which guarantees a descent of the form (4.69) for all $k \geq 0$. To show this condition, we add the difference $D_{US}(f, Ku_{k+1}) - D_{US}(f, Ku_k)$ on both sides of the inequality (4.70). Considering the definition of the data fidelity term D_{US} in (5.18) and the objective functional F in (5.19) as well as using the relation given in (4.72), we then obtain the following inequality (cf. (4.71)),

$$\begin{aligned} F(u_{k+1}) \;-\; F(u_k) \;+\; \frac{1}{\omega_k} \int_\Omega \frac{K^* \mathbf{1}_\Sigma \, (u_{k+1} \;-\; u_k)^2}{u_k} \, d\lambda & \\ \leq \int_\Sigma \left(\frac{(f \;-\; Ku_{k+1})^2}{Ku_{k+1}} \;-\; Ku_{k+1} \;-\; \frac{(f \;-\; Ku_k)^2}{Ku_k} \;+\; Ku_k \right) d\mu & \\ + \int_\Omega \left(K^* \left(\frac{f}{Ku_k} \right)^2 (u_{k+1} \;-\; u_k) \right) d\lambda \; . & \end{aligned} \tag{5.28}$$

Using an auxiliary functional $G : U(\Omega) \to \mathbb{R} \cup \{+\infty\}$ defined as

$$G(u) \;:=\; \int_\Sigma \left(\frac{(f \;-\; Ku)^2}{Ku} \;-\; Ku \right) d\mu \; ,$$

the right-hand side of (5.28) can be formally interpreted as a Taylor linearization of G of the form (4.73). In this linearization, we compute the second directional derivative $G''(u;\, w_1,\, w_2)$ using the function $\phi_{w_2}(t) \;:=\; G'(u \;+\; t\, w_2;\, w_1)$ for any $w_2 \;\in\; U(\Omega)$,

$$\begin{aligned} G''(u;\, w_1,\, w_2) \;=\; \phi'_{w_2}(t) \Big|_{t\,=\,0} \;&=\; - \int_\Sigma \frac{\partial}{\partial t} \left(\frac{f^2}{(Ku \;+\; t\, Kw_2)^2} \, Kw_1 \right) d\mu \, \Bigg|_{t\,=\,0} \\ &=\; 2 \int_\Sigma \frac{f^2 \, Kw_2 \, Kw_1}{(Ku)^3} \, d\mu \; . \end{aligned}$$

Following the proof of the damped FB-EM-REG algorithm, we obtain the descent of the objective functional F if

$$\sup_{v \,\in\, [u_k, u_{k+1}]} \int_\Sigma \frac{f^2 \, (Ku_{k+1} \;-\; Ku_k)^2}{(Kv)^3} \, d\mu \;\leq\; \frac{1 \;-\; \epsilon}{\omega_k} \int_\Omega \frac{K^* \mathbf{1}_\Sigma \, (u_{k+1} \;-\; u_k)^2}{u_k} \, d\lambda \; .$$

By solving this inequality for ω_k, we obtain the required condition (5.27) for the damping parameters (ω_k) in order to have a descent of the objective functional F. By a suitable choice of ϵ in (5.27), we additionally can guarantee that $\omega_k \leq 1$ for all $k \geq 0$.

Second and third step: Convergence of the primal and dual iterates
These partial steps can both be proven analogously to the damped FB-EM-REG algorithm in *Theorem 4.6.14*.

Fourth step: Show that $p \in \partial J(u)$
The proof of this partial step is similar to the fourth step of the damped FB-EM-REG algorithm in *Theorem 4.6.14*, except a slight difference. Due to the difference of the reconstruction half steps in (4.24) and (5.11), we have to replace the terms $\frac{f}{Ku_{k_l}}$ and $\frac{f}{Ku}$ by $\left(\frac{f}{Ku_{k_l}}\right)^2$ and $\left(\frac{f}{Ku}\right)^2$ respectively.

Fifth step: Convergence to a minimizer of the objective functional
For this partial step the same changes have to be made as for the fourth step. □

5.5 Iterative Refinement via Bregman Distance Iteration

Based on the ideas in *Section 4.7.3*, we present an iterative refinement approach for the US variational model (5.5) in this section using *inverse scale space methods* based on *Bregman distance iteration*. As a recall, such a refinement is in general desirable due to the *systematic errors* of variational regularization techniques, which cause oversmoothing effects in the reconstructions (see *Section 4.7.2*).

5.5.1 Bregman-US-FB-REC-REG Algorithm

In the following, we briefly describe the iterative refinement approach for the US data based likelihood estimation problem (5.5). The procedure is the same as in the case of the Poisson data based variational reconstruction model proposed in *Section 4.7.3* and only slight modifications in the resulting form of the iteration scheme are required. According to *Section 4.7.3*, an iterative refinement is realized by a sequence of modified US data likelihood estimation problems based on (5.5), namely (cf. (4.100))

$$u^{l+1} \in \operatorname*{arg\,min}_{\substack{u \in W(\Omega) \\ u \geq 0 \ a.e.}} \left\{ D_{US}(f, Ku) \ + \ \alpha \, D_J^{p^l}(u, u^l) \right\} , \qquad p^l \in \partial J(u^l) , \tag{5.29}$$

where $D_J^{p^l}(u, u^l)$ is the *generalized Bregman distance* defined in *Definition 4.7.1.* Due to the convexity of the Bregman distance in the first argument (see *Lemma 4.7.2*) and the continuity of the data fidelity function D_{US}, the Karush-Kuhn-Tucker conditions provide the existence of a Lagrange multiplier $\lambda \geq 0$, such that the stationary points of the functional in (5.29) need to fulfill (cf. (4.101) using the subdifferential of D_{US} given in (5.6))

$$
\begin{aligned}
0 &\in K^* \mathbf{1}_\Sigma - K^* \left(\frac{f}{K u^{l+1}} \right)^2 + \alpha \left(\partial J(u^{l+1}) - p^l \right) - \lambda , \\
0 &= \lambda \, u^{l+1} ,
\end{aligned}
$$

with $p^l \in \partial J(u^l)$. These both conditions lead to a fixed point equation of the form (cf. (4.102))

$$
0 \in u^{l+1} - \frac{u^{l+1}}{K^* \mathbf{1}_\Sigma} K^* \left(\frac{f}{K u^{l+1}} \right)^2 + \alpha \frac{u^{l+1}}{K^* \mathbf{1}_\Sigma} \left(\partial J(u^{l+1}) - p^l \right) , \tag{5.30}
$$

with $p^l \in \partial J(u^l)$ and a well defined update formula for the iterates p^l, namely

$$
p^{l+1} := p^l - \frac{1}{\alpha} \left(K^* \mathbf{1}_\Sigma - K^* \left(\frac{f}{K u^{l+1}} \right)^2 \right) \in \partial J(u^{l+1}) , \tag{5.31}
$$

with u^0 constant and $p^0 := 0 \in \partial J(u^0)$. Based on equation (5.30), we can apply the idea of the nested two step iteration (5.10) in every refinement step $l = 0, 1, \dots$ and obtain a strategy consisting of a reconstruction step

$$
u^{l+1}_{k+\frac{1}{2}} = \frac{u^{l+1}_k}{K^* \mathbf{1}_\Sigma} K^* \left(\frac{f}{K u^{l+1}_k} \right)^2 , \tag{5.32}
$$

followed by solving an adapted variational regularization problem (4.105). To transfer the shift term $\langle p^l, u \rangle$ in (4.105) to the data fidelity term, we use the scaling

$$
v^l := \frac{\alpha}{K^* \mathbf{1}_\Sigma} p^l
$$

and obtain from (5.31) the following update formula for the iterates v^l,

$$
v^{l+1} = v^l - \left(\mathbf{1}_\Omega - \frac{1}{K^* \mathbf{1}_\Sigma} K^* \left(\frac{f}{K u^{l+1}} \right)^2 \right) , \qquad v^0 = 0 , \tag{5.33}
$$

such that we can use the regularization half step (4.108) with its damped modification proposed in (4.109).

5.5.2 Stopping Rules and Pseudocode

To provide stopping criteria for the iterative Bregman distance regularization above, we proceed analogously to *Section 4.7.4*, where we proposed stopping rules for the Bregman refinement of the FB-EM-REG algorithm. In the case of an US data based problem, we have to make two modifications. First, we stop the outer Bregman refinement iteration sequence using the *generalized discrepancy principle* with the US data based data fidelity term D_{US}, i.e. at the index $l_* = l_*(\delta, f)$ where the residual $D_{US}(f, Ku^{l_*})$ reaches the noise level δ or an estimate of the noise level, i.e.

$$l_* \;=\; \max\{\, l \;\in \mathbb{N} \;:\; D_{US}(f, Ku^l) \;\geq\; \tau\delta \,\} \;, \qquad \tau \;>\; 1 \;.$$

As second modification, the stopping criterion for the error in the optimality condition (5.30) is given by (cf. (4.110))

$$opt_{k+1}^{l+1} \;=\; \left\| K^* \mathbf{1}_\Sigma \;-\; K^* \left(\frac{f}{K u_{k+1}^{l+1}} \right)^2 - K^* \mathbf{1}_\Sigma \, v^l \;+\; \alpha \, p_{k+1}^{l+1} \right\|_{2, u_{k+1}^{l+1}}^2 \;. \tag{5.34}$$

Hence, we can use *Algorithm 5.2* to solve the stepwise refinement (5.29) of the regularized US data based likelihood estimation problem (5.5).

Algorithm 5.2 *(Damped) Bregman-US-FB-REC-REG Algorithm*

---

1. ***Parameters:*** $f\,,\ \alpha > 0\,,\ \omega \in (0,1]\,,\ maxBregIts \in \mathbb{N}\,,\ \delta > 0\,,\ \tau > 1\,,$
   $maxRECIts \in \mathbb{N}\,,\ tol > 0$
2. ***Initialization:*** $l = 0\,,\ u_0^1 = u_0 := c > 0\,,\ v^0 := 0$
3. ***Iteration:***

   while $(\ D_{US}(f, K u_0^{l+1}) \geq \tau\delta$ **and** $l < maxBregIts\ )$ **do**

   $a)$ Set $k = 0$.

   while $(\ (\ k < maxRECIts\ )$ **and**
   $(\ opt_k^{l+1} \geq tol$ **or** $u_{opt_k^{l+1}} \geq tol$ **or**
   $p_{opt_k^{l+1}} \geq tol\ )\ )$ **do** $\quad\triangleright$ (5.34), (4.111)

   $i)$ Compute $u_{k+\frac{1}{2}}^{l+1}$ via reconstruction step in (5.32).

   $ii)$ Set $\omega_k^{l+1} = \omega$.

   $iii)$ Compute u_{k+1}^{l+1} via convex variational problem (4.109).

   $iv)$ $k \leftarrow k+1$

   end while

   $b)$ Compute update v^{l+1} via (5.33).

   $c)$ Set $u_0^{l+2} = u_k^{l+1}$.

   $d)$ $l \leftarrow l+1$

   end while
4. ***Return*** u_0^{l+1}

---

6 Regularization : Total Variation (TV)

In the previous *Chapters 4* and *5*, we proposed two variational frameworks with the corresponding numerical strategies using a general convex regularization functional J so far. In the following we discuss now the use of *total variation (TV) functional* as regularization energy in both frameworks. The total variation functional is popular in many problems in the calculus of variations and plays in particular an important role in several fields of mathematical image processing. In the later application, the idea of TV regularization has been firstly introduced as a denoising technique by Rudin, Osher and Fatemi in [138], and has been generalized to various other imaging tasks such as deblurring, inpainting or segmentation subsequently [49]. The main feature of these resulting regularization techniques is the efficient realization of discontinuous solutions.

6.1 Functions of Bounded Variation

In most applications, edges in an image represent important and fundamental features of an object, which will be used to analyse the available information in the image. Mathematically, edges correspond to discontinuities of a function, such that we need the possibility to represent discontinuous functions in order to obtain a useful mathematical description of an imaging problem. Unfortunately, the classical Sobolev spaces do not allow to handle such requirements, since the weak gradient of a Sobolev function is a function again. In the case of a discontinuous function, the weak derivative can be interpreted as a measure and the space $BV(\Omega)$ of *functions of bounded variation* [4, 71, 80], can represent such functions. In image processing, the space $BV(\Omega)$ was introduced by Rudin, Osher and Fatemi [138] in the field of image denoising and has successively found applications in various other imaging tasks.

In this section, we recall the definition and basic properties of the space of functions of bounded variation, which are mainly collected from [2, 4, 71, 80, 143]. In the following, we shall assume that $\Omega \subset \mathbb{R}^d$ is an open set.

Definition 6.1.1 (*Total Variation (TV)*). *Let $u \in L^1(\Omega)$. The total variation (TV) of u in Ω is defined by*

$$\int_\Omega |Du| \;=\; \sup \left\{ \int_\Omega u \operatorname{div} g \; dx \;:\; g \in C_0^1(\Omega, \mathbb{R}^d),\; \|g\|_\infty \le 1 \right\}, \tag{6.1}$$

where $C_0^1(\Omega, \mathbb{R}^d)$ denotes the space of continuously differentiable functions from Ω to $\mathbb{R}^d$ with compact support in Ω and $\operatorname{div} g = \sum_{i=1}^d \frac{\partial g_i}{\partial x_i}$. The supremum norm inequality in (6.1) means that $\sup_{x \in \Omega} |g(x)| \le 1$, i.e. $|g(x)| \le 1$ for all $x \in \Omega$.

Remark. Since the space $C_0^\infty(\Omega, \mathbb{R}^d)$ of all arbitrarily often differentiable functions with compact support in Ω is dense in $C_0^1(\Omega, \mathbb{R}^d)$, we obtain in (6.1) the same supremum if we replace $C_0^1(\Omega, \mathbb{R}^d)$ by $C_0^\infty(\Omega, \mathbb{R}^d)$.

Example 6.1.2.

(1) If u belongs to the Sobolev space $W^{1,1}(\Omega) \subset L^1(\Omega)$, then the definition of a weak derivative yields

$$\int_\Omega u \operatorname{div} g \; dx \;=\; -\int_\Omega \nabla u \cdot g \; dx \qquad \textit{for every} \qquad g \in C_0^\infty(\Omega, \mathbb{R}^d),$$

so that

$$\int_\Omega |Du| \;=\; \int_\Omega |\nabla u(x)| \; dx, \tag{6.2}$$

where $\nabla u = (\nabla_{x_1} u, \dots, \nabla_{x_d} u)$ denotes the weak gradient of u.

(2) Let u be defined in $\Omega = (-1, +1)$ as the Heaviside function, i.e.

$$u(x) \;=\; \begin{cases} 0, & \textit{if} \quad x \in (-1, 0), \\ 1, & \textit{if} \quad x \in [0, +1). \end{cases}$$

Then,

$$\int_{-1}^{+1} u\, g' \; dx \;=\; g(0) \qquad \textit{and} \qquad \int_{-1}^{+1} |Du| \;=\; 1,$$

i.e. the distributional derivative Du of u is equal to the Dirac measure δ_0 in 0.

Definition 6.1.3 (*Space of Functions of Bounded Variation*). *A function $u \in L^1(\Omega)$ has a bounded variation in Ω, if the total variation of u in Ω is finite. The set of all* *such functions with bounded variation in* Ω *is denoted by* $BV(\Omega)$, *i.e.*

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

We call $BV(\Omega)$ *the space of functions of bounded variation, which is equipped with the norm*

$$\|u\|_{BV(\Omega)} \;:=\; \|u\|_{L^1(\Omega)} \;+\; |u|_{BV(\Omega)} \;, \tag{6.3}$$

where $|\cdot|_{BV(\Omega)}$ *is a seminorm defined by*

$$|u|_{BV(\Omega)} \;:=\; \int_\Omega |Du| \;.$$

Remark 6.1.4.

(1) It can be seen from *Example 6.1.2, Item (1)*, that $W^{1,1}(\Omega) \subseteq BV(\Omega)$. In addition, the fact that the two spaces are not equal, i.e. $W^{1,1}(\Omega) \subsetneq BV(\Omega)$, can be seen from *Example 6.1.2, Item (2)*. This is due to $BV(\Omega)$ containing step functions, whose derivatives Du are distributions and hence are in particular not regular functions.

(2) The definition of total variation in (6.1) is not unique for $d \geq 2$. Depending on the definition of the supremum norm $\|g\|_\infty = \sup_{x \in \Omega} |g(x)|_{\ell^s}$ with respect to different norms on $\mathbb{R}^d$ with $1 \leq s \leq \infty$, one obtains equivalent versions of the BV seminorm $|\cdot|_{BV(\Omega)}$. More precisely, we obtain a family of total variation seminorms defined by

$$\int_\Omega |Du|_{\ell^r} \;=\; \sup \left\{ \int_\Omega u \operatorname{div} g \; dx \;:\; g \;\in\; C_0^\infty(\Omega, \mathbb{R}^d) \,,\; |g|_{\ell^s} \;\leq\; 1 \text{ on } \Omega \right\} ,$$

for $1 \leq r < \infty$ and the Hölder conjugate index s, i.e. $r^{-1} + s^{-1} = 1$. The most common formulations are the *isotropic total variation* $(r = 2)$ and the *anisotropic total variation* $(r = 1)$. For the sake of completeness, we anticipate here that the different definitions of TV have effects on the structure of solutions obtained during the TV minimization. In the case of isotropic TV, corners in the edge set will not be allowed, whereas orthogonal corners are favored by the anisotropic variant (cf. e.g. *Fig. 6.1*). For a detailed analysis, we refer e.g. to [116, 68, 21, 153].

(3) If $u \in BV(\Omega)$, then the distributional gradient $Du = (D_1 u, \ldots, D_d u)$ of u can be identified with a vector valued Radon measure, see [6, pp. 39-40] or [71, Sect. 5.1, Thm. 1]. Thus, the BV space obtains also discontinuous functions,

since in contrast to the Sobolev spaces the measure Du needs not necessarily be represented by a Lebesque measurable function.

Lemma 6.1.5. *$BV(\Omega)$ is a Banach space with the norm $\|\cdot\|_{BV(\Omega)}$ defined in* (6.3).

Proof. See [80, p. 9]. $\square$

Fig. 6.1. *Effects of TV minimization depending on the not unique definition of TV (cf. Remark 6.1.4, Item (2)).* ***First row:*** *original image.* ***Second row:*** *TV regularized results obtaining with the isotropic TV definition (left) and anisotropic one (right). We can observe that in the case of isotropic TV, corners in the edge set will not be allowed, whereas orthogonal corners are favored by the anisotropic variant.*

For many applications, the norm topology proposed in (6.3) is unfortunately too strong to study certain properties in variational methods based on TV regularization such as compactness, so that other weaker topologies are needed. In *Section 3.1.3*, we already discussed such possibilities, namely the weak and the weak* topology, which are strictly weaker as the strong norm topology. Hence, since $BV(\Omega)$ is a normed linear space, we can directly use *Definition 3.1.29* to specify the weak topology on $BV(\Omega)$. However, this topology will be used rather less due to the fact that one can say very little about the dual space of BV, so that the weak convergence is hard to characterize. Moreover, the *Banach-Alaoglu Theorem 3.1.31* provides the compactness of a set only in the weak* topology. Hence, one works with the weak* topology on $BV(\Omega)$, knowing that the BV space can actually be identified with the dual space of a separable space [4, Remark 3.12].

Definition 6.1.6 (*Weak* Topology on* BV *[4, Def. 3.11]*). *A sequence* (u_n) *in* $BV(\Omega)$ *is called weakly* convergent to some* $u \in BV(\Omega)$, *if* (u_n) *converges to* u *in* $L^1(\Omega)$*-norm and the sequence of distributional gradients* (Du_n), *interpreted as vector valued Radon measures, weakly* converges to* Du *in* Ω, *i.e.*

$$\|u_n - u\|_{L^1(\Omega)} \to 0 \qquad \text{and} \qquad \lim_{n \to \infty} \int_\Omega \phi \; dDu_n = \int_\Omega \phi \; dDu \;, \quad \forall \phi \in C_0(\Omega) \;.$$

Lemma 6.1.7. *A sequence* (u_n) *in* $BV(\Omega)$ *is weakly* convergent to some* $u \in BV(\Omega)$, *if and only if* (u_n) *is bounded in* $BV(\Omega)$*-norm and converges to* u *in* $L^1(\Omega)$*-norm, i.e.*

$$u_n \rightharpoonup^* u \qquad \Leftrightarrow \qquad \|u_n - u\|_{L^1(\Omega)} \to 0 \qquad \text{and} \qquad \sup_n \|u_n\|_{BV(\Omega)} < \infty \;.$$

Proof. See [4, Prop. 3.13]. □

In the following, we recall some basic properties of functions of bounded variation and the total variation seminorm, which will be needed in the analysis of the total variation regularization methods later.

Lemma 6.1.8 (*Convexity*). *The total variation functional* $|\cdot|_{BV(\Omega)}$ *is convex on* $BV(\Omega)$.

Proof. See [2, Thm. 2.4]. □

Remark. It is also well known that the total variation functional fails to be strictly convex, what can be simply shown by an example (see [2, Ex. 2.2]).

Lemma 6.1.9 (*Lower Semicontinuity*). *The total variation functional* $|\cdot|_{BV(\Omega)}$ *is lower semicontinuous in the* $L^1_{loc}(\Omega)$*-norm topology, i.e. for every sequence* (u_n) *of functions* *in* $BV(\Omega)$ *which converges in* $L^1_{loc}(\Omega)$ *to a function* u *holds*

$$|u|_{BV(\Omega)} \;\leq\; \liminf_{n\to\infty} |u_n|_{BV(\Omega)} \,.$$

Proof. See [71, Sect. 5.2, Thm. 1] or [80, Thm. 1.9]. □

Remark 6.1.10. Since $|\cdot|_{BV(\Omega)}$ is convex (*Lemma 6.1.8*), the strong lower semicontinuity implies the weak lower semicontinuity in $L^1(\Omega)$ (cf. [64, p. 11, Cor. 2.2]). Consequently, it follows for bounded Ω that the BV seminorm $|\cdot|_{BV(\Omega)}$ is also lower semicontinuous with respect to the weak topology on $L^p(\Omega)$ for $1 \leq p < \infty$ (cf. [2, Thm. 2.3]).

Lemma 6.1.11 (*Compactness*). *Let* $\Omega \subset \mathbb{R}^d$, $d \geq 1$, *be an open and bounded set with a Lipschitz boundary. Then,* $BV(\Omega)$ *is compactly embedded in* $L^p(\Omega)$ *for* $1 \leq p < d\,/\,(d-1)$ *(see Definition 3.1.25), i.e.*

$$BV(\Omega) \overset{c}{\hookrightarrow} L^p(\Omega) \qquad \text{for} \qquad 1 \;\leq\; p \;<\; \frac{d}{d\,-\,1}\,,$$

and is continuously embedded in $L^p(\Omega)$ *for* $p = d\,/\,(d-1)$ *(see Definition 3.1.23), i.e.*

$$BV(\Omega) \;\hookrightarrow\; L^p(\Omega) \qquad \text{for} \qquad p \;=\; \frac{d}{d\,-\,1}\,,$$

where we use the notation that $p = \infty$ *if* $d = 1$.

Proof. See [4, Cor. 3.49] or [2, Thm. 2.5]. □

Remark 6.1.12. In [2, Thm. 2.5], the authors prove that the space $BV(\Omega)$ is actually also compactly embedded in $L^p(\Omega)$ with $p = d\,/\,(d-1)$ for dimensions $d \geq 2$, however with respect to the weak topology on $L^p(\Omega)$ (see *Definition 3.1.29*).

Notice that the compactness results presented above correspond to those of functions in the Sobolev space $W^{1,1}(\Omega) \subset BV(\Omega)$, see [70, Sect. 5.7, Thm. 1; Sect. 5.6, Thm. 2]. However, note that there is also another type of compactness corresponding to the weak* topology on $BV(\Omega)$ proposed in *Definition 6.1.6*. With this topology, we obtain the following result as a direct consequence of the *Banach-Alaoglu Theorem 3.1.31*.

Lemma 6.1.13 (BV*-Weak* Compactness*). *Let* (u_n) *be a sequence in* $BV(\Omega)$ *which is uniformly bounded in* $\|\cdot\|_{BV(\Omega)}$, *then there exists a subsequence* (u_{n_j}) *and* $u \in BV(\Omega)$ *such that*

$$u_{n_j} \rightharpoonup^{*} u \qquad \text{in} \qquad BV(\Omega)\,.$$

6.2 TV Regularization in Image Processing

The concept of the space of functions of bounded variation $BV(\Omega)$ and its assigned total variation seminorm $|\cdot|_{BV(\Omega)}$ proposed in the previous section is a popular technique to solve different problems in the calculus of variations and plays an important role in several fields of mathematical image processing. The main reason for theirs popularity is the property that the BV space can represent discontinuous functions, which will even be preferred during the minimization of the TV functional. Hence, we give in the following a brief overview of the ways the TV seminorm will be used in image processing (cf. e.g. [47]) and recall the main properties of TV minimization using the standard application of image denoising.

However, we start for the moment with the problem of the image reconstruction. A commonly used and studied model in the literatur assumes that the observed data f are perturbed by additive Gaussian noise of the form (cf. *Section 2.1*),

$$f \;=\; K\bar{u} \;+\; \eta \;, \tag{6.4}$$

where $\bar{u}$ denotes the desired exact properties of an object, $K : U(\Omega) \to L^2(\Sigma)$ is a bounded (typically compact) linear operator on a Banach space $U(\Omega)$, which transforms the desired spatial information into measurement signals, and η is an additive white Gaussian noise. As already discussed in *Section 2.1*, the perturbed operator equation (6.4) is ill-posed and hence some type of regularization, which is directly related to certain a-priori information about a solution, is required to enforce a stable approximation at the desired image $\bar{u}$. In image processing, there are various tasks where one is in particular interested in the preservation of the edges in an image. Mathematically, the edges are strongly related to the discontinuities of a function, so that in the case of Gaussian perturbed data (6.4) the following variational technique based on the TV regularization functional will be used (cf. (2.5)),

$$\min_{u \,\in\, BV(\Omega)} \frac{1}{2} \|Ku \;-\; f\|^2_{L^2(\Sigma)} \;+\; \alpha\, |u|_{BV(\Omega)} \;, \qquad \alpha \;>\; 0 \;. \tag{6.5}$$

From the statistical point of view in *Section 2.2*, one uses the a-priori probability density $p(u)$ (2.8) with $J(u) \;=\; |u|_{BV(\Omega)}$. This means that images with smaller total variation (higher prior probability) are preferred in the minimization. The expected reconstructions are cartoon-like images, i.e. they will result in almost uniform (mean) intensities inside the different structures which are separated by sharp edges. An analysis of (6.5) with respect to the existence, uniqueness and stability of minimizers can be found e.g. in [2].

One of the most famous application of TV regularization in image processing is the Rudin-Osher-Fatemi (ROF) model [138] for image denoising. This model is a special case of the reconstruction problem (6.4) and (6.5) using the identity operator K,

$$\min_{u \in BV(\Omega)} \frac{1}{2} \|u - f\|^2_{L^2(\Omega)} + \alpha \, |u|_{BV(\Omega)} , \qquad \alpha > 0 , \tag{6.6}$$

where f denotes the observed noisy image. The goal of this formulation is to decompose the given noisy image $f = \bar{u} + \eta$ into a clean (exact) image $\bar{u}$ and the Gaussian noise η. However, it is in general not possible to compute the exact image $\bar{u}$ and an approximation u is wanted. In particular, one can directly observe that such a solution u approaches the noisy image f as $\alpha \to 0^+$.

The motivation for using TV in image processing is the effective suppression of noise and more significantly, the realization of homogeneous regions with mostly sharp edges. These features are in particular attractive for such applications, where the goal is to identify object shapes that are separated by sharp edges. This preservation of edges results from the fact that the TV functional essentially penalizes only the regularity of the level sets of the desired solution and does not exclude the possibility of discontinuities. This feature can be deduced from the *coarea formula.*

Lemma 6.2.1 (*Coarea Formula [80, Thm. 1.23]*). *Let $u \in BV(\Omega)$, then*

$$|u|_{BV(\Omega)} = \int_\Omega |Du| = \int_{-\infty}^{\infty} \left(\int_\Omega |D\mathbf{1}_{\{ x \in \Omega \,:\, u(x) < t \}}| \right) \, dt ,$$

where $\mathbf{1}_E$ denotes the characteristic function of the set E, defined by $\mathbf{1}_E(x) = 1$ if $x \in E$ and $\mathbf{1}_E(x) = 0$ if $x \notin E$.

Another interesting characterization of the question why the TV seminorm can preserve discontinuities, can be found in [40] with respect to the interpretation of the so-called *source condition (SC)* for total variation methods. In general, a regularization functional J is used to obtain a smoothing of an image with respect to a certain criterion, which will be determined by the choice of J and its variations. In the past, smooth, in particular quadratic, regularizations have attracted most attention, mainly due to the simplicity in analysis and computation. However, regularization functionals of the form

$$J(u) = \frac{1}{s} \int_\Omega |\nabla u|^s \, dx \qquad \text{for} \qquad 1 < s < \infty , \tag{6.7}$$

cannot yield image reconstructions with sharp edges. Since, using the classical L^2-data fidelity term as in (6.5), the regularization functionals (6.7) imply that a minimizer u

has to fulfill

$$\operatorname{div}\left(|\nabla u|^{s-2}\nabla u\right) \;=\; \frac{1}{\alpha}\,K^*(Ku\,-\,f)\;, \tag{6.8}$$

where K^* denotes the adjoint operator of K. This condition means that the smoothing process happens in two steps: in the first one, the adjoint operator K^* creates a smoothing depending on the reconstruction task represented by K. Note that this step is not present in the case of denoising problems, since K and K^* are identity operators. In the second step, the smoothing occurs by the inversion of a nondegenerate elliptic differential operator of second order on the left-hand side of (6.8), which results from the regularization functional in (6.7) only. However, this behavior changes in the case of the TV functional, which has formally the form (6.7) with $s = 1$ (see *Example 6.1.2, Item (1)*). Hence, the elliptic differential operator in the second smoothing step is then degenerate and hence effects only the level lines of the image. More precisely, the optimality condition of (6.5) is given by

$$p \;=\; \frac{1}{\alpha}\,K^*(Ku\,-\,f)\;, \qquad p\;\in\;\partial|u|_{BV(\Omega)}\;\subset\;(BV(\Omega))^*\;, \tag{6.9}$$

where ∂ denotes the subdifferential of a functional (see *Definition 3.2.6*). Here, we see that the first smoothing step by the adjoint operator K^* is actually a dual one, because it effects the subgradient p as an element of the dual space of $BV(\Omega)$. Subsequently, the second step is actually a relationship between the primal variable u and the dual variable p. In the total variation case, this leads to a significantly different behavior in the second step, since the dual variable is directly linked to the properties of the level sets. This can be shown if u is sufficiently regular with $|\nabla u| > 0$, then the subgradient is singleton given by

$$\partial|u|_{BV(\Omega)} \;=\; \{\,\kappa(u)\,\} \qquad \text{with} \qquad \kappa(u) \;=\; \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right)\;.$$

The element $\kappa(u)$ has a geometric interpretation, namely it represents the mean curvature of the level sets of u. Hence, the optimality condition (6.9) means that we can expect that the length of the level sets is minimized during TV minimization, but that it does not prohibit discontinuities in the solution.

However, despite the enormous popularity of TV minimization, it has also some unwanted effects. In the presence of noise, it tends to piecewise constant solutions, so-called *"staircasing effect"* (see *Fig. 6.2d*), which was analyzed in detail e.g. in [59] and [45]. Another deficit of TV regularization is the *systematic loss of contrast* in the reconstructions, even if the given data f are noise free (see *Fig. 6.2b* and *Fig. 6.3*). This effect is the well-known systematic error of the total variation minimization and was studied extensively in [116] and [125]. To this aspect, Meyer showed in [116] some analytic results for the ROF model (6.6) using the Banach space $G = \operatorname{div} L^\infty(\Omega, \mathbb{R}^2)$, $\Omega \subset \mathbb{R}^2$, equipped with the norm

$$\|w\|_* = \inf_{\substack{w = \operatorname{div} g \\ g \in L^\infty(\Omega,\mathbb{R}^2)}} \|g\|_{L^\infty(\Omega,\mathbb{R}^2)} = \inf_{\substack{w = \operatorname{div} g \\ g \in L^\infty(\Omega,\mathbb{R}^2)}} \operatorname*{ess\,sup}_{x \in \Omega} \sqrt{g_1^2(x) + g_2^2(x)} \,. \tag{6.10}$$

Using this space, Meyer showed that for a given image f and its ROF solution u as in (6.6), we have

$$\|f\|_* < \frac{\alpha}{2} \quad \Rightarrow \quad u = 0 \,,$$
$$\|f\|_* \geq \frac{\alpha}{2} \quad \Rightarrow \quad \|f - u\|_* = \frac{\alpha}{2} \quad \text{and} \quad \int_\Omega u\,(f - u)\,dx = \frac{\alpha}{2}\,|u|_{BV(\Omega)} \,.$$

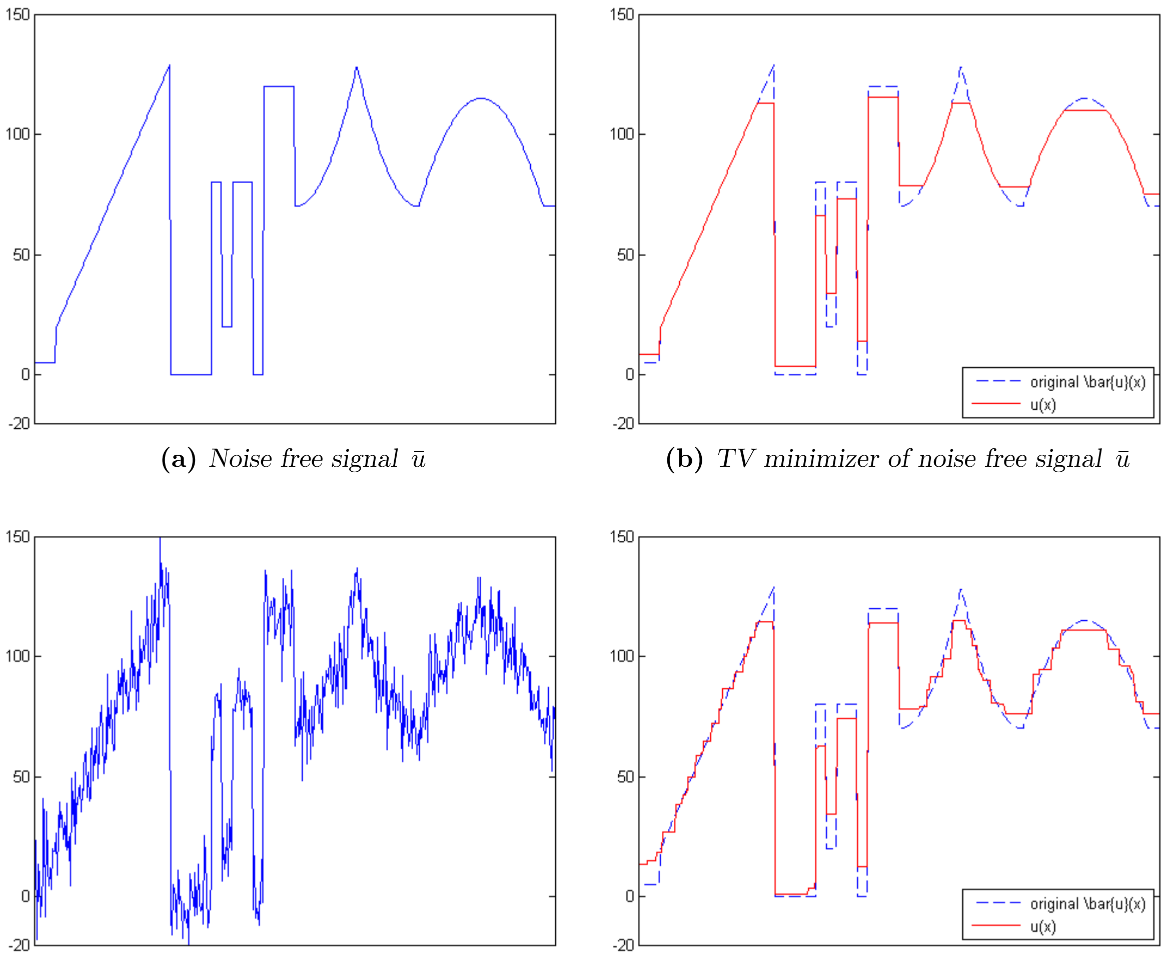
In particular, we see that the solution u vanishes completely depending on the G-norm of the image f and the regularization parameter α. Also the following quantitative result of ROF minimization was given in [116, p. 36]: Let $f = \sigma\,\mathbf{1}_{B_R(0)}$, $\sigma > 0$, be the multiple of the characteristic function of a disc with the radius $R > 0$. Then, the ROF solution is given by

$$u = \begin{cases} 0 \,, & \text{if } \frac{\sigma R}{\alpha} \leq 1 \,, \\ \left(\sigma - \frac{\alpha}{R}\right) \mathbf{1}_{B_R(0)} \,, & \text{if } \frac{\sigma R}{\alpha} \geq 1 \,. \end{cases}$$

We can observe the mentioned systematic loss of contrast in the reconstruction. More precisely, we see that u is a shrinked version of the given clean image f, where the shrinkage is proportional to the regularization parameter α. However, notice the fact that the solution u preserves the exact location of the discontinuities. In the case of a signal in one dimension, we illustrate the systematic loss of contrast of ROF minimization in *Fig. 6.3*.

In [116], Meyer also studied the aspect of *cartoon-texture decomposition* induced by the ROF model. There, he showed that the variational problem (6.6) yields a decomposition of the noise free image $f = u + v$ with $v := f - u$ into a cartoon part (or primal sketch) u and a texture part v, which obtains the oscillatory patterns (including also noise if f is noisy). An example of such a decomposition in the case of a noise free image f is shown in *Figure 6.4*.

Finally, we briefly note that based on the ROF model (6.6) several further developments were presented in the literature. Among others for instance different generalizations of the denoising problem (6.6) to general reconstruction problems based on the variational



(a) *Noise free signal* $\bar{u}$ **(b)** *TV minimizer of noise free signal* $\bar{u}$

(c) *Additive Gaussian noisy signal* f **(d)** *TV minimizer of noisy signal* f

Fig. 6.2. *Illustration of effects of TV minimization in one dimension. (a) Given noise free 1D signal* $\bar{u}$. *(b) TV minimizer of* $\bar{u}$ *using ROF model* (6.6). *Here, we observe only the systematic loss of contrast in the reconstruction. (c) Noisy signal* f *obtained by degradation of* $\bar{u}$ *with additive Gaussian noise. (d) TV minimizer of* f *using ROF model* (6.6). *Here, we observe that in the presence of noise the solution tends to a piecewise constant function in linear and quadratic parts of* $\bar{u}$ *and demonstrates the so-called "staircasing effect" of TV minimization.*

formulation (6.5) using a certain forward operator K depending on the restoration task. An abstract mathematical analysis of such methods is given e.g. in [2]. On the other side, there are several improvement suggestions with respect to the cartoon-texture decomposition of an image by replacing the L^2-data fidelity term in the ROF model (6.6) by other norms [9]. For example, Meyer proposed in [116] to use the G-norm $\|\cdot\|_*$ (6.10) as data fitting term in order to obtain a better extraction of signals with large oscillations, and thus in particular texture and noise, from the given image f. However, since this model is difficult to realize due to the form of the G-norm, several

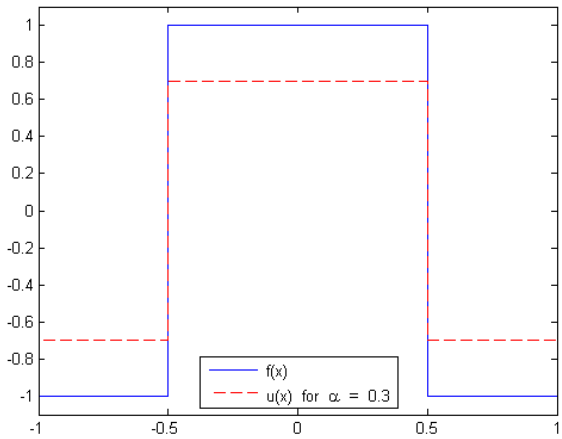


Fig. 6.3. *Illustration of systematic loss of contrast by TV minimization in one dimension.* ***Solid line:*** *Given noise free signal* $f = \bar{u}$. ***Dashed line:*** *TV minimizer* u *of* f *using ROF model* (6.6) *with* $\alpha = 0.3$. *We observe that the shrinkage is proportional to the regularization parameter* α, *but the solution* u *preserves the exact location of the discontinuities of* f.

approximations on this model were given e.g. in [161, 126, 8] and the references therein. Another possibility to modify the ROF model (6.6) was suggested in [3] and [124] to replace the L^2-norm by the L^1-norm. In [124], Nikolova has shown that this model is more effective for certain types of noise, such as salt and pepper noise. In the continuous case, the $TV - L^1$ model was studied by Chan and Esedoglu in [48] and the authors observed that this model has better behavior regarding the loss of contrast compared to the ROF model. Another very interesting approach to compensate the loss of contrast during the TV minimization has been proposed by Osher et al. in [125]. There, the authors perform a contrast enhancement and texture preservation by inverse scale space methods using the Bregman distance iteration. Note that we use the same strategy also in our Poisson framework proposed in *Section 4.7* in order to compensate the systematic errors of an arbitrary convex regularization functional.

6.3 TV Regularization in Poisson and US Speckle Frameworks

In this section, we focus our attention on the use of the TV functional as regularization energy in the Poisson and US speckle frameworks, which we proposed in *Chapters 4* and *5*.

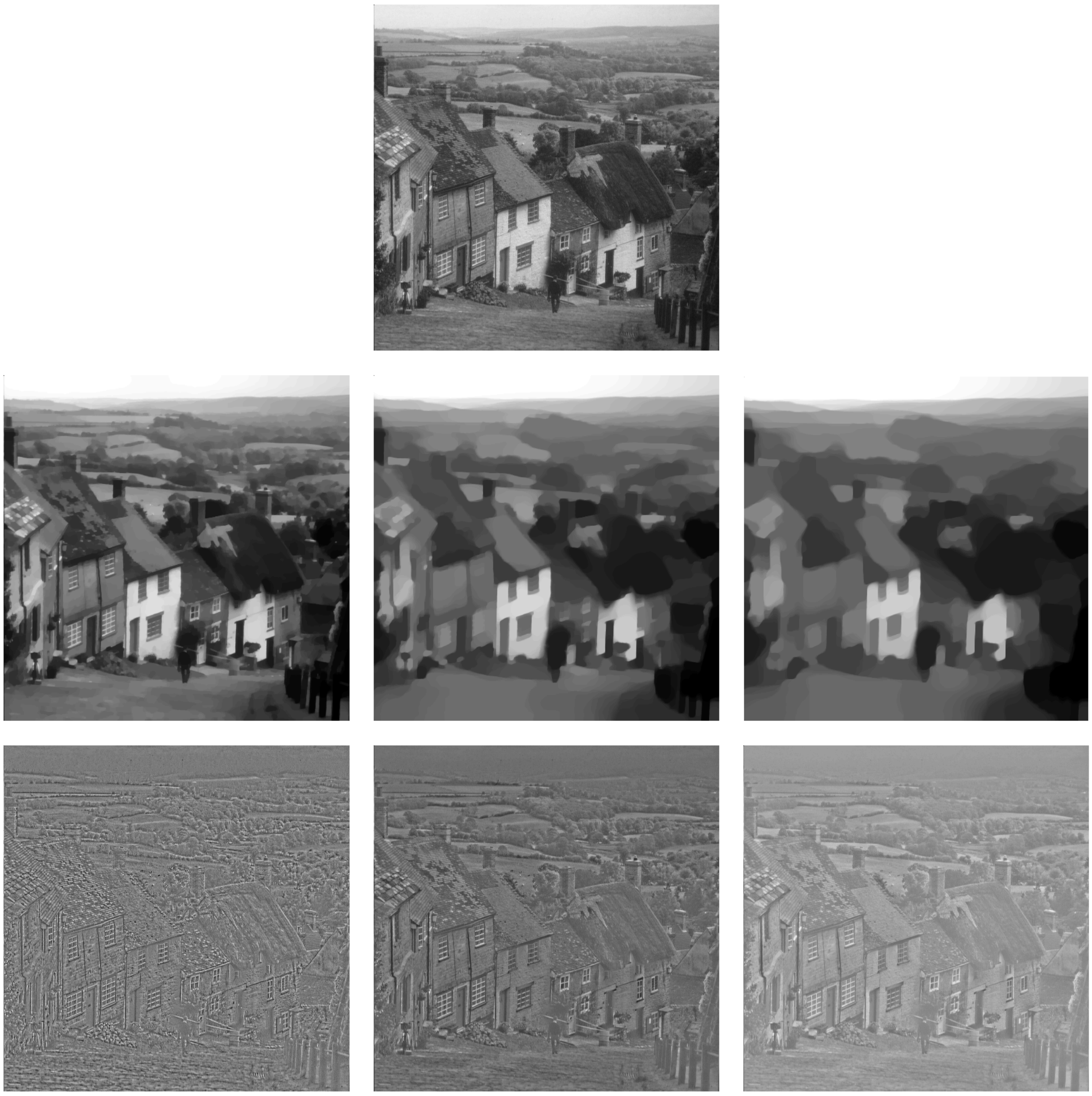


Fig. 6.4. *Illustration of the cartoon-texture decomposition induced by the ROF model* (6.6). ***First row:*** *Noise free original image* f . ***Second row:*** *Cartoon part* u *obtained by the ROF model for increasing regularization parameter* α *from left to right.* ***Third row:*** *Texture part* $v := f - u$ *corresponding to the cartoon part above.*

As already mentioned in *Section 4.1*, the application of the TV seminorm as regularization technique for Poisson distributed reconstruction problems is not new. Various methods have already been suggested for the TV regularized Poisson likelihood estimation problem (4.16),

$$\min_{\substack{u \in BV(\Omega) \\ u \geq 0 \ a.e.}} \int_\Sigma (Ku - f \log Ku) \, d\mu + \alpha \, |u|_{BV(\Omega)} , \qquad \alpha > 0 , \tag{6.11}$$

but still with some restrictions or limited successes, such as e.g. in the case of positron emission tomography (PET) [93, 127], deconvolution problems [19, 57, 72, 147], or denoising problems [104]. The main difficulties, which create these limitations, are caused by the following two features of the objective functional in (6.11). First, the *strong nonlinearity in the data fidelity term* resulting in issues in the computation of minimizers. Secondly, the strong computational efforts due to *nondifferentiability of the TV regularization functional* in the usual sense. To handle these difficulties, the authors in [72] and [147] proposed two algorithms, called *PIDAL* and *PIDSplit+*, using an augmented Lagrangian approach and the equivalent split Bregman method respectively in order to separate the reconstruction process from the regularization part. Both methods use the exact definition of TV (6.1), but require an inversion of the operator $I + K^*K$, where I is the identity operator and K^* is the adjoint of the ill-posed forward operator K. Thus, both methods are efficient only if K^*K is diagonalizable and can be inverted fast, as for instance in case of a convolution operator K via fast Fourier transform or discrete cosine transform. Additionally, in contrast to the PIDSplit+ algorithm in [147], the PIDAL algorithm in [72] ensures that Ku is nonnegative and not that the final solution u is nonnegative, which however is essential in (6.11). Another common strategy to overcome the nondifferentiability of the TV regularization functional is to use the formal definition of TV in (6.2) and to approximate it by differentiable functionals of the form

$$|u|^{\epsilon}_{BV(\Omega)} \;=\; \int_{\Omega} \sqrt{|\nabla u|^2 \,+\, \epsilon}\,, \qquad \epsilon \,>\, 0\,. \tag{6.12}$$

However, this approach creates blurring effects in the reconstructions, if ϵ is not small enough. In [14], Bardsley proposed an efficient computational method based on gradient projection and lagged-diffusivity, where the nonnegativity constraint is guaranteed via a simple projection into the feasible set. On the other hand, the methods in [57], [93] and [127] are realized as elementary modifications of the EM algorithm (see *Section 4.3*) with a fully explicit or semi-implicit treatment of TV in the iteration. A major disadvantage of these approaches is that the regularization parameter α needs to be chosen very small, since otherwise the positivity of solutions is not guaranteed and the EM based algorithm cannot be continued. Due to the additional parameter dependence on ϵ in (6.12), these algorithms are less robust.

In the case of the US speckle noise based variational regularization problem (5.4),

$$\min_{\substack{u \,\in\, W(\Omega) \\ u \,\geq\, 0 \;\; a.e.}} \int_{\Sigma} \frac{(f \,-\, Ku)^2}{Ku}\, d\mu \,+\, \alpha\, J(u)\,, \qquad \alpha \,>\, 0\,, \tag{6.13}$$

the TV regularization was to our knowledge only considered in the case of denoising problems [101, 92]. There, the authors use a gradient descent algorithm based on Euler-Lagrange equation using an approximation of TV by differentiable functionals (6.12).

In this thesis, we propose to use the robust FB-EM-REG and US-FB-REC-REG algorithm, which we introduced in *Chapters 4* and *5* using a *forward-backward (FB) splitting strategy*, in order to solve the minimization problems (6.11) and (6.13), respectively. In the context of TV regularized likelihood estimation problems (6.11) and (6.13), we rename these algorithms to FB-EM-TV and US-FB-REC-TV. In contrast to the PIDAL algorithm in [72] and the PIDSplit+ algorithm in [147], the FB-EM-TV splitting approach leads to a single step of the *expectation-maximization (EM) algorithm* (see *Section 4.3*) in the reconstruction step. Hence, we do not require any inversion of the forward operator K and can reuse this procedure for problems occurring for instance in medical tomography, such as positron emission tomography (PET) or single-photon emission computed tomography (SPECT) [167, 12]. In addition, the EM algorithm is a popular iterative method in medical imaging, microscopy, or astronomy, so that the EM algorithm existing in such applications can be used in the FB-EM-REG strategy without any additional effort. Moreover, the regularization steps in our splitting strategies lead in the case of TV regularization to the solution of a modified variant of the ROF model (6.6) with a weight in the data fidelity term. Hence, this analogy allows us to use the exact definition of the TV functional (6.1) and creates the opportunity to carry over efficient numerical schemes known for the ROF model. In *Sections 6.3.3* and *6.3.4* we present two algorithms for the weighted variant of the ROF model, namely a modified version of the *projected gradient descent algorithm* of Chambolle [41] and an *augmented Lagrangian method* which is very similar to the *alternating split Bregman algorithm* [82], and obtain in this way accurate, robust, and efficient numerical schemes. The advantage of such a numerical realization of regularization steps is that it can be performed equally well also for large regularization parameters. Thus, our proposed approaches are also applicable for problems with a low signal-to-noise ratio (SNR). Finally, using the analytical results for the FB-EM-REG algorithm in *Section 4.6* and the US-FB-REC-REG algorithm in *Section 5.4*, we will show that the assumptions made there on the regularization energy are fulfilled for the TV functional. Hence, we obtain in particular that both algorithms realize actually strictly positive reconstruction results.

6.3.1 Analytical Results

In this section, we carry out a mathematical analysis of the TV regularized estimation problems (6.11) and (6.13). We prove that these problems are well-posed, that the corresponding splitting algorithms preserve the positivity of the solution and that the damped modifications of these iteration schemes have a stable convergence behaviour. To do this, we carry over the results obtained in the general context of a convex regularization functional in *Sections 4.6* and *5.4*, and show that the assumptions on the regularization energy are fulfilled in the case of the total variation functional.

Considering the total variation regularization functional in *Sections 4.6* and *5.4*, we have to choose the image spaces $W(\Omega)$ and $U(\Omega)$ as follows,

$$W(\Omega) \;=\; BV(\Omega) \qquad \text{and} \qquad U(\Omega) \;=\; L^p(\Omega)\;, \tag{6.14}$$

where $\Omega \subset \mathbb{R}^d$, $d \geq 1$, is an open and bounded domain with Lipschitz boundary, and

$$1 \;\leq\; p \begin{cases} < \;\infty\;, & \text{if} \quad d \;=\; 1\;, \\ \leq\; \frac{d}{d-1}\;, & \text{if} \quad d \;\geq\; 2\;. \end{cases} \tag{6.15}$$

In this way, $W(\Omega)$ and $U(\Omega)$ are both Banach spaces and fulfill $W(\Omega) \subset U(\Omega)$ due to embedding results in *Lemma 6.1.11*. Moreover, following *Assumption 4.6.5 (i)*, we have to equip the space $U(\Omega)$ with a topology τ_U.

Definition 6.3.1 (*Choice of Topology* τ_U). *Due to the compact embedding results in Lemma 6.1.11 and Remark 6.1.12, we choose the topology* τ_U *as follows,*

- τ_U *is the strong norm topology on* $L^p(\Omega)$ *with* $1 \leq p < d\,/\,(d-1)$ *for* $d \geq 1$,
- τ_U *is the weak topology on* $L^p(\Omega)$ *(see Definition 3.1.29) with* $p = d\,/\,(d-1)$ *for* $d \geq 2$.

Finally, we carry over for the following analysis the definition of the objective functional F in (4.46) and (5.19), however adapted to the context of TV regularization,

$$\begin{aligned} &\min_{\substack{u \,\in\, L^p(\Omega) \\ u \,\geq\, 0 \;a.e.}} \; F_{KL}(u) \;:=\; D_{KL}(f, Ku) \;+\; \alpha\, |u|_{BV(\Omega)}\;, \\ & & \alpha \;>\; 0\;, \\ &\min_{\substack{u \,\in\, L^p(\Omega) \\ u \,\geq\, 0 \;a.e.}} \; F_{US}(u) \;:=\; D_{US}(f, Ku) \;+\; \alpha\, |u|_{BV(\Omega)}\;, \end{aligned} \tag{6.16}$$

where D_{KL} is the Kullback-Leibler (KL) functional as in *Definition 4.6.1* and D_{US} is given in (5.18).

Using the spaces $W(\Omega)$ and $U(\Omega)$ chosen in (6.14), we begin with the verification of *Assumption 4.6.5 (vi)-(viii)* with respect to the regularization functional J. In the case of the total variation regularization, i.e. $J = |\cdot|_{BV(\Omega)}$, these assumptions are all fulfilled, since the total variation functional is

- convex on $BV(\Omega)$ (see *Lemma 6.1.8*) and can be extended to a convex functional on $U(\Omega)$ as in (4.45) considering *Remark 3.2.3, Item (3)*,
- lower semicontinuous with respect to the weak topology on $L^p(\Omega)$ for $1 \leq p < \infty$ (see *Remark 6.1.10*) and hence also with respect to the topology τ_U on $U(\Omega)$,

- proper (see *Definition 3.2.2*), because $BV(\Omega)$ contains also piecewise constant functions, so that the condition on the objective functionals F_{KL} and F_{US} (6.16) in *Assumption 4.6.5 (vii)* is certainly fulfilled.

It remains to verify *Assumption 4.6.5 (viii)*, namely the sequential precompactness of the sub-level sets of TV functional with respect to the topology τ_U. To do this, we use the strategy proposed in *Remark 4.6.6, Item (5).* Hence, due to compact embedding results in *Lemma 6.1.11* and *Remark 6.1.12*, it suffices to show that the total variation functional on $U(\Omega)$ is $BV(\Omega)$-coercive. For this purpose, we first prove that the objective functionals F_{KL} and F_{US} defined in (6.16) are $BV(\Omega)$-coercive and then reduce this property to the TV functional.

Lemma 6.3.2 (*BV-Coercivity of Poisson Objective Functional*). *Let $V_\mu(\Sigma)$ and K satisfy Assumption 4.6.5 (i) - (iv). Moreover, assume that $\alpha > 0$, $f \in V_\mu(\Sigma)$ is nonnegative and $1 \leq p \leq d\,/\,(d-1)$. Additionally, assume that the operator K is bounded and it does not annihilate constant functions. Since K is linear, the latter condition is equivalent to*

$$K\mathbf{1}_\Omega \neq 0 , \tag{6.17}$$

where $\mathbf{1}_\Omega$ denotes the characteristic function on Ω. Then, the functional F_{KL} defined in (6.16) *is BV-coercive (see Definition 4.6.7), i.e. we obtain*

$$F_{KL}(u) \to +\infty \qquad \textit{whenever} \qquad \|u\|_{BV(\Omega)} \to +\infty . \tag{6.18}$$

Proof. For the proof of BV-coercivity, we derive an estimate of the form

$$\|u\|_{BV(\Omega)} \overset{(6.3)}{=} \|u\|_{L^1(\Omega)} + |u|_{BV(\Omega)} \leq c_1 \left(F_{KL}(u)\right)^2 + c_2\, F_{KL}(u) + c_3 , \tag{6.19}$$

with constants $c_1 \geq 0$, $c_2 > 0$ and $c_3 \geq 0$. Then, the desired coercivity property (6.18) follows directly from the positivity of the functional F_{KL} for all $u \in L^p(\Omega)$ with $u \geq 0$ *a.e.*

For the derivation of this estimate, we use that any $u \in BV(\Omega)$ has a decomposition of the form

$$u = w + v , \tag{6.20}$$

where

$$w = \left(\frac{\int_\Omega u \, dx}{|\Omega|}\right) \mathbf{1}_\Omega \qquad \text{and} \qquad v := u - w \quad \text{with} \quad \int_\Omega v \, dx = 0 .$$

First, we estimate $|v|_{BV(\Omega)}$ and $\|v\|_{L^1(\Omega)}$. Because constant functions have no variation, the positivity of the KL functional yields

$$\alpha\,|v|_{BV(\Omega)} \;\leq\; \alpha\,|u|_{BV(\Omega)} \;\leq\; F_{KL}(u) \qquad \Rightarrow \qquad |v|_{BV(\Omega)} \;\leq\; \frac{1}{\alpha}\,F_{KL}(u)\;.$$

This together with the Poincaré-Wirtinger inequality (see e.g. [6, Sect. 2.5.1]) yields an estimate of the L^1 norm,

$$\|v\|_{L^1(\Omega)} \;\leq\; C_1|v|_{BV(\Omega)} \;\leq\; C_1\,\frac{1}{\alpha}\,F_{KL}(u)\;, \tag{6.21}$$

where $C_1 > 0$ is a constant that depends on $\Omega \subset \mathbb{R}^d$ and d only. Now, using the decomposition (6.20) and the estimates to $|v|_{BV(\Omega)}$ and $\|v\|_{L^1(\Omega)}$, the problem (6.19) reduces to the estimation of the L^1 norm of constant functions, since

$$\begin{aligned} \|u\|_{BV(\Omega)} \;&\leq\; \|w\|_{L^1(\Omega)} \;+\; \|v\|_{L^1(\Omega)} \;+\; |v|_{BV(\Omega)} \\ &\leq\; \|w\|_{L^1(\Omega)} \;+\; (C_1\,+\,1)\,\frac{1}{\alpha}\,F_{KL}(u)\;. \end{aligned} \tag{6.22}$$

To estimate now $\|w\|_{L^1(\Omega)}$, we consider the L^1_μ distance between $Ku = Kw + Kv$ and f due to the continuous embedding of $V_\mu(\Sigma)$ in $L^1_\mu(\Sigma)$ in *Assumption 4.6.5*. With *Lemma 4.6.3 (iii)*, we obtain an upper bound,

$$\begin{aligned} \|(Kv\,-\,f)\,+\,Kw\|^2_{L^1_\mu(\Sigma)} \;&\leq\; \left(\frac{2}{3}\,\|f\|_{L^1_\mu(\Sigma)} \;+\; \frac{4}{3}\,\|Kv\,+\,Kw\|_{L^1_\mu(\Sigma)}\right)\,D_{KL}(f,Ku) \\ &\leq\; \left(\frac{2}{3}\,\|f\|_{L^1_\mu(\Sigma)} \;+\; \frac{4}{3}\,\|Kv\|_{L^1_\mu(\Sigma)} \;+\; \frac{4}{3}\,\|Kw\|_{L^1_\mu(\Sigma)}\right)\,F_{KL}(u)\;, \end{aligned}$$

and as a lower bound we obtain,

$$\begin{aligned} \|(Kv\,-\,f)\,+\,Kw\|^2_{L^1_\mu(\Sigma)} \;&\geq\; \left(\|Kv\,-\,f\|_{L^1_\mu(\Sigma)} \;-\; \|Kw\|_{L^1_\mu(\Sigma)}\right)^2 \\ &\geq\; \|Kw\|_{L^1_\mu(\Sigma)}\left(\|Kw\|_{L^1_\mu(\Sigma)} \;-\; 2\,\|Kv\,-\,f\|_{L^1_\mu(\Sigma)}\right)\;. \end{aligned} \tag{6.23}$$

Combining (6.21) with both inequalities yields

$$\begin{aligned} &\|Kw\|_{L^1_\mu(\Sigma)}\left(\|Kw\|_{L^1_\mu(\Sigma)} \;-\; 2\,\left(\|K\|\,C_1\,\frac{1}{\alpha}\,F_{KL}(u) \;+\; \|f\|_{L^1_\mu(\Sigma)}\right)\right) \\ &\quad\leq\; \left(\frac{2}{3}\,\|f\|_{L^1_\mu(\Sigma)} \;+\; \frac{4}{3}\,\|K\|\,C_1\,\frac{1}{\alpha}\,F(u) \;+\; \frac{4}{3}\,\|Kw\|_{L^1_\mu(\Sigma)}\right)\,F_{KL}(u)\;. \end{aligned} \tag{6.24}$$

This expression contains terms describing the function w only in dependence of the operator K. For the estimate of $\|w\|_{L^1(\Omega)}$ itself, we use the assumption (6.17) on the operator K. Thus, there exists a constant $C_2 > 0$ with

$$C_2 \;=\; \frac{\int_\Sigma |K\mathbf{1}_\Omega| \, d\mu}{|\Omega|} \qquad \text{and} \qquad \|Kw\|_{L^1_\mu(\Sigma)} \;=\; C_2 \, \|w\|_{L^1(\Omega)} \; . \tag{6.25}$$

This identity used in inequality (6.24) yields

$$\begin{aligned} C_2 \, \|w\|_{L^1(\Omega)} \left(C_2 \, \|w\|_{L^1(\Omega)} \;-\; 2 \left(\|K\| \, C_1 \, \frac{1}{\alpha} \, F_{KL}(u) \;+\; \|f\|_{L^1_\mu(\Sigma)} \right) \;-\; \frac{4}{3} \, F_{KL}(u) \right) \\ \leq\; \left(\frac{2}{3} \, \|f\|_{L^1_\mu(\Sigma)} \;+\; \frac{4}{3} \, \|K\| \, C_1 \, \frac{1}{\alpha} \, F_{KL}(u) \right) F_{KL}(u) \; . \end{aligned} \tag{6.26}$$

To receive an estimate of the form (6.19), we distinguish two cases:

Case 1: If

$$C_2 \, \|w\|_{L^1(\Omega)} \;-\; 2 \left(\|K\| \, C_1 \, \frac{1}{\alpha} \, F_{KL}(u) \;+\; \|f\|_{L^1_\mu(\Sigma)} \right) \;-\; \frac{4}{3} \, F_{KL}(u) \;\geq\; 1 \; , \tag{6.27}$$

then we conclude from (6.26) that

$$\|w\|_{L^1(\Omega)} \;\leq\; \frac{1}{C_2} \left(\frac{2}{3} \, \|f\|_{L^1_\mu(\Sigma)} \;+\; \frac{4}{3} \, \|K\| \, C_1 \, \frac{1}{\alpha} \, F_{KL}(u) \right) F_{KL}(u) \; ,$$

and obtain with (6.22),

$$\|u\|_{BV(\Omega)} \;\leq\; \frac{4 \, C_1 \, \|K\|}{3 \, C_2 \, \alpha} \big(F_{KL}(u) \big)^2 \;+\; \left(\frac{2}{3 \, C_2} \, \|f\|_{L^1_\mu(\Sigma)} \;+\; \frac{C_1 \;+\; 1}{\alpha} \right) F_{KL}(u) \; . \tag{6.28}$$

Case 2: If the condition (6.27) does not hold, i.e.

$$\|w\|_{L^1(\Omega)} \;<\; \frac{1}{C_2} \left(1 \;+\; 2 \left(\|K\| \, C_1 \, \frac{1}{\alpha} \, F_{KL}(u) \;+\; \|f\|_{L^1_\mu(\Sigma)} \right) + \; \frac{4}{3} \, F_{KL}(u) \right) \; ,$$

then we find from (6.22) that

$$\|u\|_{BV(\Omega)} \;\leq\; \left(\frac{2 \, \|K\| \, C_1 \, \frac{1}{\alpha} \;+\; \frac{4}{3}}{C_2} \;+\; \frac{C_1 \;+\; 1}{\alpha} \right) F_{KL}(u) \;+\; \frac{1 \;+\; 2 \, \|f\|_{L^1_\mu(\Sigma)}}{C_2} \; . \tag{6.29}$$

With the assumptions on the space $V_\mu(\Sigma)$ and the boundedness of the operator K, we have that $f \in L^1_\mu(\Sigma)$ and that $\|K\| < \infty$, so that we obtain from (6.28) and (6.29) the desired coercivity property (6.19). $\square$

Lemma 6.3.3 (*BV-Coercivity of US Objective Functional*). *Let $V_\mu(\Sigma)$ and K satisfy Assumption 5.4.2 (i) - (iv). Moreover, assume that $\alpha > 0$, $f \in V_\mu(\Sigma)$ and $1 \leq p \leq d\,/\,(d-1)$. Additionally, assume that the operator K is bounded and satisfies* (6.17). *Then, the functional F_{US} defined in* (6.16) *is BV-coercive (see Definition 4.6.7).*

Proof. The procedure of this proof is analogously to the proof of the BV-coercivity of the KL objective functional in *Lemma 6.3.2*. Due to the positivity of the functional D_{US}, we obtain analogously to (6.22) the following estimate,

$$\|u\|_{BV(\Omega)} \;\leq\; \|w\|_{L^1(\Omega)} \;+\; (C_1 \;+\; 1)\,\frac{1}{\alpha}\,F_{US}(u)\;, \tag{6.30}$$

where $C_1 > 0$ is a constant that depends on $\Omega \subset \mathbb{R}^d$ and d only. For the BV-coercivity of the functional F_{US}, we have now to estimate the L^1-norm of w. To do this, we consider the $L^1_\mu(\Sigma)$ distance between $Ku = Kw + Kv$ and f due to the continuous embedding of $V_\mu(\Sigma)$ in $L^2_\mu(\Sigma)$ in *Assumption 5.4.2 (ii)*. With Cauchy-Schwarz inequality, we obtain an upper bound,

$$\begin{aligned}\|(Kv \;-\; f) \;+\; Kw\|^2_{L^1_\mu(\Sigma)} \;&\leq\; \|Ku\|_{L^1_\mu(\Sigma)}\,D_{US}(f,Ku)\\ &\leq\; \Big(\|Kv\|_{L^1_\mu(\Sigma)} \;+\; \|Kw\|_{L^1_\mu(\Sigma)}\Big)\;F_{US}(u)\;,\end{aligned}$$

and the lower bound is given in (6.23). Combining both inequalities with (6.21) and (6.25) yields (cf. (6.26))

$$\begin{aligned}C_2\,\|w\|_{L^1(\Omega)}\left(C_2\,\|w\|_{L^1(\Omega)} \;-\; 2\left(\|K\|\,C_1\,\frac{1}{\alpha}\,F_{US}(u) \;+\; \|f\|_{L^1_\mu(\Sigma)}\right) \;-\; F_{US}(u)\right)&\\ \leq\; \|K\|\,C_1\,\frac{1}{\alpha}\,\big(F_{US}(u)\big)^2\;.&\end{aligned} \tag{6.31}$$

To receive an estimate of the form (6.19), we distinguish two cases:

Case 1: If

$$C_2\,\|w\|_{L^1(\Omega)} \;-\; 2\left(\|K\|\,C_1\,\frac{1}{\alpha}\,F_{US}(u) \;+\; \|f\|_{L^1_\mu(\Sigma)}\right) \;-\; F_{US}(u) \;\geq\; 1\;, \tag{6.32}$$

then we conclude from (6.31) that

$$\|w\|_{L^1(\Omega)} \;\leq\; \frac{1}{C_2}\,\|K\|\,C_1\,\frac{1}{\alpha}\,\big(F_{US}(u)\big)^2\;,$$

and obtain with (6.30),

$$\|u\|_{BV(\Omega)} \;\leq\; \frac{1}{C_2}\,\|K\|\,C_1\,\frac{1}{\alpha}\,\big(F_{US}(u)\big)^2 \;+\; \frac{C_1 \;+\; 1}{\alpha}\,F_{US}(u)\;. \tag{6.33}$$

Case 2: If the condition (6.32) does not hold, i.e.

$$\|w\|_{L^1(\Omega)} \; < \; \frac{1}{C_2} \left(1 \; + \; 2 \; \left(\|K\| \, C_1 \, \frac{1}{\alpha} \, F_{US}(u) \; + \; \|f\|_{L^1_\mu(\Sigma)} \right) \; + \; F_{US}(u) \right) \; ,$$

then we find from (6.30) that

$$\|u\|_{BV(\Omega)} \; \leq \; \left(\frac{2 \, \|K\| \, C_1 \, \frac{1}{\alpha}}{C_2} \; + \; \frac{C_1 \; + \; 1}{\alpha} \right) F_{US}(u) \; + \; \frac{1 \; + \; 2 \, \|f\|_{L^1_\mu(\Sigma)}}{C_2} \; . \tag{6.34}$$

With the assumptions on the space $V_\mu(\Sigma)$ and the boundedness of the operator K, we have that $f \in L^1_\mu(\Sigma)$ and that $\|K\| < \infty$, so that we obtain from (6.33) and (6.34) the desired coercivity property (6.19). □

From the BV-coercivity of the functionals F_{KL} and F_{US} in *Lemma 6.3.2* and *Lemma 6.3.3* yields that there exists a positive constant C such that a minimizer u of F_{KL} or F_{US} fulfills $\|u\|_{L^1(\Omega)} \leq C$. Hence, the following result is a direct consequence of *Lemma 6.3.2* and *Lemma 6.3.3*.

Corollary 6.3.4 (*BV-Coercivity of TV Functional*). *Suppose the assumptions of Lemma 6.3.2 or Lemma 6.3.3. Then, the functional J defined by*

$$J(u) \; := \; \begin{cases} |u|_{BV(\Omega)} \; , & \text{if} \quad \|u\|_{L^1(\Omega)} \; \leq \; C \; , \\ \infty \; , & \text{else} \; , \end{cases}$$

is BV-coercive for a positive constant C big enough.

Now, we obtain the following results as a direct consequence of *Theorems 4.6.8 - 4.6.10* and *Theorems 5.4.3 - 5.4.5*.

Theorem 6.3.5 (*Existence of Poisson Minimizers*). *Let $V_\mu(\Sigma)$ and K satisfy Assumption 4.6.5 (i) - (iv). Moreover, assume that $\alpha > 0$, $f \in V_\mu(\Sigma)$ is nonnegative and p satisfies the restrictions in* (6.15). *Additionally, assume that the operator K is bounded and satisfies* (6.17). *Then, the functional F_{KL} defined in* (6.16) *has a minimizer.*

Theorem 6.3.6 (*Existence of US Minimizers*). *Let $V_\mu(\Sigma)$ and K satisfy Assumption 5.4.2 (i) - (iv). Moreover, assume that $\alpha > 0$, $f \in V_\mu(\Sigma)$ and p satisfies the restrictions in* (6.15). *Additionally, assume that the operator K is bounded and satisfies* (6.17). *Then, the functional F_{US} defined in* (6.16) *has a minimizer.*

Theorem 6.3.7 (*Uniqueness of Poisson Minimizers*). *Let* $V_\mu(\Sigma)$ *and* K *satisfy Assumption 4.6.5 (i) - (iv). Moreover, assume that* $\alpha > 0$ *,* p *satisfies the restrictions in* (6.15) *and* $f \in V_\mu(\Sigma)$ *fulfills* $\inf_\Sigma f > 0$ *. Additionally, suppose that the operator* K *is bounded, injective and satisfies* (6.17)*. Then, the minimizer of the objective functional* F_{KL} *is unique.*

Theorem 6.3.8 (*Uniqueness of US Minimizers*). *Let* $V_\mu(\Sigma)$ *and* K *satisfy Assumption 5.4.2 (i) - (iv). Moreover, assume that* $\alpha > 0$ *,* $f \in V_\mu(\Sigma)$ *and* p *satisfies the restrictions in* (6.15)*. Additionally, suppose that the operator* K *is bounded, injective and satisfies* (6.17)*. Then, the minimizer of the objective functional* F_{US} *with* $\inf_\Omega u > 0$ *is unique.*

Theorem 6.3.9 (*Stability of Poisson Problem with respect to Perturbations in Measurements*). *We use the notations and assumptions introduced in Theorem 4.6.10. Then, the Poisson minimzation problem in* (6.16) *is stable with respect to the perturbations in the data, using the topology* τ_U *specified in Definition 6.3.1.*

Theorem 6.3.10 (*Stability of US Problem with respect to Perturbations in Measurements*). *We use the notations and assumptions introduced in Theorem 5.4.5. Then, the US minimization problem in* (6.16) *is stable with respect to the perturbations in the data, using the topology* τ_U *specified in Definition 6.3.1.*

Next, we verify the positivity preservation of the iteration sequence (u_k) obtained with the FB-EM-TV algorithm and US-FB-REC-TV algorithm, which correspond to the splitting strategies (4.24) and (5.11) with TV seminorm $|\cdot|_{BV(\Omega)}$ as regularization functional, respectively.

Lemma 6.3.11 (*Positivity of (damped) FB-EM-TV Algorithm and Poisson Denoising Strategy*). *Suppose the assumptions made in Lemmas 4.6.12 and 4.6.13 respectively, except condition* (4.58) *on the regularization functional* J *. Then, each half step of the (damped) FB-EM-TV splitting method and with it also the solution is strictly positive. The same behaviour holds also for each step of the (damped) Poisson denoising strategy* (4.36)*.*

Lemma 6.3.12 (*Positivity of (damped) US-FB-REC-TV Algorithm and US Denoising Strategy*). *Suppose the assumptions made in Lemma 5.4.6 and Corollary 5.4.7 respectively, except condition* (4.58) *on the regularization functional* J *. Then, each half step of the (damped) US-FB-REC-TV splitting method and with it also the solution is strictly positive. The same behaviour holds also for each step of the (damped) US denoising strategy* (5.15)*.*

Proof of Lemma 6.3.11 and Lemma 6.3.12. In the case of TV regularization, we have to prove that the TV seminorm $|\cdot|_{BV(\Omega)}$ fulfills condition (4.58) for the maximum principle in *Lemma 4.6.11*, i.e.

$$|v|_{BV(\Omega)} \;\leq\; |\tilde{u}|_{BV(\Omega)} \qquad \text{for} \qquad v \;=\; \min\{\max\{\tilde{u},\, a\},\, b\}\;,$$

where a and b are positive constants with $a < b$ and $\tilde{u}$ is a solution of the weighted ROF model (6.35). However, using the set

$$M \;:=\; \{\, x \,\in\, \Omega \;:\; v(x) \;=\; \tilde{u}(x)\,\} \;\subseteq\; \Omega\;,$$

we see that the function v has (due to its definition) no variation on $\Omega \setminus M$, so that we obtain

$$|v|_{BV(\Omega)} \;=\; |v|_{BV(M)} \;=\; |\tilde{u}|_{BV(M)} \;\leq\; |\tilde{u}|_{BV(\Omega)}\;.$$

□

Finally, we verify the convergence of the damped FB-EM-TV and US-FB-REC-TV splitting algorithms under appropriate assumptions on the damping parameters ω_k given in (4.65) and (5.27), respectively.

Theorem 6.3.13 (*Convergence of Damped FB-EM-TV Algorithm*). *Let $V_\mu(\Sigma)$ and K satisfy Assumption 4.6.5 (i) - (iv). Moreover, use the notations and the assumptions on data function f given in Theorem 4.6.14. Then, the objective functional F_{KL} defined in* (6.16) *is decreasing during the iteration, if the sequence of damping parameters* (ω_k) *satisfies inequality* (4.65). *In addition, if the conditions in* (4.66) *are fulfilled, then the damped FB-EM-TV algorithm converges to a minimizer of the functional F_{KL} in the sense of Theorem 4.6.14.*

Theorem 6.3.14 (*Convergence of Damped US-FB-REC-TV Algorithm*). *Let $V_\mu(\Sigma)$ and K satisfy Assumption 5.4.2 (i) - (iv). Moreover, use the notations and the assumptions on data function f given in Theorem 5.4.8. Then, the objective functional F_US defined in* (6.16) *is decreasing during the iteration, if the sequence of damping parameters* (ω_k) *satisfies inequality* (5.27). *In addition, if the conditions in* (4.66) *are fulfilled, then the damped US-FB-REC-TV algorithm converges to a minimizer of the functional F_{US} in the sense of Theorem 5.4.8.*

Proof of Theorem 6.3.13 and Theorem 6.3.14. We have to verify the additional assumptions in *Theorem 4.6.14* and *Theorem 5.4.8*.

- As already shown in the proof of *Lemma 6.3.11*, the TV functional fulfills the condition (4.58) for the maximum principle.

- It is easy to see that the TV functional is one-homogeneous, i.e. it satisfies $|\lambda u|_{BV(\Omega)} = \lambda |u|_{BV(\Omega)}$ for all $\lambda > 0$. Moreover, assumption (4.67) is directly fulfilled, since

$$\sup_{\|v\|_{BV(\Omega)} \leq 1} |v|_{BV(\Omega)} \overset{(6.3)}{\leq} \sup_{\|v\|_{BV(\Omega)} \leq 1} \|v\|_{BV(\Omega)} \leq 1 .$$

- The $U(\Omega)$-coercivity of the functional F_{KL} and F_{US} follows directly from the $BV(\Omega)$-coercivity shown in *Lemmas 6.3.2* and *6.3.3*, and the fact that $BV(\Omega)$ is continuously embedded in $U(\Omega)$ specified in (6.14), i.e. we have $\|u\|_{BV(\Omega)} \to \infty$ whenever $\|u\|_{U(\Omega)} \to \infty$. Moreover, with the choice of the space $U(\Omega)$ in (6.14), the continuous embedding of $U(\Omega)$ in $L^1(\Omega)$ is automatically fulfilled.

- With the space $U(\Omega)$ and the topology τ_U specified in (6.14) and *Definition 6.3.1* respectively, the condition (4.68) is fulfilled, if we choose $X = L^1(\Omega)$ and τ_X as the weak topology on $L^1(\Omega)$.

$\square$

6.3.2 Weighted ROF : General Form

To solve the regularized Poisson likelihood estimation problem (6.11), we proposed in *Sections 4.4.1* and *4.7.3* the (Bregman-)FB-EM-REG algorithm as a nested two step iteration strategy. The same splitting approach we also used in *Sections 5.2.1* and *5.5.1* in order to derive the (Bregman-)US-FB-REC-REG algorithm for the US speckle noise based variational regularization problem (5.4). However, we left open the question of the numerical realization of the regularization half steps (4.25), (4.27), (4.108) and (4.109) contained in both iteration strategies. Since the numerical realization of these steps depends on the structure of the chosen regularization functional, we now discuss this aspect in the case of the TV seminorm. Fortunately, all these regularization half steps have a similar form, so that we can propose a uniform numerical framework, which is also valid for the image denoising variational problems (4.37), (4.41), (5.15) and (5.17). The most general form of all the schemes above is

$$\min_{u \in BV(\Omega)} \frac{1}{2} \int_\Omega \frac{(u - q)^2}{h} + \gamma |u|_{BV(\Omega)} , \qquad \gamma > 0 , \tag{6.35}$$

with an appropriate setting of the "noise" function q, the weight function h and the regularization parameter γ. The choice of all these parameters with respect to the desired restoration method is summarized in *Table 6.1*.

We see that the variational problem (6.35) is just a modified version of the well known ROF model (6.6), with an additional weight h in the data fidelity term. This analogy

Algorithm	q	h	γ
Poisson Denoising (4.37)	f	u_k	α
Damped Poisson Denoising (4.37)	$\omega_k\, f \,+\, (1-\omega_k)\, u_k$	u_k	$\omega_k\, \alpha$
Approximated Poisson Denoising (4.41)	f	f	α
US Denoising (5.15)	$\dfrac{f^2}{u_k}$	u_k	α
Damped US Denoising (5.15)	$\omega_k\, \dfrac{f^2}{u_k} \,+\, (1-\omega_k)\, u_k$	u_k	$\omega_k\, \alpha$
Approximated US Denoising (5.17)	f	$\dfrac{f}{2}$	α
FB-EM-REG Algorithm (4.25) *US-FB-REC-REG Algorithm* (4.25)	$u_{k+\frac{1}{2}}$	$\dfrac{u_k}{K^* \mathbf{1}_\Sigma}$	α
Damped FB-EM-REG Algorithm (4.27) *Damped US-FB-REC-REG Algorithm* (4.27)	$\omega_k\, u_{k+\frac{1}{2}} \,+\, (1-\omega_k)\, u_k$	$\dfrac{u_k}{K^* \mathbf{1}_\Sigma}$	$\omega_k\, \alpha$
Bregman-FB-EM-REG Algorithm (4.108) *Bregman-US-FB-REC-REG Algorithm* (4.108)	$u_{k+\frac{1}{2}}^{l+1} \,+\, u_k^{l+1}\, v^l$	$\dfrac{u_k^{l+1}}{K^* \mathbf{1}_\Sigma}$	α
Damped Bregman-FB-EM-REG Algorithm (4.109) *Damped Bregman-US-FB-REC-REG Algorithm* (4.109)	$\omega_k^{l+1}\, u_{k+\frac{1}{2}}^{l+1} \,+\, \omega_k^{l+1}\, u_k^{l+1}\, v^l \,+\, (1-\omega_k^{l+1})\, u_k^{l+1}$	$\dfrac{u_k^{l+1}}{K^* \mathbf{1}_\Sigma}$	$\omega_k^{l+1}\, \alpha$

Table 6.1. *Overview for the setting of the functions* q , h *and parameter* γ *in* (6.35) *with respect to the different algorithms proposed in Chapters 4 and 5.*

creates the opportunity to carry over the different numerical schemes known for the ROF model, e.g. we refer to [43, 7, 42] and the references therein. Most of these computational schemes can be adapted to the weighted modification (6.35) and we consider in the following two very popular numerical realizations for TV regularized problems in image processing. In the first one, we use the exact dual TV approach (6.1) for the minimization of (6.35), which does not need any smoothing of the total variation. Then, our approach is analogous to the *projected gradient descent algorithm* of Chambolle in [41], which characterizes the subgradients of TV as divergences of vector fields with supremum norm less or equal one. The second numerical scheme will be similar to the *alternating split Bregman algorithm* [82] based on the splitting strategy in [164], where the idea is to "decouple" the L^1 and L^2 portions of the energy in the ROF model (6.6). An interesting connection of this splitting strategy was studied recently in [69] and [146], where the authors showed an equivalence to the augmented Lagrangian methods and alternating direction method of multipliers. In the case of the weighted ROF problem (6.35), we will use a slightly modified augmented Lagrangian approach of the alternating split Bregman algorithm in order to handle better the weight in the data fidelity term. Using either method, the weighted ROF problem (6.35) can be solved efficiently, obtaining accurate and robust algorithms.

6.3.3 Weighted ROF : Projected Gradient Descent Algorithm

Here we establish an iterative algorithm to compute the solution of the variational problem (6.35) using a modified variant of the projected gradient descent algorithm of Chambolle [41]. To this end, the formulation (6.35) can be written as a saddle point problem in the primal variable u and the dual variable g using the exact dual definition of the TV functional in (6.1),

$$\inf_{u \in BV(\Omega)} \sup_{\substack{g \in C_0^\infty(\Omega,\mathbb{R}^d) \\ \|g\|_\infty \leq 1}} L(u,g) := \frac{1}{2} \int_\Omega \frac{(u - q)^2}{h} + \gamma \int_\Omega u \,\mathrm{div} g \,. \tag{6.36}$$

Formally, the infimum regarding u and the supremum regarding p can be swapped. In the case of the standard ROF model (6.6), i.e. if the weight h in (6.35) is missing, this property is proved in [119]. However, this proof can also be carried over to the weighted variant (6.36) with minimal modifications. Moreover, a more precise analysis of this property for general saddle point problems is available in [64, p. 175, Prop. 2.3]. After exchanging inf and sup, the primal optimality condition for the saddle point problem (6.36) is given by

$$\frac{\partial}{\partial u} L(u,g) = 0 \qquad \Leftrightarrow \qquad u = q - \gamma \, h \,\mathrm{div} g \,. \tag{6.37}$$

Hence, if an optimal dual variable $\tilde{g}$ is available, the condition (6.37) can be used to obtain a solution of (6.36) and (6.35), i.e. the primal solution u is given by

$$u = q - \gamma \, h \,\mathrm{div} \tilde{g} \,. \tag{6.38}$$

For the computation of $\tilde{g}$, we substitute (6.37) into (6.36) and obtain a purely dual problem, which depends on g only. With terms that are constant with respect to the optimization variable and hence do not change the supremum, and under substitution of maximization by minimization of the negative functional, we obtain

$$\begin{aligned} \tilde{g} \in \operatorname*{arg\,min}_{g \in C_0^\infty(\Omega,\mathbb{R}^d)} \int_\Omega \frac{(\gamma \, h \,\mathrm{div} g - q)^2}{h} \,, \\ \text{s.t.} \quad |g(x)|_{\ell^s} - 1 \leq 0 \,, \qquad \forall x \in \Omega \,, \end{aligned} \tag{6.39}$$

where $|\cdot|_{\ell^s}$ is a vector norm on $\mathbb{R}^d$ with $1 \leq s \leq \infty$, depending on the definition of total variation for $d \geq 2$ (cf. *Remark 6.1.4, Item (2)*). In the following, we study only the most frequently used formulations, namely the *isotropic total variation* $(s = 2)$ and the *anisotropic total variation* $(s = \infty)$. Since the dual problem (6.39) is a (weighted) quadratic optimization problem with a nonlinear inequality constraint, we use the Karush-Kuhn-Tucker (KKT) conditions (cf. e.g. [89, Thm. 2.1.4]) to computate the optimal dual variable $\tilde{g}$ below.

We begin with the *isotropic problem formulation* using $s = 2$ in (6.39). In this case, the inequality constraint in (6.39) is equivalent to $|g(x)|_{\ell^2}^2 - 1 \leq 0$ for all $x \in \Omega$ and the KKT conditions yield the existence of a Lagrange multiplier $\lambda(x) \geq 0$ a.e. on Ω, such that

$$- \nabla (\gamma\, h \,\mathrm{div} g - q)(x) + \lambda(x)\, g(x) = 0 , \qquad \forall x \in \Omega , \tag{6.40}$$

and

$$\lambda(x) \left(|g(x)|_{\ell^2}^2 - 1 \right) = 0 , \qquad \forall x \in \Omega . \tag{6.41}$$

Fortunately, the multiplier λ can be specified explicitely from the complementarity condition (6.41), which yields that for any $x \in \Omega$,

$$\lambda(x) > 0 \quad \text{and} \quad |g(x)|_{\ell^2} = 1 \qquad \text{or} \qquad \lambda(x) = 0 .$$

Thus, in any case we obtain from (6.40),

$$\lambda(x) = |\lambda(x)\, g(x)|_{\ell^2} = |\nabla (\gamma\, h \,\mathrm{div} g - q)(x)|_{\ell^2} , \qquad \forall x \in \Omega ,$$

and can write (6.40) as a fixed point equation for g, obtaining the following iteration sequence,

$$g^{n+1}(x) = \frac{g^n(x) + \tau \left(\nabla (\gamma\, h \,\mathrm{div} g^n - q)(x) \right)}{1 + \tau\, |\nabla (\gamma\, h \,\mathrm{div} g^n - q)(x)|_{\ell^2}} , \qquad \forall x \in \Omega . \tag{6.42}$$

In the case of the *anisotropic problem formulation*, i.e. using $s = \infty$ in (6.39), we can proceed analogous to the isotropic case above, however with the single difference that the inequality constraint for the dual variable g in (6.39) is multi-valued for any $x \in \Omega$,

$$|g(x)| = |(g_1, \ldots, g_d)(x)|_{\ell^\infty} \leq 1 \qquad \Leftrightarrow \qquad |g_i(x)| \leq 1 \quad \text{for} \quad 1 \leq i \leq d .$$

Therefore, the KKT conditions deliver in this case the existence of Lagrange multipliers $\lambda_i(x) \geq 0$ a.e. on Ω for $1 \leq i \leq d$, such that for any $i \in \{1, \ldots, d\}$,

$$- \nabla_{x_i} (\gamma\, h \,\mathrm{div} g - q)(x) + \lambda_i(x)\, g_i(x) = 0 , \qquad \forall x \in \Omega ,$$

and

$$\lambda_i(x)\,(\,|g_i(x)|\;-\;1\,)\;=\;0\;,\qquad \forall x\;\in\;\Omega\;,$$

where ∇_{x_i} denotes the i-th component of the gradient ∇. Then, analogous to the isotropic case above, we obtain the following iteration sequence for the dual variable component g_i with $1\leq i\leq d$,

$$g_i^{n+1}(x)\;=\;\frac{g_i^n(x)\;+\;\tau\left(\nabla_{x_i}(\gamma\,h\,\mathrm{div}g^n\;-\;q)(x)\right)}{1\;+\;\tau\,|\nabla_{x_i}(\gamma\,h\,\mathrm{div}g^n\;-\;q)(x)|}\;,\qquad \forall x\;\in\;\Omega\;. \tag{6.43}$$

Finally, in a standard *discrete setting on pixels with unit step sizes* and first derivatives computed by one-sided differences, the convergence result of Chambolle in [41, Thm. 3.1] can be transferred to the weighted ROF problem (6.35). The proof based on the Banach fixed point theorem and required the condition

$$0\;<\;\tau\;\leq\;\left(\,4\,d\,\gamma\,\|h\|_{L^\infty(\Omega)}\,\right)^{-1}\;, \tag{6.44}$$

in order to obtain a contraction constant less one, where $4\,d$ is the upper bound of the discrete divergence operator. Hence, we can guarantee the convergence of (6.42) and (6.43) to a optimal solution, if the damping parameter τ satisfies the condition (6.44). Note that the weight h can be interpreted as an adaptive regularization, since the regularization parameter γ is weighted in (6.42) and (6.43) by the function h. Using these solvers, the (dual) projected gradient descent algorithm for the weighted ROF (6.35) can be now summarized as in *Algorithm 6.1.*

6.3.4 Weighted ROF : Augmented Lagrangian Method

Next, we present an efficient numerical scheme to compute the solution of the weighted ROF model (6.35) and follow the idea of the *(alternating) split Bregman algorithm* proposed by Goldstein and Osher in [82]. In contrast to the projected gradient descent algorithm above, this splitting strategy uses the formal definition of the TV seminorm (6.2),

$$|u|_{BV(\Omega)}\;\stackrel{\text{(formally)}}{=}\;\int_\Omega|\nabla u|_{\ell^r}\;, \tag{6.45}$$

where $|\cdot|_{\ell^r}$ is a vector norm on $\mathbb{R}^d$ with $1\leq r<\infty$, depending on the definition of total variation for $d\geq 2$ (cf. *Remark 6.1.4, Item (2)*). Like for the previous algorithm, we will only consider the common formulations of the *isotropic total variation* $(r=2)$ and the *anisotropic total variation* $(r=1)$.

Algorithm 6.1 *Projected Gradient Descent Algorithm for Weighted ROF*

1. ***Parameters:*** $q\,,\ h\,,\ \gamma\ >\ 0\,,\ tol\ >\ 0\,,\ d\ \geq\ 1$
2. ***Initialization:*** $n\ =\ 0\,,\ g^0\ :=\ 0\,,\ \tau\ :=\ \big(4\,d\,\gamma\,\|h\|_{L^\infty(\Omega)}\big)^{-1},\ stop\ :=\ c\ >\ tol$
3. ***Iteration:***

   while ($stop\ \geq\ tol$) **do**

   if *Isotropic TV Formulation* **then**

   $i)$ Compute g^{n+1} via (6.42).

   end if

   if *Anisotropic TV Formulation* **then**

   $i)$ Compute g^{n+1} via (6.43).

   end if

   $ii)$ Set $stop\ =\ \|\gamma\,h\,\mathrm{div} g^{n+1}\ -\ \gamma\,h\,\mathrm{div} g^n\|_{L^2(\Omega)}$.

   $iii)$ $n\ \leftarrow\ n+1$

   end while
4. ***Return*** $u\ :=\ q\ -\ \gamma\,h\,\mathrm{div} g^n$ $\qquad\qquad\triangleright$ (6.38)

The key idea of the split Bregman algorithm is to "decouple" the L^1 part (TV functional (6.45)) and L^2 part (data fitting term) of the ROF energy in (6.6) by substituting the gradient in the TV term with an auxiliary function. This idea was first proposed in [164] for L^1-regularized deconvolution problems. Based on this splitting strategy, Goldstein and Osher suggest to apply a Bregman iteration, similar to (4.97) with an appropriate setting of H_f and J, to strictly enforce the constraint condition, which results in the splitting approach. In [69] and [146], the authors showed an interesting connection of the split Bregman algorithm to the augmented Lagrangian methods (ALM) and alternating direction method of multipliers (ADMM). In the case of the weighted ROF problem (6.35), we will use a modified augmented Lagrangian approach of the alternating split Bregman algorithm in order to handle better the weight in the data fidelity term. For an overview and introduction of the ALM we refer e.g. to [73, 81, 90].

Following the decoupling idea of the split Bregman algorithm, the weighted ROF model (6.35) is equivalent to a constrained optimization problem of the form

$$\min_{u,\,\tilde{u},\,v}\ \frac{1}{2}\int_\Omega \frac{(\tilde{u}\ -\ q)^2}{h}\ +\ \gamma\int_\Omega |v|_{\ell^r} \qquad \text{s.t.} \qquad \tilde{u}\ =\ u \quad \text{and} \quad v\ =\ \nabla u\ . \qquad (6.46)$$

The difference to the split Bregman algorithm is that we introduce an additional auxiliary function $\tilde{u}$, which will simplify the handling of the weight function h in the following numerical scheme. In order to obtain an unconstrained minimization problem again, we follow the idea of the augmented Lagrangian methods and define the following augmented Lagrangian functional with respect to the problem (6.46),

$$\begin{aligned} L(u,\, \tilde{u},\, v;\, \lambda_1,\, \lambda_2) \;=\; & \frac{1}{2} \int_\Omega \frac{(\tilde{u} \,-\, q)^2}{h} \;+\; \gamma \int_\Omega |v|_{\ell^r} \\ & +\; \langle \lambda_1, v \,-\, \nabla u \rangle \;+\; \frac{\mu_1}{2} \, \| v \,-\, \nabla u \|^2_{L^2(\Omega)} \\ & +\; \langle \lambda_2, \tilde{u} \,-\, u \rangle \;+\; \frac{\mu_2}{2} \, \| \tilde{u} \,-\, u \|^2_{L^2(\Omega)} \;+\; \chi_{\tilde{u} \,\geq\, 0} \;, \end{aligned} \tag{6.47}$$

where λ_1 and λ_2 are Lagrange multipliers as well as μ_1 and μ_2 are positive relaxation parameters. To guarantee the positivity of the solution, we additionally add an indicator function $\chi_{\tilde{u} \,\geq\, 0}$ defined in (4.11). To derive a numerical scheme based on the augmented Lagrange functional in (6.47), the basic procedure is to apply the standard Uzawa algorithm (without preconditioning) [65] and to set the stepsize of the gradient ascent with respect to the Lagrange multipliers to the relaxation parameters. This leads to a splitting strategy, which iteratively minimizes the augmented Lagrangian functional with respect to the primal variables u, $\tilde{u}$ and v, and updates the Lagrange multipliers λ_1 and λ_2 subsequently, i.e. we have

$$\begin{aligned} u^{n+1} \;\in\; \arg\min_u \Big\{ & \langle \lambda_2^n, \tilde{u}^n \,-\, u \rangle \;+\; \frac{\mu_2}{2} \, \| \tilde{u}^n \,-\, u \|^2_{L^2(\Omega)} \\ & + \langle \lambda_1^n, v^n \,-\, \nabla u \rangle \;+\; \frac{\mu_1}{2} \, \| v^n \,-\, \nabla u \|^2_{L^2(\Omega)} \Big\} \;, \end{aligned} \tag{6.48}$$

$$\begin{aligned} \tilde{u}^{n+1} \;\in\; \arg\min_{\tilde{u}} \Big\{ & \frac{1}{2} \int_\Omega \frac{(\tilde{u} \,-\, q)^2}{h} \;+\; \left\langle \lambda_2^n, \tilde{u} \,-\, u^{n+1} \right\rangle \\ & + \frac{\mu_2}{2} \, \| \tilde{u} \,-\, u^{n+1} \|^2_{L^2(\Omega)} \;+\; \chi_{\tilde{u} \,\geq\, 0} \Big\} \;, \end{aligned} \tag{6.49}$$

$$\begin{aligned} v^{n+1} \;\in\; \arg\min_v \Big\{ & \gamma \int_\Omega |v|_{\ell^r} \;+\; \left\langle \lambda_1^n, v \,-\, \nabla u^{n+1} \right\rangle \\ & + \frac{\mu_1}{2} \, \| v \,-\, \nabla u^{n+1} \|^2_{L^2(\Omega)} \Big\} \;, \end{aligned} \tag{6.50}$$

$$\lambda_1^{n+1} \;=\; \lambda_1^n \;+\; \mu_1 \, (v^{n+1} \,-\, \nabla u^{n+1}) \;, \tag{6.51}$$

$$\lambda_2^{n+1} \;=\; \lambda_2^n \;+\; \mu_2 \, (\tilde{u}^{n+1} \,-\, u^{n+1}) \;. \tag{6.52}$$

The efficiency of this strategy is now strongly dependent on the question, how fast we can solve each of the subproblems (6.48), (6.49) and (6.50). Since the minimization problem (6.48) is now "decoupled" from the L^1-norm, it is also differentiable with the following optimality condition,

$$(\mu_2 I \;-\; \mu_1 \Delta)\, u^{n+1} \;=\; \lambda_2^n \;+\; \mu_2\, \tilde{u}^n \;-\; \operatorname{div}(\lambda_1^n \;+\; \mu_1\, v^n)\;, \tag{6.53}$$

where I is the identity operator and Δ denotes the Laplace operator. As we will see in (6.64) and (6.65), this problem can be solved efficiently in the discrete setting using the *discrete cosine transform (DCT)*,

$$u^{n+1} \;=\; DCT^{-1}\left(\frac{DCT\big(\,\lambda_2^n \;+\; \mu_2\, \tilde{u}^n \;-\; \operatorname{div}(\lambda_1^n \;+\; \mu_1\, v^n)\,\big)}{\mu_2 \;+\; \mu_1\, \hat{k}} \right)\;, \tag{6.54}$$

where $\hat{k}$ represents the negative Laplace operator $-\Delta$ in the discrete cosine space and DCT^{-1} denotes the inverse discrete cosine transform. Moreover, the minimization problem (6.49) is also differentiable and can be actually computed by an explicit formula of the form

$$\tilde{u}^{n+1} \;=\; \begin{cases} \dfrac{q \;+\; h\,(\mu_2\, u^{n+1} \;-\; \lambda_2^n)}{I \;+\; \mu_2\, h}\;, & \text{if} \quad \dfrac{q \;+\; h(\mu_2\, u^{n+1} \;-\; \lambda_2^n)}{I \;+\; \mu_2\, h} \;\geq\; 0\;, \\ 0\;, & \text{else}\;. \end{cases} \tag{6.55}$$

For the minimization problem with respect to v in (6.50), we differentiate between the isotropic and anisotropic formulation of total variation in (6.45). The simplest case is the *anisotropic* TV definition with $r \;=\; 1$, where (6.50) is equal to

$$\begin{aligned} v^{n+1} \;\in\; \arg\min_{v} \Bigg\{ \gamma \sum_{i\,=\,1}^{d} \int_\Omega |v_i| \;+\; \sum_{i\,=\,1}^{d} \int_\Omega (\lambda_1^n)_i\, (v_i \;-\; \nabla_{x_i} u^{n+1}) \\ +\; \frac{\mu_1}{2} \sum_{i\,=\,1}^{d} \|v_i \;-\; \nabla_{x_i} u^{n+1}\|_{L^2(\Omega)}^2 \Bigg\}\;, \end{aligned} \tag{6.56}$$

where ∇_{x_i} denotes the i-the component of the gradient ∇. The unique minimizer of this problem is given explicitely by a simple one-dimensional shrinkage formula of the form

$$\begin{aligned} v_i^{n+1}(x) \;=\; \operatorname{sgn}\Big(\big(\nabla_{x_i} u^{n+1} \;-\; (1/\mu_1)\,(\lambda_1^n)_i\big)(x) \Big) \\ \max\Big(\big|\big(\nabla_{x_i} u^{n+1} \;-\; (1/\mu_1)\,(\lambda_1^n)_i\big)(x)\big| \;-\; (\gamma/\mu_1)\,,\, 0 \Big) \end{aligned} \tag{6.57}$$

for any $x \,\in\, \Omega$ and $1 \,\leq\, i \,\leq\, d$. This shrinkage is extremely efficient and requires only a few operations per element of v^{n+1}. For the *isotropic* TV definition with $r \;=\; 2$ in

(6.45), the minimization problem (6.50) is equal to

$$v^{n+1} \in \arg\min_{v} \left\{ \gamma \int_{\Omega} \sqrt{v_1^2 + \cdots + v_d^2} + \sum_{i=1}^{d} \int_{\Omega} (\lambda_1^n)_i \, (v_i - \nabla_{x_i} u^{n+1}) + \frac{\mu_1}{2} \sum_{i=1}^{d} \| v_i - \nabla_{x_i} u^{n+1} \|_{L^2(\Omega)}^2 \right\} .$$

Since the variables $v_1, \ldots, v_d$ are not decoupled here as they were in the anisotropic case (6.56), we need another treatment fo this problem. Fortunately, we can still solve this minimization problem explicitly using a generalized shrinkage formula presented in [164],

$$v_i^{n+1}(x) = \frac{\big(\nabla_{x_i} u^{n+1} - (1/\mu_1)\,(\lambda_1^n)_i\big)(x)}{\big|\big(\nabla u^{n+1} - (1/\mu_1)\,\lambda_1^n\big)(x)\big|_{\ell^2}} \max\Big(\big|\big(\nabla u^{n+1} - (1/\mu_1)\,\lambda_1^n\big)(x)\big|_{\ell^2} - (\gamma/\mu_1)\,,\, 0 \Big) \tag{6.58}$$

for any $x \in \Omega$ and $1 \leq i \leq d$, where the convention $(0/0) \cdot 0 = 0$ is used. Using these solvers, the augmented Lagrangian method for the weighted ROF problem (6.35) can be now summarized as in *Algorithm 6.2*.

Discrete Laplace Inversion via Cosine Transform

In the augmented Lagrangian method above, we have seen that the minimization problem (6.48) with respect to the image variable u leads to the inversion of the Laplace operator shown in (6.53). In (6.54), we proposed a solution of this problem using the discrete cosine transform. In the following, we verify this proposal and assume initially that we want to solve the Poisson equation,

$$f = -\Delta u \,, \tag{6.59}$$

with f and u satisfying Neumann boundary conditions. Then, using the discrete setting as proposed in *Definition 2.1.2*, a discrete finite differences approximation of (6.59) on d dimensional regular grid of $N_1 \times \cdots \times N_d$ points is given by

$$f_{i_1,\ldots,i_d} = 2 \left(\frac{1}{h_1^2} + \cdots + \frac{1}{h_d^2} \right) u_{i_1,\ldots,i_d} - \frac{1}{h_1^2} \left(u_{i_1+1,\ldots,i_d} + u_{i_1-1,\ldots,i_d} \right) - \cdots - \frac{1}{h_d^2} \left(u_{i_1,\ldots,i_d+1} + u_{i_1,\ldots,i_d-1} \right) , \tag{6.60}$$

Algorithm 6.2 *Augmented Lagrangian Method for Weighted ROF*

1. ***Parameters:*** $q\,,\ h\,,\ \gamma\ >\ 0\,,\ \mu_1\ >\ 0\,,\ \mu_2\ >\ 0\,,\ tol\ >\ 0$
2. ***Initialization:*** $n\ =\ 0\,,\ \tilde{u}^0\ :=\ q\,,\ v^0\ :=\ 0\,,\ \lambda_1^0\ :=\ 0\,,\ \lambda_2^0\ :=\ 0\,,\ stop\ :=\ c\ >\ tol$
3. ***Iteration:***

while ($stop\ \geq\ tol$) **do**

i) Compute u^{n+1} via (6.54).

ii) Compute $\tilde{u}^{n+1}$ via (6.55).

if *Isotropic TV Regularization* **then**

iii) Compute v^{n+1} via (6.58).

end if

if *Anisotropic TV Regularization* **then**

iii) Compute v^{n+1} via (6.57).

end if

iv) Update λ_1^{n+1} via (6.51).

v) Update λ_2^{n+1} via (6.52).

vi) Set $stop\ =\ \|\tilde{u}^{n+1}\ -\ \tilde{u}^n\|_{L^2(\Omega)}$.

vii) $n\ \leftarrow\ n+1$

end while

4. ***Return*** $\tilde{u}^n$

where $h_k\ =\ \frac{1}{N_k}\,,\ k\ =\ 1,\ldots d$, denotes the stepsize of the image grid in the k-th direction. Due to the Neumann boundary conditions , we can rewrite f and u in terms of the inverse discrete cosine transform. This transform has the following form in the case of a function g with Neumann boundary conditions,

$$g_{i_1,\ldots,i_d}\ =\ \sum_{p_1\,=\,0}^{N_1\,-\,1}\cdots\sum_{p_d\,=\,0}^{N_d\,-\,1}\ \beta_{p_1}^{N_1}\ \cdots\ \beta_{p_d}^{N_d}\ \hat{g}_{p_1,\ldots,p_d}$$
$$\cos\left(\frac{\pi\,(2i_1+1)\,p_1}{2N_1}\right)\ \cdots\ \cos\left(\frac{\pi\,(2i_d+1)\,p_d}{2N_d}\right)$$

with

$$\beta_{p_k}^{N_k}\ =\ \begin{cases}\frac{1}{\sqrt{N_k}}\,, & \text{if}\quad p_k\ =\ 0\,,\\ \sqrt{\frac{2}{N_k}}\,, & \text{if}\quad 0\ <\ p_k\ \leq\ N_k\ -\ 1\,,\end{cases}$$

for all $0 \leq i_k \leq N_k - 1$ with $1 \leq k \leq d$. Hence, replacing f and u in (6.60) by the representation of the inverse discrete cosine transform yields

$$
\begin{aligned}
\hat{f}_{p_1,\ldots,p_d} \prod_{j=1}^{k} \cos\left(\frac{\pi\,(2i_j+1)\,p_j}{2N_j}\right) \\
= 2\left(\frac{1}{h_1^2} + \cdots + \frac{1}{h_d^2}\right) \hat{u}_{p_1,\ldots,p_d} \prod_{j=1}^{k} \cos\left(\frac{\pi\,(2i_j+1)\,p_j}{2N_j}\right) \\
- \sum_{k=1}^{d} \frac{1}{h_k^2}\, \hat{u}_{p_1,\ldots,p_d} \prod_{j \neq k} \cos\left(\frac{\pi\,(2i_j+1)\,p_j}{2N_j}\right) \\
\left(\cos\left(\frac{\pi\,(2i_k+3)\,p_k}{2N_k}\right) + \cos\left(\frac{\pi\,(2i_k-1)\,p_k}{2N_k}\right)\right), \qquad (6.61)
\end{aligned}
$$

for all $0 \leq p_k \leq N_k - 1$ with $1 \leq k \leq d$. To simplify this equality, we use the addition theorem $\cos(x \pm y) = \cos(x)\cos(y) \mp \sin(x)\sin(y)$ and can rewrite (6.61) to

$$
\cos\left(\frac{\pi\,(2i_k+3)\,p_k}{2N_k}\right) + \cos\left(\frac{\pi\,(2i_k-1)\,p_k}{2N_k}\right) = 2\cos\left(\frac{\pi\,(2i_k+1)\,p_k}{2N_k}\right)\cos\left(\frac{2\,\pi\,p_k}{2\,N_k}\right)
$$

for all $1 \leq k \leq d$, so that we obtain

$$
\hat{f}_{p_1,\ldots,p_d} = \left(2\left(\frac{1}{h_1^2} + \cdots + \frac{1}{h_d^2}\right) - \sum_{k=1}^{d} \frac{2}{h_k^2} \cos\left(\frac{2\,\pi\,p_k}{2\,N_k}\right)\right) \hat{u}_{p_1,\ldots,p_d} .
$$

Finally, using the relation $\cos(2x) = 1 - \sin^2(x)$, the discrete cosine coefficients $\hat{u}_{p_1,\ldots,p_d}$ of solution u of the Poisson equation (6.59) can be represented by

$$
\hat{u}_{p_1,\ldots,p_d} = \frac{\hat{f}_{p_1,\ldots,p_d}}{\hat{k}_{p_1,\ldots,p_d}}, \qquad 0 \leq p_k \leq N_k - 1, \qquad k \in \{1, \ldots, d\}, \qquad (6.62)
$$

with the following representation of the negative Laplace operator $-\Delta$ in the discrete cosine space,

$$
\hat{k}_{p_1,\ldots,p_d} := 4 \sum_{k=1}^{d} \left(\frac{\sin\left(\frac{\pi\,p_k}{2\,N_k}\right)}{h_k}\right)^2 . \qquad (6.63)
$$

The computational difficulty of (6.62) is that we have to divide by zero, if $p_k = 0$ for all $k \in \{1, \ldots, d\}$. However, in the optimality condition (6.53), we have to invert a Laplace operator of the form

$$
(\mu_2 I - \mu_1\,\Delta)\,u = f . \qquad (6.64)
$$

Hence, denoting with DCT the discrete cosine transform operator, we can rewrite this equation as

$$\begin{aligned} DCT(f) &= DCT(\mu_2 \, u \; - \mu_1 \, \Delta u) \\ &= \mu_2 \, DCT(u) \; + \; \mu_1 \, DCT(-\Delta u) \; = \; \mu_2 \, DCT(u) \; + \mu_1 \, DCT(u) \, \hat{k} \; , \end{aligned}$$

so that the solution u of (6.64) is given by

$$u \; = \; DCT^{-1} \left(\frac{DCT(f)}{\mu_2 \; + \; \mu_1 \, \hat{k}} \right) \; , \tag{6.65}$$

where DCT^{-1} denotes the inverse discrete cosine transform operator. We see also that in contrast to (6.62), the right-hand side of (6.65) is now completely unproblematic to solve.

7 Regularization : Nonlocal Total Variation (NL-TV)

In this chapter, we consider a recently proposed *nonlocal (NL) extension* of the total variation (TV) functional (6.1) and use this new approach as regularization energy in the context of the Poisson framework, which we proposed in *Section 4*. Here, the notion nonlocal means that any point in an image can interact directly with any other point in the whole image domain. The main idea of nonlocal extension is based on the definition of *nonlocal derivative operators*, with a view to realize additional prior information derived from the object itself. In the continuous setting, such nonlocal operators have first been proposed by Gilboa and Osher in [79], using a variant of the gradient and divergence definitions in a discrete setting on weighted graphs given in context of semi-supervised machine learning by Zhou and Schölkopf in [171, 172].

7.1 Introduction

In *Section 6.2* we have seen the problem of the cartoon-texture decomposition of the Rudin-Osher-Fatemi model (6.6), which shows that TV minimization is not able to preserve texture and fine structures in an image (cf. *Fig. 6.4*). This effect is caused by the regularity assumption of the TV formulation on the image model, namely that the image has a simple geometrical description consisting of a set of connected sets (objects) with smooth contours (edges). Additionally, the model assumes that the image is smooth inside single objects and has discontinuous jumps across the boundaries. Therefore, TV regularization is optimal to reduce the noise and to reconstruct the main geometrical configuration in an image. However, it fails to preserve texture, details and fine structures, because they behave in all aspects like noise and thus cannot be distinguished from noise.

In the following, we present a quite different approach, which has been proposed recently by Gilboa and Osher in [79]. This strategy extends the TV functional to a nonlocal variant using the definition of nonlocal derivative operators based on a nonlocal weight function (graph). The notion *nonlocal* means that any point can directly interact with any other point in the whole image domain, where the intensity of the interaction is depending on the value of the weight function. This weight function should represent the similarity of the two points and should be significant, if both points are similar in an appropriate measure. Therefore, the expectation is that such an approach is able to process both structures (geometrical parts) and texture within the same framework, due to the identification of recurring structures in the whole image. To introduce this strategy in the following, we begin briefly with the discussion of *local smoothing filters*, in order to clarify the difference to the nonlocal ones. Subsequently, we present the approach of *neighborhood filters*, which will form the basis of the nonlocal idea. Finally, due to the deficiencies of neighborhood filters in the presence of noise, we introduce the *nonlocal means (NL-means) algorithm*, a robust and stable extension of these filters. In the following, we will only give a short repetition of these methods und refer to [33] for a detailed discussion.

7.1.1 Local Denoising Methods

The denoising techniques commonly used in image processing are of local structure, meaning that the methods involve only a small spatially neighborhood around a point to denoise the value at this point. The most popular example using this strategy is the *Gaussian smoothing filter* with a fast decaying Gaussian kernel. In addition, due to the local definition of derivatives, methods based on (weak) partial derivatives are also local techniques. To these belong, for instance, approaches based on partial differential equations, such as anisotropic diffusion techniques [6, 49, 129, 165], as well as TV regularization methods using functionals (6.1) and (6.2). Hence, caused by the fundamental assumption that the noise is oscillatory and the image is smooth or piecewise smooth, local methods are effective to separate smooth parts from oscillatory ones due to the local perspective of the image. However, many geometrically fine structures are as oscillatory as noise, so that such methods cannot distinguish noise from fine structures and will remove them both. Additionally, in many cases, local methods create new artefacts such as “blurring”, the “staircasing effect”, the “checkerboard effect”, “wavelet outliers” and many other. To illustrate this behaviour of local methods, we briefly discuss the *Gaussian smoothing filter* and *anisotropic diffusion filters* in the following.

The Gaussian smoothing filter belongs to the class of *local smoothing filters*, which are based on the idea that the gray or color values of an image are similar in a local spatial neighborhood. Thus, a local averaging process should be able to reduce the random perturbations (noise) in an image. Actually, this strategy reduces the noise, but with the potential problem of oversmoothing. This problem concerns in particular the edges of an object, since there the assumption of locally similar gray values is most strongly violated. For instance, in the case of a black-white edge, it will be averaged to a continuous descent of the level values and appears optically blurred. The *frequency domain filters*, which have a different point of view than the local smoothing filters, show a similar behaviour. These methods perform the denoising process in the frequency space and damp high frequencies, which are characteristical for the noise. Such an approach is called *low-pass filter*, because only the low frequencies remain unchanged. The similarity to local smoothing filters lies in the fact that the absence of high frequencies induces a small local variation of gray level values. Hence, damping of high frequencies leads to outliers and artefacts at the edges.

The main deficiency of local smoothing filters is the problem of oversmoothing at the edges. To improve the performance of such methods, *anisotropic diffusion filters* [165, 129] perform the convolution process only in the direction orthogonal to the gradient of a function and avoid therefore the blurring effects at the edges. Hence, diffusion filters handle edges very well, but fail on flat regions creating some artefacts (cf. e.g. [33]).

7.1.2 Neighborhood Filters

The local denoising methods discussed above perform the restoration process using only a spatial proximity of points and cannot handle discontinuities (edges) and flat zones in the same framework. Hence, we recall in the following another class of filters, so-called *neighborhood filters (NFs)*, which damp this behaviour. The NFs perform an averaging process, but they use a new definition of *neighborhood*. In contrast to local smoothing filters, NFs are not just defining a neighborhood as a spatial closeness of points, but also take the gray level values of an image into account.

The more general continuous form of NFs is given as follows: Let f be the given noisy image and ξ is a reference function. Then, the neighborhood filter solution NF_ξ of the function f at point $x \in \Omega$, with $\Omega \subset \mathbb{R}^d$ open and bounded, is defined by

$$NF_\xi(f)(x) \;=\; \frac{1}{C(x)} \int_\Omega w_\xi(x,y)\, f(y)\; dy \;, \tag{7.1}$$

where $C(x) \;=\; \int_\Omega w_\xi(x,y)\; dy$ is the normalization factor. The kernel function w_ξ determines the actual form of the filter, depending in particular on the reference image ξ and its gray level values. The value $w_\xi(x,y)$ represents the similarity of points x

and y with respect to an appropriate measure and should be significant if x and y are similar. Finally, in image denoising tasks, the reference image is usually chosen as the given noisy image f. However, it is in general better to choose ξ as close as possible to the unknown true image in order to introduce relevant informati in the weight function w_ξ regarding image structures.

The idea of NFs go back to the works of Yaroslavsky (*Yaroslavsky filter*) [168] and Lee (*sigma filter*) [105] using the following kernel form in (7.1),

$$w_\xi(x,y) \;=\; \begin{cases} 1\,, & \text{if} \quad |\xi(y) - \xi(x)| \,\le\, \delta \quad \text{and} \quad |y - x| \,\le\, \rho\,, \\ 0\,, & \text{else}\,, \end{cases} \tag{7.2}$$

where δ and ρ are positive parameters. The name sigma filter results from the fact that Lee chose the parameter δ as 2σ, where σ is assumed to be the standard deviation of additive white Gaussian noise in the image. The idea of this choice is that any point with the gray level value difference greater than 2σ most likely comes from a different population and should be excluded from the average. However, the Yaroslavsky and sigma filter are less popular than the more recently proposed *SUSAN filter* [151] and *bilateral filter* [156]. The difference between these two and (7.2) is that instead considering a hard ball restriction of radius h and ρ, SUSAN and bilateral filters use a weight function of the form

$$w_\xi(x,y) \;=\; e^{-\frac{|\xi(y)-\xi(x)|^2}{\delta^2}}\, e^{-\frac{|y-x|^2}{\rho^2}}\,, \tag{7.3}$$

where δ and ρ act now as filtering parameters. However, in practice there is few difference between approaches (7.2) and (7.3).

We see in (7.2) and (7.3) that the averaging process of the NFs is not only based on a spatial neighborhood, but also on a neighborhood of gray and color level values. In particular, the latter feature represents the crucial difference to the local smoothing filters. Additionally, we obtain the following characterization of the NFs observing the weight functions (7.2) and (7.3): For small ρ and large δ, the NFs behave as local smoothing filters. For large ρ and small δ, we actually observe a nonlocal averaging process over similar gray level values, meaning that any point y in the image domain can be used to estimate the gray level value at point x. Therefore, certain self-similarity structures in an image can be utilized to reduce the variance of the noise.

The main advantage of the nonlocal strategy, in comparison to the local smoothing filters, is that NFs allow to process both structures and texture in the same framework. For instance, inside a homogeneous region, the gray level values differ slightly from each other in the case of additive Gaussian noise and the NFs (7.1) with (7.2) and (7.3) compute a smoothed region using arithmetic or Gaussian mean, respectively. On the other side, if we have an edge separating two regions and the gray value difference between these regions is larger than δ, then the NFs compute averages over points y belonging to the same region as the reference point x. Therefore, these algorithms do not lead to blurring effects at the edges. However, NFs have the disadvantage that they can yield undesirable blocky structures, the so-called "staircasing effect", in the restored images [35]. This fact can be explained by the strong relation of NFs to nonlinear diffusion partial differential equations [13, 33, 34, 150].

7.1.3 NL-means Algorithm

Recently, a generalization of the neighborhood filters above has been proposed by Buades, Coll and Morel in [33], calling the algorithm *nonlocal means (NL-means)*. This efficient model consists of denoising a gray level value at a point by averaging point values with similar structures (patches). Thus, not only the local gray level values of points are used to define similarity, but rather the values of a window around these points will be compared to define a similarity neighborhood. Mathematically, the idea of the NL-means algorithm is written as an averaging process (7.1) with the following weight function,

$$w_\xi(x,y) \;=\; e^{-\frac{(G_\sigma * |\xi(y+\cdot)-\xi(x+\cdot)|^2)(0)}{\delta^2}}\,, \tag{7.4}$$

where h is a positive filtering parameter and G_σ a Gaussian convolution kernel with standard deviation σ and

$$(G_\sigma * |\xi(y+\cdot)-\xi(x+\cdot)|^2)(0) \;=\; \int_{\mathbb{R}^d} G_\sigma(t)\,|\xi(y+t)-\xi(x+t)|^2\,dt\,.$$

The definition of the weight function (7.4) shows that this function is significant only if the window around y has similar structures as the corresponding window around x. Hence, the NL-means algorithm is very efficient in reducing noise, while preserving contrast in natural images and redundant structures such as texture.

The idea of similarity windows has been proposed first by Efros and Leung in the context of texture synthesis [62]. The idea there was to search for similar image patches in a sample image and to fill in holes in another image using the center value of found patches. There are also further works, which base on the utilization of the image values in a window around a point in order to take the advantage of self-similarity of natural images. For instance, the works of Kervrann and Boulanger use an adaptive and patch-based approach to denoise an image [96, 97, 98]. Similar idea of NL-means, but with distinct differences, are also the statistical neighborhood approaches used in an universal denoiser called *DUDE* of Weissman et al. [166] and in the *UINTA algorithm* of Awate and Whitaker [10]. A brief summary of both methods is given for example in [33].

7.2 Nonlocal Variational Framework

The concepts of the neighborhood filters and the NL-means algorithm presented in *Sections 7.1.2* and *7.1.3* are both of nonlocal nature. These methods use an averaging process over similar gray level values, where similar point values can be located arbitrarily far away from each other in the whole image domain. The particular advantage of the NL-means algorithm, in comparison to the usually used local (filtering) methods, is that it is very efficient to reduce noise, while preserving relevant information along an edge, regular texture patterns and contrast of natural images. However, the NL-means algorithm is used in the filtering theory and it is not clear as such a nonlocal approach can be utilized in the context of energy minimization and variational regularization theory, in particular as it can be generalized to other imaging tasks than denoising.

7.2.1 Variational Understanding of Nonlocal Filtering Methods

A first variational understanding of nonlocal filtering methods was given by Kindermann, Osher and Jones in [99] as a minimization functional with nonlocal correlation terms. There, the authors interpreted the usage of a neighborhood filter as a single step of solving a fixed-point equation, which based on the optimality condition of a certain nonlocal functional. However, due to the nonconvexity of functionals occurring in [99], Gilboa and Osher proposed in [78] an alternative convex nonlocal quadratic functional of weighted differences,

$$J_w(u) \;=\; \frac{1}{4}\int_{\Omega\times\Omega} (u(x) \;-\; u(y))^2\, w_\xi(x,y)\; dx\, dy\;, \tag{7.5}$$

where the weight function $w_\xi \,:\, \Omega \,\times\, \Omega \,\rightarrow\, \mathbb{R}_{\geq 0}$ is nonnegative and symmetric, i.e. $w_\xi(x,y) \;=\; w_\xi(y,x)$. In order to utilize the superior properties of the NL-means algorithm, the authors proposed to use the weight function (7.4). In addition, Gilboa and Osher showed in [78] that the functional (7.5) can be viewed as a continuous generalization of graphs and is related to concepts from spectral graph theory [50, 118]. Namely, the linear operator associated with the Euler-Lagrange equation of the functional (7.5) is closely related to the graph Laplacian. Hence, the Euler-Lagrange descent flow of

(7.5) can be interpreted as a nonlocal diffusion process, which is able to achieve the superior filtering properties of the NL-means algorithm in [33]. Later, Gilboa et al. given a further generalization of (7.5) in [77] using a more general convex framework,

$$J_w(u) = \frac{1}{2} \int_{\Omega \times \Omega} \phi(|u(x) - u(y)|) \, w_\xi(x,y) \; dx \, dy \;, \tag{7.6}$$

where the function $\phi : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is nonnegative, convex and fulfills $\phi(0) = 0$.

7.2.2 Nonlocal Operators of Gilboa and Osher

The approaches proposed in (7.5) and (7.6) provide a first variational framework of nonlocal regularization functionals, however they do not allow a systematic and coherent extension of local regularization energies to nonlocal ones. Hence, based on the definition of gradient and divergence operators on weighted graphs by Zhou and Schölkopf in [171, 172], Gilboa and Osher proposed in [79] a uniform variational framework using continuous graphs and nonlocal derivative operators. This novel approach allows to define new types of regularization functionals in image processing and other areas. In particular, it provides a nonlocal extension of all derivative based methods in inverse problems. Moreover, caused by the nonlocal structure of the weighted graph, this framework allows to adapt penalization energies to the geometry of the underlying functions, which one wants to recover. Note that also Bougleux et al. proposed in [25, 26] a regularization framework on weighted graphs for image and mesh filtering using similar operators in the discrete setting.

To introduce the framework of Gilboa and Osher in [79], we consider a nonnegative and symmetric weight function $w_\xi : \Omega \times \Omega \to \mathbb{R}_{\geq 0}$, as e.g. defined in (7.4), on a reference image ξ. Then, the *nonlocal gradient* $\nabla_w u(x)$ at $x \in \Omega$ is defined as the vector of all partial derivatives $\nabla_w u(x, \cdot)$, such that

$$\nabla_w u(x,y) := (u(y) - u(x)) \sqrt{w_\xi(x,y)} \;, \qquad \forall y \in \Omega \;. \tag{7.7}$$

Subsequently, the graph divergence operator of a vector $v : \Omega \times \Omega \to \mathbb{R}$ can be defined by the standard adjoint relation with the gradient operator, i.e.

$$\langle \nabla_w u, v \rangle := - \langle u, \mathrm{div}_w v \rangle \qquad \forall u : \Omega \to \mathbb{R} \;, \qquad \forall v : \Omega \times \Omega \to \mathbb{R} \;,$$

which leads to the following definition of the *nonlocal divergence* $\mathrm{div}_w v(x)$ at $x \in \Omega$,

$$(\mathrm{div}_w v)(x) = \int_\Omega (v(x,y) - v(y,x)) \sqrt{w_\xi(x,y)} \; dy \;. \tag{7.8}$$

Then, the *Laplacian* of a graph is defined by

$$\Delta_w u(x) \;:=\; \frac{1}{2}\,\mathrm{div}_w(\nabla_w u(x)) \;=\; \int_\Omega (u(y) \,-\, u(x))\, w_\xi(x,y)\; dy \;,$$

where a factor $\frac{1}{2}$ is needed to be consistent with the standard definition of the Laplacian. The definition of the graph Laplacian in this way has the well known properties of the Laplace operator, namely it is self-adjoint and negative semi-definite, i.e. it holds

$$\langle \Delta_w u, u \rangle \;=\; \langle u, \Delta_w u \rangle \qquad \text{and} \qquad \langle \Delta_w u, u \rangle \;=\; -\,\langle \nabla_w u, \nabla_w u \rangle \;\le\; 0 \;.$$

7.2.3 Nonlocal Total Variation (NL-TV) Functional

The nonlocal derivative operators of Gilboa and Osher, introduced in *Section 7.2.2*, have the advantage that they provide a systematic and coherent framework, which allows an accurate extension of non-smooth energies, such as total variation, to a nonlocal formulation. Using the definition of the nonlocal gradient and divergence operator in (7.7) and (7.8), the *nonlocal TV (NL-TV) functional* $|\cdot|_{NL-BV(\Omega)}$ can be defined similarly to the local case in (6.1).

Definition 7.2.1 (*Nonlocal Total Variation (NL-TV)*). *The nonlocal total variation of a function* u *in* Ω *is defined by*

$$|u|_{NL-BV(\Omega)} \;=\; \sup \left\{ \int_\Omega u\, \mathrm{div}_w\, g\, dx \;:\; g \;\in\; C(\Omega \times \Omega, \mathbb{R}), \quad \|g\|_\infty \;\le\; 1 \right\} \;, \tag{7.9}$$

where $C(\Omega \times \Omega, \mathbb{R})$ *denotes the space of continuous functions from* $\Omega \times \Omega$ *to* $\mathbb{R}$. *The supremum norm inequality in* (7.9) *means that* $\sup_{x,\,y\,\in\,\Omega} |g(x,y)| \;\le\; 1$.

Remark 7.2.2.

(1) Similar to (6.2), one obtains a formal characterization of the nonlocal TV functional using the nonlocal gradient operator ∇_w,

$$|u|_{NL-BV(\Omega)} \;=\; \int_\Omega |\nabla_w u(x)|\, dx \;. \tag{7.10}$$

(2) In *Remark 6.1.4, Item (2)*, we have seen that the definition of total variation in (6.1) is not unique for the image dimension $d \;\ge\; 2$. The same behavior remains in the case of the nonlocal TV definition in (7.9) and (7.10). Namely, depending on the choice of the inner norm in the supremum inequality in (7.9) or the vector norm in (7.10), we obtain an isotropic or anisotropic formulation of the nonlocal TV functional.

7.2.4 Nonlocal Regularization for Inverse Problems

In variational regularization theory, the adaptive properties of nonlocal regularization functionals can be utilized by replacing the common local regularization functionals by nonlocal ones. However, the main difficulty of nonlocal regularization strategies is the good estimation of the weight function w_ξ, in particular in the case of inverse problems, where the given data f usually lie in a different space than the desired approximation u. In some inverse problems like denoising or deconvolution, the nonlocal weight graph w_ξ can be directly estimated from the noisy image or the given measurements. For many other problems, the observation f cannot be used directly to estimate the regularization graph and other approaches are necessary to overcome this problem. A first strategy was proposed by Lou et al. [109] in the case of deconvolution and tomographic reconstructions. There, the main idea is to precompute a crude solution of inverse problems by a suitable fast image restoration method and to use this reconstruction as the reference image ξ for the computation of the weight function w_ξ. However, a more accurate approach should be to choose the reference image ξ so that it is as close as possible to the desired true object. Hence, in the general setup of variational regularization problems (4.92), it is more appropriate to consider the problem,

$$\min_u \; H_f(Ku - f) \;+\; \alpha \, J_w(u) \; , \qquad \text{s.t.} \quad w_\xi \;=\; w_u \; ,$$

where J_w denotes a nonlocal regularization functional, as in (7.5) or (7.6), with respect to the weight graph w_ξ with reference image ξ. This formulation means that we perform an optimization problem with respect to the desired image u and the optimal graph w_u simultaneously. Since a direct numerical solution of this problem is difficult to compute, Peyré et al. proposed in [130] to update the weight graph during the reconstruction process using a forward-backward operator splitting technique [51]. The same approach was also given by Zhang et al. in [170], however using a (preconditioned) Bregmanized operator splitting strategy.

7.3 Nonlocal Operators on Directed Graphs

The main challenge of nonlocal regularization strategies presented in *Section 7.2* is the development of efficient numerical solvers, in particular in the case of high-dimensional inverse problems. The main reason lies in the high complexity of the weight graph w_ξ, which allows a high number of possible interactions between points in the image domain Ω. Therefore, in the case of fully nonlocal approaches, we expect a maximal memory complexity of $|\Omega|^2$ and a high computational time, caused by the possibility that any point can directly interact with any other point in the image. Hence, a variety of approaches have been proposed recently to reduce these problems. A usual approach to improve computational time and storage efficiency is the so-called *semi-local approach* (cf. e.g. [33], [78], [79], [170], ...), which uses for each point $x \in \Omega$ a small search window $\Omega_w \subset \Omega$ centered at x. Subsequently, only similarity weights between x and the points in Ω_w will be computed. Another possibility to reduce the complexity of the nonlocal weight graph is to eliminate the computation of weights for points with dissimilar neighborhoods, which requires a fast preselection of similar patches. In [113], Mahmoudi and Sapiro use a local average of gray values in a certain neighborhood and gradients to preclassify the image patches and thereby to reduce the number of weight computations. A similar strategy was also used by Brox, Kleinschmidt and Cremers in [29], arranging the data in a cluster tree.

In [78, 77, 79, 170], a further strategy was proposed to improve computational time and storage efficiency. For each point $x \in \Omega$ a fixed number M of best neighbors with highest weight values in the semi-local searching window Ω_w are included in the neighborhood, where $M << |\Omega_w|$. However, the use of this simplification strategy destroys in general the symmetry of the weight function w_ξ, which is an essential assumption in the nonlocal operator framework of Gilboa and Osher in *Section 7.2.2.* In the context of graph theory, the symmetry destruction of the weight function w_ξ leads to a directed structure of the weight graph w_ξ. Hence, we propose in the following the use of nonlocal derivative operators on directed graphs, i.e. the weight function w_ξ does not need to be symmetric. Subsequently, we show that our framework is an extension of the nonlocal framework of Gilboa and Osher in [79] in the sense, that up to normalization factors, they will coincide in the case of a symmetric (undirected) weight function. Finally, we also show that in the discrete setting the local gradient and divergence operators can be viewed as a special case of the proposed framework.

7.3.1 Continuous Formulation of Directed Graphs

In the following, we use a variant of the gradient and divergence definitions on directed graphs given by Hein et al. in [86]. However, in our case we introduce a completely *continuous framework*, which can be transferred easily to the discrete setting later. Hence, let $\Omega \subset \mathbb{R}^d$ and $w : \Omega \times \Omega \to \mathbb{R}_{\geq 0}$ be a nonnegative weight function, which *does not need to be symmetric*. The weight value $w(x, y)$ can be interpreted as a positive measure between the points x and y. Despite the continuous formulation in the following, we try to retain the standard notions of graph theory and denote a pair of points

$(x, y) \in \Omega \times \Omega$ as an edge, if and only if $w(x, y) > 0$. Note that we consider each edge as an ordered pair of points (x, y), representing a directed connection from x to y with weight $w(x, y)$. Subsequently, the set of all such edges is denoted by $E \subset \Omega \times \Omega$, i.e.

$$E \ := \ \{ (x,y) \ \in \ \Omega \times \Omega \ : \ w(x,y) \ > \ 0 \} \ . \tag{7.11}$$

Hence, in an analogous way to graph theory, we denote the triple $G = (\Omega, E, w)$ as a *weighted directed graph* consisting of a set of points $\Omega \subset \mathbb{R}^d$ and a set of directed edges E characterized by the weight function w. In particular, note that due to the definition of the set of edges E in (7.11), it holds

$$w(x,y) \ = \ 0 \qquad \text{if and only if} \qquad (x,y) \ \notin \ E \ . \tag{7.12}$$

Finally, we define the *outgoing* and *ingoing degree functions* $d^{out}, d^{in} : \Omega \to \mathbb{R}_{\geq 0}$ of a point $x \in \Omega$ as

$$d^{out}(x) \ := \ \int_{N(x)} w(x,y) \ dy \qquad \text{and} \qquad d^{in}(x) \ := \ \int_{\Omega} w(y,x) \ dy \ ,$$

where the notation $N(x)$ denotes the neighborhood of a point x, consisting of points $y \in \Omega$ connected to x by a directed edge $(x, y) \in E$, i.e.

$$N(x) \ := \ \{ y \ \in \ \Omega \ : \ w(x,y) \ > \ 0 \} \ . \tag{7.13}$$

Additionally, we assume that $d^{out}(x) + d^{in}(x) > 0$ for all $x \in \Omega$, meaning that each point in Ω has at least one in- or outgoing edge, i.e. there are no isolated points in the graph G.

7.3.2 Hilbert Spaces of Functions on Directed Graphs

Let $G = (\Omega, E, w)$ be a weighted directed graph. In the following, we denote with $H(\Omega)$ the space of functions $u : \Omega \to \mathbb{R}$, with assigning a real value $u(x)$ to each point $x \in \Omega$. Moreover, there are also functions defined on the edges E of the graph, such as the weight function w. Hence, let $H(E)$ be the space of real-valued functions $v : E \to \mathbb{R}$ defined on the edges E of the graph G. Subsequently, we define the *inner products* for both function spaces $H(\Omega)$ and $H(E)$. For the functions on Ω, we use a weighted version of the standard L^2 inner product of the form

$$\langle u, \tilde{u} \rangle_{H(\Omega)} \ := \ \frac{1}{C_{H(\Omega)}} \int_{\Omega} u(x) \, \tilde{u}(x) \, \chi(x) \ dx \ , \tag{7.14}$$

where $C_{H(\Omega)} := \int_\Omega \chi(x)\, dx$ is a normalization factor and

$$\chi(x) \;=\; \chi^{out}\Big(d^{out}(x)\Big) \;+\; \chi^{in}\Big(d^{in}(x)\Big) \tag{7.15}$$

a measure on the neighborhood of the point x with $\chi^{out}, \chi^{in} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$. In addition, we assume that $\chi^{out}(0) = \chi^{in}(0) = 0$ and χ^{out}, χ^{in} are both strictly positive on $\mathbb{R}_{>0}$. According to (7.14), we define a norm on $H(\Omega)$ induced by the inner product as $\|u\|_{H(\Omega)} = \langle u, u\rangle_{H(\Omega)}^{1/2}$.

The inner product for functions on the edge set E is defined as follows,

$$\langle v, \tilde{v}\rangle_{H(E)} \;:=\; \frac{1}{C_{H(E)}} \int_{\Omega\times\Omega} v(x,y)\, \tilde{v}(x,y)\, \phi(w(x,y))\; dx\, dy\;, \tag{7.16}$$

where $C_{H(E)} := \int_{\Omega\times\Omega} \phi(w(x,y))\, dxdy$ is a normalization factor and $\phi : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a function with $\phi(0) = 0$ and is strictly positive on $\mathbb{R}_{>0}$. Actually, with the assumptions on the function ϕ and property (7.12), we integrate only over the set of edges E in (7.16), such that we can rewrite (7.16) to

$$\begin{aligned} \langle v, \tilde{v}\rangle_{H(E)} \;&=\; \frac{1}{C_{H(E)}} \int_E v(x,y)\, \tilde{v}(x,y)\, \phi(w(x,y))\; dx\, dy\;, \\ &=\; \frac{1}{C_{H(E)}} \int_\Omega \int_{N(x)} v(x,y)\, \tilde{v}(x,y)\, \phi(w(x,y))\; dy\, dx\;. \end{aligned} \tag{7.17}$$

Note that the elements of the function space $H(E)$ are vector-fields and we are interested in a norm of a vector. Hence, let $1 \leq p < \infty$ and $v \in H(E)$, then the *p-norm* of a vector-field v at $x \in \Omega$ is defined as

$$|v|_{\ell^p}(x) \;:=\; \left(\int_{N(x)} |v(x,y)|^p\, \phi(w(x,y))\; dy\right)^{1/p}\;. \tag{7.18}$$

Moreover, we also define the *maximum norm* of a vector-field $v \in H(E)$ at a point $x \in \Omega$ as

$$|v|_{\ell^\infty}(x) \;:=\; \sup_{y\,\in\, N(x)} |v(x,y)|\, \phi(x,y)\;.$$

Now, it is easy prove that the inner products (7.14) and (7.16) are well defined. Since both inner products are defined via additional weight functions χ and ϕ, we obtain a family of inner products on $H(\Omega)$ and $H(E)$. Hence, we rename the Hilbert spaces $H(\Omega)$ and $H(E)$ to $H(\Omega,\chi)$ and $H(E,\phi)$, respectively. Finally, note that the elements of the function space $H(E,\phi)$ can be interpreted as a flow on the edges, such that the function value on an edge (x,y) corresponds to the "mass" flowing from the point x to the point y (per unit time) (cf. [86]).

7.3.3 Definition of Nonlocal Operators on Directed Graphs

Based on the continuous formulation of directed graphs in *Section 7.3.1* and the definition of Hilbert spaces in *Section 7.3.2*, we will now introduce a continuous variant of the gradient and divergence operators on directed graphs given by Hein et al. in [86].

Definition 7.3.1 (*Gradient Operator on Directed Graphs*). *Let* $G = (\Omega, E, w)$ *be a weighted directed graph. Then, the* gradient operator $\nabla_w : H(\Omega, \chi) \to H(E, \phi)$ *of a function* $u \in H(\Omega, \chi)$ *is defined on a directed edge* $(x, y) \in E$ *as follows,*

$$(\nabla_w u)(x, y) := (u(y) - u(x))\,\gamma(w(x, y)) \ , \qquad \forall (x, y) \in E \ , \tag{7.19}$$

where $\gamma : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ *is strictly positive on* $\mathbb{R}_{>0}$ *with* $\gamma(0) = 0$ *. Hence, the* nonlocal gradient $\nabla_w u(x)$ *at a point* $x \in \Omega$ *is defined as the vector of all partial derivatives* $(\nabla_w u)(x, \cdot\,)$ *with respect to the set of edges* $(x, y) \in E$ *, i.e.*

$$\nabla_w u(x) := (\nabla_w u)(x, y) \ , \qquad \forall y \in N(x) \ .$$

Definition 7.3.2 (*Divergence Operator on Directed Graphs*). *Let* $G = (\Omega, E, w)$ *be a weighted directed graph. Then, the* divergence operator $\mathrm{div}_w : H(E, \phi) \to H(\Omega, \chi)$ *of a vector-field* $v \in H(E, \phi)$ *is defined by the standard adjoint relation with the gradient operator with respect to the inner products introduced in* (7.14) *and* (7.16)*, i.e.*

$$\langle \nabla_w u, v \rangle_{H(E)} := -\,\langle u, \mathrm{div}_w v \rangle_{H(\Omega)} \ , \qquad \forall u \in H(\Omega, \chi) \ , \qquad \forall v \in H(E, \phi) \ . \tag{7.20}$$

Proposition 7.3.3. *The* divergence $(\mathrm{div}_w v)(x)$ *, defined in* (7.20)*, at a point* $x \in \Omega$ *is given by*

$$\begin{aligned}(\mathrm{div}_w v)(x) = \frac{C_{H(\Omega)}}{\chi(x)\, C_{H(E)}} \Bigg(&\int_{N(x)} v(x, y)\,\gamma(w(x, y))\,\phi(w(x, y))\; dy \\ &- \int_{\Omega} v(y, x)\,\gamma(w(y, x))\,\phi(w(y, x))\; dy \Bigg) \ ,\end{aligned} \tag{7.21}$$

where $\chi(x)$ *is defined in* (7.15) *and the constants* $C_{H(\Omega)}$ *and* $C_{H(E)}$ *are given as follows,*

$$C_{H(\Omega)} = \int_{\Omega} \chi(x)\; dx \qquad \textit{and} \qquad C_{H(E)} = \int_{\Omega} \int_{N(x)} \phi(w(x, y))\; dy\, dx \ . \tag{7.22}$$

Proof. From the expressions of the inner product on $H(E, \phi)$ in (7.16) and (7.17), as well as from the definition of the edge derivative in (7.19), the left-hand side of (7.20) is written as,

$$
\begin{aligned}
\langle \nabla_w u, v \rangle_{H(E)} &= \frac{1}{C_{H(E)}} \left(\int_{\Omega \times \Omega} u(y)\, v(x,y)\, \gamma(w(x,y))\, \phi(w(x,y))\; dx\, dy \right. \\
&\qquad \left. - \int_\Omega \int_{N(x)} u(x)\, v(x,y)\, \gamma(w(x,y))\, \phi(w(x,y))\; dy\, dx \right) \\
&= \frac{1}{C_{H(E)}} \left(\int_\Omega u(y) \left(\int_\Omega v(x,y)\, \gamma(w(x,y))\, \phi(w(x,y))\; dx \right) dy \right. \\
&\qquad \left. - \int_\Omega u(x) \left(\int_{N(x)} v(x,y)\, \gamma(w(x,y))\, \phi(w(x,y))\; dy \right) dx \right) .
\end{aligned}
$$

We consider the first term on the right-hand side and rename x and y by y and x, respectively. With this renaming, we obtain

$$
\begin{aligned}
\langle \nabla_w u, v \rangle_{H(E)} &= \frac{1}{C_{H(E)}} \left(\int_\Omega u(x) \left(\int_\Omega v(y,x)\, \gamma(w(y,x))\, \phi(w(y,x))\; dy \right. \right. \\
&\qquad \left. \left. - \int_{N(x)} v(x,y)\, \gamma(w(x,y))\, \phi(w(x,y))\; dy \right) dx \right) \\
&\overset{(7.20)}{=} - \langle u, \operatorname{div}_w v \rangle_{H(\Omega)} \\
&\overset{(7.14)}{=} - \frac{1}{C_{H(\Omega)}} \int_\Omega u(x)\, (\operatorname{div}_w v)(x)\, \chi(x)\; dx \; .
\end{aligned}
$$

Then, the result (7.21) is obtained by taking $u(x) = 1$ for all $x \in \Omega$. □

Remark. As mentioned in [86], the first term on the right-hand side of (7.21) can be interpreted as the outgoing flow, whereas the second term can be seen as the ingoing flow.

Definition 7.3.4 (*Graph Laplacian for Directed Graphs*). *Let* $G = (\Omega, E, w)$ *be a weighted directed graph. Then, the* graph Laplacian $\Delta_w : H(\Omega,\chi) \to H(\Omega,\chi)$ *of a function* $u \in H(\Omega,\chi)$ *is defined by*

$$
\Delta_w u \;:=\; \operatorname{div}_w(\nabla_w u) \; . \tag{7.23}
$$

Proposition 7.3.5. *The graph Laplacian* $\Delta_w u(x)$ *, defined in* (7.23)*, at a point* $x \in \Omega$ *is given by*

$$\begin{aligned} \Delta_w u(x) \;=\; \frac{C_{H(\Omega)}}{\chi(x)\, C_{H(E)}} \Bigg(&\int_{N(x)} (u(y) \;-\; u(x))\, \gamma^2(w(x,y))\, \phi(w(x,y))\ dy \\ &+ \int_{\Omega} (u(y) \;-\; u(x))\, \gamma^2(w(y,x))\, \phi(w(y,x))\ dy \Bigg) \ , \end{aligned} \tag{7.24}$$

where $\chi(x)$ *is defined in* (7.15) *and the constants* $C_{H(\Omega)}$ *and* $C_{H(E)}$ *are given in* (7.22)*.*

Proof. The explicit form (7.24) results directly from the definition of the graph Laplacian in (7.23), using the definitions of the gradient and divergence operator in (7.19) and (7.21), respectively. □

Lemma 7.3.6. *The graph Laplacian* Δ_w *defined in* (7.23) *is self-adjoint and negative semi-definite, i.e. it holds*

$$\langle \tilde{u}, \Delta_w u \rangle_{H(\Omega)} \;=\; \langle \Delta_w \tilde{u}, u \rangle_{H(\Omega)} \ , \qquad \forall u,\, \tilde{u} \;\in\; H(\Omega, \chi)\ ,$$

and

$$\langle \Delta_w u, u \rangle_{H(\Omega)} \;\leq\; 0 \ , \qquad \forall u \;\in\; H(\Omega, \chi)\ .$$

Proof. Both properties result directly from the definitions of the graph Laplacian in (7.23) and the divergence operator in (7.20). □

7.3.4 Special Case : Undirected Graphs

Let $G = (\Omega,\, E,\, w)$ be an undirected weighted graph. In this case, the weight function w is symmetric, i.e. $w(x,y) = w(y,x)$. The symmetry of the weight function implies that whenever there is a directed edge from x to y, there is also a directed edge from y to x with the same weight value. Hence, this implies that there is no difference between out- and ingoing edges and we obtain $d^{out} \equiv d^{in}$, such that we denote the degree function by d with

$$d(x) \;=\; \int_{N(x)} w(x,y)\ dy \;\overset{(7.12)}{=}\; \int_{\Omega} w(x,y)\ dy\ , \qquad \forall x \;\in\; \Omega\ .$$

Consequently, we also have $\chi^{out} = \chi^{in}$ in (7.15), such that there is a unique function χ as weight in the inner product (7.14).

Since there is no difference between out- and ingoing edges and due to the properties of the functions γ and ϕ ($\gamma(0) = \phi(0) = 0$), the divergence $(\mathrm{div}_w v)(x)$ (7.21) of a vector field $v \in H(E,\phi)$ at a point $x \in \Omega$ simplifies in the case of an undirected graph as follows,

$$\begin{aligned}(\mathrm{div}_w v)(x) &= \frac{C_{H(\Omega)}}{\chi(x)\, C_{H(E)}} \int_{N(x)} (v(x,y) - v(y,x))\, \gamma(w(x,y))\, \phi(w(x,y))\, dy \\ &= \frac{C_{H(\Omega)}}{\chi(x)\, C_{H(E)}} \int_{\Omega} (v(x,y) - v(y,x))\, \gamma(w(x,y))\, \phi(w(x,y))\, dy \; .\end{aligned} \tag{7.25}$$

With the same argument, we also obtain a simplification of the graph Laplacian $\Delta_w u(x)$ (7.24) of a function $u \in H(\Omega,\chi)$ at a point $x \in \Omega$ in the case of an undirected graph as follows,

$$\begin{aligned}\Delta_w u(x) &= \frac{2\, C_{H(\Omega)}}{\chi(x)\, C_{H(E)}} \int_{N(x)} (u(y) - u(x))\, \gamma^2(w(x,y))\, \phi(w(x,y))\, dy \\ &= \frac{2\, C_{H(\Omega)}}{\chi(x)\, C_{H(E)}} \int_{\Omega} (u(y) - u(x))\, \gamma^2(w(x,y))\, \phi(w(x,y))\, dy \; .\end{aligned} \tag{7.26}$$

With a suitable choice of the functions χ, γ, and ϕ, one now can obtain from equation (7.26) the different expressions of the Laplace operator in graph theory, such as the normalized and combinatorial Laplacian (cf. [86]).

7.3.5 Special Case : Nonlocal Operators of Gilboa and Osher

In *Section 7.2.2*, we introduced a framework of nonlocal derivative operators proposed by Gilboa and Osher in [79]. In this framework, the authors considered a symmetric weight function w, i.e. an undirected weighted graph, to define the nonlocal gradient and divergence operators, as well as the nonlocal graph Laplacian. In the following, we show that the framework in *Section 7.2.2* represents a special case of the operators (7.19), (7.25) and (7.26), using a suitable choice of functions χ, γ and ϕ. For this purpose, we use the simplifications of derivative operators in *Section 7.3.4* and set

$$\gamma(w(x,y)) = \sqrt{w(x,y)} \qquad \text{and} \qquad \phi(w(x,y)) = 1$$

for all $(x,y) \in E$, with E being defined as in (7.11). Additionally, we use a uniform measure χ (7.15) on the neighborhood of a point of the form

$$\chi(x) = 1 \; , \qquad \forall x \in \Omega \; .$$

Hence, we obtain from (7.19), (7.25) and (7.26) the following formulations of nonlocal derivative operators at a point $x \in \Omega$,

$$\begin{aligned}
\nabla_w u(x) &= (u(y) - u(x)) \sqrt{w(x,y)} , \qquad \forall y \in N(x) , \\
(\mathrm{div}_w v)(x) &= \frac{|\Omega|}{|E|} \int_\Omega (v(x,y) - v(y,x)) \sqrt{w(x,y)} \, dy , \\
\Delta_w u(x) &= \frac{2\,|\Omega|}{|E|} \int_\Omega (u(y) - u(x))\, w(x,y) \, dy .
\end{aligned}$$

Up to normalization factors, these operators correspond to the framework of Gilboa and Osher mentioned in *Section 7.2.2*.

7.3.6 Discretization of Nonlocal Operators on Directed Graphs

In this section, we discretize the continuous formulation of directed graphs in *Section 7.3.1* and consider the discrete versions of the nonlocal operators on directed graphs proposed in *Section 7.3.3*.

Using the discrete setting proposed in *Definition 2.1.2* and the continuous formulation of directed graphs in *Section 7.3.1*, we denote the triple $G = (V, E, w)$ as a *discrete weighted directed graph* consisting of a finite set V of $N_1 \cdots N_d$ vertices (pixels) and a finite set $E \subset V \times V$ of weighted directed edges. In this context, we denote by u_x, $x = (i_1, \ldots, i_d)$, the value of a pixel $x \in V$ for $1 \leq i_k \leq N_k$ and $k = 1, \ldots, d$. Moreover, let w_{xy} be the discrete version of the weight function $w(x,y)$. Analogously to (7.13), we use the neighbor set notation N_x defined as

$$N_x := \{ y \in V : w_{xy} > 0 \} .$$

Then, the outgoing d^{out} and ingoing d^{in} degrees measure the sum of the out- and ingoing weights of a vertex,

$$d_x^{out} = \sum_{y \in N_x} w_{xy} \qquad \text{and} \qquad d_x^{in} = \sum_{y \in V} w_{yx} , \qquad \forall x \in V . \tag{7.27}$$

Moreover, corresponding to *Section 7.3.2*, we denote with $H(V,\chi)$ and $H(E,\phi)$ the Hilbert spaces of the real -valued functions on the vertices V and edges E respectively, associated with the following inner products (cf. (7.14) and (7.17)),

$$\begin{aligned}
\langle u, \tilde{u} \rangle_{H(\Omega)} &:= \frac{1}{C_{H(\Omega)}} \sum_{x \in V} u_x \, \tilde{u}_x \, \chi_x , \qquad u, \tilde{u} \in H(V,\chi) , \\
\langle v, \tilde{v} \rangle_{H(E)} &:= \frac{1}{C_{H(E)}} \sum_{x \in V} \sum_{y \in N_x} v_{xy} \, \tilde{v}_{xy} \, \phi(w_{xy}) , \qquad v, \tilde{v} \in H(E,\phi) .
\end{aligned} \tag{7.28}$$

In the discrete setting, the function χ is defined as in (7.15) by

$$\chi_x \;:=\; \chi^{out}\Big(d_x^{out}\Big) \;+\; \chi^{in}\Big(d_x^{in}\Big) \;, \qquad \forall x \;\in\; V \;,$$

and the normalization constants $C_{H(\Omega)}$ and $C_{H(E)}$ are given analogously to (7.22),

$$C_{H(\Omega)} \;:=\; \sum_{x \,\in\, V} \chi_x \qquad \text{and} \qquad C_{H(E)} \;:=\; \sum_{x \,\in\, V} \; \sum_{y \,\in\, N_x} \phi(w_{xy}) \;. \tag{7.29}$$

Then, the *weighted discrete gradient operator* ∇_w of a function $u \,\in\, H(V, \chi)$ at a vertex $x \,\in\, V$ is given corresponding to *Definition 7.3.1* as

$$\nabla_w u_x \;:=\; (\nabla_w u)_{xy} \;=\; (u_y \;-\; u_x)\, \gamma(w_{xy}) \;, \qquad y \;\in\; N_x \;. \tag{7.30}$$

Moreover, the *weighted discrete divergence operator* div_w of a vector-field $v \,\in\, H(E, \phi)$ at a vertex $x \,\in\, V$ is given corresponding to *Definition 7.3.2* as

$$(\mathrm{div}_w v)_x \;:=\; \frac{C_{H(\Omega)}}{\chi_x \, C_{H(E)}} \left(\sum_{y \,\in\, N_x} v_{xy}\, \gamma(w_{xy})\, \phi(w_{xy}) \;-\; \sum_{y \,\in\, V} v_{yx}\, \gamma(w_{yx})\, \phi(w_{yx}) \right) \;. \tag{7.31}$$

Finally, the *discrete graph Laplacian* Δ_w on a directed graph of a function $u \,\in\, H(V, \chi)$ at a vertex $x \,\in\, V$ is given corresponding to *Definition 7.3.4* as

$$\begin{aligned} \Delta_w u_x \;:=\; \frac{C_{H(\Omega)}}{\chi_x \, C_{H(E)}} \Bigg(& \sum_{y \,\in\, N_x} (u_y \;-\; u_x)\, \gamma^2(w_{xy})\, \phi(w_{xy}) \\ & + \sum_{y \,\in\, V} (u_y \;-\; u_x)\, \gamma^2(w_{yx})\, \phi(w_{yx}) \Bigg) \;. \end{aligned} \tag{7.32}$$

7.3.7 Special Case : Discrete Local Derivative Operators

In the following, we show that the discrete local gradient and divergence operators also can be put into the framework proposed in *Section 7.3.3.* For this purpose, we use the discrete versions of nonlocal operators on directed graphs proposed in *Section 7.3.6* and simplify these to the discrete local derivative setting.

Using the discrete setting of the framework in Section 7.3.6, we obtain the discrete local gradient and divergence operator by a suitable choice of functions χ, γ and ϕ. For the sake of convenience, we consider in the following only the case of a two dimensional image, but the approach is extendable to arbitrary dimensions in a straight-forward way.

The *discrete finite forward difference approximation* of the gradient is given by

$$\nabla u_{i_1,i_2} \;=\; \left((\nabla u)^1_{i_1,i_2},\, (\nabla u)^2_{i_1,i_2} \right) \;=\; \left(\frac{u_{i_1+1,i_2} \,-\, u_{i_1,i_2}}{h_1} \,,\, \frac{u_{i_1,i_2+1} \,-\, u_{i_1,i_2}}{h_2} \right) ,$$

where $h_1 = \frac{1}{N_1}$ and $h_2 = \frac{1}{N_2}$ denote the stepsizes of the image grid in the first and second direction. Hence, regarding the form of the weighted gradient in (7.30), we set for a fixed vertex $x = (i_1, i_2)$ the value $\gamma(w_{xy})$ as

$$\gamma(w_{xy}) \;:=\; \begin{cases} \frac{1}{h_1} \,, & \text{if} \quad y \;=\; (i_1+1, i_2) \;, \\ \frac{1}{h_2} \,, & \text{if} \quad y \;=\; (i_1, i_2+1) \;, \\ 0 \,, & \text{else} \;. \end{cases} \tag{7.33}$$

In particular, due to the assumption in *Definition 7.3.1* that the function γ is strictly positive on $\mathbb{R}_{>0}$ with $\gamma(0) = 0$, we obtain from (7.33) that the weight w_{xy} is strictly positive if and only if $y = (i_1+1, i_2)$ or $y = (i_1, i_2+1)$, such that the neighbor set N_x of a vertex $x = (i_1, i_2)$ is given by

$$N_x \;=\; \{\, y \;=\; (i_1+1, i_2) \quad \text{and} \quad y \;=\; (i_1, i_2+1) \,\} \;. \tag{7.34}$$

This result implies that each vertex has two edges only, namely one edge in the first direction and one in the second one. Hence, we can define the weighted graph w at each pixel $x = (i_1, i_2)$ as

$$w_{xy} \;:=\; \begin{cases} w_1 \;>\; 0 \,, & \text{if} \quad y \;=\; (i_1+1, i_2) \;, \\ w_2 \;>\; 0 \,, & \text{if} \quad y \;=\; (i_1, i_2+1) \;, \\ 0 \,, & \text{else} \;. \end{cases}$$

This implies that the out- and ingoing degrees (7.27) of a vertex coincide, i.e.

$$d_x^{out} \;=\; d_x^{in} \;=\; w_1 \;+\; w_2 \;, \qquad \forall x \;\in\; V \;,$$

where the right-hand side is in particular independent from x, such that

$$\chi \;:=\; \chi_x \;=\; \chi^{out}\Big(w_1 \;+\; w_2\Big) \;+\; \chi^{in}\Big(w_1 \;+\; w_2\Big) \;, \qquad \forall x \;\in\; V \;.$$

Therefore, the normalization constants in (7.29) simplify to

$$C_{H(\Omega)} \;=\; \chi\,|V| \qquad \text{and} \qquad C_{H(E)} \;=\; \big(\phi(w_1) \;+\; \phi(w_2)\big)\,|V| \;.$$

Consequently, we obtain from (7.32) the following form of the Laplace operator,

$$
\begin{aligned}
\Delta_w u_{i_1,i_2} \;&=\; \frac{1}{C} \left(\frac{u_{i_1+1,i_2} \;-\; u_{i_1,i_2}}{h_1^2}\, \phi(w_1) \;+\; \frac{u_{i_1,i_2+1} \;-\; u_{i_1,i_2}}{h_2^2}\, \phi(w_2) \right. \\
&\qquad \left. +\; \frac{u_{i_1-1,i_2} \;-\; u_{i_1,i_2}}{h_1^2}\, \phi(w_1) \;+\; \frac{u_{i_1,i_2-1} \;-\; u_{i_1,i_2}}{h_2^2}\, \phi(w_2) \right) \\
&=\; \frac{1}{C} \left(\frac{u_{i_1+1,i_2} \;-\; 2\,u_{i_1,i_2} \;+\; u_{i_1-1,i_2}}{h_1^2}\, \phi(w_1) \right. \\
&\qquad \left. +\; \frac{u_{i_1,i_2+1} \;-\; 2\,u_{i_1,i_2} \;+\; u_{i_1,i_2-1}}{h_2^2}\, \phi(w_2) \right) \;,
\end{aligned}
\tag{7.35}
$$

with $C \;:=\; \phi(w_1) \;+\; \phi(w_2)$. For the characterization of the divergence operator (7.31), note that the vector-field $v \;\in\; H(E, \phi)$ is a function on the edge set E . However, as we can see in (7.34), each vertex has two edges only, such that for a fixed pixel $x \;=\; (i_1, i_2)$ the vector v_{xy} is two-valued of the form

$$
v_{xy} \;=\; \left(v_{x,(i_1+1,i_2)},\, v_{x,(i_1,i_2+1)} \right) \;=:\; \left(v^1_{i_1,i_2},\, v^2_{i_1,i_2} \right) \;.
$$

Consequently, we obtain from (7.31) the following form of the divergence operator,

$$
\begin{aligned}
(\mathrm{div}_w v)_{i_1,i_2} \;&=\; \frac{1}{C} \left(\frac{v^1_{i_1,i_2}}{h_1}\, \phi(w_1) \;+\; \frac{v^2_{i_1,i_2}}{h_2}\, \phi(w_2) \right. \\
&\qquad \left. -\; \frac{v^1_{i_1-1,i_2}}{h_1}\, \phi(w_1) \;-\; \frac{v^2_{i_1,i_2-1}}{h_2}\, \phi(w_2) \right) \\
&=\; \frac{1}{C} \left(\frac{v^1_{i_1,i_2} \;-\; v^1_{i_1-1,i_2}}{h_1}\, \phi(w_1) \;+\; \frac{v^2_{i_1,i_2} \;-\; v^1_{i_1,i_2-1}}{h_2}\, \phi(w_2) \right) \;.
\end{aligned}
\tag{7.36}
$$

In formulations (7.35) and (7.36), we obtain the standard discrete local divergence and Laplace operator, if we set $\phi \;\equiv\; 1$ on $\mathbb{R}_{>0}$.

7.4 NL-TV Regularization in Poisson and US Speckle Frameworks

In *Section 7.2*, we introduced a variational regularization strategy in image processing, which can realize additional prior information derived from an image itself. The main characteristic of this novel approach is the use of a nonlocal weighted graph, which identifies similar structures in an image and allows to adapt the penalization to the geometry of underlying functions, which one wants to recover. Based on such a graph, Gilboa and

Osher proposed in [79] a nonlocal operator framework (see *Section 7.2.2*), which allows to extend non-smooth energies, such as total variation, to a nonlocal variant. In order to take advantage of the adaptive properties of nonlocal regularization strategies in the context of inverse problems with Poisson and US speckle corrupted data, we use in this section the nonlocal TV (NL-TV) functional (7.9) as regularization energy in the frameworks proposed in *Chapters 4* and *5*. Finally, we mention that the use of the NL-TV regularization in the combination with the Kullback-Leibler data fidelity term was already proposed by Steidl and Teuber in [154], however in the context to remove multiplicative Gamma noise in the images.

We now consider the NL-TV regularized likelihood estimation problems (4.16) and (5.5), and propose to use the nested two step iteration schemes, which we introduced in *Chapters 4* and *5* using a forward-backward splitting approach. As already discussed in the case of the local TV functional in *Section 6.3.2*, the open question in these two step iteration schemes remains the numerical realization of the regularization half steps (4.25), (4.27), (4.108) and (4.109). Using the definition of the NL-TV functional $|\cdot|_{NL-BV(\Omega)}$ in (7.9) or (7.10), the general form of all these regularization half steps is given analogously to (6.35) by

$$\min_u \ \frac{1}{2} \int_\Omega \frac{(u \ - \ q)^2}{h} \ + \ \gamma \, |u|_{NL-BV(\Omega)} \ , \qquad \gamma \ > \ 0 \ , \tag{7.37}$$

with an appropriate setting of the "noise" function q, the weight function h and the regularization parameter γ, as summarized in *Table 6.1*. The variational problem (7.37) is a modified version of the NL-ROF model proposed in [79], with weight h in the data fidelity term. To solve the standard NL-ROF model, extensions of Chambolle's projected gradient descent algorithm [41] and alternating split Bregman algorithm [82] have been proposed in [79] and [170], respectively. The only difference of these nonlocal extensions compared to the local versions of algorithms is that one has to replace the local gradient and divergence operators by nonlocal ones. Hence, based on these approaches, the modified projected gradient descent algorithm of Chambolle proposed in *Section 6.3.3* and the augmented Lagrangian method proposed in *Section 6.3.4* for solving the weighted ROF model (6.35) can be extended to solve the weighted NL-ROF problem (7.37).

In the following, we use the nonlocal operator framework proposed in *Section 7.3*, where we defined the nonlocal gradient and divergence operators in the more general case of directed graphs.

7.4.1 Weighted NL-ROF : Projected Gradient Descent Algorithm

Using the dual definition of the NL-TV functional $|\cdot|_{NL-BV(\Omega)}$ in (7.9), we can proceed analogously to *Section 6.3.3* to provide a numerical scheme for (7.37). Namely, the primal solution u is given similarly to (6.38) by

$$u \;=\; q \;-\; \gamma\, h \operatorname{div}_w \tilde{g} \;, \tag{7.38}$$

where div_w is defined in (7.21). Subsequently, depending on the isotropic or anisotropic formulation of the NL-TV functional (cf. *Remark 7.2.2, Item (2)*), we obtain the following iteration schemes to compute the optimal dual variable $\tilde{g}$ in (7.38). In the case of the *isotropic NL-TV formulation*, we obtain for a fixed point $x \in \Omega$ (cf. (6.42)),

$$g^{n+1}(x,y) \;=\; \frac{g^n(x,y) \;+\; \tau \big(\nabla_w \,(\gamma\, h \operatorname{div}_w g^n \;-\; q)\big)(x,y)}{1 \;+\; \tau\, |\nabla_w \,(\gamma\, h \operatorname{div}_w g^n \;-\; q)|_{\ell^2}(x)} \;, \qquad \forall y \;\in\; N(x) \;, \tag{7.39}$$

where ∇_w and div_w are defined in (7.19) and (7.21), and the vector norm $|\cdot|_{\ell^2}$ is given by (7.18). In the case of the *anisotropic NL-TV formulation*, we obtain the following iteration for a fixed point $x \in \Omega$ (cf. (6.42)),

$$g^{n+1}(x,y) \;=\; \frac{g^n(x,y) \;+\; \tau \big(\nabla_w(\gamma\, h \operatorname{div}_w g^n \;-\; q)\big)(x,y)}{1 \;+\; \tau\, |\big(\nabla_w(\gamma\, h \operatorname{div}_w g^n \;-\; q)\big)(x,y)|} \;, \qquad \forall y \;\in\; N(x) \;. \tag{7.40}$$

Convergence Study of Iteration Scheme

As already mentioned, Gilboa and Osher presented in [79] a nonlocal extension of the projected gradient descent algorithm of Chambolle [41] to solve the NL-ROF model, i.e. (7.37) with $h \;\equiv\; 1$. There, the authors also proved in the discrete setting the convergence of the algorithm to the global minimizer, if

$$0 \;<\; \tau \;\leq\; \frac{1}{\|\operatorname{div}_w\|^2} \;.$$

In the following, we resolve $\|\operatorname{div}_w\|^2$ in the context of discrete nonlocal operators on directed graphs, which we proposed in *Section 7.3.6*. Subsequently, we show that in the setting of discrete local gradient and divergence operators, we obtain with our framework the original bound $\|\operatorname{div}_w\|^2 \;\leq\; 8$ of Chambolle in [41].

Lemma 7.4.1. *With the notations and definitions of the discrete nonlocal operators in Section 7.3.6, the projected gradient descent algorithm of Chambolle [41] to solve the*

NL-ROF model, i.e. (7.37) *with* $h \equiv 1$, *converges for*

$$0 < \tau \leq \frac{C_{H(E)}}{4\, C_{H(\Omega)}\, C_{max}} \tag{7.41}$$

with

$$C_{max} := \max_{x \in V} \left\{ \frac{\sum_{y \in N_x} \gamma^2(w_{xy})\, \phi(w_{xy})}{\chi_x}, \frac{\sum_{y \in V} \gamma^2(w_{yx})\, \phi(w_{yx})}{\chi_x} \right\} . \tag{7.42}$$

Proof. Let $v \in H(E, \phi)$, then

$$\begin{aligned}
\|\mathrm{div}_w v\|^2_{H(\Omega)} &\overset{(7.28)}{=} \frac{1}{C_{H(\Omega)}} \sum_{x \in V} (\mathrm{div}_w v)^2 \chi_x \\
&\overset{(7.31)}{=} \frac{C_{H(\Omega)}}{C^2_{H(E)}} \sum_{x \in V} \left(\sum_{y \in N_x} v_{xy}\, \gamma(w_{xy})\, \phi(w_{xy}) - \sum_{y \in V} v_{yx}\, \gamma(w_{yx})\, \phi(w_{yx}) \right)^2 \frac{\chi_x}{\chi_x^2} .
\end{aligned}$$

Subsequently, using $(a - b)^2 \leq 2\,(a^2 + b^2)$ and the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
\|\mathrm{div}_w v\|^2_{H(\Omega)} \leq\ & \frac{2\, C_{H(\Omega)}}{C^2_{H(E)}} \sum_{x \in V} \Bigg(\left(\sum_{y \in N_x} v^2_{xy}\, \phi(w_{xy}) \right) \left(\sum_{y \in N_x} \gamma^2(w_{xy})\, \phi(w_{xy}) \right) \\
& + \left(\sum_{y \in V} v^2_{yx}\, \phi(w_{yx}) \right) \left(\sum_{y \in V} \gamma^2(w_{yx})\, \phi(w_{yx}) \right) \Bigg) \frac{1}{\chi_x} \\
\overset{(7.42)}{\leq}\ & \frac{2\, C_{H(\Omega)}\, C_{max}}{C^2_{H(E)}} \sum_{x \in V} \left(\sum_{y \in N_x} v^2_{xy}\, \phi(w_{xy}) + \sum_{y \in V} v^2_{yx}\, \phi(w_{yx}) \right) .
\end{aligned}$$

The property $\phi(0) = 0$ and the renaming of x and y by y and x, respectively, yields

$$\sum_{x \in V} \sum_{y \in V} v^2_{yx}\, \phi(w_{yx}) = \sum_{x \in V} \sum_{y \in N_x} v^2_{xy}\, \phi(w_{xy}) .$$

Hence, we have

$$\|\mathrm{div}_w v\|^2_{H(\Omega)} \leq \frac{4\, C_{H(\Omega)}\, C_{max}}{C^2_{H(E)}} \sum_{x \in V} \sum_{y \in N_x} v^2_{xy}\, \phi(w_{xy}) \overset{(7.28)}{=} \frac{4\, C_{H(\Omega)}\, C_{max}}{C_{H(E)}} \|v\|^2_{H(E)} .$$

□

Remark.

- Due to the dependence of the right-hand side of (7.41) on the functions γ, ϕ and χ, we cannot propose an explicit bound on τ. Hence, the bound τ depends on the special choice of these functions and the underlying weighted graph w.

- In *Section 7.3.7*, we discussed the choice of functions γ, ϕ and χ, to obtain the standard discrete local gradient and divergence operators in our framework. Using the definitions from there, we obtain from (7.41) the following bound on τ in the two dimensional discrete local setting,

$$0 \ < \ \tau \ \leq \ \frac{1}{2\left(\frac{1}{h_1^2} \ + \ \frac{1}{h_2^2}\right)} \ ,$$

  where h_1 and h_2 denote the stepsizes of the image grid in the first and second dimension. In the case of unit step sizes, this bound simplifies to $\tau \ \leq \ \frac{1}{4}$, which does not coincide with the original bound $\tau \ \leq \ \frac{1}{8}$ of Chambolle in [41]. The reason is that the inner products defined in [41] are unweighted and not normalized, in contrast to the definition of the inner products (7.28) in our framework. Hence, if we formally set $\chi \ \equiv \ 1$ and $\phi \ \equiv \ 1$ on $\mathbb{R}_{>0}$, as well as $C_{H(\Omega)} \ = \ C_{H(E)} \ = \ 1$, then we obtain from (7.41) the following bound on τ,

$$0 \ < \ \tau \ \leq \ \frac{1}{4\left(\frac{1}{h_1^2} \ + \ \frac{1}{h_2^2}\right)} \ ,$$

  which simplifies in the case of unit stepsizes to $\tau \ \leq \ \frac{1}{8}$.

- In the case of the weighted NL-ROF model (7.37), the convergence proof of Gilboa and Osher in [79] for the NL-ROF problem can be transferred to the weighted version. Hence, we can guarantee the convergence of the algorithm to an optimal solution with *Lemma 7.4.1*, if the damping parameter τ satisfies

$$0 \ < \ \tau \ \leq \ \frac{C_{H(E)}}{4\, C_{H(\Omega)}\, C_{max}\, \gamma \, \|h\|_{L^\infty(\Omega)}} \ . \tag{7.43}$$

The (dual) projected gradient descent algorithm for the weighted NL-ROF model (7.37) can be now summarized as in *Algorithm 7.1*.

7.4.2 Weighted NL-ROF : Augmented Lagrangian Method

Using the definition of the NL-TV functional $|\cdot|_{NL-BV(\Omega)}$ in (7.10), we can proceed analogously to *Section 6.3.4* to provide a numerical scheme for the weighted NL-ROF problem (7.37). Thus, the subproblem with respect to u^{n+1} consists in solving the following linear equation (cf. (6.53)),

$$(\mu_2 I \ - \ \mu_1 \Delta_w)\, u^{n+1} \ = \ \lambda_2^n \ + \ \mu_2\, \tilde{u}^n \ - \ \mathrm{div}_w(\lambda_1^n \ + \ \mu_1\, v^n) \ , \tag{7.44}$$

where div_w and Δ_w are defined in (7.21) and (7.24). Since the graph Laplacian Δ_w is negative semi-definite (see *Lemma 7.3.6*) and the operator $\mu_2 I \ - \ \mu_1\, \Delta_w$ with weight

Algorithm 7.1 *Projected Gradient Descent Algorithm for Weighted NL-ROF*

1. ***Parameters:*** $q\,,\ h\,,\ \gamma\ >\ 0\,,\ \sigma\ >\ 0\,,\ \delta\ >\ 0\,,\ tol\ >\ 0$
2. ***Initialization:*** $n\ =\ 0\,,\ g^0\ :=\ 0\,,\ stop\ :=\ c\ >\ tol$

   $\tau\ :=\ C_{H(E)}\,/\,(4\,C_{H(\Omega)}\,C_{max}\,\gamma\,\|h\|_{L^\infty(\Omega)})$ ▷ (7.43), (7.41)
3. ***Iteration:***

   a) Compute weighted graph w_ξ via (7.4) with reference image $\xi\ =\ q$.

   while (*stop* $\geq$ *tol*) **do**

   if *Isotropic TV Formulation* **then**

   i) Compute g^{n+1} via (7.39).

   end if

   if *Anisotropic TV Formulation* **then**

   i) Compute g^{n+1} via (7.40).

   end if

   ii) Set *stop* $=\ \|\gamma\,h\,\mathrm{div}_w g^{n+1}\ -\ \gamma\,h\,\mathrm{div}_w g^n\|_{L^2(\Omega)}$.

   iii) $n\ \leftarrow\ n+1$

   end while
4. ***Return*** $u\ :=\ q\ -\ \gamma\,h\,\mathrm{div}_w g^n$ ▷ (7.38)

function w is diagonal dominant, u^{n+1} in (7.44) can be solved by using a Gauss-Seidel algorithm. Moreover, the minimization problem with respect to $\tilde{u}^{n+1}$ can be computed by using an explicit formula given by (6.55). Similar as in *Section 6.3.4*, the vector-field v^{n+1} is obtained by applying a shrinkage operator. In the case of the *anisotropic NL-TV formulation*, the vector $v^{n+1}(x,\cdot)$ at a point $x\ \in\ \Omega$ is given similar to (6.57) by

$$\begin{aligned} v^{n+1}(x,y)\ &=\ \mathrm{sgn}\Big(\big(\nabla_w u^{n+1}\ -\ (1/\mu_1)\,\lambda_1^n\big)(x,y)\Big) \\ &\qquad\qquad \max\Big(\,\big|\big(\nabla_w u^{n+1}\ -\ (1/\mu_1)\,\lambda_1^n\big)(x,y)\big|\ -\ (\gamma/\mu_1)\,,\,0\Big) \end{aligned} \tag{7.45}$$

for all $y\ \in\ N(x)$, where ∇_w is defined in (7.19). In the case of the *isotropic NL-TV formulation*, the vector $v^{n+1}(x,\cdot)$ at a point $x\ \in\ \Omega$ is given similar to (6.58) by

$$\begin{aligned} v^{n+1}(x,y)\ &=\ \frac{\big(\nabla_w u^{n+1}\ -\ (1/\mu_1)\,\lambda_1^n\big)(x,y)}{\big|\big(\nabla_w u^{n+1}\ -\ (1/\mu_1)\,\lambda_1^n\big)\big|_{\ell^2}(x)} \\ &\qquad\qquad \max\Big(\,\big|\big(\nabla_w u^{n+1}\ -\ (1/\mu_1)\,\lambda_1^n\big)\big|_{\ell^2}(x)\ -\ (\gamma/\mu_1)\,,\,0\Big) \end{aligned} \tag{7.46}$$

for all $y \in N(x)$, where the vector norm $|\cdot|_{\ell^2}$ is given by (7.18). Finally, the updates of the Lagrange multipliers are given similar to (6.51) and (6.52) by

$$\lambda_1^{n+1} = \lambda_1^n + \mu_1 (v^{n+1} - \nabla_w u^{n+1}) \,, \tag{7.47}$$

$$\lambda_2^{n+1} = \lambda_2^n + \mu_2 (\tilde{u}^{n+1} - u^{n+1}) \,. \tag{7.48}$$

The augmented Lagrangian method for the weighted NL-ROF model (7.37) can be now summarized as in *Algorithm 7.2*.

Algorithm 7.2 *Augmented Lagrangian Method for Weighted NL-ROF*

1. ***Parameters:*** $q\,,\ h\,,\ \gamma > 0\,,\ \mu_1 > 0\,,\ \mu_2 > 0\,,\ \sigma > 0\,,\ \delta > 0\,,\ tol > 0$
2. ***Initialization:*** $n = 0\,,\ \tilde{u}^0 := q\,,\ v^0 := 0\,,\ \lambda_1^0 := 0\,,\ \lambda_2^0 := 0\,,\ stop := c > tol$
3. ***Iteration:***

   a) Compute weighted graph w_ξ via (7.4) with reference image $\xi = q$.

   while ($stop \geq tol$) **do**

   i) Compute u^{n+1} via (7.44) using Gauss-Seidel algorithm.
   ii) Compute $\tilde{u}^{n+1}$ via (6.55).

   if *Isotropic TV Regularization* **then**
   iii) Compute v^{n+1} via (7.46).
   end if

   if *Anisotropic TV Regularization* **then**
   iii) Compute v^{n+1} via (7.45).
   end if

   iv) Update λ_1^{n+1} via (7.47).
   v) Update λ_2^{n+1} via (7.48).
   vi) Set $stop = \|\tilde{u}^{n+1} - \tilde{u}^n\|_{L^2(\Omega)}$.
   vii) $n \leftarrow n+1$

   end while
4. ***Return*** $\tilde{u}^n$

8 Reconstruction : Results in PET and US Imaging

In the following we will illustrate the performance of the numerical schemes, which we proposed in *Chapters 4* and *5*, by 2D and 3D reconstructions on synthetic and real data in positron emission tomography and medical ultrasound imaging.

8.1 Positron Emission Tomography (PET)

Positron emission tomography (PET) is a biomedical imaging technique, which enables to visualize biochemical and physiological processes, such as glucose metabolism, blood flow or receptor concentrations (see e.g. [167, 160, 12]). This modality is mainly applied in *nuclear medicine* and can be used for instance to detect tumors, to locate areas of the heart affected by coronary artery disease and to identify brain regions influenced by drugs. Therefore, PET is categorized as a *functional imaging technique* and differs from methods such as X-ray computed tomography (CT) that depicts priori anatomy structures. The data acquisition in PET is based on weak radioactively marked pharmaceuticals, so-called *tracers*, which are injected into the blood circulation, and bindings dependent on the choice of the tracer to the molecules to be studied. Used markers are suitable radio-isotopes, which decay by emitting a positron, which annihilates almost immediately with an electron. The resulting emission of two photons will then detected by the tomograph device. Due to the radioactive decay, measured data can be modeled as an *inhomogeneous Poisson process* with a mean given by the *X-ray transform* of the spatial tracer distribution [123, Sect. 3.2]. The X-ray transform maps a function on $\mathbb{R}^d$ into the set of its line integrals [123, Sect. 2.2]. More precisely, if $\theta \in S^{d-1}$ and $x \in \theta^\perp$, then the X-ray transform $\bar{K}$ may be defined by

$$(\bar{K}u)(\theta, x) \;=\; \int_{\mathbb{R}} u(x \,+\, t\,\theta)\; dt \;,$$

and corresponds to the integral of u over the straight line through x with direction θ. Up to notation, in the two dimensional case the X-ray transform coincides with the more popular Radon transform, which maps a function on $\mathbb{R}^d$ into the set of its hyperplane integrals [123, Sect. 2.1]. If $\theta \in S^{d-1}$ and $s \in \mathbb{R}$, then the Radon transform can be defined by

$$(\bar{K}u)(\theta, s) \;=\; \int_{x\,\cdot\,\theta\,=\,s} u(x)\;dx \;=\; \int_{\theta^\perp} u(s\,\theta \,+\, y)\;dy\;, \tag{8.1}$$

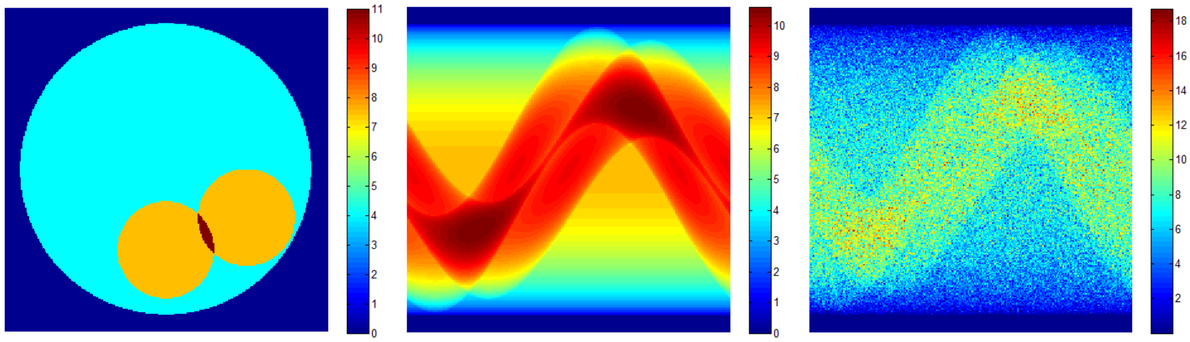
and corresponds in the two dimensional case to the integral of u over the straight line represented by a direction θ and a distance to origin s.

In the following sections, we illustrate the performance of the (Bregman-)FB-EM-REG algorithm proposed in *Sections 4.4.1* and *4.7.3* using synthetic and real data in PET. Here, we use total variation regularization and augmented Lagrangian method described in *Section 6.3.4* in order to solve the weighted ROF problem (6.35) occuring in the regularization half steps of our splitting strategy. In the augmented Lagrangian method, we have seen that the inversion of the Laplace operator equation in (6.53) can be solved efficiently using the discrete cosine transform (6.54). To realize this iteration step, we use a MATLAB implementation of the multidimensional (inverse) discrete cosine transform of A. Myronenko [121].

8.1.1 2D Synthetic Results

In this section we compute reconstruction results using synthetic 2D PET data $f \in \mathbb{R}^{257\times 256}$ (see *Fig. 8.1c*) simulated for a simple object $\bar{u} \in \mathbb{R}^{256\times 256}$ (see *Fig. 8.1a*). The data are obtained via a Monte-Carlo simulation for $s \in [-1,1]$ sampled at 257 samples and $\theta \in [0, 2\pi]$ sampled at 256 samples in (8.1), using one million simulated events.

In *Fig. 8.2* we present EM reconstructions for different numbers of iterations following algorithm (4.15) with data f illustrated in *Fig. 8.1c*. We can observe that early stopping in *Fig. 8.2a* leads to a natural regularization, however with blurring effects and inhomogeneities in the whole object. A higher number of iterations leads to sharper results, as in *Fig. 8.2b*, however the reconstructions suffer more and more from the undesired "checkerboard effect", as in *Fig. 8.2c*, due to the convergence of results to positions of single decay events. In *Fig. 8.2d* we additionally display the expected monotone descent of the objective functional in (4.13) for 1000 EM iterations. Finally, we present in *Fig. 8.2e* the typical behavior of EM iterates for ill-posed problems as described in [135]. Namely, the (metric) distance, here Kullback-Leibler, between the iterates and the exact solution decreases initially before it increases as the noise is



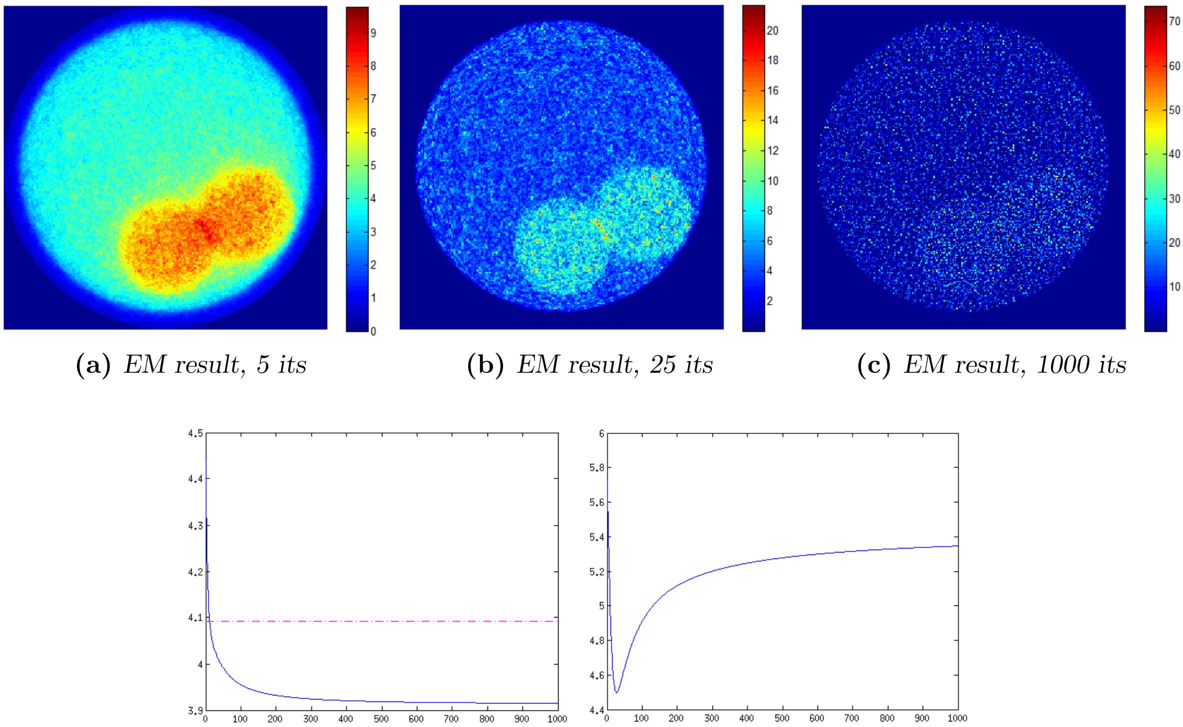
(a) *Original object* **(b)** *Analytical data* **(c)** *Simulated measurements*

Fig. 8.1. *Synthetic 2D PET data. (a) Exact object* $\bar{u} \in \mathbb{R}^{256 \times 256}$. *(b) Exact Radon data* $\bar{f} = \bar{K}\bar{u} \in \mathbb{R}^{257 \times 256}$ *with* $s \in [-1, 1]$ *sampled at* 257 *samples and* $\theta \in [0, 2\pi]$ *sampled at* 256 *samples in* (8.1). *(c) Simulated PET measurements* f *via a Monte-Carlo simulation with* s *and* θ *as in (b), using one million events.*

amplified during the iteration process. The minimal distance in *Fig. 8.2e* is reached approximately after 25 iterations.

In *Fig. 8.3* we illustrate reconstruction results obtained with the FB-EM-TV algorithm (4.24) using different regularization parameters α. In comparison to the EM results in *Fig. 8.2*, the regularized EM algorithm reduces noise and oscillations very well, and reconstructs successful the main geometrical configurations of the desired image in *Fig. 8.1a*, despite the low SNR of the given data in *Fig. 8.1c*. In *Fig. 8.3a* the reconstruction is under-smoothed, whereas in *Fig. 8.3c* the computed image is over-smoothed. A visually resonable reconstruction is illustrated in *Fig. 8.3b*. Moreover, different statistical results for the FB-EM-TV reconstruction in *Fig. 8.3b* are plotted in *Fig. 8.4*. As expected, we observe a decreasing behavior of the objective functional values and Kullback-Leibler distances to the given measurements f and exact image $\bar{u}$.

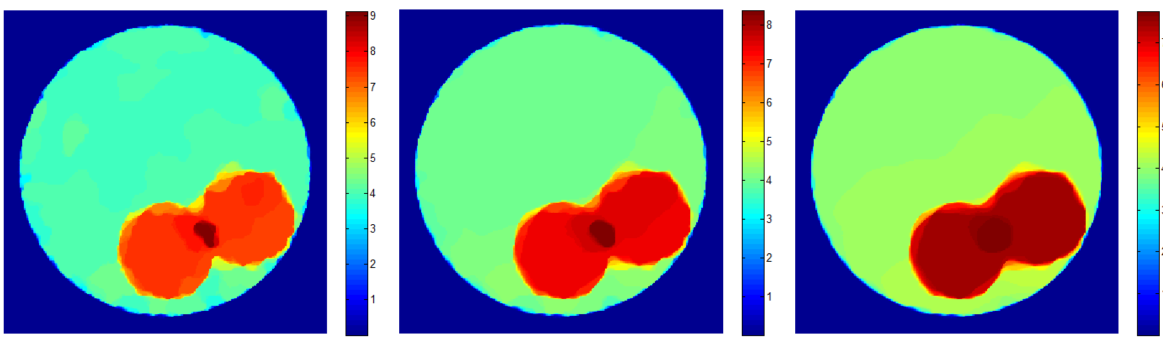
In *Fig. 8.5* we present reconstruction results for different refinement steps of the Bregman-FB-EM-TV algorithm proposed in *Section 4.7.3*. Corresponding to the characteristic of inverse scale space methods, we observe that the results will be improved with increasing iteration number with respect to the systematic error of chosen regularization functional. In the case of the total variation regularization, the systematic error is the reduction of contrast, which will be refined by Bregman distance iteration, as we can observe in the maximal intensity of reconstructions in *Fig. 8.5*. Moreover, in *Fig. 8.5f* we plot the Kullback-Leibler distance between the Bregman iterates and the exact solution. There, we can see that the reconstruction result at 7th refinement step has the smallest distance to the original image $\bar{u}$.



(a) *EM result, 5 its* (b) *EM result, 25 its* (c) *EM result, 1000 its*

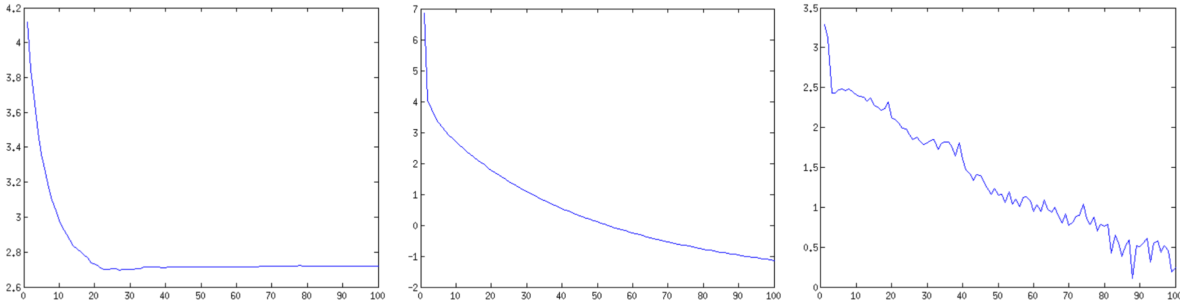
(d) $D_{KL}(f, \bar{K}u_k)$ (e) $D_{KL}(u_k, \bar{u})$

Fig. 8.2. *Synthetic 2D PET data from Fig. 8.1: EM reconstructions. (a)-(c) Reconstruction results obtained with the EM algorithm* (4.15) *and stopped at different iteration numbers. (d) Kullback-Leibler distances* D_{KL} *between given measurements* f *and transformed EM iterates* $\bar{K}u_k$ *for* 1000 *iterations (solid line), as well as between* f *and exact Radon data* $\bar{K}\bar{u}$ *(dash-dot line). (e) Kullback-Leibler distance between EM iterates* u_k *and exact object* $\bar{u}$ *for* 1000 *iterations.*

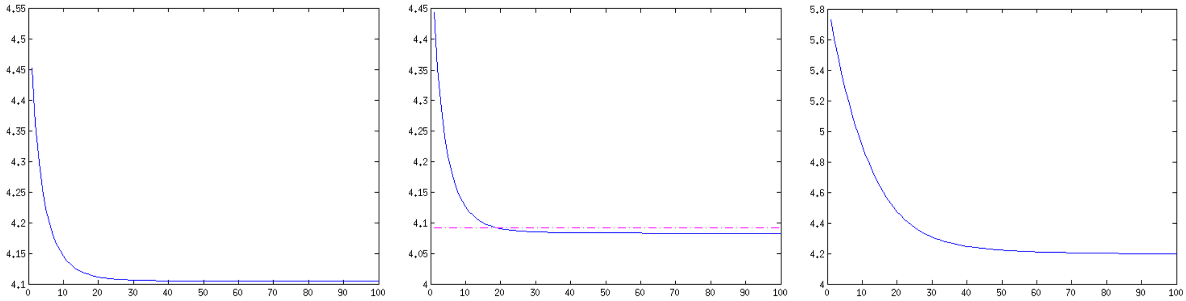


(a) *FB-EM-TV,* $\alpha = 0.0003$ (b) *FB-EM-TV,* $\alpha = 0.0005$ (c) *FB-EM-TV,* $\alpha = 0.0008$

Fig. 8.3. *Synthetic 2D PET data from Fig. 8.1: FB-EM-TV reconstructions. (a) - (c) Reconstruction results obtained with the FB-EM-TV splitting algorithm* (4.24) *using different regularization parameters* α .



(a) opt_k (4.32) **(b)** u_{opt_k} (4.33) **(c)** p_{opt_k} (4.33)



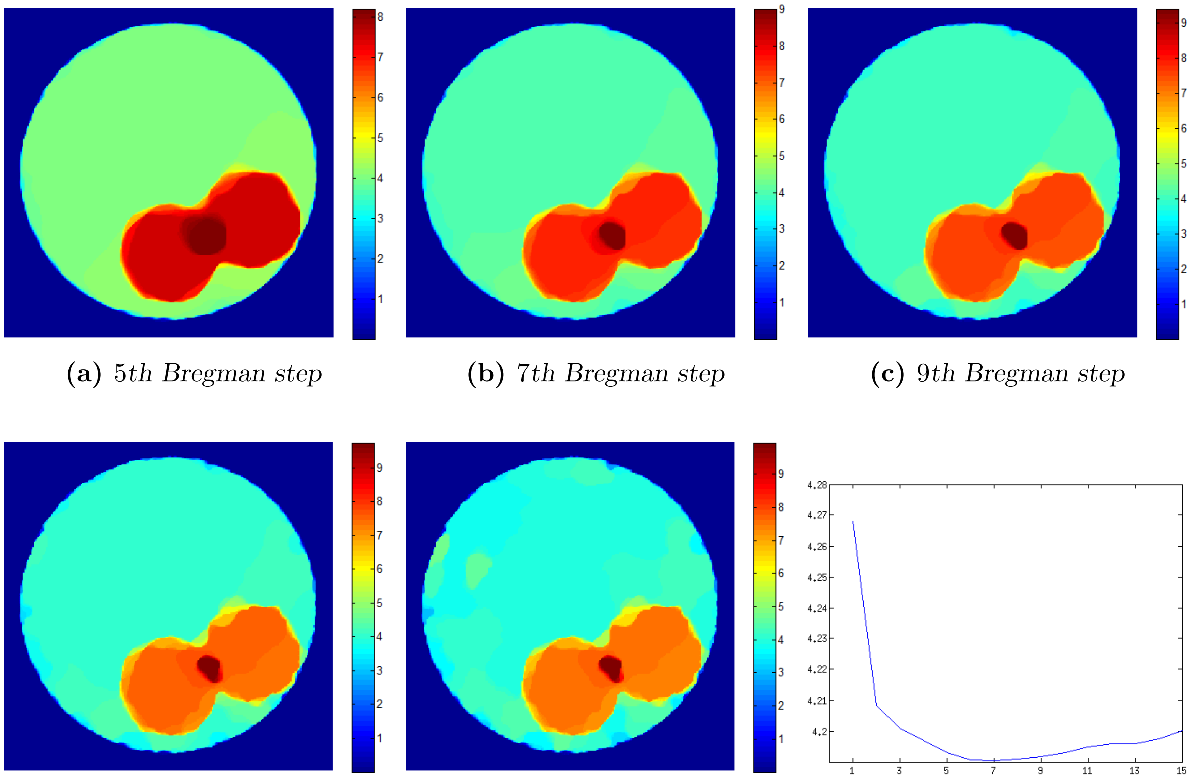
(d) $D_{KL}(f, \bar{K}u_k) + \alpha\, |u_k|_{BV}$ **(e)** $D_{KL}(f, \bar{K}u_k)$ **(f)** $D_{KL}(u_k, \bar{u})$

Fig. 8.4. *Synthetic 2D PET data from Fig. 8.1: different statistics for the result in Fig. 8.3b with 100 FB-EM-TV iterations. (a) - (c) Stopping rules proposed in Section 4.4.4. (d) Values of the objective functional. (e) Kullback-Leibler distances between given measurements f and transformed FB-EM-TV iterates $\bar{K}u_k$ (solid line), as well as between f and exact Radon data $\bar{K}\bar{u}$ (dash-dot line). (e) Kullback-Leibler distance between FB-EM-TV iterates u_k and exact object $\bar{u}$.*

8.1.2 2D Real Data Results

In *Fig. 8.6* we illustrate the performance of the FB-EM-TV algorithm by evaluation of cardiac $H_2\,^{15}O$ measurements obtained with positron emission tomography. This tracer is used in the nuclear medicine for the quantification of myocardial blood flow [142]. However, this quantification needs a segmentation of myocardial tissue, left and right ventricle [142, 20], which is extremely difficult to realize due to very low SNR of $H_2\,^{15}O$ data. Hence, to obtain the tracer intensity in the right and left ventricle, we take a fixed 2D layer in two different time frames.

The tracer intensity in the right ventricle is illustrated in *Fig. 8.6a*, whereby the tracer intensity in the left ventricle is presented in *Fig. 8.6b*, using measurements 25 seconds and 45 seconds after tracer injection in the blood circulation respectively. To illustrate the SNR problem, we present in *Fig. 8.6 (left)* reconstructions with the classical EM algorithm. As expected, the results suffer from unsatisfactory quality and are impossible



(a) 5*th Bregman step* **(b)** 7*th Bregman step* **(c)** 9*th Bregman step*

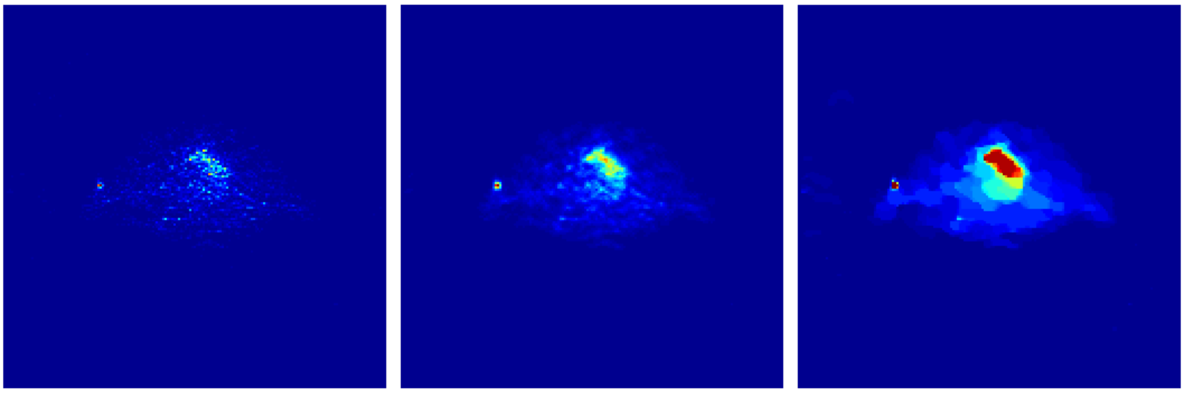
(d) 11*th Bregman step* **(e)** 13*th Bregman step* **(f)** $D_{KL}(u^l, \bar{u})$

Fig. 8.5. *Synthetic 2D PET data from Fig. 8.1: Bregman-FB-EM-TV reconstructions. (a) - (e) Reconstruction results at different refinement steps of the Bregman-FB-EM-TV algorithm proposed in Section 4.7.3. (f) Kullback-Leibler distance between Bregman-FB-EM-TV iterates* u^l *and exact object* $\bar{u}$ *for 15 Bregman iterations.*

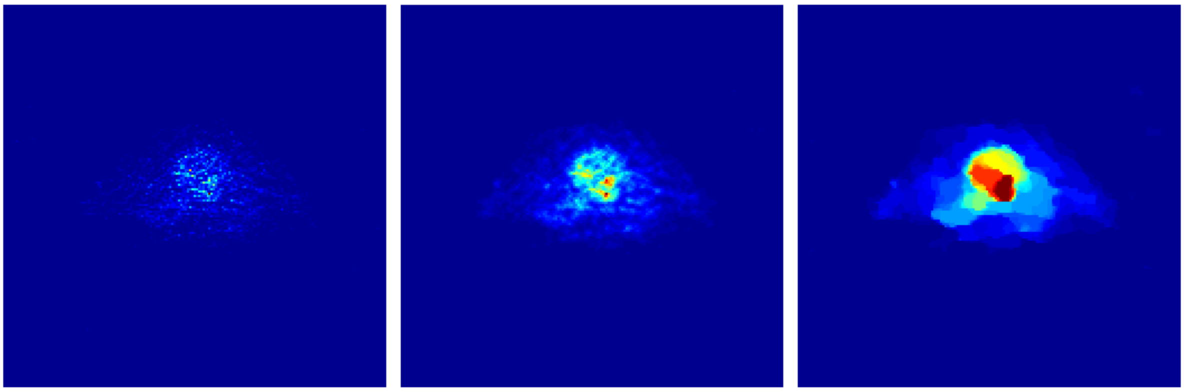
to interpret. Hence, we take EM reconstructions with Gaussian smoothing (*Fig. 8.6 (middle)*) as references. The results in *Fig. 8.6 (right)* show the reconstructions with the proposed FB-EM-TV algorithm. We can see that the results with the FB-EM-TV algorithms are well suited for further use, such as segmentation for quantification of myocardial blood flow, despite the very low SNR of $H_2\,^{15}O$ data [20].

8.1.3 3D Real Data Results

In this section we present 3D reconstruction results generated with the (Bregman-)FB-EM-TV algorithm using cardiac ^{18}F-FDG measurements obtained with PET. The measurements and corresponding 3D EM algorithm for the reconstruction process were provided by K. Schäfers and T. Kösters (EIMI, WWU Münster). The ^{18}F-FDG tracer is an important radiopharmaceutical in nuclear medicine and is used for measuring glucose



(a) *Right ventricle: EM, Gaussian smoothed EM and FB-EM-TV results (from left to right)*

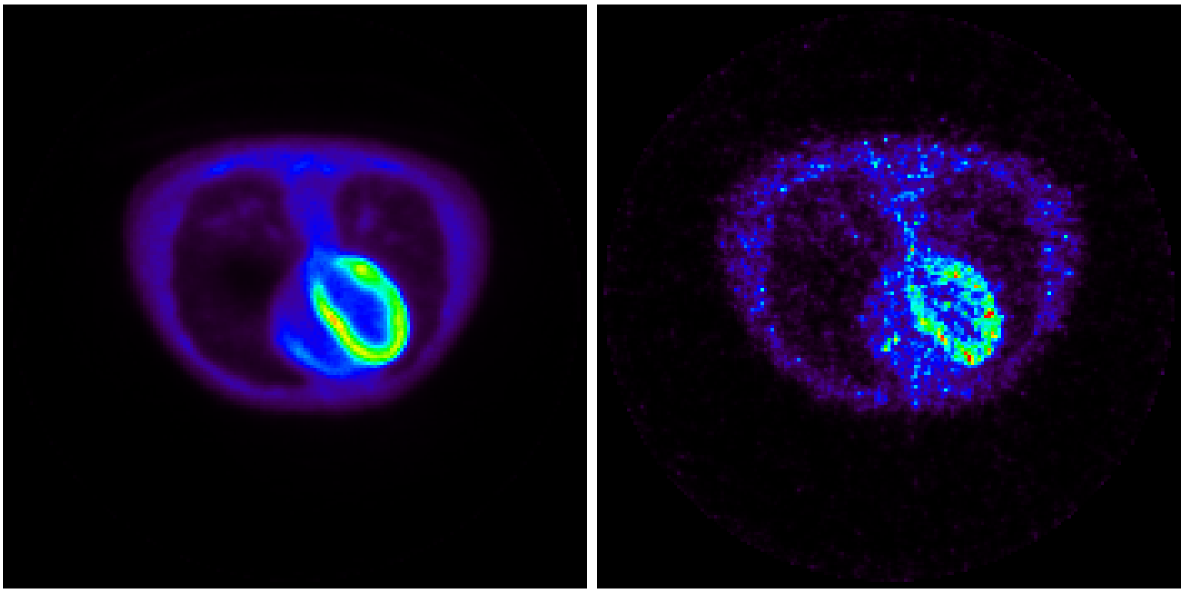


(b) *Left ventricle: EM, Gaussian smoothed EM and FB-EM-TV results (from left to right)*

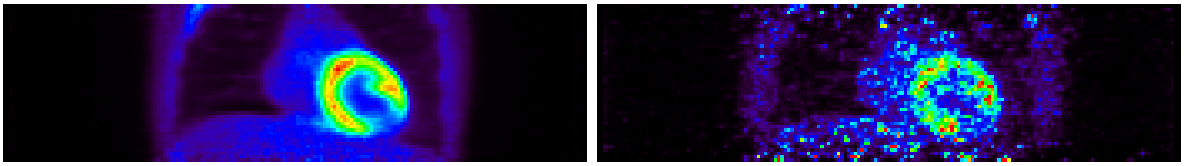
Fig. 8.6. *Cardiac $H_2{}^{15}O$ PET measurements: tracer intesity results of different reconstruction methods in two different time frames. (a) Tracer intensity in the right ventricle using measurements 25 seconds after tracer injection in the blood circulation. (b) Tracer intensity in the left ventricle using measurements 45 seconds after tracer injection in the blood circulation.*

metabolism, e.g. in brain, heart or tumors. In the following, in order to illustrate the 3D data set, we take a fixed transversal, coronal and sagittal slice of reconstructions. In *Fig. 8.7 (left)* we display a Gaussian smoothed EM reconstruction after a data acquisition of 20 minutes as a ground truth for very high count rates. To simulate low count rates, we take the measurements after the first 5 seconds only. The corresponding Gaussian smoothed EM reconstruction is illustrated in *Fig. 8.7 (right)*.

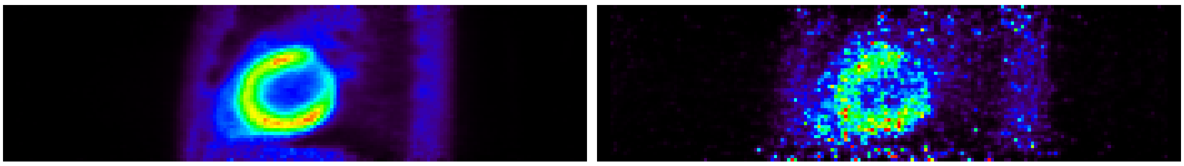
In *Fig. 8.8* we show reconstruction results obtained with the FB-EM-TV algorithm (left) and its extension via Bregman distance regularization (right) using measurements after 5 seconds acquisition time of the data. There, we can observe that the major structures of the object are well reconstructed by both approaches also for low count rates. However, as expected, the structures in the Bregman-FB-EM-TV result can be identified better than in the standard FB-EM-TV reconstruction. In particular, this aspect can be observed well in *Fig. 8.9*, where we present scaled versions of both reconstructions in order to allow a quantitative comparison. In *Fig. 8.9*, the reconstructions from in *Fig. 8.8* are scaled to the maximum intensity of the EM result in *Fig. 8.7 (left)* obtained with measurements after 20 minutes data acquisition. There, we can observe that the result with the Bregman-FB-EM-TV algorithm has more realistic quantitative values than the reconstruction with the standard FB-EM-TV algorithm.



(a) *Transversal view: 20 minutes (left) and 5 seconds (right) data acquisition time*

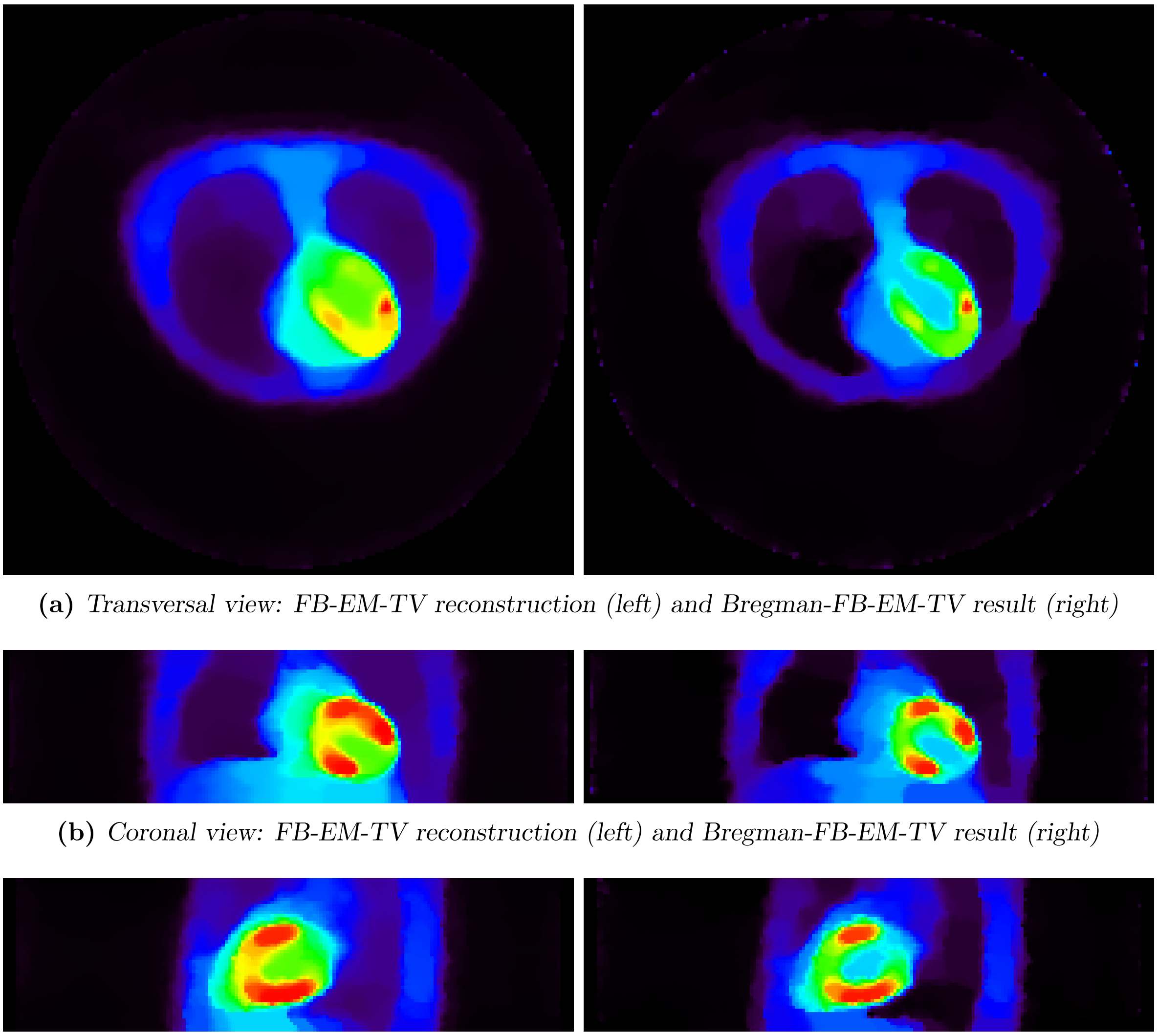


(b) *Coronal view: 20 minutes (left) and 5 seconds (right) data acquisition time*



(c) *Sagittal view: 20 minutes (left) and 5 seconds (right) data acquisition time*

Fig. 8.7. *Cardiac* ^{18}*F-FDG 3D PET measurements: tracer intensity results obtained with the EM algorithm* (4.15) *for different count rates.* ***Left:*** *EM reconstruction, 20 iterations, with Gaussian smoothing any 10th step after 20 minutes data acquisition.* ***Right:*** *As left but after 5 seconds data acquisition. Additionally, the reconstruction is scaled to the maximum intensity of the result on the left-hand side due to the strong presence of noise outside of region of interest.*



(a) *Transversal view: FB-EM-TV reconstruction (left) and Bregman-FB-EM-TV result (right)*

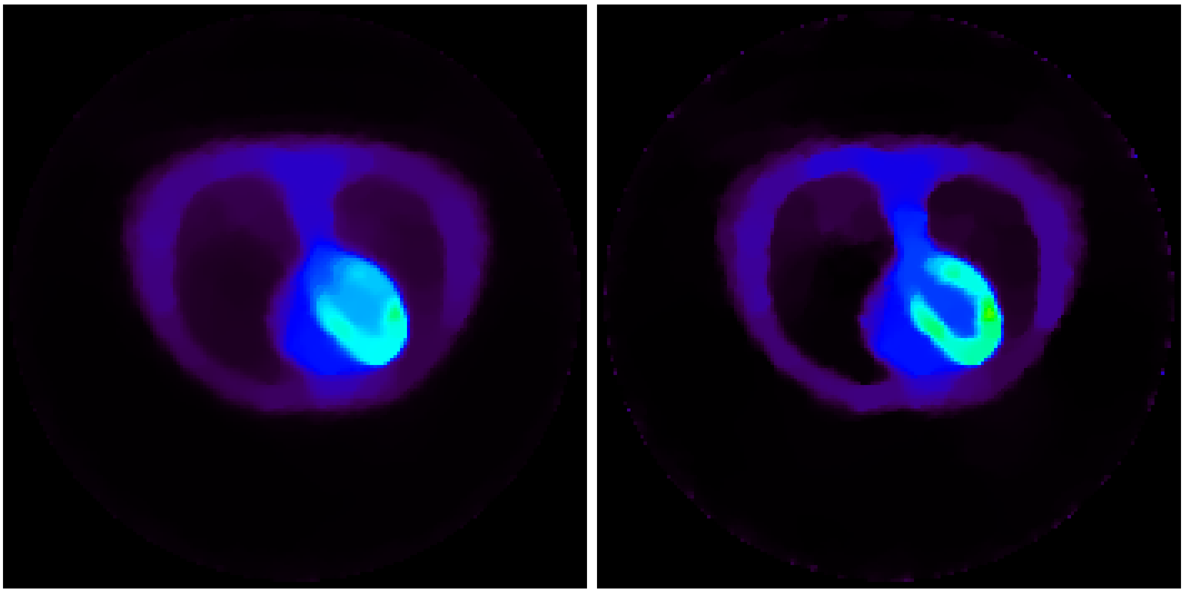
(b) *Coronal view: FB-EM-TV reconstruction (left) and Bregman-FB-EM-TV result (right)*

(c) *Sagittal view: FB-EM-TV reconstruction (left) and Bregman-FB-EM-TV result (right)*

Fig. 8.8. *Cardiac ^{18}F-FDG 3D PET measurements: tracer intensity results obtained with the (Bregman-)FB-EM-TV algorithm for measurements after 5 seconds data acquisition.* ***Left:*** *Reconstruction with the FB-EM-TV algorithm* (4.24), *20 iterations.* ***Right:*** *Reconstruction with the Bregman-FB-EM-TV algorithm proposed in Section 4.7.3 at 6th refinement step.*

8.2 Poisson Noise : TV vs. NL-TV Regularization

In this section we compare the performance of TV and NL-TV regularization in restoration problems with Poisson noise. In the case of NL-TV regularization, we use the projected gradient descent algorithm (with isotropic NL-TV formulation) proposed in *Section 7.4.1* in order to solve the weighted NL-ROF problem (7.37), which occurs in the regularization half steps of the FB-EM-REG algorithm. The algorithm is implemented in C with a MEX-interface to MATLAB and is parallelized with OpenMP. In the current



(a) *Transversal view: FB-EM-TV reconstruction (left) and Bregman-FB-EM-TV result (right)*

(b) *Coronal view: FB-EM-TV reconstruction (left) and Bregman-FB-EM-TV result (right)*

(c) *Sagittal view: FB-EM-TV reconstruction (left) and Bregman-FB-EM-TV result (right)*

Fig. 8.9. *Cardiac ^{18}F-FDG 3D PET measurements: quantitative comparison between Bregman- and FB-EM-TV reconstructions for measurements after 5 seconds data acquisition.* ***Left and right:*** *Results from Fig. 8.8 are scaled to the maximum intensity of ground truth in Fig. 8.7 (left).*

version of the algorithm, we use the NL-means weights (7.4), however with $G_\sigma \equiv 1$, in order to compute the weighted graph w_ξ. Moreover, in every regularization half step of the FB-EM-REG algorithm we recompute the weighted graph using the current "noisy" image as reference image ξ. The current version of the algorithm performs the fully nonlocal approach without using a search window around a pixel. Hence, in order to improve computational time of the dual iteration of the projected gradient descent algorithm and to reduce the complexity of the weighted graph, we only take a fixed number M of best neighbors with highest weight values for any pixel in an image. Using this approach, the symmetry of the weighted graph will be destroyed such that we have a directed structure of the graph. Finally, in order to implement the algorithm efficiently, we use priority queues and heap structures for the choice of the M best neighbors of every pixel. The corresponding functions to sort these structures are partially taken form [132] and [145].

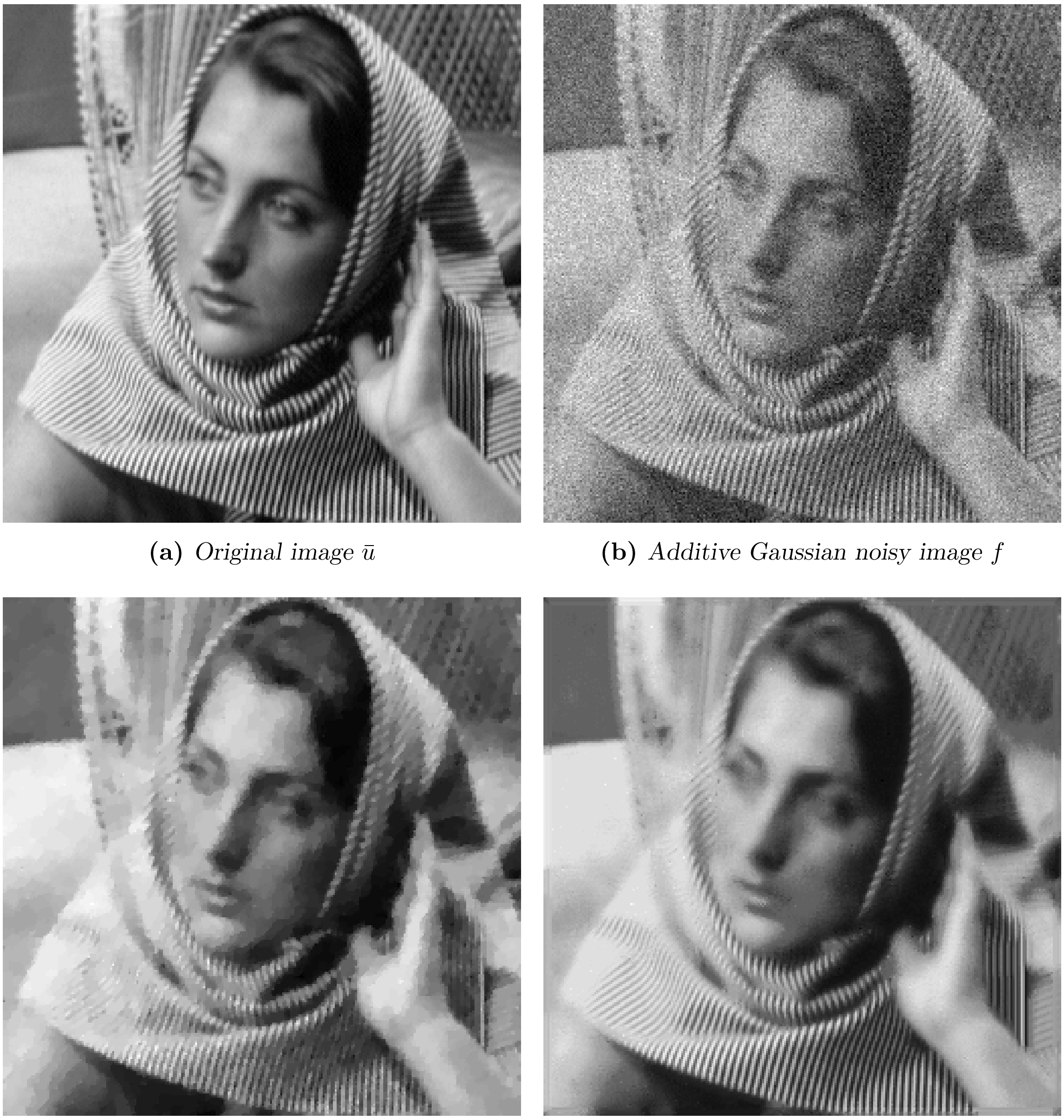
At the beginning we represent in *Fig. 8.10* a comparison between TV and NL-TV regularization in the case of the standard additive Gaussian noise. As expected, the denoised image by the NL-ROF model in *Fig. 8.10d* is significantly better than the result obtained by the standard ROF model in *Fig. 8.10c*. In particular, texture and recurring structures are well preserved by the nonlocal approach.

In *Fig. 8.11* we show the comparison between TV and NL-TV regularization in the case of Poisson noise in the image. For the denoising process, we use the exact Poisson denoising strategy proposed in *Section 4.5.1*. Note that for the computation of the nonlocal weighted graph, we use the NL-means weights (7.4) which are more suitable for additive Gaussian noise in an image from the statistical point of view. Nevertheless, using these weights the Poisson denoised image by the NL-TV regularization functional in *Fig. 8.11d* has better preserved texture and recurring structures compared with TV regularized result in *Fig. 8.11c*.

Finally, motivated by the result in *Fig. 8.11* that the standard NL-means weights are also suitable for Poisson noise, we illustrate in *Fig. 8.12* reconstruction results with the FB-EM-TV and FB-EM-NL-TV algorithm for synthetic PET data simulated on a part of 'Barbara' image. The measurements in *Fig. 8.12c* are obtained by using the forward Radon operator as in *Fig. 8.1*. As expected, the reconstruction result with the standard EM algorithm (4.15) in *Fig. 8.12d* suffers from unsatisfactory quality, despite a low level of noise in the data. In *Fig. 8.12e* and *Fig. 8.12f* reconstruction results with the FB-EM-REG algorithm are presented using TV and NL-TV regularization, respectively. We can see that the FB-EM-NL-TV algorithm delivers visually better result than the local variant, in particular the major structures in the image are preserved better. However, also the nonlocal approach cannot reconstruct texture and fine details present in the original object in *Fig. 8.12a*. The reason is that the forward Radon operator already destroys these high-frequency informations.

8.3 Medical Ultrasound (US) Imaging

Ultrasound (US) imaging is one of the most important techniques in the field of medical diagnostic, which enables a real-time examination of patients in almost all medical fields.



(a) *Original image* $\bar{u}$

(b) *Additive Gaussian noisy image* f

(c) *ROF result* u_{TV}

(d) *NL-ROF result* u_{NL-TV}

Fig. 8.10. *Comparison between TV and NL-TV regularization in the case of additive Gaussian noise. (a) Original image* $\bar{u}$ *(a part of the 'Barbara' image) in the intervall* $[0,1]$. *(b)* $\bar{u}$ *is degraded by additive Gaussian noise with variance* $\sigma^2 = 0.006$. *(c) Denoising result* u_{TV} *with ROF model* (6.6). *(d) Denoising result* u_{NL-TV} *with NL-ROF model* (7.37) $(h \equiv 1)$ *using patch size* 9×9 *and* $M = 10$. *The regularization parameters in (c) and (d) are chosen so that* $\|f - u_{TV}\|_{L^2(\Omega)} \approx \|f - u_{NL-TV}\|_{L^2(\Omega)} \left(> \|f - \bar{u}\|_{L^2(\Omega)} \right)$.

In particular, since US imaging is a low-risk and painless application, it can be used in sensible areas, such as in prenatal care or examination of kidneys and heart. The data acquisition in US imaging is based on the propagation of acoustic waves (ultrasound

(a) *Original image* $\bar{u}$

(b) *Poisson noisy image* f

(c) *FB-EM-TV result* u_{TV}

(d) *FB-EM-NL-TV result* u_{NL-TV}

Fig. 8.11. *Comparison between TV and NL-TV regularization in the case of Poisson noise. (a) Original image* $\bar{u}$ *(a part of the 'Barbara' image) in the intervall* $[0,1]$ *. (b)* $\bar{u}$ *is degraded by Poisson noise in the form that* $\bar{u}$ *is first scaled up by a factor* 100 *, subsequently degraded by Poisson noise and finally scaled back with the same factor. (c) Denoising result* u_{TV} *with TV regularization functional. (d) Denoising result* u_{NL-TV} *with NL-TV regularization functional using patch size* 9×9 *and* $M = 10$ *. In both results, we use the Poisson denoising strategy* (4.37) *with* $\omega_k = 1$ *for all* $k \geq 0$ *and* 5 *iteration steps. The regularization parameters in (c) and (d) are chosen so that* $D_{KL}(f, u_{TV}) \approx D_{KL}(f, u_{NL-TV})$ $(> D_{KL}(f, \bar{u}))$ *.*

(a) *Original image $\bar{u}$* **(b)** *Exact Radon data $\bar{f}$* **(c)** *Poisson noisy data f*

(d) *EM result u_{EM}* **(e)** *FB-EM-TV result u_{TV}* **(f)** *FB-EM-NL-TV result u_{NL-TV}*

Fig. 8.12. *Simulated 2D PET data for a part of 'Barbara' image: results of different reconstruction methods. (a) Original image $\bar{u}$. (b) Exact Radon data $\bar{f} = \bar{K}\bar{u}$ with Radon forward operator $\bar{K}$ from Fig. 8.1. (c) $\bar{f}$ is slightly degraded by Poisson noise. (d) Reconstruction result obtained with the EM algorithm* (4.15). *(e) Reconstruction result obtained with the FB-EM-TV algorithm* (4.24). *(f) Reconstruction result obtained with the FB-EM-NL-TV algorithm* (4.24) *using patch size 9×9 and $M = 10$. In (d) - (f), we always perform* 100 *iteration steps. The regularization parameters in (e) and (f) are chosen so that $D_{KL}(f, \bar{K}u_{TV}) \approx D_{KL}(f, \bar{K}u_{NL-TV}) \left(> D_{KL}(f, \bar{f})\right)$.*

waves) in the body, which will be reflected on the boundary layers between two different mediums. The echos, which arise during this process, will be registered and the echo intensity will be measured, which will be interpreted in order to generate an image of reflections. However, the scattering of the ultrasound beam from tissue inhomogeneities leads to so-called *interference effects*, which will be visible in an image as an acoustic noise called *speckle* and causes the main degradation of the image quality.

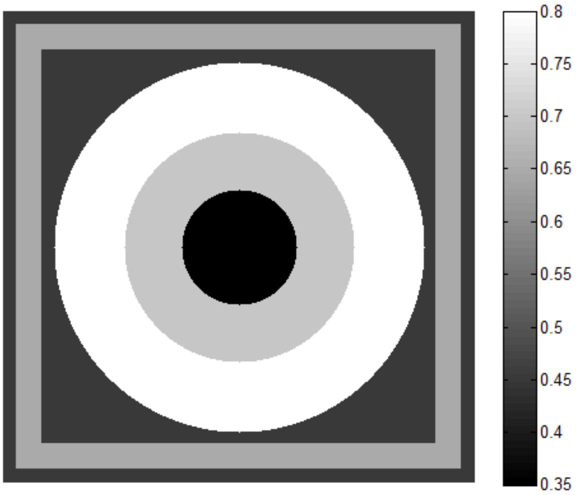
In the following sections, we illustrate the performance of the US-FB-REC-REG algorithm for image denoising problems, which are more relevant issues in medical ultrasound imaging, using synthetic and real data. We use the denoising strategy proposed in *Section 5.3*, and consider total variation regularization and the augmented Lagrangian method described in *Section 6.3.4* in order to solve the weighted ROF problems in (5.15).

8.3.1 2D Synthetic Results

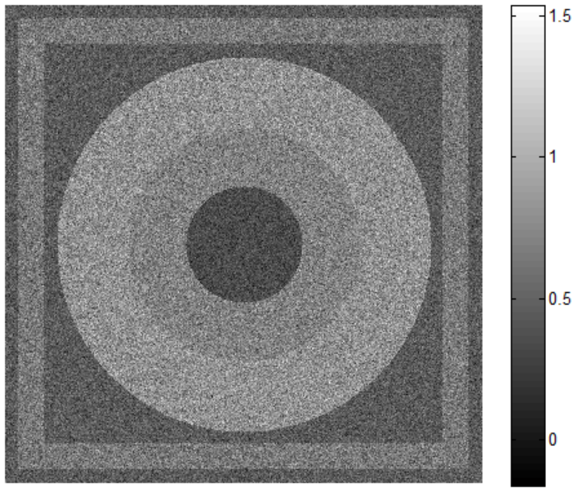
In this section we illustrate the performance of the US speckle noise denoising strategy proposed in *Section 5.3* for a synthetic 2D image presented in *Fig. 8.13a*, where the corresponding US speckle noisy image is given in *Fig. 8.13b*. For this purpose, we compare in *Fig. 8.13* the denoising results obtained with the standard ROF model (6.6) and the US speckle noise adapted strategy (5.15). There, we can see that both approaches deliver relative similar results, excepting the edge distortion of the inner circle in the case of ROF result in *Fig. 8.13c*. This aspect is restored significantly better in the result obtained with the US denoising strategy in *Fig. 8.13d*. Finally, we notice that in contrast to the Poisson framework, where the damping strategy (4.26) is only needed in the case of very high regularization parameter, here the damping approach is always required.

8.3.2 2D Real Data Results

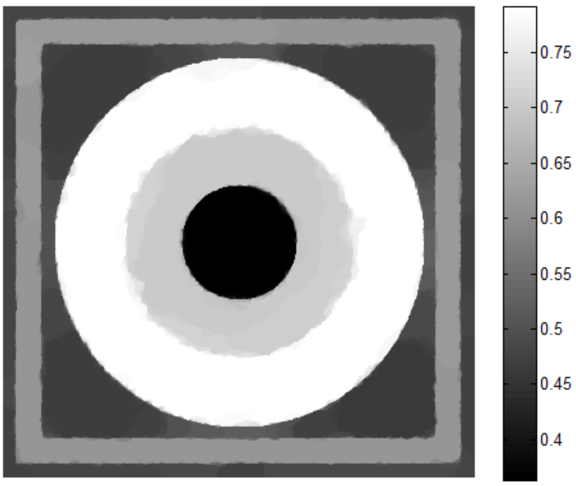
In *Fig. 8.14* we show denoising results for real 2D ultrasound images. The data were provided by J. Stypmann (UKM, Münster). In *Fig. 8.14a* and *Fig. 8.14b* two cardiac data sets are illustrated, in *Fig. 8.14c* a liver data set. There, we can observe that the denoising results on the right-hand side of *Fig. 8.14* seem to be less suited for diagnostic tasks, however are promising for a subsequently automatic segmentation of different anatomical structures in an US image.



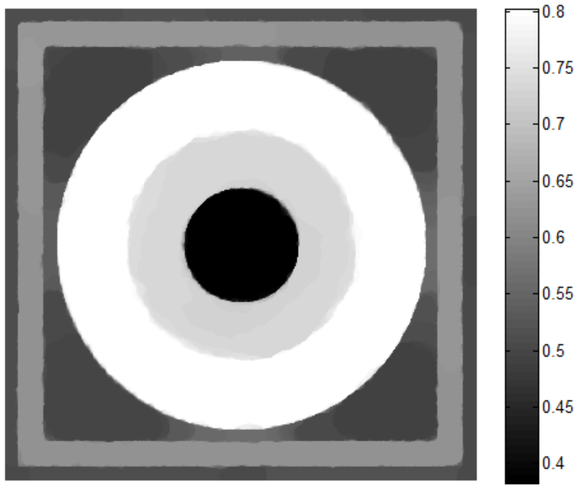
(a) *Original image* $\bar{u}$



(b) *US speckle noisy image* f

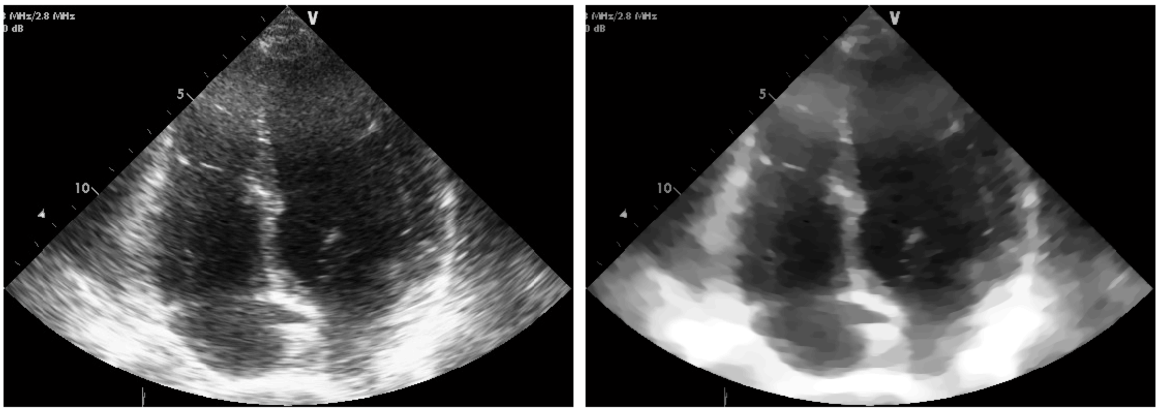


(c) *ROF result* u_{ROF}

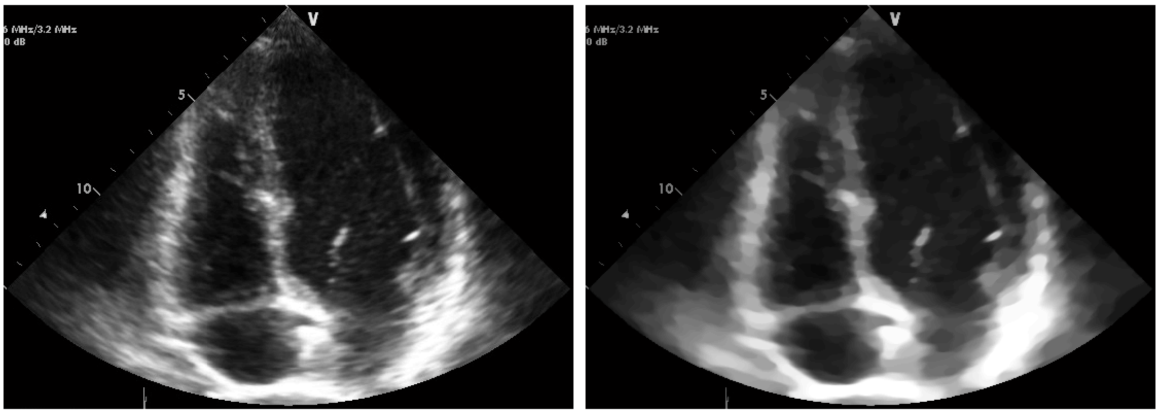


(d) *US denoising scheme result* u_{US}

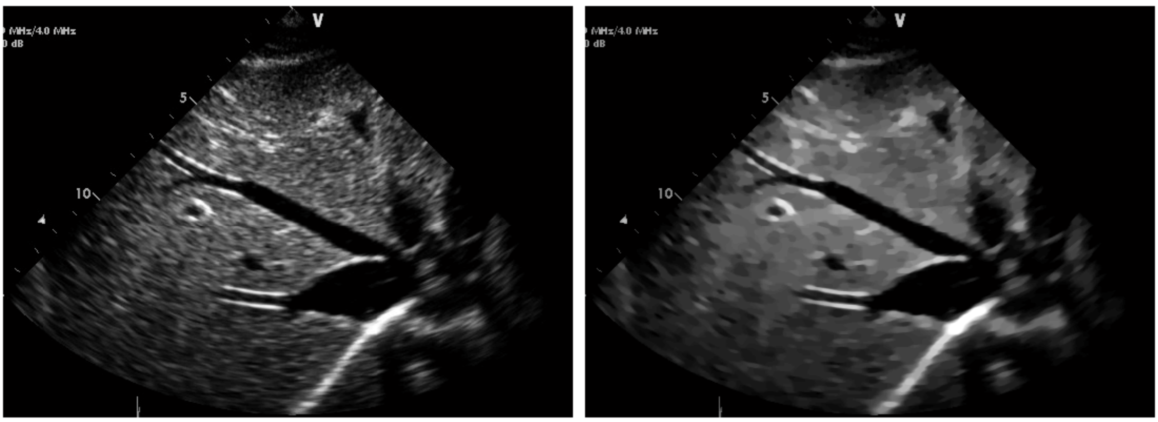
Fig. 8.13. *Synthetic 2D US data: comparison between standard ROF model and US speckle noise adapted model. (a) Original synthetic image* u *. (b)* $\bar{u}$ *degraded by US speckle noise of the form* $f = \bar{u} + \sqrt{\bar{u}}\,\eta$ *, where* η *is a Gaussian distributed random variable with expected value* 0 *and variance* $\sigma^2 = 0.05$ *. (c) Denoising result* u_{ROF} *obtained with the standard ROF model* (6.6)*. (d) Denoising result* u_{US} *obtained with the US denoising strategy* (5.15) *using* 20 *iteration steps with* $\omega_k = 0.1$ *for all* $k \geq 0$ *. Regularization parameters in (c) and (d) are chosen so that* $D_{US}(f, u_{ROF}) \approx D_{US}(f, u_{US})$ $(> D_{US}(f, \bar{u}))$ *, where the functional* D_{US} *is defined in* (5.18)*.*



(a) *US data of left and right ventricle with atria*



(b) *US data of left and right ventricle with atria*



(c) *US data of liver with hepatic veins*

Fig. 8.14. *Real 2D US data: denoising results of different data sets. (a) - (b) US data sets of left and right ventricle with atria for two different frequency settings of the ultrasound device. (c) US data set of liver with hepatic veins.* ***Left:*** *Noisy measurements.* ***Right:*** *Denoising results obtained with the denoising strategy* (5.15) *using* 10 *iteration steps and* $\omega_k = 0.1$ *for all* $k \geq 0$.

Bibliography

[1] J. G. Abbott and F. L. Thurstone, *Acoustic speckle: Theorie and experimental analysis*, Ultrason. Imag., 1 (1979), pp. 303–324. 101

[2] R. Acar and C. R. Vogel, *Analysis of bounded variation penalty methods for ill-posed problems*, Inverse Problems, 10 (1994), pp. 1217–1229. 19, 42, 122, 125, 126, 127, 131

[3] S. Alliney, *A property of the minimum vectors of a regularizing functional defined by means of the absolute norm*, IEEE Trans. Signal Process., 45 (1997), pp. 913–917. 132

[4] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford Mathematical Monographs, Oxford University Press, 2000. 121, 122, 125, 126

[5] G. Aubert and J.-F. Aujol, *A variational approach to removing multiplicative noise*, SIAM J. Appl. Math., 68 (2008), pp. 925–946. 13, 14

[6] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, vol. 147 of Applied Mathematical Sciences, Springer, 2002. 29, 30, 56, 65, 123, 138, 158

[7] J.-F. Aujol, *Some first-order algorithms for total variation based image restoration*, J. Math. Imaging Vis., 34 (2009), pp. 307–327. 145

[8] J.-F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle, *Image decomposition into a bounded variation component and an oscillating component*, J. Math. Imaging Vis., 22 (2005), pp. 71–88. 132

[9] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher, *Structure-texture image decomposition - modeling, algorithms, and parameter selection*, Int. J. Comput. Vis., 67 (2006), pp. 111–136. 131

[10] S. P. Awate and R. T. Whitaker, *Unsupervised, information-theoretic, adaptive image filtering for image restoration*, IEEE Trans. Pattern Anal. Mach. Intell., 28 (2006), pp. 364–376. 162

[11] M. Bachmayr, *Iterative total variation methods for nonlinear inverse problems*, master's thesis, Johannes Kepler University, Linz, 2007. 85

[12] D. L. Bailey, D. W. Townsend, P. E. Valk, and M. N. Maisey, eds., *Positron Emission Tomography: Basic Sciences*, Springer, 2005. 41, 135, 183

[13] D. Barash, *A fundamental relationship between bilateral filtering, adaptive smoothing, and the nonlinear diffusion equation*, IEEE Trans. Pattern Anal. Mach. Intell., 24 (2002), pp. 844–847. 161

[14] J. M. Bardsley, *An efficient computational method for total variation-penalized Poisson likelihood estimation*, Inverse Problems and Imaging, 2 (2008), pp. 167–185. 43, 134

[15] ——, *A theoretical framework for the regularization of Poisson likelihood estimation problems*, Inverse Problems and Imaging, 4 (2010), pp. 11–17. 59, 69

[16] J. M. Bardsley and J. Goldes, *Regularization parameter selection methods for ill-posed Poisson maximum likelihood estimation*, Inverse Problems, 25 (2009), p. 095005. 55

[17] J. M. Bardsley and N. Laobeul, *Tikhonov regularized Poisson likelihood estimation: theoretical justification and a computational method*, Inverse Problems in Science and Engineering, 16 (2008), pp. 199–215. 42, 59

[18] ——, *An analysis of regularization by diffusion for ill-posed Poisson likelihood estimations*, Inverse Problems in Science and Engineering, 17 (2009), pp. 537–550. 42, 59

[19] J. M. Bardsley and A. Luttman, *Total variation-penalized Poisson likelihood estimation for ill-posed problems*, Adv. Comput. Math., 31 (2009), pp. 35–59. 42, 59, 134

[20] M. Benning, T. Kösters, F. Wübbeling, K. Schäfers, and M. Burger, *A nonlinear variational method for improved quantification of myocardial blood flow*

using dynamic $H_2{}^{15}O$ PET, in Nuclear Science Symposium Conference Record, 2008, pp. 4472–4477. 187, 188

[21] B. Berkels, M. Burger, M. Droske, O. Nemitz, and M. Rumpf, *Cartoon extraction based on anisotropic image classification*, in Vision, Modeling, and Visualization Proceedings, 2006, pp. 293–300. 123

[22] M. Bertero, P. Boccacci, G. Talenti, R. Zanella, and L. Zanni, *A discrepancy principle for Poisson data*, Inverse Problems, 26 (2010), p. 105004. 55

[23] M. Bertero, H. Lanteri, and L. Zanni, *Iterative image reconstruction: a point of view*, in Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation Therapy (IMRT), Y. Censor, M. Jiang, and A. Louis, eds., vol. 7 of Publications of the Scuola Normale, CRM series, 2008, pp. 37–63. 10, 11, 13, 14

[24] J. M. Borwein and A. S. Lewis, *Convergence of best entropy estimates*, SIAM J. Optim., 1 (1991), pp. 191–205. 61

[25] S. Bougleux, A. Elmoataz, and M. Melkemi, *Discrete regularization on weighted graphs for image and mesh filtering*, in Proceedings of the 1st International Conference on Scale Space and Variational Methods in Computer Vision, LNCS 4485, Springer, 2007, pp. 128–139. 163

[26] ——, *Local and nonlocal discrete regularization on weighted graphs for image and mesh processing*, Int. J. Comput. Vis., 84 (2009), pp. 220–236. 163

[27] K. Bredies, *A forward-backward splitting algorithm for the minimization of non-smooth convex functionals in Banach space*, Inverse Problems, 25 (2009), p. 015005. 53

[28] L. M. Bregman, *The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming*, USSR Comp. Math. and Math. Phys., 7 (1967), pp. 200–217. 86

[29] T. Brox, O. Kleinschmidt, and D. Cremers, *Efficient nonlocal means for denoising of textural patterns*, IEEE Trans. Image Process., 17 (2008), pp. 1083–1092. 166

[30] C. Brune, *4D Imaging in Tomography and Optical Nanoscopy*, PhD thesis, Institute for Computational and Applied Mathematics, University of Münster, June 2010. http://wwwmath.uni-muenster.de/num/publications/2010/Bru10/. 6

[31] C. Brune, A. Sawatzky, and M. Burger, *Bregman-EM-TV methods with*

application to optical nanoscopy, in Proceedings of the 2nd International Conference on Scale Space and Variational Methods in Computer Vision, vol. 5567 of LNCS, Springer, 2009, pp. 235–246. 6

[32] ——, *Primal and dual Bregman methods with application to optical nanoscopy*, Int. J. Comput. Vis., 92 (2011), pp. 211–229. 6, 92, 99, 100

[33] A. Buades, B. Coll, and J. M. Morel, *A review of image denoising algorithms, with a new one*, Multiscale Model. Simul., 4 (2005), pp. 490–530. 158, 159, 161, 162, 163, 166

[34] ——, *Neighborhood filters and PDE's*, Numer. Math., 105 (2006), pp. 1–34. 161

[35] ——, *The staircasing effect in neighborhood filters and its solution*, IEEE Trans. Image Process., 15 (2006), pp. 1499–1505. 161

[36] C. B. Burckhardt, *Speckle in ultrasound B-mode scans*, IEEE Trans. Sonic Ultrason., 25 (1978), pp. 1–6. 101

[37] M. Burger, K. Frick, S. Osher, and O. Scherzer, *Inverse total variation flow*, Multiscale Model. Simul., 6 (2007), pp. 366–395. 85, 88, 90, 91

[38] M. Burger, G. Gilboa, S. Osher, and J. Xu, *Nonlinear inverse scale space methods*, Comm. Math. Sci., 4 (2006), pp. 179–212. 85, 88, 91, 96

[39] M. Burger and S. Osher, *Convergence rates of convex variational regularization*, Inverse Problems, 20 (2004), pp. 1411–1421. 86

[40] M. Burger, E. Resmerita, and L. He, *Error estimation for Bregman iterations and inverse scale space methods in image restoration*, Computing, 81 (2007), pp. 109–135. 88, 128

[41] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vis., 20 (2004), pp. 89–97. 135, 145, 146, 148, 177, 178, 180

[42] ——, *Total variation minimization and a class of binary MRF models*, in Energy Minimization Methods in Computer Vision and Pattern Recongnition, vol. 3757 of LNCS, Springer, 2005, pp. 136–152. 145

[43] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, vol. 9 of Radon Series Comp. Appl. Math., De Gruyter, 2010, ch. An Introduction to Total Variation for Image Analysis, pp. 263–340. 145

[44] A. Chambolle, R. A. De Vore, N.-Y. Lee, and B. J. Lucier, *Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage*, IEEE Trans. Image Process., 7 (1998), pp. 319–335. 56

[45] A. Chambolle and P. L. Lions, *Image recovery via total variation minimization and related problems*, Numer. Math., 76 (1997), pp. 167–188. 129

[46] R. H. Chan and K. Chen, *Multilevel algorithm for a Poisson noise removal model with total-variation regularization*, Int. J. Comput. Math., 84 (2007), pp. 1183–1198. 56

[47] T. Chan, S. Esedoglu, F. Park, and A. Yip, *Handbook of Mathematical Models in Computer Vision*, Springer, 2006, ch. Total Variation Image Restoration: Overview and Recent Developments, pp. 17–31. 127

[48] T. F. Chan and S. Esedoglu, *Aspects of total variation regularized L^1 function approximation*, SIAM J. Appl. Math., 65 (2005), pp. 1817–1837. 132

[49] T. F. Chan and J. Shen, *Image Processing and Analysis: Variational, PDE, Wavelet and Stochastic Methods*, SIAM, 2005. 56, 121, 158

[50] F. R. K. Chung, *Spectral Graph Theory*, vol. Conference Board of the Mathematical Sciences of Regional Conference Series in Mathematics, American Mathematical Society, 1997. 162

[51] P. Combettes and V. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul., 4 (2005), pp. 1168–1200. 53, 165

[52] P. L. Combettes and J.-C. Pesquet, *A proximal decomposition method for solving convex variational inverse problems*, Inverse Problems, 24 (2008), p. 065014. 53

[53] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1958. 14

[54] I. Csiszar, *Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems*, Ann. Statist., 19 (1991), pp. 2032–2066. 60

[55] C. L. De Vito, *Functional Analysis*, Pure and Applied Mathematics, a Series of Monographs and Textbooks, Academic Press, 1978. 31

[56] A. P. Dempster, N. M. Laird, and D. B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm*, J. Royal Stat. Soc., Series B, 39 (1977), pp. 1–38. 41, 46, 47

[57] N. Dey, L. Blanc-Féraud, C. Zimmer, P. Roux, Z. Kam, J.-C. Olivio-Marin, and J. Zerubia, *3D microscopy deconvolution using Richardson-Lucy algorithm with total variation regularization*, Tech. Report 5272, Institut National de Recherche en Informatique et en Automatique, 2004. 4, 11, 40, 42, 43, 44, 134

[58] I. S. Dhillon and J. A. Tropp, *Matrix nearness problems with Bregman divergences*, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 1120–1146. 88

[59] D. C. Dobson and F. Santosa, *Recovery of blocky images from noisy and blurred data*, SIAM J. Appl. Math., 56 (1996), pp. 1181–1198. 129

[60] D. L. Donoho and I. M. Johnstone, *Adapting to unknown smoothness via wavelet shrinkage*, J. Am. Stat. Assoc., 90 (1995), pp. 1200–1224. 56

[61] J. Douglas and H. H. Rachford, *On the numerical solution of heat conduction problems in two and three space variables*, Trans. Americ. Math. Soc., 82 (1956), pp. 421–439. 53

[62] A. A. Efros and T. K. Leung, *Texture synthesis by non-parametric sampling*, in Proceedings of the 7th IEEE International Conference on Computer Vision, vol. 2, 1999, pp. 1033–1038. 161

[63] P. P. B. Eggermont, *Maximum entropy regularization for Fredholm integral equations of the first kind*, SIAM J. Math. Anal., 24 (1993), pp. 1557–1576. 60

[64] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*, vol. 1 of Studies in Mathematics and Its Applications, North-Holland Publishing Company, 1976. 21, 31, 33, 35, 37, 38, 126, 146

[65] H. C. Elman and G. H. Golub, *Inexact and preconditioned Uzawa algorithms for saddle point problems*, SIAM J. Numer. Anal., 31 (1994), pp. 1645–1661. 150

[66] R. Engelking, *General Topology*, vol. 6 of Sigma Series in Pure Mathematics, Heldermann Verlag, 1989. 21, 22, 23

[67] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Mathematics and Its Applications, Kluwer Academic Publisher, 2000. 12, 96

[68] S. Esedoglu and S. J. Osher, *Decomposition of images by the anisotropic Rudin-Osher-Fatemi model*, Comm. Pure Appl. Math., 57 (2004), pp. 1609–1626. 123

[69] J. E. Esser, *Primal Dual Algorithms for Convex Models and Applications to*

Image Restoration, Registration and Nonlocal Inpainting, PhD thesis, University of California, Los Angeles, 2010. 54, 145, 149

[70] L. C. Evans, *Partial Differential Equations*, vol. 19 of Graduate Studies in Mathematics, American Mathematical Society, 1998. 126

[71] L. C. Evans and R. F. Gariepy, *Measure Theory and Fine Properties of Functions*, Studies in Advanced Mathematics, CRC Press, 1992. 121, 122, 123, 126

[72] M. A. T. Figueiredo and J. Bioucas-Dias, *Deconvolution of Poissonian images using variable splitting and augmented lagrangian optimization*, in IEEE Workshop on Statistical Signal Processing, Cardiff, 2009. 42, 54, 134, 135

[73] M. Fortin and R. Glowinski, *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, vol. 15 of Studies in Mathematics and its Applications, Elsevier Science Publishers B.V., 1983. 149

[74] D. Gabay, *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, vol. 15 of Studies in Mathematics and its Applications, Elsevier Science Publishers B.V., Amsterdam, 1983, ch. Applications of the method of multipliers to variational inequalities, pp. 299–331. 74

[75] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*, J. Appl. Stat., 20 (1993), pp. 25–62. 16, 56

[76] S. Geman and D. E. McClure, *Bayesian image analysis: an application to single photon emission tomography*, Statistical Computation Section, American Statistical Association, (1985), pp. 12–18. 16

[77] G. Gilboa, J. Darbon, S. Osher, and T. Chan, *Nonlocal convex functionals for image regularization*, CAM Report 06-57, UCLA, 2006. 163, 166

[78] G. Gilboa and S. Osher, *Nonlocal linear image regularization and supervised segmentation*, Multiscale Model. Simul., 6 (2007), pp. 595–630. 162, 166

[79] ——, *Nonlocal operators with applications to image processing*, Multiscale Model. Simul., 7 (2008), pp. 1005–1028. 157, 158, 163, 166, 172, 177, 178, 180

[80] E. Giusti, *Minimal Surfaces and Functions of Bounded Variation*, vol. 80 of Monographs in Mathematics, Birkhäuser, 1984. 121, 122, 124, 126, 128

[81] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and Operator-*

Splitting Methods in Nonlinear Mechanics, vol. 9 of Studies in Applied Mathematics, SIAM, 1989. 149

[82] T. Goldstein and S. Osher, *The split Bregman method for L^1-regularized problems*, SIAM J. Imaging Sci., 2 (2009), pp. 323–343. 135, 145, 148, 177

[83] C. W. Groetsch, *The theory of Tikhonov regularization for Fredholm equations of the first kind*, vol. 105 of Research Notes in Mathematics, Pitman Advanced Publishing Program, 1984. 12, 40

[84] ——, *Inverse Problems in the Mathematical Sciences*, Vieweg Verlag, 1993. 12

[85] C. W. Groetsch and O. Scherzer, *Non-stationary iterated Tikhonov-Morozov method and third-order differential equations for the evaluation of unbounded operators*, Math. Meth. Appl. Sci., 23 (2000), pp. 1287–1300. 88, 90

[86] M. Hein, J.-Y. Audibert, and U. von Luxburg, *Graph Laplacians and their convergence on random neighborhood graphs*, Journal of Machine Learning Research, 8 (2007), pp. 1325–1365. 166, 168, 169, 170, 172

[87] S. W. Hell, *Toward fluorescence nanoscopy*, Nature Biotechnology, 21 (2003), pp. 1347–1355. 4, 11, 40, 41

[88] F. M. Henderson and A. J. Lewis, *Principles and Applications of Imaging Radar: Manual of Remote Sensing*, vol. 2, Wiley and Sons, 1998. 11

[89] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms I*, vol. 305 of Grundlehren der mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences), Springer Verlag, 1993. 47, 51, 57, 65, 94, 104, 147

[90] K. Ito and K. Kunisch, *Lagrange Multiplier Approach to Variational Problems and Applications*, vol. 15 of Advances in Design and Control, SIAM, 2008. 149

[91] A. N. Iusem, *Convergence analysis for a multiplicatively relaxed EM algorithm*, Mathematical Methods in the Applied Sciences, 14 (1991), pp. 573–593. 47

[92] Z. Jin and X. Yang, *A variational model to remove the multiplicative noise in ultrasound images*, J. Math. Imaging Vis., 39 (2011), pp. 62–74. 4, 40, 107, 108, 134

[93] E. Jonsson, S. C. Huang, and T. Chan, *Total variation regularization in positron emission tomography*, CAM Report 98-48, UCLA, 1998. 42, 43, 134

[94] D. Kaplan and Q. Ma, *On the statistical characteristics of log-compressed Rayleigh signals: Theoretical formulation and experimental results*, A. Acoust. Soc. Am., 95 (1994), pp. 1396–1400. 102

[95] J. L. Kelley, *General Topology*, vol. 27 of Graduate Texts in Mathematics, Springer, 1955. 23, 24

[96] C. Kervrann, *An adaptive window approach for image smoothing and structures preserving*, in Proceedings of the 8th European Conference on Computer Vision, vol. 3023 of LNCS, Springer, 2004, pp. 132–144. 162

[97] C. Kervrann and J. Boulanger, *Optimal spatial adaption for patch-based image denoising*, IEEE Trans. Image Process., 15 (2006), pp. 2866–2878. 162

[98] ——, *Unsupervised patch-based image regularization and representation*, in Proceedings of the 9th European Conference on Computer Vision, vol. 3954 of LNCS, Springer, 2006, pp. 555–567. 162

[99] S. Kindermann, S. Osher, and P. W. Jones, *Deblurring and denoising of images by nonlocal functionals*, Multiscale Model. Simul., 4 (2005), pp. 1091–1115. 162

[100] K. C. Kiwiel, *Proximal minimization methods with generalized Bregman functions*, SIAM J. Control Optim., 35 (1997), pp. 1142–1168. 86

[101] K. Krissian, R. Kikinis, C.-F. Westin, and K. Vosburgh, *Speckle-constrained filtering of ultrasound images*, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, 2005, pp. 547–552. 13, 107, 134

[102] 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 (2005), pp. 2500–2513. 4, 40

[103] M. M. Lavrentiev, A. V. Avdeev, M. M. Lavrentiev, Jr., and V. I. Priimenko, *Inverse Problems of Mathematical Physics*, Inverse and Ill-Posed Problems Series, VSP, 2003. 11, 12

[104] T. Le, R. Chartrand, and T. J. Asaki, *A variational approach to reconstructing images corrupted by Poisson noise*, J. Math. Imaging Vis., 27 (2007), pp. 257–263. 42, 56, 57, 134

[105] J.-S. Lee, *Digital image smoothing and the sigma filter*, Computer Vision, Graphics and Image Processing, 24 (1983), pp. 255–269. 160

[106] H. Liao, F. Li, and M. K. Ng, *Selection of regularization parameter in total variation image restoration*, J. Opt. Soc. Am. A, 26 (2009), pp. 2311–2320. 55

[107] P. L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., 16 (1979), pp. 964–979. 43, 53, 54

[108] J. Llacer and J. Núñez, *Iterative maximum likelihood and Bayesian algorithms for image reconstruction in astronomy*, in The Restoration of Hubble Space Telescope Images, R. L. White and R. J. Allen, eds., The Space Telescope Science Institute, Baltimore, Md, USA, 1990, pp. 62–69. 4, 40

[109] Y. Lou, X. Zhang, S. Osher, and A. Bertozzi, *Image recovery via nonlocal operators*, J. Sci. Comput., 42 (2010), pp. 185–197. 165

[110] T. Loupas, W. N. McDicken, and P. L. Allan, *An adaptive weighted median filter for speckle suppression in medical ultrasonic images*, IEEE Trans. Circuits Syst., 36 (1989), pp. 129–135. 102, 107

[111] L. B. Lucy, *An iterative technique for the rectification of observed distributions*, Astronomical Journal, 79 (1974), pp. 745–754. 41, 46

[112] A. Luttman, *A theoretical analysis of L^1 regularized Poisson likelihood estimation*, Inverse Prob. Sci. Eng., 18 (2010), pp. 251–264. 42, 69

[113] M. Mahmoudi and G. Sapiro, *Fast image and video denoising via nonlocal means of similar neighborhoods*, IEEE Signal Processing Letters, 12 (2005), pp. 839–842. 166

[114] R. E. Megginson, *An Introduction to Banach Space Theory*, vol. 183 of Graduate Texts in Mathematics, Springer, 1998. 21, 23, 25, 26, 27, 29, 30, 31

[115] R. Meise and D. Vogt, *Einführung in die Funktionalanalysis*, Aufbaukurs Mathematik, Vieweg, 1992. 24

[116] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures*, vol. 22 of University Lecture Series, American Mathematical Society, Boston, MA, USA, 2001. 85, 92, 123, 130, 131

[117] J. Modersitzki, *Numerical Methods for Image Registration*, Numerical Mathematics and Scientific Computation, Oxford University Press, 2004. 9

[118] B. Mohar, *The Laplacian spectrum of graphs*, in Graph Theory, Combinatorics, and Applications, vol. 2, Wiley, 1991, pp. 871–898. 162

[119] J. Müller, *Parallel total variation minimization*, master's thesis, Institute for Computational and Applied Mathematics, University of Münster, 2008. 146

[120] J. R. Munkres, *Topology: A First Course*, Prentice-Hall, 1975. 23

[121] A. Myronenko, January 2011. https://sites.google.com/site/myronenko/software. 184

[122] M. Nagorni and S. W. Hell, *4Pi-confocal microscopy provides three-dimensional images of the microtubule network with 100- to 150-nm resolution*, J. Struct. Biol., 123 (1998), pp. 236–247. 44

[123] F. Natterer and F. Wübbeling, *Mathematical Methods in Image Reconstruction*, SIAM Monographs on Mathematical Modeling and Computation, 2001. 11, 43, 46, 47, 84, 183, 184

[124] M. Nikolova, *A variational approach to remove outliers and impulse noise*, J. Math. Imaging Vis., 20 (2004), pp. 99–120. 132

[125] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, *An iterative regularization method for total variation-based image restoration*, Multiscale Model. Simul., 4 (2005), pp. 460–489. 85, 88, 91, 92, 95, 96, 99, 130, 132

[126] S. Osher, A. Solé, and L. Vese, *Image decomposition and restoration using total variation minimization and the H^{-1} norm*, Multiscale Model. Simul., 1 (2003), pp. 349–370. 132

[127] V. Y. Panin, G. L. Zeng, and G. T. Gullberg, *Total variation regulated EM algorithm [SPECT reconstruction]*, IEEE Trans. Nucl. Sci., 46 (1999), pp. 2202–2210. 42, 43, 134

[128] G. B. Passty, *Ergodic convergence to a zero of the sum of monotone operators in hilbert spaces*, J. Math. Anal. Appl., 72 (1979), pp. 383–390. 43, 54

[129] P. Perona and J. Malik, *Scale-space and edge detection using anisotropic diffusion*, IEEE Trans. PAMI, 12 (1990), pp. 629–639. 89, 158, 159

[130] G. Peyré, S. Bougleux, and L. Cohen, *Non-local regularization of inverse problems*, in Proceedings of the 10th European Conference on Computer Vision, LNCS 5304, Springer, 2008, pp. 57–68. 165

[131] R. Plato, *On the discrepancy principle for iterative and parametric methods to solve linear ill-posed equations*, Numer. Math., 75 (1996), pp. 99–120. 96

[132] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, 3rd ed., 2007. 193

[133] M. Renardy and R. C. Rogers, *An Introduction to Partial Differential Equations*, vol. 13 of Texts in Applied Mathematics, Springer, 2nd ed., 2003. 27

[134] E. Resmerita and R. S. Anderssen, *Joint additive Kullback-Leibler residual minimization and regularization for linear inverse problems*, Math. Meth. Appl. Sci., 30 (2007), pp. 1527–1544. 49, 59, 60, 61, 64, 69

[135] E. Resmerita, H. W. Engl, and A. N. Iusem, *The expectation-maximization algorithm for ill-posed integral equations: a convergence analysis*, Inverse Problems, 23 (2007), pp. 2575–2588. 47, 48, 184

[136] W. H. Richardson, *Bayesian-based iterative method of image restoration*, J. Opt. Soc. Am., 62 (1972), pp. 55–59. 41, 46

[137] R. T. Rockafellar, *Convex Analysis*, Princeton Landmarks in Mathematics, Princeton University Press, 1970. 31

[138] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268. 19, 42, 56, 92, 121, 128

[139] W. Rudin, *Functional Analysis*, McGraw-Hill Series in Higher Mathematics, McGraw-Hill Book Company, 1973. 28

[140] A. Sawatzky, C. Brune, J. Müller, and M. Burger, *Total variation processing of images with Poisson statistics*, in Proceedings of the 13th International Conference on Computer Analysis of Images and Patterns, vol. 5702 of LNCS, Springer, 2009, pp. 533–540. 6

[141] 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, 2008, pp. 5133–5137. 6

[142] K. P. Schäfers, T. J. Spinks, P. G. Camici, P. M. Bloomfield, C. G. Rhodes, M. P. Law, C. S. R. Baker, and O. Rimoldi, *Absolute quantification of myocardial blood flow with $H_2{}^{15}O$ and 3-dimensional PET: An experimental validation*, J. Nucl. Med., 43 (2002), pp. 1031–1040. 187

[143] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen, *Variational Methods in Imaging*, vol. 167 of Applied Mathematical Sciences, Springer, 2009. 21, 29, 56, 86, 88, 122

[144] O. Scherzer and C. W. Groetsch, *Inverse scale space theory for inverse problems*, in Scale-Space and Morphology in Computer Vision, vol. 2106 of LNCS, Springer, 2001, pp. 317–325. 88, 90, 91

[145] R. Sedgewick, *Algorithms in C: Fundamentals, Data Structures, Sorting, Searching*, Addison-Wesley Professional, 3rd ed., 1997. 193

[146] S. Setzer, *Splitting Methods in Image Processing*, PhD thesis, University of Mannheim, 2009. 53, 54, 145, 149

[147] S. Setzer, G. Steidl, and T. Teuber, *Deblurring Poissonian images by split Bregman techniques*, J. Vis. Commun. Image R., 21 (2010), pp. 193–199. 42, 54, 134, 135

[148] L. A. Shepp and Y. Vardi, *Maximum likelihood reconstruction for emission tomography*, IEEE Trans. Med. Imaging, 1 (1982), pp. 113–122. 4, 11, 40, 41, 44, 46, 47

[149] H.-C. Shin, R. Prager, H. Gomersall, N. Kingsbury, G. Treece, and A. Gee, *Estimation of speed of sound in dual-layered media using medical ultrasound image deconvolution*, Ultrasonics, 50 (2010), pp. 716–725. 102

[150] A. Singer, Y. Shkolnisky, and B. Nadler, *Diffusion interpretation of nonlocal neighborhood filters for signal denoising*, SIAM J. Imaging Sci., 2 (2009), pp. 118–139. 161

[151] S. M. Smith and J. M. Brady, *SUSAN - a new approach to low level image processing*, Int. J. Comput. Vision, 23 (1997), pp. 45–78. 160

[152] 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 (1993), pp. 1014–1023. 4, 40

[153] G. Steidl and T. Teuber, *Anisotropic smoothing using double orientations*, in Proceedings of the 2nd International Conference on Scale Space and Variational Methods in Computer Vision, LNCS 5567, Springer, 2009, pp. 477–489. 123

[154] ——, *Removing multiplicative noise by Douglas-Rachford splitting methods*, J. Math. Imaging Vis., 36 (2010), pp. 168–184. 177

[155] D. M. Strong, J.-F. Aujol, and T. F. Chan, *Scale recognition, regularization parameter selection, and Meyer's G norm in total variation regularization*, Multiscale Model. Simul., 5 (2006), pp. 273–303. 55

[156] C. Tomasi and R. Manduchi, *Bilateral filtering for gray and color images*, in Proceedings of the 6th IEEE International Conference on Computer Vision, 1998, pp. 839–846. 160

[157] P. Tseng, *Applications of a splitting algorithm to decomposition in convex programming and variational inequalities*, SIAM J. Control Optim., 29 (1991), pp. 119–138. 43, 53, 74

[158] M. Tur, K. C. Chin, and J. W. Goodman, *When is speckle noise multiplicative?*, Applied Optics, 21 (1982), pp. 1157–1159. 4, 13, 40, 102

[159] T. A. Tuthill, R. H. Sperry, and K. J. Parker, *Deviations from Rayleigh statistics in ultrasonic speckle*, Ultrason. Imag., 10 (1988), pp. 81–89. 102

[160] Y. Vardi, L. A. Shepp, and L. Kaufman, *A statistical model for positron emission tomography*, J. Am. Stat. Assoc., 80 (1985), pp. 8–20. 4, 40, 41, 47, 183

[161] L. A. Vese and S. J. Osher, *Modeling textures with total variation minimization and oscillating patterns in image processing*, J. Sci. Comput., 19 (2003), pp. 553–572. 132

[162] C. R. Vogel, *Computational Methods for Inverse Problems*, Frontiers in Applied Mathematics, SIAM, 2002. 55

[163] R. F. Wagner, S. W. Smith, J. M. Sandrik, and H. Lopez, *Statistics of speckle in ultrasound B-scans*, IEEE Trans. Sonic Ultrason., 30 (1983), pp. 156–163. 102

[164] Y. Wang, J. Yang, W. Yin, and Y. Zhang, *A new alternating minimization algorithm for total variation image reconstruction*, SIAM J. Imaging Sci., 1 (2008), pp. 248–272. 145, 149, 152

[165] J. Weickert, *Anisotropic Diffusion in Image Processing*, Teubner, 1998. 89, 158, 159

[166] T. Weissman, E. Ordentlich, G. Seroussi, S. Verdu, and M. J. Weinberger, *Universal discrete denoising: known channel*, IEEE Trans. Inf. Theory, 51 (2005), pp. 5–28. 162

[167] M. N. Wernick and J. N. Aarsvold, eds., *Emission Tomography: The Fundamentals of PET and SPECT*, Elsevier Academic Press, 2004. 4, 11, 40, 41, 44, 135, 183

[168] L. P. Yaroslavsky, *Digital Picture Processing: An Introduction*, Springer, 1985. 160

[169] K. Yosida, *Functional Analysis*, Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen, Band 123, Springer, 4th ed., 1974. 25, 27, 28, 29, 30

[170] X. Zhang, M. Burger, X. Bresson, and S. Osher, *Bregmanized nonlocal regularization for deconvolution and sparse reconstruction*, SIAM J. Imaging Sci., 3 (2010), pp. 253–276. 165, 166, 177

[171] D. Zhou and B. Schölkopf, *A regularization framework for learning from graph data*, in Proceedings of the ICML Workshop on Statistical Relational Learning and Its Connections to Other Fields, 2004, pp. 132–137. 157, 163

[172] ——, *Regularization on discrete spaces*, in Pattern Recognition, Proceedings of the 27th DAGM Symposium, vol. 3663 of LNCS, Springer, 2005, pp. 361–368. 157, 163

Lebenslauf

Name	Alex Sawatzky
Geburtsdatum	06.08.1982
Geburtsort	Orechow-Log / Russland
Familienstand	ledig
Eltern	Alfred Sawatzky Walentina Sawatzky, geb. Didenko
Schulbildung	Hermann-Leeser-Realschule, von 1995 – 1999 in Dülmen Hans-Böckler-Schule, von 1999 – 2002 in Münster
Abitur	am 22.06.2002 in Münster
Studium	Mathematik mit Nebenfach Informatik; Westfälische Wilhelms-Universität Münster, von 2002 – 2007
Promotionsstudiengang	Mathematik Westfälische Wilhelms-Universität Münster, von 2007 – 2011
Prüfungen	Diplom im Fach Mathematik am 24.05.2007 an der Westfälischen Wilhelms-Universität Münster
Tätigkeiten	Studentische Hilfskraft von 2005 – 2007
Beginn der Dissertation	Juni 2007, Institut für Numerische und Angewandte Mathematik, Betreuer Prof. Dr. Martin Burger